

**Assessing the impact and effectiveness of anti-trafficking interventions in the  
criminal justice sector:**

**A discussion paper on issues, obstacles and opportunities**

## Table of contents

|   |    |
|---|----|
| <i>Acknowledgments</i> .....  | 3  |
| <i>Key concepts and terms</i> .....                                       | 4  |
| <i>Executive summary</i> .....  | 5  |
| Introduction .....  | 10 |
| 1: Understanding and defining success.....                                | 12 |
| 2: Attributing and understanding change .....                             | 18 |
| 3: Data availability and accessibility .....                              | 20 |
| 4. Data quality: issues of method, validity and reliability.....          | 23 |
| 5: Identifying unintended consequences .....                              | 25 |
| 6: The cost of evaluations: dealing with budget and time constraints..... | 29 |
| Towards the future.....   | 31 |
| <i>Bibliography</i> .....   | 33 |

## **Acknowledgments**

This discussion paper was written by Anne Gallagher and Rebecca Surtees, with the assistance of Andrea Zakarias. The authors wish to thank those who agreed to be interviewed for so generously sharing their experiences and insights.

The discussion paper was supported by the Australian Agency for International Development (AusAID) through the Asia Regional Trafficking in Persons Project. However, the views expressed should not be taken to reflect those of the Australian Government or of AusAID.

## Key concepts and terms

**Criminal justice sector:** Laws, policies, institutions and practices related to the investigation, prosecution and adjudication of crimes as well as the sentencing and treatment of offenders. In the context of trafficking in persons, the term “criminal justice sector” is used to identify areas of focus or work in the above areas that do not deal primarily with victim support, assistance and repatriation / reintegration, or with prevention of trafficking. Typical “criminal justice sector” interventions include support for reform of criminal laws; training of police, prosecutors and judges; and institutional reform (e.g. capacity building of institutions such as specialist investigation units, prosecutorial offices and courts, and development of procedures and protocols governing investigations and prosecutions). Sometimes external criminal justice interventions will extend to supporting the process of investigation and even prosecution.

**Evaluation<sup>1</sup>:** The systematic and objective assessment of an on-going or completed project, programme or policy; its design, implementation and results. The aim of an evaluation is generally to determine the relevance and fulfilment of objectives, development efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors. An **impact evaluation** will focus specifically on measuring the positive and negative, primary and secondary long-term effects produced by a development intervention. Such an evaluation would measure direct and indirect impacts as well as those that are both intended and unintended. An evaluation focusing on **effectiveness** would measure the merit, worth or significance of an activity, policy or programme.

**Trafficking in persons:** Generally understood to refer to the process by which individuals are placed in situations of exploitation from which they cannot escape. The international legal definition<sup>2</sup> comprises three elements: (i) an *action* such as transporting or harbouring; (ii) secured through *means* such as force, deception and abuse of authority; (iii) for the purpose of *exploitation* such as forced labour and sexual exploitation. The international legal definition of trafficking in children excludes the means element, i.e., for trafficking in children to occur, it is sufficient that one of the stipulated actions has been undertaken, accompanied by an intention to exploit.

---

<sup>1</sup> Adapted from Development Assistance Committee (DAC) (2002), *Glossary of key terms in evaluation and results based management*: Paris: Organisation for Economic Co-operation and Development (OECD) DAC [hereafter DAC 2002].

<sup>2</sup> As set out in Article 2 of the United Nations Protocol to Prevent, Suppress and Punish Trafficking in Persons, Especially Women and Children supplementing the United Nations Convention against Transnational Organized Crime, Dec. 12, 2000, UN Doc. A/RES/55/25 (Annex II), entered into force Dec. 25, 2003.

## **Executive summary**

Over the past decade, global concern about human trafficking has prompted massive investment into anti-trafficking interventions by intergovernmental organisations, states and civil society. While initial waves of intervention took place in a performance evaluation vacuum, there have been growing calls for greater transparency and accountability within the anti-trafficking sector, including through rigorous impact evaluation. Such calls are fully justified. The human rights imperatives that underpin anti-trafficking work and the significant investment of public resources demand that interventions demonstrate accountability, results and beneficial impact. However, the question of how this can happen in practice is more complicated and there has been relatively little analysis of the practical issues and challenges that may arise in efforts to evaluate anti-trafficking work. In the specialised area of criminal justice responses to trafficking, such analysis is virtually non-existent, despite the increasing attention and resources that are being directed to this aspect of the anti-trafficking response.

Impact and effectiveness evaluation in the context of international development is both complicated and contested. Multiple theories and elaborate methodologies abound, and these can present a daunting impediment to those seeking practical guidance on the essential problem of determining “what works”. This discussion paper has sought to cut through some of the complexities by focusing attention on several areas that are directly implicated in measuring anti-trafficking interventions in the criminal justice sector. It is the result of a comprehensive review of relevant literature and recent evaluation reports, as well as interviews with key players in the field.

### **Understanding and defining success**

Determinations of success (or failure) in relation to a particular criminal justice intervention are likely to vary according to who one consults and their role in the intervention. The discussion paper identifies a range of interested parties including the recipient government / agency; implementing partner organisation; donors; involved criminal justice practitioners (e.g. investigators, prosecutors, judges); trafficking victims and victim support agencies; suspects; evaluators; technical experts; and silent stakeholders such as an external government that is seeking to promote a particular response to trafficking. Each of these parties may assess the success (or failure) of an intervention using different standards and criteria. For example, a priority issue for a recipient government may be whether the intervention helped to ameliorate external criticisms of its response. For a donor, the question of whether the intervention has secured clear, unambiguous results that can be quantified, measured and reported is likely to be critical to its view of success. On the other hand, victims and victim support agencies, may judge an intervention on very different criteria, such as whether it facilitated victim removal from exploitation and their protection from further harm.

Within these stakeholder groups, there may be internal differences as to what constitutes a successful outcome and, even at the individual level, views on what constitutes “success” may change over time and in response to different context. For example, some victims may view the prosecution of their exploiters, even after a long and difficult process, as a positive and empowering result. For others, their involvement in the criminal justice process may have been so traumatic and disempowering that a conviction becomes irrelevant. In practice, it is clear that the perspectives of some stakeholders are prioritised over others. Donors and implementing agencies often exercise disproportionate influence over the assessment process and its findings. Unsurprisingly, it is the views of the most direct and easily accessible stakeholders that are routinely sought and taken into account. This may be one

reason why victims are rarely involved in impact assessments of anti-trafficking interventions in the criminal justice sector. While a strong impact assessment should evidently take into account different views and positions, it is also important to accept that not all stakeholders are equal and not all perspectives are, or should be, of equal weight. Perhaps the most that can be done is to ensure that the criteria for success are transparent and logical; that all major stakeholders are consulted; and that different perceptions of the programme and its success are openly acknowledged in the evaluation, even when this diverges from the general assessment.

Ensuring that criteria for success are transparent and logical is another challenge. Many anti-trafficking interventions in the criminal justice sector set success criteria at a high level of ambition and abstraction or, conversely, at a very low output level. Both approaches enable implementers to exercise strong control over the “success story”. There is a strong preference, among donors and others, for quantifiable indicators such as number of arrests, prosecutions, officers trained, etc. Views on success and attached criteria are often influenced by underlying assumptions that may not be specifically articulated or acknowledged. These include the assumption that increases in prosecutions and convictions are a reliable indicator of success; that declines in prevalence indicate positive change; and that certain commonly preferred indicators such as behavioural and attitudinal change can in fact be measured. None of these assumptions has a strong evidence base and all deserve to be interrogated.

### **Attributing and understanding change**

A critical aspect of evaluation is not only to assess what change has been made, but also the extent to which this can be attributed to the specific programme or intervention being evaluated. In the case of anti-trafficking work, attribution is complicated by the fact that trafficking and responses to trafficking are both situated within much larger, complex economic, political and social systems that are themselves subject to numerous pressures and influences. In short, the criminal justice response to trafficking does not take place in a vacuum; it is informed by a wide range of factors that are beyond the control of a single project or programme. Measuring training-related attitudinal and behavioural change among law enforcement officials, for example, requires a pre-intervention baseline that in most circumstances will be impossible to accurately establish. Even with such a baseline, a range of other factors outside the intervention could impact on the nature and rate of change. Important external influences identified by the study include: the broader criminal justice environment (laws, institutions and practices); corruption; related and concurrent interventions; and political pressure – both internal and external. A rigorous approach to evaluation requires the acknowledgement of factors influencing change and the allocation of an appropriate weighting to these contributing factors.

### **Data availability and accessibility**

Much of the data needed to evaluate criminal justice responses to trafficking is scarce or non-existent. This severely compromises the capacity of criminal justice interventions to collect baseline information against which future developments can be assessed. Even where good data does exist, it is not always accessible to evaluators or even to project staff for monitoring and other purposes. As a consequence, the task of measuring performance and change is rendered difficult or impossible.

Many factors impact on the availability and accessibility of evaluation data. Often politics and national priorities play an important role. Governments may fail to collect information that is not considered to be important to them. Criminal justice agencies may suppress or manipulate data to avoid criticism or the exposure of corrupt practices. Data accessibility can also be controlled to justify a certain evaluation outcome. In all countries, the criminal justice culture generally does not encourage the sharing of information with outsiders including those in a position to critically evaluate their practices. Often a lack

of cooperation and trust will explain why certain data is not made available to external players. Short-term interventions may never develop the relationships of trust that permit the sharing of sensitive information. The issue of capacity is also relevant. Data collection, maintenance and analysis are expensive and time-consuming, often requiring high levels of expertise. In many countries, these resources are scarce or simply unavailable. Legal processes and the institutionalised disconnect between agencies (such as investigators and prosecutors) can also contribute to fragmented data: 'creating fault lines in the system itself through which the data falls and cannot be found'.<sup>3</sup>

In some cases, the criminal justice data that would be of most interest to an evaluation cannot be made available because it is protected by law, or should not be made available if its release could be prejudicial to privacy, security or ethical concerns. These are important concerns and, provided they are used legitimately, need to be respected by external parties. In this regard it is important to acknowledge the existence of a worrying double standard: donors and external bodies regularly expect and demand access to information that would never be made available in their own countries.

### **Data quality: issues of method, validity and reliability**

If evaluation design and implementation are compromised by missing data or poor analysis, the findings will be inadequate and likely irrelevant. Currently the quality of trafficking-related data is very weak. Threats to the quality of data in the context of criminal justice responses are multiple and complex. Lack of comparability is one problem: differences in laws and criminal justice practices mean that data is generally not comparable between countries and even less comparable at regional or global levels. In some countries, internal or external pressure to improve performance can compromise comparability of data sets by inflating investigation or prosecution statistics with cases that are not in fact trafficking. Further, the common practice of different agencies collecting information independent of each other results in different data sets or intersecting datasets that cannot easily be consolidated, particularly in an environment where information sharing and case tracking is inadequate or non-existent.

It is essential that those engaged in evaluation of anti-trafficking interventions acknowledge the potential impact of bias and selection effects on the validity of findings related to outcomes and impact. For example, the tendency of criminal justice authorities to focus their attention on soft targets (e.g. low-level recruiters and brokers rather than exploiters), and thereby to create selection effects and biases in available data, appears to be widespread. Current criminal justice approaches to trafficking that tend to focus on one 'type' of victim or one 'type' of exploitation, mean that reported cases are also not likely to be representative of the actual situation. As a result, evaluation findings based on reported cases can, at best, be applied to this particular group and not extrapolated to the broader universe of unreported cases of trafficking. Respondent bias can also distort data. For example, criminal justice agencies may control access to key documents such as court records or case files in an effort to present a particular kind of case or result that is not necessarily representative of the actual situation.

A mix of quantitative and qualitative indicators appears to be needed to effectively measure the outcome and impact of any trafficking related intervention in the criminal justice sector. Numbers and statistics are important in providing a sense of the scale of change or improvement. However, they are not very helpful in explaining why a particular change or improvement took place. Qualitative inquiry can capture more ephemeral but even more important changes such as those related to behaviour or organisational culture.

---

<sup>3</sup> Interview, Bangkok, Thailand, 2 September 2010.

## Identifying unintended consequences

It is a generally accepted principle that evaluations conducted within the context of international development assistance should capture all effects of an intervention – intended and unintended, negative and positive. In short, knowledge about what *does not* work is as valuable as knowledge about what does. This principle has particular resonance in the context of anti-trafficking interventions in the criminal justice sector, where the potential for negative consequences has been shown to be very high. Despite this acknowledged risk, evaluations of anti-trafficking interventions rarely consider negative or unexpected consequences. There is some indication that the situation is changing but attention to negative impacts generally remains incidental and superficial. For example, while a few evaluation reports examined for this discussion paper expressed an initial commitment to considering unanticipated positive and negative results, these commitments did not flow through to the reports' findings or recommendations.

The lack of attention to negative or unintended results is not unique to the anti-trafficking sector and insights from other contexts can be useful to understanding why this happens and how it can be addressed. For example, it appears that the traditional evaluation structure, which assesses performance against pre-determined indicators, will naturally focus on *expected* change and leave aside the unexpected and unintended. The existence of strong institutional disincentives to adverse findings should not be discounted in seeking to understand why negative impacts are not sought or welcomed. In short, there are usually few winners among donors and implementing agencies when an expensive intervention is shown to have produced negative externalities. Political and strategic considerations will affect the willingness of those involved in encouraging, funding or otherwise sponsoring trafficking-related interventions to play down potential problems associated with their preferred approaches.

## The cost of evaluations: dealing with budget and time constraints

While different interventions and evaluation models will require different levels of resources, it can be expected that any serious and comprehensive evaluation will be both time and resource intensive. This is particularly true for criminal justice interventions where the “gold standard” approach of tracking cases through the system requires massive investment in time, resources and relationship-building. While it is certainly possible to undertake evaluations within “real world” budget constraints, it is necessary to acknowledge that this will not be without consequence. Budgetary considerations affect every aspect of an evaluation, including the number of interviews that can be conducted; the type of data that can be collected; the ability to combine quantitative and qualitative data collection and analysis; the size and professional experience of the research team; the level of analysis that can be conducted; and the cost of translation and interpretation. In addition, the question of who, amongst target groups and stakeholders, should be involved in the evaluation process (and thereby contribute to defining “success”) is often considered and decided with reference to cost and time considerations. A threshold, different for every evaluation, will be reached where cost limitations threaten the validity of evaluation findings.

Of the interventions examined for this discussion paper, it appears that, while substantial resources were invested in design, monitoring, and output measurement, far less was applied to measuring and evaluating outcomes and longer-term impacts. However, striking a balance between intervention costs and evaluation costs can be tricky. Beyond certain very general principles (e.g. that the level of resources allocated to an evaluation should be proportionate to the scale of the intervention and its anticipated impact), it is difficult to be specific about how much is “too much” and “too little”. That being said, it is essential to recognise the potential losses when evaluations do not take place at all, or are

undertaken at too low a level. Findings from evaluations can help to avoid costly mistakes and negative consequences – as well as to ensure that effective approaches are captured and replicated.

Recent calls for more rigor and transparency in evaluating the success or failure of anti-trafficking interventions are both welcome and overdue. Criminal justice responses to trafficking, often reactive, unplanned, and at high risk of negative consequences, are in particular need of scrutiny. The present paper affirms the need for discussions about evaluating anti-trafficking interventions to move from criticism and complaint about what is not being done well (or at all) to serious, informed engagement with obstacles and constraints and how these can be addressed. A key challenge that has emerged from the present research is the lack of a common vision of what constitutes an effective criminal justice response to trafficking. It is the collective responsibility of donors, implementing agencies and others with power and influence, to work towards articulation of such a vision and to ensure accountability of those who are provided the resources to promote its realisation.

# Assessing the impact and effectiveness of anti-trafficking interventions in the criminal justice sector: issues, obstacles and opportunities

## Introduction

Global concern about human trafficking has nurtured great legal and normative change. Over the past decade it has also prompted massive investment into anti-trafficking interventions by intergovernmental organisations, states and civil society. Initial waves of intervention took place in a performance evaluation vacuum. This was noted by, amongst others, the United States (US) Government Accountability Office which, in 2006, criticised the absence of ‘measurable goals and associated indicators to evaluate the overall effectiveness of [US] efforts to combat trafficking abroad’ and echoed the State Department’s Inspector-General in calling for ‘performance indicators to compare progress in combating trafficking from year to year’.<sup>4</sup> Time, experience and heightened expectations about what can be achieved have fed a demand for tools and mechanisms to make sense of the problem and validate what is being done by, for example, measuring the true extent of trafficking or evaluating the absolute and comparative worth of an individual State’s response or a particular intervention. A rapid rise in the formulation and application of “success indicators” is one manifestation of the new environment within which anti-trafficking is being discussed and targeted.<sup>5</sup>

Equally relevant are the increasingly frequent calls for greater transparency and accountability within the anti-trafficking sector, including through rigorous impact evaluation.<sup>6</sup> It is not difficult to sustain a strong argument that anti-trafficking interventions, including those in the criminal justice sector, should be carefully monitored and evaluated. Certainly the human rights imperatives that underpin anti-trafficking work and the significant investment of public resources demand that interventions demonstrate accountability, results and beneficial impact. The question of how this can happen in practice is more complicated and there has been relatively little analysis of the practical issues and challenges that may arise in efforts to evaluate anti-trafficking work. It is the authors’ view that

---

<sup>4</sup> United States Government Accountability Office (USGAO), *Human Trafficking: Better Data, Strategy, and Reporting Needed to Enhance U.S. Anti-Trafficking Efforts Abroad*, p. 25 (USGAO, 2006) [hereafter USGAO 2006].

<sup>5</sup> Anne Gallagher and Janie Chuang, ‘The Use of Indicators to Measure Government Responses to Human Trafficking’, in Kingsbury and Merry eds. *Indicators as a Technology of Global Governance* (Oxford University Press, forthcoming 2012).

<sup>6</sup> Ginny Baumann, *Measuring the Effectiveness of Anti-Slavery Work* (Free the Slaves, 2006); Gergana Danailova-Trainor ‘Evaluation of Counter-Trafficking Programmes and Research’ in Christine Aghazarm et al, *Human Trafficking: New Directions for Research*, pp. 44-56 (International Organization for Migration (IOM), 2008) [hereafter Danailova-Trainor 2008]; Mike Dottridge, ‘Research needs concerning the monitoring, evaluation and impact assessment of both research about human trafficking and projects and programmes to address human trafficking’, in Christine Aghazarm et al, *Human Trafficking: New Directions for Research*, pp. 57-71 (IOM, 2008) [hereafter Dottridge 2008]; Mike Dottridge, *Measuring Responses to Trafficking in Human beings in the European Union: An Assessment Manual* (European Commission, 2007); Global Alliance Against Traffic in Women (GAATW), ed., *Collateral Damage: The Impact of Anti-Trafficking Measures on Human Rights Around the World* (GAATW, 2007) [hereafter GAATW 2007]; GAATW Caroline Hames, Fleur Dewar and Rebecca Napier-Moore, *Feeling good about feeling bad ... A global review of evaluation in anti-trafficking initiatives*, p. 5 (GAATW, 2010) [hereafter GAATW 2010]; Jo Ann Intili et al., *An Evaluation Framework for USAID-Funded TIP Prevention and Victim Protection Programs*, p. 20 (Social Transition Team, Office of Democracy, Governance and Social Transition of the United States Agency for International Development (USAID/E&E/DGST), 2009) [hereafter Intili et al. 2009]; IOM, *Handbook on performance indicators for counter-trafficking projects*, pp. 8, 17, 49 (IOM, 2008) [hereafter IOM 2008]; Ann Jordan, ‘More funding for impact assessments?’, Rights Work: [www.rightswork.org](http://www.rightswork.org), online publication, 15 October 2010; Ruth Rosenberg, *Best Practices for Programming to Protect and Assist Victims of Trafficking in Europe and Eurasia*, pp. 34, 69 (USAID/E&E/DGST, 2008) [hereafter Rosenberg 2008]; Rebecca Surtees, *Re/integration of trafficked persons: developing monitoring and evaluation mechanisms*, p. 19 (KBF and NEXUS Institute, 2009) [hereafter Surtees 2009]; USGAO, *Human Trafficking: Monitoring and Evaluation of International Projects Are Limited, but Experts Suggest Improvements*, pp.10-11, 36 (USGAO, 2007); Erin Williamson et al., *Where is the Research on Human Trafficking and the Evaluation of Anti-Trafficking Efforts?*, pp. 3, 6 (ICF International, 2008).

discussions of anti-trafficking evaluations should identify and acknowledge these obstacles and constraints, and consider what might be done to accommodate or overcome them.

Impact and effectiveness evaluation in the context of international development is both complicated and contested. Multiple theories and elaborate methodologies abound, and these can present a daunting impediment to those seeking practical guidance on the essential problem of determining “what works”. This discussion paper has sought to cut through some of the complexities by focusing attention on several basic topics that are directly implicated in measuring anti-trafficking interventions in the criminal justice sector. These include: 1) understanding and defining success; 2) accounting for internal and external factors; 3) issues concerning availability and accessibility of criminal justice data; 4) concerns about data quality for evaluations; 5) taking account of unintended consequences; and 6) balancing the need for evaluations against their often heavy costs. This analysis is by no means exhaustive. There is currently insufficient expertise and practice in this sector to support a rigorous exploration of how impact evaluation is being conducted and how it should be strengthened.

With these caveats, the authors set for themselves the modest goal of contributing, in a practical way, to current discussions about how trafficking related interventions, specifically those in the criminal justice sector, can be effectively evaluated. The discussion paper affirms the need for impact and effectiveness evaluation, while identifying and exploring the challenges and obstacles that are inevitably implicated. To that extent it should be relevant not only for practitioners and policy makers who work directly on anti-trafficking criminal justice interventions, but also to donors who fund these initiatives. While the paper’s focus is on one aspect of the trafficking response, many of the identified issues are equally relevant to the task of evaluating interventions in other areas, such as anti-trafficking prevention and protection.

The discussion paper is based on a review of relevant literature including a selection of recent evaluation reports of interventions with a significant criminal justice focus, as well as confidential interviews with twelve individuals currently working in the anti-trafficking field with professional backgrounds in criminal justice (law enforcement, prosecution, the judiciary), international law, human rights, development, and monitoring and evaluation. The findings and conclusions also reflect the experiences of the authors in designing, managing, implementing and evaluating anti-trafficking interventions in different regions, including Europe, the former Soviet Union and Asia.

## Part 1: Understanding and defining success

Determining whether a project or intervention has been successful is the ultimate aim of any evaluation. This exercise is rarely simple or direct. Determinations of success (or failure) are likely to vary according to who one consults and their role in the intervention, as well as, most importantly, the criteria against which success is measured and the assumptions that are built into that criteria.

### *Who's success?*

All anti-trafficking criminal justice interventions involve multiple stakeholders, each of whom will be impacted differently and may therefore be expected to have different (and even conflicting) views on what constitutes “success”. Externally supported interventions increase the range of stakeholders who will have a perspective on whether and why the intervention has been successful. This group potentially includes the recipient government / agency; implementing partner organisation; donors; involved criminal justice practitioners (e.g. investigators, prosecutors, judges); trafficking victims and victim support agencies; suspects; evaluators; technical experts; and silent stakeholders.

A **recipient government**, for example, may consider whether the intervention reflected well on the target agency and whether it is possible to claim national credit for any successes. The extent to which the intervention addressed or deflected criticisms of the governments’ response to trafficking could also be an important factor, as could its contribution to national priorities and plans. For an **implementing organisation**, success factors could include consideration of whether the intervention raised the profile of the organisation in a positive way or whether the funding agency is satisfied. The extent to which the intervention contributed to broader organisational goals and strategies could also be relevant to perceptions of success. **Donors** may reflect on whether the intervention has secured clear, unambiguous results that can be quantified, measured and reported and whether it can be justified as representing value for money. In evaluating success, donors may also question whether the intervention strengthened important political or strategic relationships (e.g. with a partner country or regional institution) or whether it could be replicated elsewhere. **Criminal justice practitioners** within the target agency involved as partners may consider whether there has been any measurable improvement against institutional success criteria such as increases in victim rescues, arrests, prosecutions, convictions and involvement of victim-witnesses. They may also be interested in whether perpetrators have been deterred from committing future offences and whether there are any negative effects of the intervention on other aspects of their work.

**Victims and victim support agencies** can be expected to adopt very different criteria in determining the success of a criminal justice intervention. These could include whether the intervention facilitated accurate and timely identification of victims, their removal from exploitation and their protection from further harm; whether victims received the assistance and protection they required; and whether the intervention led to greater criminal justice awareness of the rights of trafficked persons and translated into appropriate treatment. Other relevant success criteria for victim support agencies could include whether the legal process was comprehensible to the victim, and whether it supported access to benefits such as compensation and any right to stay in the destination country.

If the views of **suspects** were considered important to judgments of success, relevant criteria could include whether the intervention supported fair trial rights and proportionate sentencing, and whether correct procedures were followed in the legal process. Additional and different perspectives may include those of the **evaluators** whose primary focus would likely be on the extent to which the

intervention secured its stated aims, performed against its predetermined indicators and demonstrated value for money. **Technical experts** and others attached to the implementing agency could have a different view of what constitutes success, which may (or may not), intersect and overlap with some of the views outlined above.

Finally, in this area, it is possible to identify **silent stakeholders** who, while not formally associated with the intervention, nevertheless make their own judgment of success and potentially sway the perspectives of others. A conspicuous example in the present context is the US Government, which engages in a controversial but highly influential annual evaluation of State responses to trafficking.<sup>7</sup> The US evaluation mechanism uses a range of success criteria that, particularly in application, prioritises stronger criminal justice responses, evidenced by institution building and increases in prosecutions and convictions. This process may well affect the perspective of key stakeholders, such as recipient states and criminal justice agencies, on the success of any external intervention. Its role as a major donor in the area of criminal justice responses to trafficking provides the US Government with an additional avenue through which to advocate its vision of success.

### ***Dealing with conflicting and divergent opinions on success***

Within these stakeholder groups, there may be internal differences as to what constitutes a successful outcome and, even at the individual level, views on what constitutes “success” may change over time and in response to different context. Some victims, for example, may view the conviction of their exploiter, even after a long and personally difficult legal process, as a positive (and even empowering) result. For others the process may be so profoundly disorienting and disempowering that even a successful prosecution cannot assuage their dissatisfaction. Importantly, victim support agency views of success may not align with those of trafficking victims. Agencies may value the criminal justice process differently than individual trafficking victims (e.g. prioritising prosecutions or compensation claims). In addition, victim support agencies, like other stakeholders, are subject to a range of institutional and programmatic considerations that will not necessarily resonate with their clients. The issue of shelter detention for victims of trafficking provides a relevant example, with a recent study confirming a sharp divergence in views, between support agencies and victims regarding the value and impact of shelter detention.<sup>8</sup>

Discussions and document reviews highlight that the perspectives of certain stakeholders are often prioritised in evaluating criminal justice interventions in the counter-trafficking sector while other perspectives are marginalised or discounted altogether. For example, in many externally funded interventions, monitoring and evaluation is very much a donor-driven process. Donors exercise a tight grip over the structure, composition and implementation of evaluations, not least to ensure that their views and interests are given due attention. In relation to other interventions, it is the implementing agency that controls the evaluation process – a clear conflict of interest that is rarely remarked on. Donor or implementing agency control over the evaluation process can translate into a capacity to

---

<sup>7</sup> In 2000, the United States Congress passed the Victims of Trafficking and Violence Protection Act (TVPA) requiring its State Department to issue annual reports describing ‘the nature and extent of severe forms of trafficking in persons’ and assessing governmental efforts across the world to combat such trafficking against criteria established by United States law. The TVPA lays down “minimum standards” for the elimination of trafficking as well as detailed criteria for evaluating the performance of states. The reports use a ranking system to classify all states reviewed into four tiers of anti-trafficking compliance. Any bottom-tier state, being one that does not comply with the minimum standards and that is not making significant efforts to do so, may be subject to a range of economic sanctions. United States Department of State, *Trafficking in Persons Reports: 2001-2011* (United States Department of State, (2001)-(2011)) [hereafter US Department of State (2001)-(2011)].

<sup>8</sup> Anne Gallagher and Elaine Pearson, *The High Cost of Freedom: A Legal and Policy Analysis of Shelter Detention for Victims of Trafficking*, 32 Human Rights Quarterly 73 (2010). [hereafter Gallagher and Pearson 2010].

ensure that inconvenient evidence-based conclusions are downplayed or ignored.<sup>9</sup>

Research for the present article confirmed that it is the views of the most direct and easily accessible beneficiaries (e.g. criminal justice officials and agencies receiving support) that are most consistently sought out and taken into account. This can skew an impact assessment, particularly when the claims of such groups are not tested against the views of others. As has been noted in a parallel context, programme beneficiaries can be reluctant to criticise an intervention because of appreciation for assistance provided and/or a fear that such assistance could be withdrawn if negatively assessed.<sup>10</sup> In none of the reviewed evaluations were the views of victims or suspects sought or considered. Of course, this does not mean that such perspectives are completely ignored. For example, most of the technical experts interviewed explained that their assessment of the success or failure of a project, activity or single case was tied to their perception of how the intervention impacted victims and/or suspects. It is essential to acknowledge however, that this perception may differ substantially from how victims or suspects themselves assess success or failure.

Results of an intervention can be interpreted differently, depending on individual perspective, background and expertise, with widely diverging assessments of what constitutes a good outcome. Several practitioners interviewed cited the example of an externally supported cross-border cooperation mechanism that was linked to a significant number of victim rescues. While the donor in this case regarded the victim rescues to be a strong indicator of success, technical experts within the project had a different, more nuanced view. They pointed out that this figure revealed nothing about: the number of victims who were not rescued (including those who continued to be exploited with the knowledge of counterparts); the extent to which “rescues” translated into victims being offered and accepting assistance; the cost of the cooperation process relative to alternative strategies; the time and effort put into making it work; and the apparent inability of the cooperation process to facilitate the identification and arrest of suspects.

Another example of conflicting perspectives of success relates to trafficking prosecutions. For criminal justice practitioners, a prosecution that results in the conviction of a trafficker may reasonably be viewed as a success. However, for the victim-witness, that same prosecution may have compounded previous trauma and created little or no perception of benefit. Under these circumstances, it is very likely that trafficked persons and criminal justice agencies will feel very differently about the impact and value of a criminal justice intervention that seeks to enhance the capacity of investigators and prosecutors to bring victims into the criminal justice process as witnesses.

### ***Against what criteria is success determined?***

In some cases, the criteria for success of an anti-trafficking criminal justice intervention are not articulated at all. More typically, particularly in recent years, such criteria are formal and explicit; attached to the relevant project or programme as objectives, together with predetermined indicators. Of course, this does not prevent unarticulated criteria from influencing an evaluation. For example, while pre-determined indicators are unlikely to extend to the question of whether the intervention has served a donor’s political or strategic interests, this measure of success may nevertheless be reflected in the evaluation process and its conclusions. However, at least in principle, it is the intervention’s formal indicators that determine whether or not success has been achieved.

---

<sup>9</sup> Dottridge 2008, pp. 65-66.

<sup>10</sup> Anette Brunovskis and Rebecca Surtees, *Leaving the past behind? When trafficking victims decline assistance*, p. 105 (Fafu and VNEXUS Institute, 2007) [hereafter Brunovskis and Surtees 2007].

Several experts interviewed for this study identified the lack of clearly defined end-points as a major problem with criminal justice interventions in the anti-trafficking sector. One participant, for example, noted the lack of agreement on what constitutes progress: 'It's not that we can't measure impact because of our starting point, it's because we don't actually know our end point, where we're trying to get to'.<sup>11</sup> Other problems identified in the literature and confirmed through interviews and examination of evaluation reports include the adoption of unrealistic or unclear objectives, formulated in such a way as to make evaluation of outcomes or impact virtually impossible. Examples cited included high-sounding objectives (e.g. "to eliminate trafficking") or those that are extremely vague (e.g. to "develop more effective responses"). Interviewees also pointed to an apparently widespread preference for weak or easily attainable goals such as whether the programme has been implemented as planned, rather than whether it is able to demonstrate change and impact.<sup>12</sup> This preference for process-related goals over those focused on outcomes can enable stakeholders to retain control over the "success story" in a highly uncertain environment by ensuring that evaluative measurement extends only to those aspects that are under the direct control of the project or intervention.

Definitions of success in the area of criminal justice responses to trafficking almost invariably fall back on what can be measured, hence the preference for quantifiable indicators such as number of arrests or prosecutions, number of officials trained, instances of cross-border investigation cooperation, number of special interview suites established, number of relevant treaties ratified, etc. One criminal justice specialist expressed frustration that 'results have become so important – and the more concrete they are, the more they are appreciated'<sup>13</sup>, even when the basis of these results may be weak or even flawed. Another highlighted how this preference for the measurable can have negative implications in an environment where 'even bad data is [considered] better than no data at all'.<sup>14</sup> One evaluation expert with experience in anti-trafficking interventions criticised the strong donor preference for (often poor quality) quantification over a more sophisticated and robust mix of qualitative and quantitative methods, noting the recent re-orientation, of one donor, 'to measuring things and crunching numbers and improving things through metrics'.<sup>15</sup> The many problems associated with an approach to evaluation that prioritises and rewards only what can be measured are identified and considered at various points throughout this study.

### ***Interrogating underlying values and assumptions***

Views on the success of an intervention will often depend on underlying values and assumptions. Criminal justice interventions in the field of counter-trafficking are rife with untested assumptions about the value of a particular approach, about stakeholders and beneficiaries that are rarely explicitly articulated. Any credible evaluation should extend to considering whether the underlying assumptions of an intervention are valid.<sup>16</sup> Several of these assumptions, as they relate to how the success of an intervention is measured, are considered further below.

---

<sup>11</sup> Interview (telephone), 18 November 2010.

<sup>12</sup> Rosenberg notes that this is also the case in almost all of the protection programs reviewed for assessment of trafficking programs in the Europe and Eurasia region: Rosenberg 2008, p. 34.

<sup>13</sup> Focus Group Discussion, Bangkok, Thailand, 19 November 2009.

<sup>14</sup> See, for example, the response of a major United Nations agency to data quality problems in the context of its global trafficking report: 'A poor indicator is better than no indicator as long as it is not represented as more than it is. Over time, the collection of information from so many different perspectives can, in aggregate, make up for many of the deficiencies of the data itself. Our global data set, reviewed time and again, can indeed tell us something more about the trends and patterns of the problem. This information is vital so that, in a world of limited resources, efforts can be focused for maximum effect': United Nations Office on Drugs and Crime (UNODC), *Global Report on Trafficking in Persons*, p. 13 (UNODC, 2009) [hereafter UNODC 2009].

<sup>15</sup> Interview (telephone), 11 November 2010.

<sup>16</sup> Surtees 2009, p. 19.

*Assumption 1: Increases in trafficking related prosecutions and/or convictions is a strong and reliable indicator of success*

As noted previously, an increase in trafficking-related prosecutions is widely viewed to be indicative of a more effective criminal justice response to trafficking.<sup>17</sup> Certainly this indicator should not be dismissed outright. The failure of States to arrest and prosecute exploiters has likely contributed to the high levels of impunity currently enjoyed by traffickers and to the widespread denial of justice to victims.

However, an increase in measurable criminal justice activity (arrests, prosecutions and convictions) is a crude and potentially misleading success indicator. For example, changes in prosecution rates may be attributable to new trafficking laws, with offences previously charged under other laws, such as sexual assault, pimping, immigration fraud, now prosecuted as trafficking. They may also reflect changes in trafficking prevalence as well as the way data is collected. Most importantly, numerical information provides no insight into the quality of prosecutions and convictions. A number of respondents interviewed for this study argued that the emphasis on increased prosecutions and convictions has driven many of the unintended consequences identified at Part 5 below, including a misidentification of people smugglers or marriage brokers as traffickers in order to boost the number of trafficking cases pursued. Fortunately, there appears to be growing acknowledgement of the need to focus on the quality of prosecutions and convictions. For example, success indicators developed by one regional criminal justice project consider whether the procedural guarantees for a fair trial were provided and whether convictions result in adequate and proportionate penalties.<sup>18</sup> The International Organization for Migration's manual of performance indicators suggests convictions as an indicator of success - with the caveat that this assumes the definition of trafficking is in line with international standards and the existence of a credible justice system that respects due process.<sup>19</sup>

*Assumption 2: Declines in prevalence of trafficking indicates positive change*

While the prevalence (of crime) is not a routine indicator for criminal justice interventions<sup>20</sup>, there appears to be a growing assumption that a decline in the prevalence of trafficking (presumably evidenced by the number of identified victims) is a useful indicator of success. The US *Trafficking in Persons Report*, for example, implies a prevalence measurement in establishing, as success criteria, "whether [the government under assessment] achieves appreciable progress in eliminating severe forms of trafficking when compared to ...the previous year".<sup>21</sup>

The use of prevalence as an indicator is problematic on a number of levels. It does not take into account well-known obstacles and pitfalls associated with quantifying the extent of the trafficking problem.<sup>22</sup> It also fails to acknowledge that even if prevalence could be accurately measured, changes in prevalence could be attributed to a range of factors beyond the rate at which trafficking is actually

---

<sup>17</sup> US Department of State (2001)-(2011); UNODC 2009; IOM 2008.

<sup>18</sup> Asia Regional Trafficking in Persons Project (ARTIP) *Monitoring and Evaluation Plan*, pp. 4, 13 (not published, 2009) [hereafter ARTIP 2009].

<sup>19</sup> IOM 2008, p. 38.

<sup>20</sup> Governance and Social Development Resource Centre (GSDRC), *Helpdesk Research Report: Prevalence as an Indicator for Anti-forced Labour and Anti-sex Trafficking Projects*, p. 2 (GSDRC, 2008).

<sup>21</sup> United States Department of State, *Trafficking in Persons Report: 2010*, p. 367 (United States Department of State, 2010).

<sup>22</sup> See, for example, USGAO 2006; Ernesto Savona and Sonia Stefanizzi, eds., *Measuring Human Trafficking Complexities and Pitfalls* (Springer Science, 2007); David Feingold, 'Trafficking in Numbers: The Social Construction of Human Trafficking Data' in Peter Andreas and Kelly Greenhill, eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*, pp. 46-74 (Cornell University Press, 2010); Habib, Ali, *Data Collection on Victims of Human Trafficking: An Analysis of Various Sources*, 6(1) *Journal of Human Security* 55 (2010).

occurring. For example, increased attention given to trafficking could result in higher levels of prevalence being recorded. Similarly, increases in the number of identified and assisted victims could well be a partial measure of a country's efforts to tackle trafficking.<sup>23</sup>

Further, decreased prevalence may not necessarily signal success but rather the adaptation of trafficking systems to anti-trafficking responses. Consider a recent attempt to use prevalence of child prostitution to measure the success of a criminal justice anti-trafficking intervention targeting child trafficking.<sup>24</sup> The evaluation research sought to map prevalence but did not interrogate underlying assumptions by, for example, questioning whether a measurable change in prevalence could be attributed to other factors, such as law enforcement pressures that resulted in victims being shifted to new or less visible locations.

The assumption about prevalence also works in reverse: that increases in prevalence are indicative of failure. However, as one study of impact evaluation has noted, success can also mean doing *less badly*.<sup>25</sup> It is not unreasonable to assume that, at least in some parts of the world, the rate at which individuals are being trafficked is increasing. Under such circumstances, the best that an intervention may be able to hope for is to slow down that rate.

*Assumption 3: Impact and key success indicators (such as attitudinal change) can be measured accurately and within the life of an intervention*

Assumptions and expectations about what can be measured (and when) are often unrealistic. The increasingly strident calls for trafficking related interventions to demonstrate impact provide a case in point. Several of those interviewed questioned whether impact could in fact be measured, at least within the life of a project or immediately afterwards. The issue of data quality and accessibility is central to the feasibility of impact assessment and is considered in further detail below. Timing is another critical aspect. Trafficking cases can take years from reception of a complaint to case resolution in court. The results and impacts of criminal justice interventions that seek to influence criminal justice responses can only fully be measured once the entire process has been completed. When project objectives are cast, in the words of one evaluation expert: 'at the upper end of the logic model', impact becomes 'virtually impossible to measure unless you put some sort of longitudinal study in place which lasts for a couple of decades'.<sup>26</sup>

The typical evaluation reflects a single moment in time, a snapshot of the outcomes or impact of an intervention at the time the evaluation took place. This reality must be weighed against the fact that the criminal justice response to trafficking is process-based – with cases moving through different stages, often over many years. Assessments of success or failure can be expected to change over time. For example, whereas a victim witness may feel positive about the criminal justice process in the early stages, this may wane if the process goes on for many months and years without a result. The effects of a high quality investigation may be effectively cancelled out by a poorly prepared prosecution or a trial marked by corruption. The perception of success in such a case would inevitably change depending on the stage at which it was evaluated.

Many criminal justice interventions seek to change the attitudes and behaviour of criminal justice

---

<sup>23</sup> Rebecca Surtees, *Second Annual Report on Victims of Trafficking in South-Eastern Europe*, p. 25 (IOM, 2005).

<sup>24</sup> Andrew Jones et al., *An Evaluation of the International Justice Mission's "Project Lantern": Assessment of Five-year Impact and Change in the Public Justice System* (International Justice Mission, 2010).

<sup>25</sup> Center for Global Development (CGD), *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, p. 20 (CGD, 2006) [hereafter CGD 2006].

<sup>26</sup> Interview (telephone), 11 November 2010.

personnel, usually through capacity building activities such as training and mentoring. It is important to acknowledge that changes in attitude and behaviour, as well as in organisational culture, do not happen quickly. Any evidence that might be available to demonstrate change of this kind will likely only come from the way cases are investigated, prosecuted and adjudicated over an extended time period. In short, the *impact* of capacity building interventions, such as training, is unlikely to be immediately discernible. One criminal justice adviser highlighted this obstacle, noting that changes in behaviour attributable to skills training in which he was involved were not detectable to counterparts or external observers until at least several years after such training commenced.

Time is not the only constraint and the practical obstacles to measuring changes at these levels should not be underestimated. Practitioners interviewed for this study pointed out the near-impossibility of measuring such attitudinal and behavioural change within the confines of a typical project. The difficulties are neatly illustrated by one project that has invested heavily in following up law enforcement trainees in an effort to assess change in workplace behaviour and thereby move beyond reliance on the one-sided and frequently inconclusive feedback received from trainees themselves. Project staff involved in this evaluation experienced a range of practical difficulties in monitoring workplace performance. Supervisors were rarely available for consultation and, regardless, generally lacked the knowledge and skills to be able to adequately assess the performance and capacities of their subordinates on the relevant issues. In addition to these problems, the basic data required to measure changes in performance or behaviour was often inaccessible or not available in a form that could be analysed.

## **Part 2: Attributing and understanding change**

A critical aspect of evaluation is not only to assess what change has been made, but also the extent to which this can be attributed to the specific programme or intervention being evaluated. This is usually done by adopting a “counterfactual” approach that compares the situation before the intervention with the situation afterwards.<sup>27</sup>

In the case of anti-trafficking work, attribution is complicated by the fact that trafficking and responses to trafficking are both situated within much larger, complex economic, political and social systems that are themselves subject to numerous pressures and influences. In short, the criminal justice response to trafficking does not take place in a vacuum; it is informed by a wide range of factors that are beyond the control of a single project or programme. As such, attributing change or impact to a specific intervention can be extremely difficult. Equally challenging are the important tasks of untangling the threads of what causes change, how and when; and establishing the appropriate weight to be given to different contributing factors.

Consider, for instance, a project that focuses on training law enforcement officials to investigate trafficking cases. Identification and attribution of change related to knowledge does not need to be particularly difficult: participants can be tested prior to training and again, once training is completed. Any changes detected can be measured and reasonably attributed to the intervention. However, when it comes to attributing change in skills, attitude and behaviour, the situation becomes much more complex. For example, how is it possible to establish a pre-intervention baseline in relation to these areas against which to measure change? Even if such a baseline could be established, post-intervention assessment and attribution of change will be difficult for many of the reasons identified

---

<sup>27</sup> GAATW 2010, p. 14.

previously: such change takes time to happen and even longer to manifest itself; data sources are likely to be incomplete and may be biased; and, as discussed further below, any changes that are detected may reflect a range of factors and influences outside the intervention.

The broader criminal justice environment in which the intervention takes place, (including laws, institutions and practices), can dramatically affect the impact and effectiveness of any intervention. For example, the success of a project that seeks to encourage the participation of victims and witnesses in criminal justice proceedings will depend heavily on the presence or absence of structures and procedures that support the cooperation of foreign victim witnesses such as reflection periods and temporary residence permits. The way in which trafficking is defined in law; the manner in which it is criminalised; and the legal rights, obligations and responsibilities of various players can all be expected to have quite significant – and perhaps difficult to anticipate – influences on the outcome of criminal justice interventions. In some countries it is the police, not the prosecutor, who decide which cases will be pursued in court. Interventions aiming to enhance the skills of prosecutors are likely to have only limited impact in environments that constrain the space within which prosecutors operate. In many countries, judges are not truly independent: for example, their decisions are influenced or even controlled by the executive branch of government. Where judicial independence is severely compromised, efforts to improve judicial capacity to adjudicate trafficking cases fairly and in accordance with established laws, will be similarly constrained.

The nature and extent of corruption-related criminal justice weaknesses will be another important factor in determining the capacity of any intervention to support genuine and sustainable change. Corruption can influence every aspect of the trafficking cycle. It can prevent correct identification of victims and the arrest of perpetrators; facilitate interference with witnesses; compromise prosecutions; and influence sentencing. Corruption also impacts the availability, reliability and validity of critical information that is needed to evaluate an intervention. For example, law enforcement agencies may be reluctant to share investigation files or may not keep accurate case records to avoid corruption coming to light. Judges may close court proceedings or seal court documents in cases where corruption occurs. These factors can make it very difficult to fairly assess and attribute change.

A single criminal justice intervention is often implemented in conjunction or parallel with similar interventions that may have different time frames, approaches and goals, complicating the task of attributing change to a specific intervention.<sup>28</sup> At best, it will be difficult to disentangle the outcomes and impact of interventions with similar goals. At worst, the results of one intervention can be challenged or even cancelled out by another that is working toward different goals and/or employing different strategies. One practitioner interviewed provided the example of a project in which he was involved declining to directly fund cross-border investigations of trafficking cases. The project team in this case reasoned that such an approach is unsustainable and potentially damaging of efforts to secure long-term funding for such investigations from the national police agency. The funding was subsequently provided by another project, which no doubt saw the provision of such assistance as an easy and measurable “win”.

Finally, internal and external pressures on States to respond to trafficking in a particular way are often present but rarely acknowledged. The impact of these pressures can be ignored or under-appreciated, complicating efforts to attribute change and success to a particular intervention. A useful example is provided by the annual US *Trafficking in Persons Report*, which, as noted above, evaluates and ranks the performance of individual states with respect to their efforts to prevent trafficking, prosecute offenders and protect victims. Recent studies have highlighted the influence of the report on the

---

<sup>28</sup> Danailova-Trainor 2008, p. 46.

behaviour of States.<sup>29</sup> Certainly the evaluation of any anti-trafficking intervention should consider the possibility that detected changes (e.g. a stronger focus on prosecution of traffickers) may be more accurately attributed to a criticisms emanating from the Trafficking in Persons Reports (e.g. that the government concerned had failed to actively prosecute trafficking) than to an intervention that sought to promote effective prosecution of trafficking cases.

### **Part 3: Data availability and accessibility**

Central to any evaluation is the availability of and access to valid, reliable and accurate data. Availability refers to whether the necessary data actually exists and has been collected. Accessibility refers to whether the data is obtainable by evaluators in a useable format.

Much of the data needed to evaluate criminal justice responses to trafficking simply does not exist. In many countries, the systematic collection of data is not part of the day-to-day operations of criminal justice agencies and institutions. Data that is collected is often not easy for evaluators to process – for example, it is contained in hand-written case files; different case numbers are used by different agencies / institutions; it requires translation, etc. As a consequence, vital information (e.g. about rates of trafficking, number, type, quality of investigations, arrests, prosecutions, convictions, protection orders, etc.) is scarce, unreliable and not always verifiable. This severely compromises the capacity of criminal justice interventions to collect baseline information against which future developments can be assessed.

Even where good data does exist, it is not always accessible to evaluators, or even to project staff for monitoring and other purposes. Lack of accessibility can have complex causes that may involve, practical, legal or ethical aspects. While such difficulties are often genuine, they can also be used as an excuse for failing to engage in rigorous evaluation (e.g. conducting process rather than outcome evaluation) and even for failing to evaluate at all.<sup>30</sup>

Some of the many factors that have an impact on relevant criminal justice data availability and accessibility are outlined below. Often these factors are related and mutually reinforcing, serving to interfere with both data availability and access and, consequently, with efforts to evaluate criminal justice anti-trafficking interventions.

#### ***Political will and priorities***

---

<sup>29</sup> Anne Gallagher, *Improving the Effectiveness of the International Law of Human Trafficking: A Vision for the Future of the US Trafficking in Persons Reports*, 12 Human Rights Review (2011) [hereafter Gallagher 2011]; H. Richard Friman, 'Numbers and Certification: Assessing Foreign Compliance in Combating Narcotics and Human Trafficking' in Peter Andreas and Kelly Greenhill, eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*, pp. 75-109 (Cornell University Press, 2010); Kay Warren, 'The Illusiveness of Counting 'Victims' and the Concreteness of Ranking Countries: Trafficking in Persons from Colombia to Japan' in Peter Andreas and Kelly Greenhill, eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*, pp. 110-26 (Cornell University Press, 2010).

<sup>30</sup> Consider the report of a government-led evaluation of the Australian response to trafficking, which noted that: '[d]ue to its illicit nature, the number of people trafficked is difficult to quantify. Consequently, in seeking funds to implement the Action Plan, the AFP [Australian Federal Police] was unable to accurately estimate the impact that TSETT [its specialist trafficking task force] would have on trafficking in persons': Australian National Audit Office, *Management of the Australian Government's Action Plan to Eradicate Trafficking in Persons*, p. 19 (Commonwealth of Australia, 2009).

The availability and accessibility of data can be heavily dependent on political will and priorities. For example, a country of origin for trafficked persons that is focusing on safe return and reintegration may not prioritise the generation and dissemination of criminal justice data. If agencies within a particular country are under pressure to improve performance they may suppress or even manipulate data to avoid criticism and/or ensure high success rates. Political dynamics can also directly impact the availability and accessibility of data required for evaluation of an intervention. Some evaluations are commissioned with “political” motives in mind – for example, to bolster support for a favoured intervention or to provide a justification for making changes to an intervention or its staffing. Further, as discussed at Part 1 above, different stakeholders may hold strongly divergent opinions about the value of a particular intervention. Persons or institutions who are opposed to an evaluation being conducted or who have a vested interest in promoting a particular outcome may be able to pre-empt it or obstruct access to data, acceptance of evaluation results or continuation of an evaluation contract.<sup>31</sup>

### ***Corruption***

Corruption, tied very closely to the political issues outlined above, can have a direct and profound impact on data availability and accessibility.<sup>32</sup> One respondent recounted a situation where programme staff, working with the Ministry of Justice to gain access to court files, faced enormous resistance from the judiciary, widely believed to be engaged in corruption in trafficking in persons cases. The impasse ended, rather dramatically, in the judiciary ordering that no case files could be accessed by outsiders, not even by staff of the relevant Ministry.

### ***Trust, cooperation and coordination***

Availability of and access to data is often inhibited by lack of cooperation and coordination. In some cases, the problem will exist between individuals and entities within the state structure itself, for example, between investigators and prosecutors. Many countries do not have an integrated and coherent system for collecting data from the initiation of an investigation through prosecution, conviction and sentencing. In such an environment, lack of cooperation limits the ability to evaluate the process as a whole, as well as the impact of a particular intervention. The problem of poor coordination and collaboration can also exist between State entities and external stakeholders, such as NGOs and experts working within a particular intervention. In many cases, governments are not comfortable sharing potentially sensitive information with outsiders or those in a position to critically evaluate their response. In some cases, as noted below, there may also be legal prohibitions on sharing such data, linked to issues of data protection. In short-term project cycles, the likelihood of overcoming these barriers is very low. Several criminal justice experts interviewed for this study shared their experience-based view that interventions of a longer duration stand a better chance of developing relationships of trust that permit the sharing of sensitive information.

### ***Limited capacity to collect, maintain and share data***

Data collection, maintenance and analysis are expensive and time-consuming, often requiring high levels of expertise. In many countries, these resources are scarce or simply unavailable. As a result, the task of measuring performance and change is difficult or impossible. Even when information is collected, it may not be standardised and managed in a way that renders it comprehensible, and

---

<sup>31</sup> Michael Bamberger and Jim Rugh, *Real World Evaluation. Working Under Budget, Time, Data and Political Constraints*, p.7 (American Evaluation Association: Professional Development Workshop Session 21, 2008) [hereafter Bamberger and Rugh 2008].

<sup>32</sup> It also has the potential to impact data reliability, a point which is discussed in more detail in the section below.

thereby, accessible, whether to internal staff for monitoring purposes or outside evaluators. This can reflect the communication and coordination problems outlined above. Not uncommonly, lines of reporting and communication between central and provincial levels may be weak, even more so when trying to collect data down to district or village level. In some countries that have undergone decentralisation there may not be even be an obligation to report to the central government. One law enforcement specialist outlined how legal processes and the institutionalised disconnect between agencies often translates into fragmented data. He noted that in one country the work of initial investigators is completely separated from the work of interrogators and prosecutors on the same case. Cases are assigned different numbers at each step of the criminal justice process: providing numerous opportunities for data to be mislaid, 'creating fault lines in the system itself through which the data falls and cannot be found'.<sup>33</sup>

Criminal justice agencies are often aware of data weaknesses and interested in working towards more systematic collection of trafficking-related data. One law enforcement specialist argued that many within the criminal justice sector feel that such data collection is important and in their professional interest: 'I don't think the commitment is the issue; it's just that there are no systems in place'.<sup>34</sup>

### ***Legal and ethical issues***

Often, the criminal justice data that would be of most interest to an evaluation cannot or should not be made available. In some situations, accessibility will be restricted by national laws and regulations. In other situations, a specific law or regulation may not apply but the sharing or use of certain data may raise ethical concerns. Evaluations may also be governed by legal and ethical codes imposed by the implementing organisation or the funding agency. All of those involved in evaluation of criminal justice interventions must conform to relevant legal and ethical restrictions even when this limits access to relevant evaluation data.<sup>35</sup> In the criminal justice context, legal and ethical concerns can sometimes, but not always, be met through strategies such as assurances of anonymity and confidentiality, or through the sharing of de-personalised cases or aggregate data sets.

### ***Lack of transparency within the criminal justice sector***

In every country, the prevailing criminal justice culture will rarely contribute to an open and easy flow of information. To some extent this may reflect the appropriate legal and ethical restrictions around information sharing outlined above. Generally however, concepts such as openness and transparency are not always valued or prioritised, as various criminal justice specialists explained:

I think gaining access to information is extremely difficult in the criminal justice system and that's not just true in [this country or this region], it's true in [my home country also]. Prosecutors and judges don't generally lay open their files to outsiders and sometimes there are laws that prevent it actually, privacy laws. And there's a sense in law enforcement that it could compromise investigations, it could compromise the safety of individuals. There is a reluctance to be criticised or hold yourself open to be criticised. There are political considerations. So I think just gaining access to the information [is a problem], assuming you had the resources to really review them and analyse them... and the expertise to be able to figure out what's going on in these cases.<sup>36</sup>

Of course, a desire to control information is not unique to criminal justice agencies. A reluctance to

---

<sup>33</sup> Interview, Bangkok, Thailand, 2 September 2010.

<sup>34</sup> Interview, Bangkok, Thailand, 2 September 2010.

<sup>35</sup> Surtees 2009.

<sup>36</sup> Interview, Bangkok, Thailand, 1 August 2010.

share information is unfortunately characteristic of many stakeholders in the counter-trafficking sector.

## **Part 4. Data quality: issues of method, validity and reliability**

Reliable, valid, accurate and relevant data is critical to the credibility and usefulness of any evaluation of any trafficking intervention. If the evaluation design and implementation are compromised by missing data or poor analysis, the findings will be inadequate and likely irrelevant.<sup>37</sup> Currently the quality of trafficking-related data (both specific to criminal justice data and more generally) is very weak. Many countries do not collect even the most basic information and even the best efforts tend to provide only a partial picture of the situation. Those involved in implementing or evaluating anti-trafficking can be remarkably sanguine about such limitations, as illustrated by the response of a United Nations (UN) agency to data quality problems in the context of its global trafficking report:

A poor indicator is better than no indicator as long as it is not represented as more than it is. Over time, the collection of information from so many different perspectives can, in aggregate, make up for many of the deficiencies of the data itself. Our global data set, reviewed time and again, can indeed tell us something more about the trends and patterns of the problem. This information is vital so that, in a world of limited resources, efforts can be focused for maximum effect.<sup>38</sup>

The following analysis, focusing on threats to the quality of data, builds on the previous discussions about obstacles to data availability and accessibility. While far from conclusive, this discussion is intended to provide some helpful indication of what data might be important for such evaluations and limitations of data sets commonly used in criminal justice evaluations.

### ***Standardization and comparability of data sets***

Criminal justice data related to trafficking is generally not comparable between countries and even less comparable at regional or global levels. This is at least partly due to significant differences in how States understand and criminalise trafficking. For example, individuals identified as irregular migrants in one country may be registered as victims of trafficking in another. Cases that are registered as labour trafficking in one country may be identified as instances of recruitment fraud or labour rights violations in another. The lack of standardised and comparable data sets compromises robust evaluation. Sometimes these problems can occur within a country. Even where laws are clear, immigration agencies that have not been trained to identify victims of trafficking may register trafficking cases as instances of illegal migration. Law enforcement agencies that have only dealt with trafficking into sexual exploitation may make similar mistakes when confronted with factory or farm-based exploitation. In some countries, internal or external pressure to improve performance can compromise comparability of data sets by inflating investigation or prosecution statistics with cases that are not in fact trafficking.

### ***Different institutions; different data systems***

Trafficking cases can implicate different agencies that collect data independent of each other, resulting in different data sets or intersecting datasets that cannot easily be consolidated. In the criminal justice arena, data necessarily moves between agencies as the case progresses. In many countries,

---

<sup>37</sup> Intili et al. 2009, p. 20.

<sup>38</sup> UNODC 2009.

information sharing and case tracking is inadequate or non-existent, with police, prosecutors and the courts adopting different approaches to the recording and classification of case information. Such weaknesses make the tracking of cases time-consuming, expensive, and sometimes impossible. This is significant because case tracking provides valuable information on the functioning of the criminal justice system from a number of different angles and perspectives. The use of different systems by different agencies also risks duplication of data, resulting in a single case being counted more than once and thereby compromising the quality and reliability of the data set.

### **Data biases and selection effects**

All data is invariably subject to biases and selection effects. Evaluations should acknowledge the potential impact of such factors on the validity of findings related to outcomes and impact. A recent major UN report on trafficking serves to illustrate the nature and extent of bias and selection effects in relation to criminal justice data, and the consequent risk of misinterpretation. A key finding of the study was that: 'women play a key role as perpetrators of human trafficking. In Europe, for example, women make up a larger share of those convicted for human trafficking offences than for most other forms of crime'.<sup>39</sup> The report does not sufficiently acknowledge the likely selection effect resulting from the operational reality that women are overrepresented amongst low-level brokers and recruiters, the typical soft target of law enforcement.

The tendency of criminal justice authorities to focus their attention on **soft targets** and thereby to create selection effects and biases in available data appears to be widespread. There is considerable anecdotal evidence available to support a contention that those most at risk of apprehension and prosecution are the "facilitators", rather than the ones who benefit most directly from the exploitation of others.<sup>40</sup> That tendency impacts the quality and value of available information.

The acknowledged low levels of accurate victim identification mean that reported, and thereby available, cases are not likely to be **representative** of the actual situation. A combination of factors (sometimes including the national legal framework) is responsible for the widespread tendency to focus criminal justice attention on foreign female victims of sexual exploitation. It is also important to note that many trafficking victims consciously choose not to self-identify and report their trafficking experience.<sup>41</sup> These limitations mean that evaluation findings based on reported cases can, at best, be applied to this particular group and not extrapolated to the broader universe of unreported cases of trafficking. Such evaluations are accordingly not capable of identifying the gaps in the criminal justice response that are relevant to this much broader group.

**Respondent bias** can distort data, as well as influence perceptions of success as discussed at Part 1 above. For example, criminal justice agencies may control access to key documents such as court records or case files in an effort to present a particular kind of case or result that is not necessarily representative of the actual situation. Practitioners involved in case analysis and trial observation for one project acknowledged that the information received and situations observed could not be confirmed as representative of the general caseload. This bias is also relevant in terms of data collected from stakeholders who have benefitted from an intervention and may be concerned that disclosing problems and failures would reflect badly on them as professionals / institutions and may compromise the provision of further support.

---

<sup>39</sup> UNODC 2009, p. 10.

<sup>40</sup> Association of South East Asian Nations, *Progress Report on Criminal Justice Responses to Trafficking in Persons in the ASEAN Region* (ASEAN, 2011).

<sup>41</sup> Brunovskis and Surtees 2007.

Both **quantitative and qualitative indicators** are needed to effectively measure the outcome and impact of any trafficking related intervention. The importance of qualitative measurements lies in the fact that numbers and statistics may provide a sense of the scale of change or improvement, but do not explain why this change or improvement took place.<sup>42</sup> For example, a reduction in trafficker prosecutions could mean very different things: that fewer people are committing this type of offence, or that the police are failing to apprehend individuals committing this type of offence.<sup>43</sup> One M&E specialist interviewed for this study highlighted the value of qualitative inquiry in evaluating criminal justice responses to trafficking, noting that this was particularly important in capturing changes in organizational culture. Other criminal justice practitioners expressed the view that typical quantitative indicators of success (e.g. number of officials trained, number of specialist units established, number of instances of cross-border cooperation), rarely provided the insight necessary to measure positive change.

## Part 5: Identifying unintended consequences

It is a generally accepted principle that evaluations conducted within the context of international development assistance should capture all effects of an intervention – intended and unintended, negative and positive. This is considered particularly important in relation to impact evaluations where the stakes and expected returns are higher.<sup>44</sup> The gap between an in-principle commitment to discovering unintended, particularly negative consequences, and the reality on the ground is considerable. Of course, this is not a problem limited to development cooperation. In almost every area of scientific and social enquiry, it is the anticipated, positive findings that receive the most attention. Experiments, initiatives and programs that are considered successful are much more likely to be promoted by participating institutions and published in academic journals.<sup>45</sup> The propensity to favour success distorts reality by, for example, providing an undeservedly optimistic assessment of what has happened and what is possible.

This tendency has particular resonance for anti-trafficking interventions. The contemporary campaign against trafficking has been tainted by misgivings that efforts to address trafficking may be motivated by other, hidden agendas, or could camouflage social or political campaigns against marginalized or stigmatized groups such as migrant workers, illegal migrants, smuggled migrants, refugees and sex workers. Over the past several years, many studies and reports have identified serious negative consequences attached to some trafficking responses. Examples drawn from this growing body of literature include the detention of trafficking victims in jails, immigration centres, and shelter facilities; the prosecution of trafficked persons for status offences such as illegal immigration or illegal work or for offences that relate to having been trafficked such as prostitution; discriminatory emigration restrictions; and ill-conceived raids and rescues that compromise the rights and safety of victims.<sup>46</sup> While such

---

<sup>42</sup> Surtees 2009, p. 26.

<sup>43</sup> ARTIP 2009, p. 3.

<sup>44</sup> The OECD, for example, has defined impact in the context of evaluations as: “[p]ositive and negative, primary and secondary long term effects by [of] a development intervention, directly or indirectly, intended or unintended” (DAC 2002, p. 24).

<sup>45</sup> CGD 2006, p. 28.

<sup>46</sup> Brunovskis and Surtees 2007; GAATW 2007; Gallagher and Pearson 2010; Rebecca Surtees, *Listening to victims. Experiences of identification, return and assistance in South-Eastern Europe* (International Centre for Migration Policy Development (ICMPD) and NEXUS Institute, 2007); Noy Thrupkaew, ‘Beyond Rescue’, *The Nation*, 26 October 2009, available at: <http://www.theinvestigativefund.org/reporters/noythrupkaew/>; Noy Thrupkaew, ‘The crusade against sex trafficking’, *The Nation*, 5 October 2009, available at: <http://www.theinvestigativefund.org/reporters/noythrupkaew/>.

results have been commonly cited as “collateral damage”, this conceptualization of the problem may be misleading to the extent that it implies that the negative impact of certain laws, policies and practices was both unforeseen and regretted.

The principle that adequate evaluation includes identification and exploration of unintended negative or positive consequences appears to be a sound one. However, it does raise several important questions. Specifically, against what criteria, and by whom, is an unintended consequence identified as negative or positive? In addition, against what criteria is a consequence identified as unintended? The relevance of these questions, as well as their potential complexity, can be illustrated by examples drawn from discussions with criminal justice practitioners.

The first example relates to the training of investigators, a common criminal justice intervention in the anti-trafficking sector. Under political or other pressures, managers will sometimes move these investigators on to other crime types that are of higher priority. From the perspective of the intervention, this is an unanticipated result that negatively impacts the development of long-term investigative capacity within the specific area of counter-trafficking. An external evaluator, however, may look at the situation more broadly, concluding that an increase in the overall quality of national investigative capacity, particularly in relation to serious crimes, is an unintended, but positive consequence. The recipient law enforcement agency may hold another, even more positive view; that trafficking was privileged at the expense of organizational effectiveness and that well-trained, competent investigators are now being deployed where their skills and experience are most needed. Indeed, from the law enforcement perspective, the outcome may not even be unforeseen. The decision of a law enforcement agency to accept an offer of external training in one crime type may be influenced by an assessment of broader institutional needs.

Other examples indicate that unintended consequences may not – or should not – be open to multiple or flexible interpretations. Failure to support a conviction with adequate evidence, inappropriate sentencing, and violation of fair trial rights in trafficking prosecutions appear to be common in countries that are burdened with an underdeveloped or otherwise dysfunctional criminal justice system. One technical expert spoke of observing the trial of an individual being prosecuted for trafficking offences. The project in which he was working had provided specialist training to the police investigators involved in the case. There was also considerable external pressure on the government to demonstrate it was responding effectively to trafficking. No evidence was presented at the trial to support the prosecutor’s allegation of an intention to exploit, no evidence had been collected from the destination country and no victims were present to be examined. The trial lasted just 45 minutes before the judges retired, coming back to court 15 minutes later with a pre-typed judgment that found the accused guilty of trafficking. As the respondent noted, “It was a done deal, it was just a show court”. International fair trial standards were clearly not met in this case and, to the extent that the intervention contributed to the case coming to court, this was a negative consequence of that intervention. In terms of numbers and statistics however, it may well have been recorded as an anticipated and successful consequence of external support.

### ***Lack of attention to negative or unexpected consequences***

Despite the acknowledged potential for risk as well as benefit, evaluations of anti-trafficking interventions rarely consider negative or unexpected consequences. For example, a major, critical report on monitoring and evaluation of anti-trafficking programs issued by the US Government Accountability Office in 2007 is silent on the issue, as is the 2009 UNODC Global Report on Trafficking.

It was not until eight years after its inception, in 2009, that the US Trafficking in Persons Report addressed negative impacts of anti-trafficking responses such as victim detention and criminalization. The country assessments, which are the heart of the report, still fail to address this aspect in any meaningful way.<sup>47</sup> An International Organization for Migration handbook on performance indicators for anti-trafficking interventions acknowledges that stakeholders include those negatively affected by interventions but makes no other reference to unexpected or negative consequences.<sup>48</sup> An evaluation framework for USAID-funded trafficking in persons programs includes those negatively affected by interventions in their definitions of “impact” and “stakeholder” but otherwise makes no reference to negative or unexpected consequences.<sup>49</sup> While some of the evaluation terms of reference and reports examined for this study indicated that they would be considering positive and negative results attributable to or correlated with interventions and investments, these commitments did not flow through to the reports’ findings or recommendations.

### ***Why the lack of attention to negative or unexpected consequences?***

It is important to explore why negative or unintended results of anti-trafficking interventions have been glossed over or ignored. As noted previously, this is a problem that is much wider than the focus of the present study. For that reason, it is especially valuable to consider whether there are any features of the anti-trafficking sector or of criminal justice interventions in this sector that may exacerbate or otherwise affect the pervasive tendency to prefer good news over bad, positive results over negative ones.

One possible explanation (that would apply to development cooperation generally) is that evaluations of anti-trafficking interventions are almost always structured around the intervention’s established objectives, usually reduced to a set of predetermined indicators. As discussed in detail at Part 1 above, expectations of what the intervention should be achieving are themselves based on certain assumptions that are not open to interrogation. A typical criminal justice intervention, for example, may seek to develop a specialist capacity within the national prosecutor’s office. A standard evaluation will consider whether the identified steps to reach that goal (i.e. awareness-raising, skills training, designation of specialists, institution-building, etc.) have been achieved. It will not likely extend to enquiring whether specialization of the prosecutorial function is desirable; or whether the allocation of human and financial resources to the crime of trafficking has negatively affected the ability of the prosecutor’s office to pursue other serious crimes. In short, pre-determined indicators work to exclude the unforeseen, whether positive or negative: “[b]y definition, they focus on expected change and agreed-upon areas of change, leaving aside the unexpected and unintended”.<sup>50</sup> The strict application of predetermined indicators in country assessments conducted by the U.S. government has been cited as a reason why consideration of negative consequences has never reached beyond the introductory section of the U.S. TIP Report.<sup>51</sup>

It is also important to consider whether detecting and analysing unexpected negative consequences requires a different level of **technical or contextual knowledge** than is needed to assess performance against objectives and indicators. An example is provided by a recent legislative change in one country of South East Asia that has blurred the distinction between trafficking and migrant smuggling. Those interviewed on this issue expressed the view that the legislative change had been prompted by a range of external political pressures on that country to address migrant smuggling *and* improve its prosecution

---

<sup>47</sup> Gallagher 2011.

<sup>48</sup> IOM 2008.

<sup>49</sup> Intili et al. 2009.

<sup>50</sup> Chris Roche, *Impact Assessment for Developing Agencies: Learning to Value Change* (Oxfam, 1999).

<sup>51</sup> Gallagher 2011.

rates in trafficking cases. The anticipated consequence of this new law is greater criminal justice activity in the areas of both smuggling and trafficking. However, it is also expected that trafficking victims are less likely to be accurately identified and afforded their rights and protections. In the words of one senior practitioner: “They will just be dealt with as illegal migrants instead of victims of serious crime”. Only those with a sophisticated understanding of the legal framework and insight into its practical application are in a position to appreciate and expose these potential negative impacts.

To what extent do **incentives** (and disincentives) affect whether negative unintended consequences of anti-trafficking interventions are identified and examined? A study that considered why impact assessments are so rare notes institutional disincentives to adverse findings:

Government agencies involved in social development programs or international assistance need to generate support from taxpayers and donors. Since impact evaluations can go any way—demonstrating positive, zero, or negative impact—a government or organization that conducts such research runs the risk of findings that undercut its ability to raise funds. Policymakers and managers also have more discretion to pick and choose strategic directions when less is known about what does or does not work. This can even lead organizations to pressure researchers to soften or modify unfavourable studies or simply to suppress the results—despite the fact that knowledge of what does not work is as useful as knowledge of what does.<sup>52</sup>

**Strategic preferences** can affect the willingness of those involved in encouraging, funding or otherwise sponsoring trafficking-related interventions to play down potential problems associated with their preferred approaches. Within the criminal justice arena, there is, as noted previously, a strong preference for quantifiable results in terms of arrest, prosecution and conviction. Indications that increase in arrests and prosecutions may be due to inappropriate targeting or that increased convictions have been secured at the expense of free trial guarantees is unlikely to be solicited or welcomed. **Political considerations** can also play an important role. A donor that is engaged in a struggle against migrant smuggling may be tempted to turn a blind eye to a recipient state that is using its trafficking laws and procedures to prosecute smugglers. Research conducted for this study indicates that donors are sometimes uninterested in negative impacts of the anti-trafficking interventions that they are supporting. Several practitioners cited instances in which donors were provided confidential information on matters of concern, such as failure of counterparts to provide for a fair trial, disproportionate penalties of persons convicted through unjust processes, and the prosecution of migrant smuggling as human trafficking. No reaction was forthcoming.

Those commissioning evaluations may pressure evaluators to ignore or suppress negative findings. It is unlikely that even the most “independent” evaluator is completely impervious to the preference of their clients for a verdict that their money has been well and wisely spent. In addition, evaluators are usually heavily reliant on those closest to the intervention to provide information and insight. This is particularly the case in relation to negative or unexpected impacts that may not be as easily identified or shared as positive achievements against predetermined indicators. In typical criminal justice interventions, it is project staff and counterparts who are the most important sources of information about unexpected, negative consequences. These persons may be unwilling to pass on negative findings for fear it will reflect badly on them or that it will result in loss of employment or project-related benefits.

The situation does appear to slowly be changing. Some recent evaluations and reports about evaluations emphasize the importance of measuring unintended harmful effects of interventions.<sup>53</sup> While attention to this issue at the operational level continues to be very superficial, there are a few

---

<sup>52</sup> CGD 2006.

<sup>53</sup> ARTIP 2009; Dottridge 2008; GAATW 2007; GAATW 2010; Surtees 2009.

exceptions. For example, one regional criminal justice project reports having used its knowledge of negative consequences to commence a dialogue with counterparts about specific issues such as inappropriate targeting of offenders and failure to respect procedural guarantees of a fair trial. Project training materials have also been adjusted to ensure that criminal justice officials are aware of the potential for negative impacts and able to monitor their work accordingly.

## **Part 6: The cost of evaluations: dealing with budget and time constraints**

A recent study of evaluation in the anti-trafficking sector found that it is practical constraints such as time and resources, rather than a lack of interest or commitment, that leads many stakeholders to neglect impact evaluation. The same study confirmed that resources for evaluating initiatives are often quite “ungenerous” and commonly the first casualty when budgets are cut.<sup>54</sup> It is understandable that donors and programmers seek to assess the outcome and impact of interventions quickly and economically. However, as explored at various points throughout this study, those goals are generally incompatible with what is needed for an effective evaluation of an anti-trafficking intervention in the criminal justice sector. Measuring the outcomes and impact of such interventions requires analysis of change over time, often over an extended period of years, or even decades. In addition, the data needed to measure this change must be sufficiently rigorous to support methodologically sound conclusions. These realities mean that while different interventions and evaluation models will require different levels of resources, it can be expected that any serious and comprehensive evaluation will be both time and resource intensive.

Consider, for example, the resources required to track trafficking cases through the criminal justice system. Case analysis is widely agreed to be an important aspect of evaluation: perhaps even the ‘gold standard’ in relation to criminal justice interventions. However, as practitioners interviewed for this study observed, the raw data required to secure an understanding of how cases are proceeding or have been concluded is invariably very difficult to access and verify. Substantial effort must be put into negotiating access with counterparts. Tracking cases usually also involves expensive translation of voluminous documentation, and lengthy and wide-ranging interviews with relevant officials to clean and validate data sets as well as fill in missing information. Additional efforts, expertise, and resources are required to analyse case files within the context of the national legal system and to ensure that the inevitable sample selection biases are identified and taken into account.

### ***Resources vs. reliability***

While it is certainly possible to undertake evaluations within “real world” budget constraints, it is necessary to acknowledge that this will not be without consequence. Budgetary considerations affect every aspect of an evaluation, including the number of interviews that can be conducted; the type of data that can be collected; the ability to combine quantitative and qualitative data collection and analysis; the size and professional experience of the research team; the level of analysis that can be conducted; and the cost of translation and interpretation.<sup>55</sup> A threshold, different for every evaluation, will be reached where cost limitations threaten the validity of evaluation findings. Impact evaluations are especially vulnerable to validity threats incurred through time and budgetary constraints. These can include threats to the quality of evaluation design and implementation, threats to statistical analysis,

---

<sup>54</sup> GAATW 2010, pp. 15, 20.

<sup>55</sup> Michael Bamberger, Jim Rugh and Linda Mabry, *RealWorld Evaluation Working Under Budget, Time, Data, and Political Constraints*, p. 1 (Sage Publications, 2006) [hereafter Bamberger, Rugh and Mabry 2006].

theoretical coherence and adequacy of the counterfactual and the extent to which findings can be generalized.<sup>56</sup>

### ***The cost of including different stakeholders***

The question of who, amongst target groups and stakeholders, should be involved in the evaluation process is often considered and decided with reference to cost (and time) considerations. Decisions about whose views should be included are critical, not least because, as discussed at Part 1 above, they will ultimately determine how and by whom success is defined.

Access to some stakeholders may be easier, less costly and less time consuming to secure than others. Direct project counterparts such as criminal justice practitioners and victim support agencies could fall into this first group while accessing government agencies and officials can, depending on the context, be more challenging. The involvement of trafficked persons and suspects in any evaluation is likely to be very expensive because sufficient time must be allowed to identify a representative sample of the target group, secure informed consent and develop the rapport necessary for safe, ethical and useful interviewing. Involvement of victims and/or suspects in an evaluation can also incur additional costs relating to transport, interpretation and security. It is not surprising that when time and budgets are constrained, it is these more complicated (and seemingly indirect) stakeholders that are routinely omitted from consultations. The possibility that the more difficult-to-reach groups are in any event accorded a lower priority, and that resource constraints are used as an excuse to exclude them, should not be discounted.<sup>57</sup> Certainly, as noted earlier, few if any, evaluations of criminal justice interventions have engaged directly with trafficked persons or suspects.

### ***Balancing evaluation and programme costs***

Of the interventions examined for this study, it appears that, while substantial resources were invested in design, monitoring, and output measurement, far less was applied to measuring and evaluating outcomes and longer-term impacts. It is often argued that the expense of comprehensive evaluations is too high relative to the overall costs of the intervention. Certainly, as noted previously, trade-offs are inevitable. While outcome and impact evaluation is clearly important in establishing the effectiveness and appropriateness of an intervention (and in identifying good, replicable examples<sup>58</sup>), such exercises drain time, energy and funding away from implementation.

What then constitutes the optimum balance between intervention costs and evaluation costs? While this question is far from settled, there is general agreement on certain basic points. For example: that minor or short-term interventions may not be amenable to evaluation but that any substantial intervention should be evaluated; that evaluations must be specific to the intervention and setting; and that the level of resources allocated to an evaluation should be proportionate to the scale of the intervention and its anticipated impact.

Beyond these general points, it is difficult to be specific about how much is “too much” and “too little”. Different approaches are likely to generate different criteria. Some have argued that the solution lies in embedding evaluations more firmly in the programme cycle. Others have suggested separating impact evaluations from the intervention itself – acknowledging that measuring impact of anti-trafficking

---

<sup>56</sup> Bamberger, Rugh and Mabry 2006, p. 22.

<sup>57</sup> Bamberger and Rugh 2008, p. 7.

<sup>58</sup> CGD 2006, p. 16.

interventions serves a broader public good and, therefore, deserves separate resourcing.<sup>59</sup> An extension of this line of argument considers the cost of the evaluation, not relative to the intervention, but rather relative to the value of the knowledge that will be generated from the evaluation. In the case of trafficking, this value can be measured both in terms of avoiding the harms associated with inappropriate or wasteful interventions and in reaching more people with proven approaches.

### ***The costs of not evaluating***

While a good deal is said about the cost of evaluations, far less discussion takes place around the potential losses when evaluations do not take place or are undertaken at too low a level. Failure to evaluate interventions properly or at all may lead to wasted resources in the long term by enabling ineffective approaches and methods to continue unchecked. Weak or ineffective interventions that are continued, even replicated can have unintended and negative impacts on programme beneficiaries.<sup>60</sup> Findings from evaluations can help to avoid costly mistakes. The risk of wasting funds on ineffective (and even harmful) programs is particularly acute for programs that are expected to be replicated in other countries or settings. The value of impact evaluations takes on particularly urgency in programs where real harm can be done through incorrect or non-intervention<sup>61</sup>, not least in the anti-trafficking sphere where failure to effectively prosecute traffickers and protect trafficked persons has very real and long-term implications for trafficked persons and society at large. Justifiable questions have been asked about whether organisations should be implementing (or replicating) programs that have not been adequately evaluated.<sup>62</sup>

### **Towards the future**

Recent calls for more rigor and transparency in evaluating the success or failure of anti-trafficking interventions are both welcome and overdue. While the need for quality impact evaluation is felt across the development spectrum, the demand in the area of criminal justice responses to trafficking appear to be especially acute. Human exploitation through trafficking has only recently been identified as a priority issue for States and the international community. There is growing recognition that using the criminal justice system to end impunity for traffickers and secure justice for their victims is an important component of an effective national response. To date however, States have been developing and adapting their criminal justice responses on the run, often under strong political pressure, and principally through trial and error. Donors and implementing agencies have been operating in a similarly reactive and unplanned way. This has increased the risk of bad decisions and negative consequences. Interventions are rarely subject to careful evaluation and, as a result, the assumptions on which those interventions are based are not being tested and lessons learned in one context are not being shared in others.

Of course, it is much easier to call for more quality in impact evaluation than it is to actually deliver. This discussion paper has affirmed the need for discussions about evaluating anti-trafficking interventions to move from criticism and complaint about what is not being done well (or at all) to serious, informed engagement with obstacles and constraints and how these can be addressed. A different approach to this issue also requires commitment to a new level of openness and transparency. Donors and implementing agencies in particular have a responsibility to contribute to an environment of learning

---

<sup>59</sup> CGD 2006, pp. 23-24; Danailova-Trainor 2008.

<sup>60</sup> CGD 2006, p. 16.

<sup>61</sup> CGD 2006, p. 16, 20.

<sup>62</sup> CGD 2006, p. 3; Surtees 2009, p. 49.

and self-reflection that will help ensure good practices are replicated and poor ones discarded.

One of the main obstacles to effective impact evaluation revealed by the research for this article lies in the failure of states, the international community and the anti-trafficking sector to specify and communicate a common vision of what constitutes an effective criminal justice response to trafficking. The absence of a unified vision manifests itself in the lack of clearly defined outcomes that is typical of most, if not all criminal justice interventions. It also exacerbates inherent complications such as the presence of multiple, divergent, stakeholders; difficulties in accessing reliable and verifiable data; and the risk of unintended negative consequences. The absence of an overarching vision allows mediocre or even harmful interventions to flourish and good work to go unrecognized and unrewarded. It is a collective responsibility to work towards articulation of this vision and to ensure accountability of those who are provided the resources to promote its realization.

## Bibliography

- Ali, H.M. (2010) 'Data Collection on Victims of Human Trafficking: An Analysis of Various Sources.' *Journal of Human Security*, Vol. 6, No. 1: 55-69.
- Australian National Audit Office (2009) *Management of the Australian Government's Action Plan to Eradicate Trafficking in Persons*. Australian Capital Territory, Commonwealth of Australia.
- Asia Regional Trafficking in Persons Project (ARTIP) (2009) *Monitoring and Evaluation Plan*. (not published).
- Association of South East Asian Nations (ASEAN) (2011) *Progress Report on Criminal Justice Responses to Trafficking in Persons in the ASEAN Region*. Jakarta: ASEAN.
- Bamberger, M., Rugh, J. and Mabry, L (2006) *RealWorld Evaluation Working Under Budget, Time, Data, and Political Constraints*. California: Sage Publications.
- Bamberger, M. and Rugh, J. (2008) *Real World Evaluation. Working Under Budget, Time, Data and Political Constraints*. American Evaluation Association: Professional Development Workshop Session 21.
- Baumann, G. (2006) *Measuring the Effectiveness of Anti-Slavery Work*. Washington D.C.: Free the Slaves.
- Brunovskis, A. and R. Surtees (2008) 'Agency or illness: conceptualization of trafficking victims' choices and behaviours in the assistance system.' *Gender, technology and development*. 12(1). – not cited.
- Brunovskis, A. and Surtees, R (2007) *Leaving the past behind? When trafficking victims decline assistance*. Oslo: Fafo and Vienna: NEXUS Institute.
- Center for Global Development (CGD) (2006) *When Will We Ever Learn? Improving Lives Through Impact Evaluation*. Washington D.C.: CGD:
- Development Assistance Committee (DAC) (2002) *Glossary of key terms in evaluation and results based management*. Paris: Organisation for Economic Co-operation and Development DAC.
- Danailova-Trainor, G. (2008) 'Evaluation of Counter-Trafficking Programmes and Research' in Aghazarm, C. et al, *Human Trafficking: New Directions for Research*. Geneva: International Organization for Migration: 44-56.
- Dottridge, M. (2008) 'Research needs concerning the monitoring, evaluation and impact assessment of both research about human trafficking and projects and programmes to address human trafficking' in Aghazarm, C. et al, *Human Trafficking: New Directions for Research*. Geneva: International Organization for Migration: 57-71.
- Dottridge, M. (2007) *Measuring Responses to Trafficking in Human beings in the European Union: An Assessment Manual*. Brussels: European Commission.
- Feingold, D. (2010) 'Trafficking in Numbers: The Social Construction of Human Trafficking Data' in Andreas, P. and Greenhill, K eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*. Ithaca: Cornell University Press: 46-74.
- Friman, H.R. (2010) 'Numbers and Certification: Assessing Foreign Compliance in Combating Narcotics and Human Trafficking' in Andreas, P. and Greenhill, K. eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*. Ithaca: Cornell University Press: 75-109.

Gallagher, A. and Chuang, C. (forthcoming 2012) 'The Use of Indicators to Measure Government Responses to Human Trafficking' in Kingsbury, Davis and Merry eds. (forthcoming 2012) *Indicators as a Technology of Global Governance*. Oxford University Press.

Gallagher, A. and Pearson, E. (2010) 'The High Cost of Freedom: A Legal and Policy Analysis of Shelter Detention for Victims of Trafficking.' *Human Rights Quarterly*, Vol. 32: 73-114.

Gallagher, A. (2011) 'Improving the Effectiveness of the International Law of Human Trafficking: A Vision for the Future of the US Trafficking in Persons Reports.' *Human Rights Review*, Vol. 12.

Global Alliance Against Traffic in Women (GAATW) (2007), ed. *Collateral Damage: The Impact of Anti-Trafficking Measures on Human Rights Around the World*. Bangkok: GAATW.

GAATW Hames, C., Dewar, F. and Napier-Moore, R. (2010) *Feeling good about feeling bad ... A global review of evaluation in anti-trafficking initiatives*. Bangkok: GAATW.

Governance and Social Development Resource Centre (GSDRC) (2008) *Helpdesk Research Report: Prevalence as an Indicator for Anti-forced Labour and Anti-sex Trafficking Projects*. GSDRC.

Intili, J. et al. (2009) *An Evaluation Framework for USAID-Funded TIP Prevention and Victim Protection Programs*. Washington D.C.: Social Transition Team, Office of Democracy, Governance and Social Transition of the United States Agency for International Development (USAID/E&E/DGST).

International Organization for Migration (IOM) (2008) *Handbook on performance indicators for counter-trafficking projects*. Geneva: IOM.

Jordan, A (2010) 'More funding for impact assessments?' *Rights Work*: [www.rightswork.org](http://www.rightswork.org), online publication, 15 October.

Roche, C. (1999) *Impact Assessment for Developing Agencies: Learning to Value Change*. Great Britain: Oxfam.

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.

Savona, E. and Stefanizzi, S. eds. (2007) *Measuring Human Trafficking Complexities and Pitfalls*. New York: Springer Science.

Surtees, R. (2009) *Re/integration of trafficked persons: developing monitoring and evaluation mechanisms*. Brussels: KBF and Washington: NEXUS Institute.

Surtees, R. (2007) *Listening to victims. Experiences of identification, return and assistance in South-Eastern Europe*. Vienna: International Centre for Migration Policy Development (ICMPD) and NEXUS Institute.

Surtees, R. (2005) *Second Annual Report on Victims of Trafficking in South-Eastern Europe*. Geneva: IOM.

Thrupkaew, N. (2009a) 'Beyond Rescue' *The Nation*, 26 October. Available at: <http://www.theinvestigativefund.org/reporters/noythrupkaew/>

Thrupkaew, N. (2009b) 'The crusade against sex trafficking' *The Nation*, 5 October, Available at: <http://www.theinvestigativefund.org/reporters/noythrupkaew/>.

UNIFEM (referring to Roche 1999 cited in UNIFEM) – not cited

United Nations Office on Drugs and Crime (UNODC) (2009) *Global Report on Trafficking in Persons*. Vienna: UNODC.

United States Department of State (2001)-(2011), *Trafficking in Persons Reports: 2001-2011*. Washington D.C.: United States Department of State.

United States Government Accountability Office (USGAO) (2006) *Human Trafficking: Better Data, Strategy, and Reporting Needed to Enhance U.S. Anti-Trafficking Efforts Abroad*. Washington D.C.: USGAO.

USGAO (2007) *Human Trafficking: Monitoring and Evaluation of International Projects Are Limited, but Experts Suggest Improvements*. Washington D.C.: USGAO.

UNOHCHR (2002) *Recommended Principles and Guidelines on Human Rights and Human Trafficking*. New York: United Nations: E/2002/68/Add.1. – not cited

Warren, K. (2010) 'The Illusiveness of Counting 'Victims' and the Concreteness of Ranking Countries: Trafficking in Persons from Colombia to Japan' in Andreas, P. and Greenhill, K. eds., *Sex, Drugs and Body Counts: The Politics of Numbers in Global Crime and Conflict*. Ithaca: Cornell University Press: 110-26.