



Combatting Human Trafficking since Palermo: What Do We Know about What Works?

Katharine Bryant & Todd Landman

To cite this article: Katharine Bryant & Todd Landman (2020) Combatting Human Trafficking since Palermo: What Do We Know about What Works?, Journal of Human Trafficking, 6:2, 119-140, DOI: [10.1080/23322705.2020.1690097](https://doi.org/10.1080/23322705.2020.1690097)

To link to this article: <https://doi.org/10.1080/23322705.2020.1690097>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 01 Mar 2020.



Submit your article to this journal [↗](#)



Article views: 5073



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 2 View citing articles [↗](#)



Combatting Human Trafficking since Palermo: What Do We Know about What Works?

Katharine Bryant^a and Todd Landman^b

^aManager-Global research, Minderoo Foundation's Walk Free Initiative, Perth, Western Australia; ^bPolitical Science, School of Politics and International Relations, Faculty of Social Sciences and Executive Director, Rights Lab, University of Nottingham, Nottingham, UK

ABSTRACT

In 2016, there were an estimated 40.3 million victims of modern slavery in the world, more than were enslaved during the Transatlantic Slave Trade. Since the adoption of the 2000 *UN Trafficking Protocol*, numerous efforts from inter-governmental agencies, governmental agencies, international non-governmental organizations (INGOs), and domestic non-governmental organizations (NGOs) have strived to combat the phenomena of human trafficking through legal-institutional means, direct interventions, and programs of support for those exploited. This anti-trafficking work has paid varying degrees of attention to the principles and methods of monitoring, evaluation, and impact assessment, but has often been subject to the end of project evaluations. Similar to findings of reviews of evaluations in the international development sector, evaluations of anti-trafficking programming have primarily focused on assessing the progress of project implementation and the achievement of outputs, rather than tracking the achievement of outcomes or impact. This is further complicated by the hidden nature of human trafficking and the trauma experienced by human-trafficking victims. As a consequence, despite some evidence of raised awareness and increased levels of funding, organizations are still struggling to demonstrate impact and discern what works to combat human trafficking. This article analyses the evaluations of counter-trafficking programming produced since the *Protocol* to draw conclusions regarding the lessons learned from these interventions and the methods used to monitor and evaluate human-trafficking programs. By highlighting gaps, this article provides a series of suggestions on how to better track progress and impact toward the elimination of modern slavery.

KEYWORDS

Measurement; baseline; monitoring; evaluation; impact assessment; attribution; contribution; theory of change; causal inference

Introduction

Nearly 20 years on from the ratification of the *Protocol to Prevent, Suppress, and Punish Trafficking in Persons Especially Women and Children (the Trafficking Protocol)* (United Nations Convention against Transnational Organized Crime and the Protocols Thereto (CTOC), 2004), it is time to take stock and ask if we are any closer to answering the question, “what works to combat human trafficking?” There were an estimated 40.3 million people in forced labor and forced marriage in the world in 2016 (International Labour Organization, Walk Free Foundation, & International Organization for Migration, 2017). While it is difficult to estimate how many of these individuals were also victims of human trafficking, these remain the most accurate figures estimating related forms of exploitation, representing more people in slavery than during the Transatlantic Slave Trade (Voyages, 2019).

CONTACT Katharine Bryant  kbryant@minderoo.com.au  PO Box 3155, Broadway Nedlands, WA 6009

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/uhmt.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

There have been numerous efforts to combat human trafficking and other types of exploitation at the international, regional, domestic, and local levels pursued by a variety of agencies and organizations. These groups combat human trafficking through legal-institutional means (international law and domestic statutory legal frameworks), direct interventions, and programs of support for those affected by exploitation. Taken together, these concerted efforts have paid varying degrees of attention to the principles and methods of monitoring, evaluation, and impact assessment.

Systematic review studies are one method to offer insight as to what is effective across comparable programs. Walk Free conducted one such review in 2015, called the promising practices database, identifying 179 evaluations of modern slavery programming and interventions to tackle related phenomena such as forced genital mutilation/cutting (FGM/C), child labor, and early marriage. The findings from this and similar reviews of anti-trafficking interventions so far show us that the quality of evaluations needs to improve before a clear understanding of what works can become a reality (Bryant & Joudo, 2018; Davy, 2015; 2016; Hames, Dewar, & Napier-Moor, 2010; van der Laan, Smit, Busschers, & Aarten, 2011).

To take stock of anti-trafficking interventions and monitoring and evaluation methods, this article analyses the findings of a promising practices database (Bryant & Joudo, 2018), drawing on 90 evaluations of programs to combat human trafficking produced between 2000 and 2015. This article highlights some of the shortcomings of monitoring and evaluation in anti-trafficking programming, some of which are common to international development more broadly and some of which are specific to anti-human trafficking interventions. For example, many evaluations have focused on assessing the progress of project implementation and the achievement of outputs, rather than tracking the achievement of outcomes or impact, while the concealed nature of human trafficking makes it difficult to establish baselines against which to track progress. As a consequence, organizations are still struggling to demonstrate impact and discern what works to combat human trafficking and modern slavery.¹

However, while concrete answers to the question “what works” remain elusive, a thematic analysis of the 90 evaluations points to some lessons learned regarding specific interventions and methods for conducting monitoring and evaluation. These practices, while not proven, point to some interesting “promising” practices of both interventions and methods of evaluation to reflect upon some twenty years since the ratification of the *UN Trafficking Protocol*.

A Note on Definitions and Terminology

This article refers to human trafficking as defined by the 2000 *UN Trafficking Protocol*, but recognizes that there has been a shift in recent years to refer to human trafficking as “modern slavery.” Human trafficking remains the focus of this article, but where modern slavery is referenced, it is used as an umbrella term that encompasses many different forms of exploitation, including slavery, human trafficking, forced labor, debt bondage, forced or servile marriage, and the sale and exploitation of children. In this sense, we follow the International Labor Organization (ILO) and Walk Free definition that “essentially, it [modern slavery] refers to situations of exploitation that a person cannot refuse or leave because of threats, violence, coercion, abuse of power, or deception” (International Labour Organization, Walk Free Foundation, & International Organization for Migration, 2017; Walk Free Foundation, 2018).

Regarding monitoring, evaluation and impact assessment: international development projects, including those tackling human trafficking, should include a well specified set of objectives, milestones

¹In November 2018, the Gilder Lehrman Center for the Study of Slavery, Resistance and Abolition held at Yale University held an international conference “Fighting Modern Slavery: What Works?” which brought together academics, NGOs, government agencies, and private sector organizations to examine different efforts to combat modern slavery. The conference covered much ground, but it was clear that despite the increased attention to the issue of modern slavery, much more work is needed to understand how and in what way anti-slavery interventions can make a demonstrable impact.

within the overall framework, key performance indicators (KPIs), outcomes, and impact, along with a budget broken down into categories of expenditure over time. The process of monitoring is therefore dedicated to the periodic collection of information on the different aspects of the project to see if it is on course to deliver its main outcomes and impact, and whether its activities are within budget. Evaluation is related to monitoring, but adds an additional focus on project relevance, efficiency, effectiveness, impact, and sustainability, as well as demonstrable evidence that can be assessed and analyzed (OECD/DAC, 1991). An evaluation is thus an assessment of the whole project cycle, and can be done with an inception phase, a mid-term review, and a final review. It focuses on both the achievements of the project and whether its activities had positive, negative, and/or unintended consequences. Alternatively, while few have been conducted in the anti-trafficking field, impact assessments² are designed to analyze the impact of purposive interventions; impact is defined as “positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended.” (OECD/DAC, 1991) and focuses on the “long term and sustainable changes introduced by a given intervention” (Blankenberg, 1995) rather than the immediate consequences. As such, impact differs from the outputs or the outcomes of an intervention and should be separately assessed, even though the three elements may significantly affect each other.

Materials and Methods

The promising practices database is a systematic review of evaluations of anti-modern slavery programming and related phenomena from 2000 to 2015; a 2019 update of the database has recently begun.³ The starting point for the development of the database was to seek an answer to the question, “what works to combat modern slavery?” The theory was that since there have been many interventions and organizations involved in the fight against modern slavery since the ratification of the *UN Trafficking Protocol*, there must be some common lessons that can be drawn regarding what works.

The anti-modern slavery field is diverse and operates at different levels with various expertises. Inter-governmental agencies, such as the ILO and the United Nations Office on Drugs and Crime (UNODC) work to set standards, promulgate international legal instruments, and advocate for ending human trafficking through their formal institutional structures. These actions are reinforced by the 2015 Sustainable Development Goals (SDGs), including target 8.7,⁴ 16.2,⁵ 5.3,⁶ and 10.7⁷ (UN General Assembly, 2015). Government agencies take concrete steps to translate these international conventions and SDGs into legislative and policy action. An estimated 136 countries, for example, have enacted anti-trafficking legislation since the early 2000s (Minderoo Foundation, 2019).⁸ International non-governmental organizations (INGOs) work to produce an independent evidence-base, awareness raising, advocacy, and programmatic delivery to help lift people out of human trafficking, and provide sustainable

²The logic of impact assessments has evolved initially from Environmental Impact Assessment (EIA), to Social Impact Assessment (SIA) to Human Rights Impact Assessment (HRIA) See Götzmann, Vanclay, & Seier (2016); Harrison and Stephenson (2010); Landman (2006); Vanclay (2003).

³The promising practices database covers all forms of modern slavery. This article contains a summary of the methodology used to develop the database, however, given the focus of this special issue, only those evaluations categorized as covering human trafficking are reviewed in the results section. A paper reviewing all evaluations related to modern slavery and related phenomena, including a more detailed description of the methodology, can be found on the Minderoo Foundation website. The database is current as of 31st December 2015. See <https://www.minderoo.com.au/promising-practices/>

⁴“Take immediate and effective measures to eradicate forced labor, end modern slavery and human trafficking and secure the prohibition and elimination of the worst forms of child labor, including recruitment and use of child soldiers, and by 2025 end child labor in all its forms.”

⁵“End abuse, exploitation, trafficking and all forms of violence against and torture of children.”

⁶“Eliminate all harmful practices, such as child, early and forced marriage and female genital mutilation.”

⁷“Facilitate orderly, safe, regular and responsible migration and mobility of people, including through the implementation of planned and well-managed migration policies.”

⁸Of 183 countries included in the 2019 Measurement, Action, Freedom report, released by the Minderoo Foundation’s Walk Free Initiative. See www.globalslaveryindex.org

solutions for survivors to be re-integrated into the economy or society of their respective (or, due to human trafficking's often transnational nature new) countries. Other organizations have an exclusive focus on supply chains as part of larger corporate social responsibility, sustainability, and business and human rights approaches, where the "Ruggie Principles" (OHCHR, 2011) are adapted to include the problem of modern slavery in business. At the local level, many secular⁹ and faith-based civil society organizations (CSOs), as well as social movement organizations and networks¹⁰ work with local governmental authorities, including law enforcement, to share information, raise awareness, advocate, and provide support in the community for human-trafficking victims.

Given prior and existing activity, and as new actors such as the Freedom Fund and the Global Fund to End Modern Slavery enter into the anti-modern slavery policy area, the promising practices database aimed to draw out the lessons learned from the work conducted to date. Previous reviews of evaluations have identified a limited number of robust evaluations and impact assessments conducted in the anti-trafficking field (Davy, 2016; van der Laan et al., 2011). However, most projects and interventions to combat trafficking and related forms of exploitation are subject to log frame programmatic documents, donor reporting, mid-term reviews, and end of project evaluations. By broadening our search to include end of project evaluations, as well as any impact assessments that exist, it was hoped that lessons could be determined in terms of what works, and just as importantly, what does not work to combat modern slavery.

Throughout 2015, a team of researchers based at Walk Free conducted systematic searches of gray and academic literature to identify these evaluations. An evaluation was defined broadly to capture donor reports and end of project evaluations, using the definition: "evaluation measures progress towards outputs, or change in outcomes, or an assessment of an impact, of a development programme, policy, or intervention".¹¹ Various academic databases and international organization websites and databases were searched using the search criteria. Search terms included truncated versions of "traffic" and "evaluation", or "slavery" and "evaluation" or "forced labour" and "evaluation". Further evaluations were provided to the team by international organization partners after summaries were identified in relevant databases. Knowing that the number of evaluations would be limited, but that lessons could be learned from evaluations of interventions supporting groups vulnerable to modern slavery, search terms were expanded to include refugees, IDPs, migration, and forced genital mutilation/cutting (FGM/C).¹²

Initial searches and review of titles and abstracts led to the inclusion of 410 potential evaluations, reduced to a total of 344 for review after the removal of duplicates. Another 165 were then removed, leaving a total of 179. These 165 were removed because they were descriptions of programmes, literature reviews, lists of good practice determined by other organizations, mid-term evaluations, formative (or pre-assessment) evaluations, not in English, a summary of a larger document which was not available, annual reports rather than evaluations, or did not include an explicit methodology of how the evaluations were conducted. Explicit methodologies were defined as the inclusion of a methodology section or a description of the actions taken by the individual or team to conduct the evaluation. Lists of good practice determined by other organizations were removed as these had inconsistent or contradictory criteria from the work being conducted of what constituted a good practice. Systematic reviews of evaluations, or reviews of reviews were also excluded from the final database.¹³ The final 179 evaluations are housed in an endnote file and can also be downloaded from the Minderoo Foundation website.¹⁴

⁹The Global Modern Slavery Directory provides a list of 2,978 NGOs working on modern slavery by country. As of 6th August 2019, available at: <http://www.globalmodernslavery.org/>.

¹⁰The Rights Lab at the University of Nottingham and the Independent Anti-Slavery Commissioner mapped multi-agency anti-slavery partnerships across the UK and found that most areas of the UK have some form of partnership in place. Interactive map available at: <http://iascmap.nottingham.ac.uk/>

¹¹Adapted from 3ie definition of impact evaluation. <http://www.3ieimpact.org/en/evidence/impact-evaluations/>

¹²See Appendix 1 for the full search criteria.

¹³See Appendix 2 for the inclusion criteria for the final database.

¹⁴See <https://www.minderoo.com.au/promising-practices/>

Once these evaluations were housed in the database, they were classified according to term lists organized under the following categories:

- type of modern slavery,
- sector of exploitation,
- target populations,
- country/ region,
- type of program and activities,
- independent/ internal,
- evaluation methodology,
- did the program meet its objectives?

Each evaluation also included a free text write up of the program objectives and evaluation findings.

The development of term lists was an iterative process that drew upon the content of the evaluations and predetermined terms that could be used to search the final database.¹⁵ For example, the development of the term lists for program type and activities drew on existing anti-modern slavery frameworks and the types of interventions as described by the evaluations. There is not an agreed international framework to describe interventions to tackle modern slavery. Combatting modern slavery draws heavily on trafficking frameworks, which predominately follow a criminal justice approach, as defined by the “3 P” framework of protection, prevention, and prosecution in the *Trafficking Protocol* (CTOC, 2004). The “3 P” framework is further iterated by Section 108 of the 2000 U.S. Trafficking Victims Protection Act (TVPA) and its subsequent reauthorizations (US Department of State, 2018) and frameworks and toolkits developed by UNODC (UNODC, 2008, 2009). However, not all modern slavery interventions follow a criminal justice approach, therefore, categorizing program type began by identifying the activities described by the evaluations and then grouping these activities into types of programs based on their commonalities. After testing a sample of the evaluations using these term lists, two members of the research team independently categorized all 179 evaluations. The final term list for types of intervention was:

- supporting government,
- service delivery and coordination,
- research,
- business transparency,
- economic empowerment,
- risk-based prevention.

A similar exercise by the interagency Coordination Group against Trafficking in persons (ICAT) in 2016 identified five intervention areas, which overlap with the Walk Free categories (*in brackets in the below list*). These were:

- (1) Increase identifications, increase relevance and quality of support, strengthen victim-centered responses, improve compensation. (*service delivery and coordination*)
- (2) Address context-based risk, develop safer migration pathways and actions, strengthen social protection, (*risk-based prevention*)
- (3) Improve data and data systems, increase learning, develop quality standards, (*research*)
- (4) Reduce enabling environment, shrink markets for goods/ services PBTL, clean supply chains, strengthen financial systems barriers, (*business transparency/ economic empowerment*)

¹⁵All term lists are available in Bryant and Joudo (2018).

- (5) Strengthen legal frameworks and norms, disrupt more trafficking networks, increase (strategic) prosecutions and confiscations. (*supporting government*) (ICAT, 2016)

Terms to assess the evaluations' methodology were adapted from the Maryland Scientific Methods Scale. While the meta-evaluation literature has not reached a definitive consensus on the best means to compare the quality of methodology in outcome evaluations, the Maryland Scientific Methods Scale is the leading approach (Farrington, Gottfredson, Sherman, & Welsh, 2002). It is a simple 5-point scale developed in the field of criminology to assist the assessment of the scientific validity of criminological interventions. It has the benefit of being a simplified framework to assess the methodology employed by evaluators, prioritizing internal validity, establishing causal order and eliminating external variables. As Table 1 illustrates, the minimum level of methodology for results to be considered reliable is level 3 – evaluation designed with pre- and posttest measures with a comparable control. The highest level of validity is attained by testing the intervention with a randomized control trial (RCT) (level 5).

The use of this scale is not without controversy, in particular, because of the debate on the ethics and reasonableness of RCTs in development work (Burrell, 2012; Harkins, 2017). As described above, interventions in the anti-modern slavery sector are also diverse and cannot always be assessed using a criminological approach. A modified scale was used in the development of the promising practices database to suit this diversity better, borrowing from the Maryland Scale, but recognizing that pre- and post-assessments and RCTs are expensive and difficult to implement for anti-modern slavery work. The final methodology categories added in important factors for these evaluations, including participative elements and qualitative methods. There is scope for further work on a modified scale to apply to evaluations in anti-modern slavery work.

Results

Ninety evaluations in the database were categorized as “human trafficking”. This occurred when “human trafficking” or “trafficking in persons” was included within the document regardless of whether the phenomena described constituted human trafficking according to definitions outlined in the *Trafficking Protocol*. Often the evaluation would not also include the original program document or only the briefest of descriptions of what the program aimed to tackle, so it was difficult to verify if the program was or was not attempting to combat human trafficking.

Of these 90 evaluations, the majority were conducted in Europe and Central Asia ($n = 72$), followed by Asia Pacific ($n = 52$) and Africa ($n = 36$) (refer to Figure 1). The fewest were conducted

Table 1. Maryland scientific methods scale.

Maryland Scientific Methods Scale ¹⁶			
	<i>Method of assessment</i>	<i>Characteristics</i>	<i>Reliability scale</i>
Level 1	Posttest measures	Fails to rule out threats to internal validity and fails to establish causal order	No confidence in methodology
Level 2	Pre- and posttest measures	Causal order established, but fails to rule out many threats to internal validity	No confidence in methodology
Level 3	Pre- and posttest measures with comparable control conditions	Rules out many threats to internal validity. Presents problems with selection effects and regression to the mean	Minimum level for results to be considered valid and to draw conclusions about efficacy
Level 4	Pre- and posttest measures with comparable control conditions, controlling for variables	Better statistical control of extraneous influences on the outcome. Deals with selection and regression threats.	Can draw conclusions from results
Level 5	Randomized control trials	Requires a sufficiently large number of units. Has the highest internal validity.	High confidence in results

¹⁶Adapted from Farrington et al. (2002).

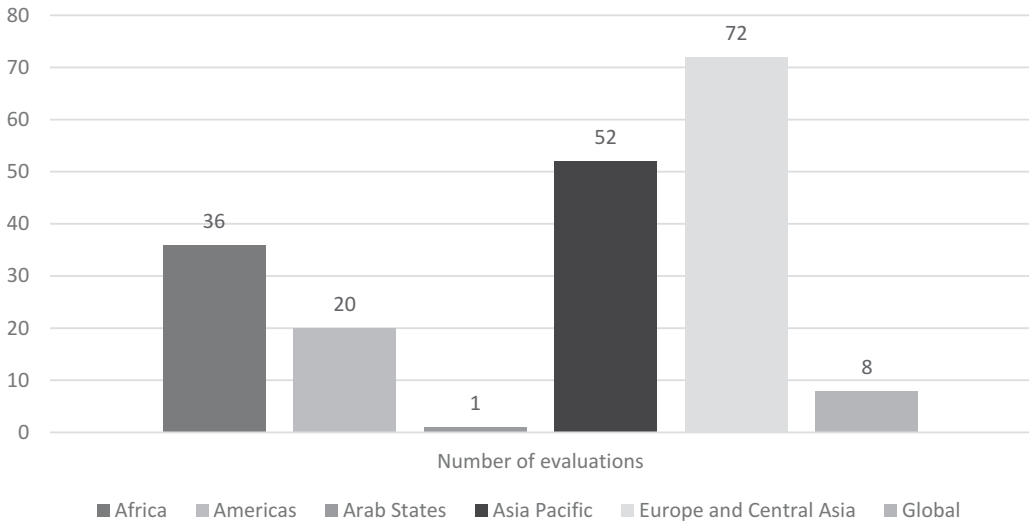


Figure 1. Regional breakdown of trafficking interventions.

in Arab States ($n = 1$) and the Americas ($n = 20$).¹⁷ Eight evaluations were tagged as “global” and referred to evaluations of global programs or global frameworks to combat human trafficking.

As evident in Figure 2, The number of evaluations identified has increased since the early 2000s, with peak numbers of evaluations conducted in 2006 (11) and 2012 (18). The low numbers of evaluations in 2015 are likely due to publication lag, as searches were completed by late 2015.

The most common types of interventions, as depict in Figure 3, were those that supported the government ($n = 60$), provided service delivery and coordination ($n = 53$), and risk-based prevention ($n = 44$), or a combination of all three ($n = 20$). These arguably align with the anti-human trafficking “3 P” framework of prosecution, protection, and prevention.

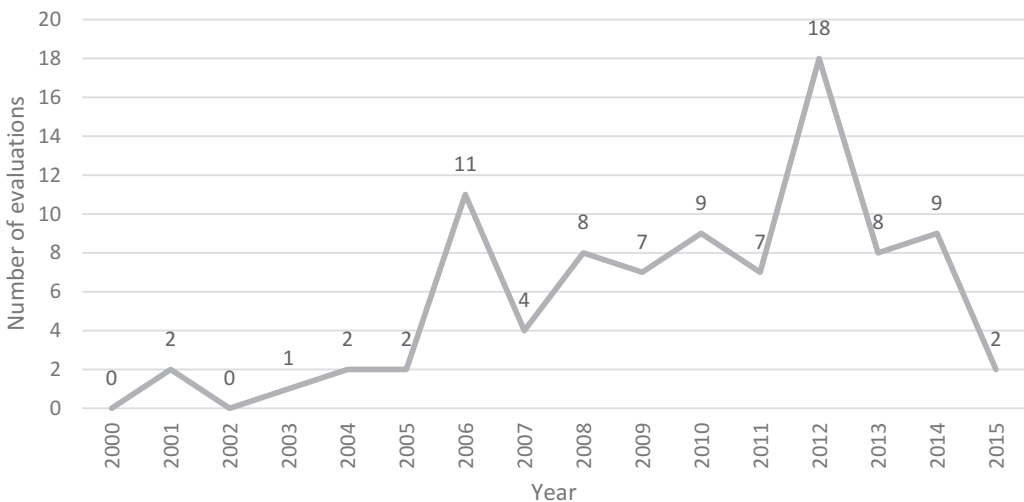


Figure 2. Number of evaluations since 2000.

¹⁷Numbers do not equal 179. Evaluations and programmes often covered multiple countries and therefore multiple regions.

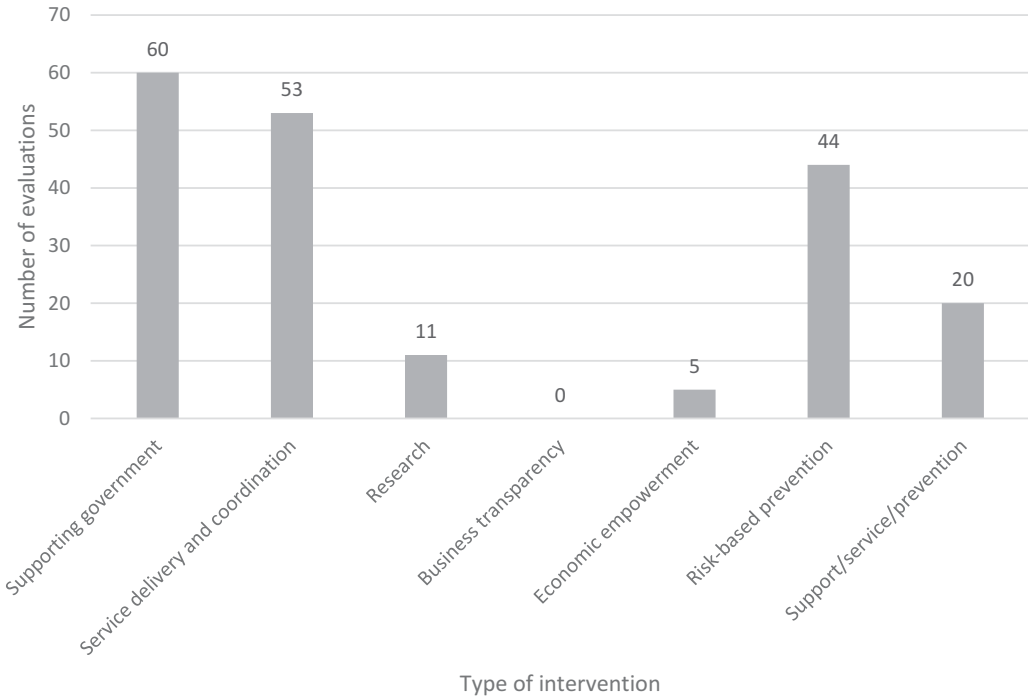


Figure 3. Number of evaluations by type of intervention.

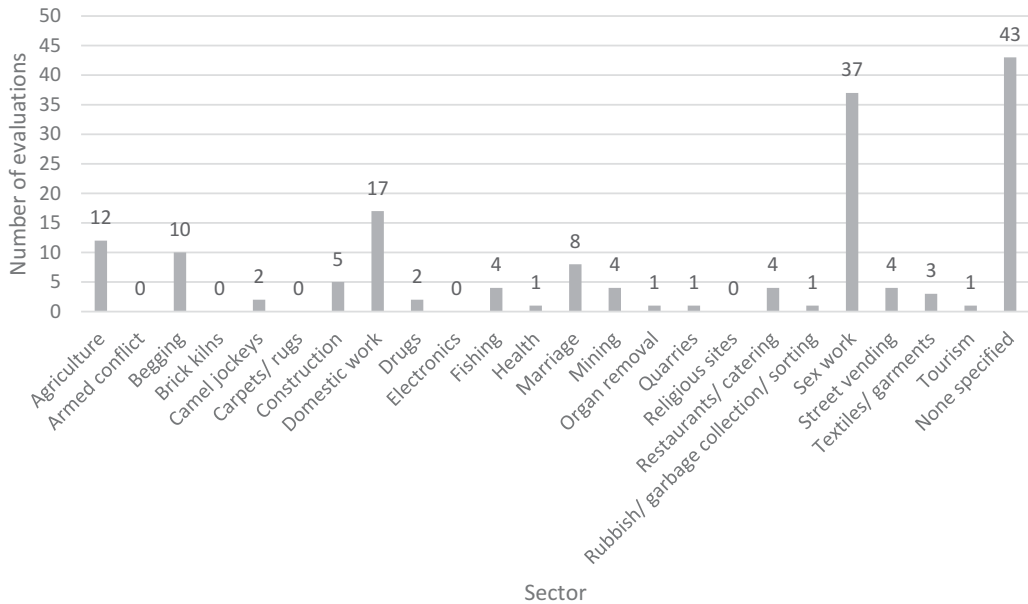


Figure 4. Number of evaluations by sector.

The interventions aimed to tackle exploitation in the following sectors: sex work ($n = 37$), domestic work ($n = 17$), agriculture ($n = 12$) and begging ($n = 10$), among others (refer to Figure 4). Forty-three evaluations did not specify which sector the program was focusing on,

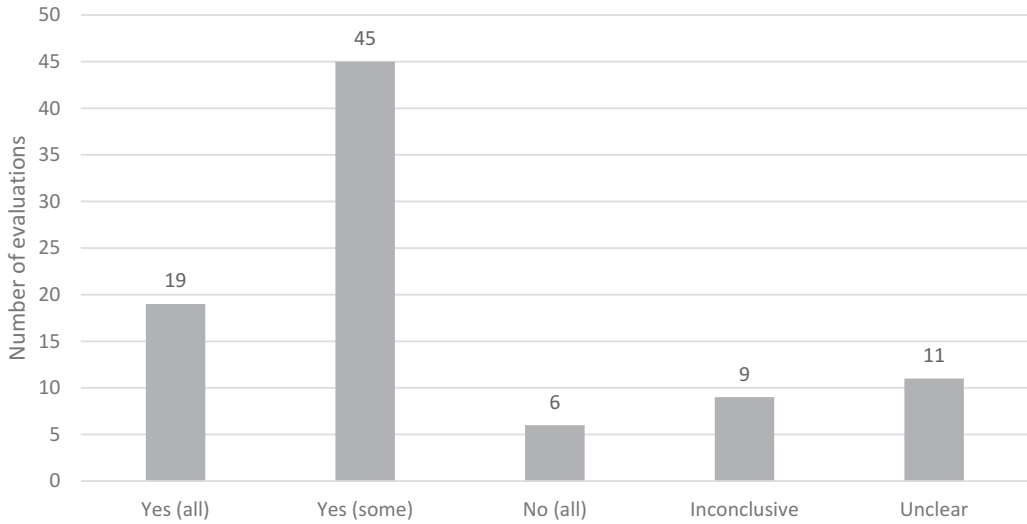


Figure 5. Did the project/ program achieve its objectives?

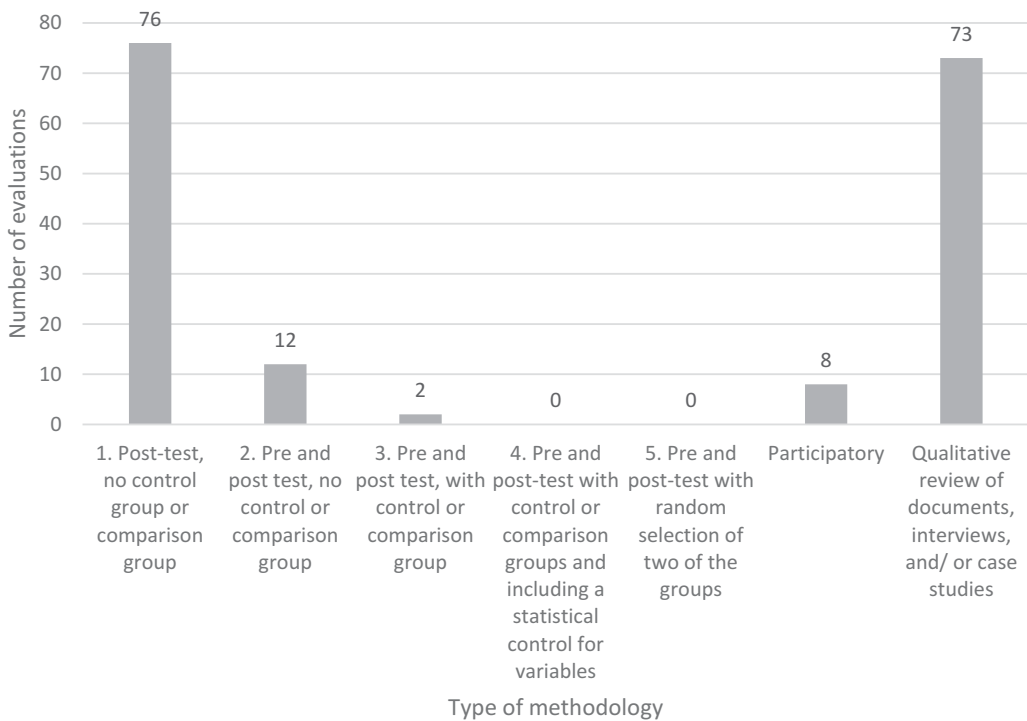


Figure 6. Evaluation methodology by number of evaluations.

but instead stated that the interventions aimed to tackle human trafficking more generally. The categorization of different sectors was fairly generous and was based on the description of any sector within the evaluation, often found in the description of the program and its objectives.

Evident in [Figure 5](#), half the evaluations ($n = 45$) concluded that the program was successful, with the achievement of some program objectives or outcomes. The term “Inconclusive” was used when the

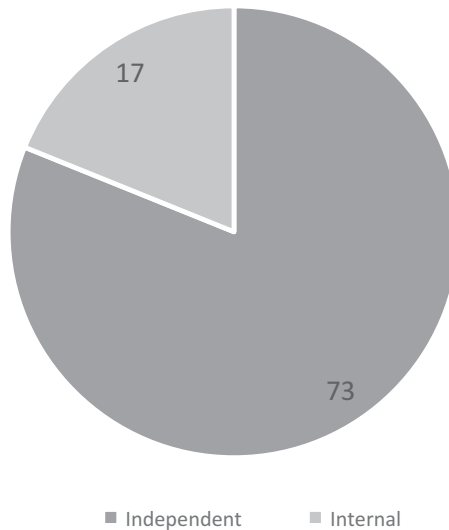


Figure 7. Independent versus internal evaluation?

author of the evaluation determined that the results were inconclusive, while “Unclear” was used when the research team categorizing the evaluations were unable to determine if objectives had or had not been met.

The standard of evaluation was fairly low (refer to [Figure 6](#)), with only two evaluations reaching level three on the Maryland scale; none were categorized above level three. The vast majority were tagged as posttest, without a control or comparison group, and with a qualitative review of documents, interviews and/or case studies produced during program implementation.

Finally, the majority of evaluations conducted were independent (refer to [Figure 7](#)), which is to say, they were conducted by a team or individual independent of the organization that implemented the program.

Discussion

Evaluation in Practice in the Anti-trafficking Field

The promising practices database and the current literature on monitoring and evaluation of anti-trafficking interventions establish patterns in current practice from which we can learn important lessons. Several systematic reviews have been conducted in the past decade, throwing light on the practice of evaluation in the sector and offering insight into gaps in our knowledge (Bryant & Joudo, 2018; Davy, 2015; 2016; Hames et al., 2010; van der Laan et al., 2011). These studies support many of the conclusions of this paper.

First, it is apparent that the number of evaluations of anti-trafficking interventions are on the rise, but that evaluation is still seen as optional. While signs point to a growing practice of conducting evaluations, there is a dearth of publicly available evaluations of anti-trafficking programs and interventions – often they are kept internal or not conducted at all (Davy, 2016). Similar findings appear even when the end of project evaluations are included. Knowing that there could be more than 1,500 active organizations¹⁸ operating in the anti-trafficking and anti-modern slavery policy area, but that only 179 evaluations were identified is telling. This finding remains valid even with the

¹⁸Data taken from Polaris Global Modern Slavery Directory. Figures represent number of organizations which list English as an operating language as of 6 August 2019.

inclusion of more recent evaluations. The initial searches for the 2019 update to the database found an additional 11 anti-trafficking evaluations in academic and international organization databases covering the period 2016–2019.¹⁹

Further, many of the full-text 179 evaluations were not publicly available, with either only short summaries available on websites. Those that could be accessed after direct requests were made to organizations for specific evaluations were included in the database, while far more could not be located and were subsequently removed. If there are to be lessons learned from previous anti-trafficking interventions, then the full versions of these documents should be made public.

Second, the evaluations appear not to meet established standards of evaluation, which can adversely affect their reliability and pose a barrier to evidence-based anti-trafficking work. In 73 instances, the anti-trafficking evaluations consisted of qualitative review of program documents, interviews with key stakeholders, and individual case studies. While there are benefits to the use of qualitative methods in evaluation (Patton, 2002), in practice some of these 73 evaluations were of poor quality in terms of methodology and content. Few evaluations, for example, triangulated their analyses or evaluated the reliability of documents or interviewees. The evaluations were also opaque regarding the methodology used or how their key findings were determined. There were some instances in the broader promising practices database outside of anti-trafficking interventions where an executive summary determined that a program was a success, which was not supported by a closer inspection of the full report. This limits the validity of the findings and the lessons that can be shared from these studies.

In recent years there are indications that the quality of anti-trafficking evaluations is improving. Of the 11 additional evaluations identified, five were level 3 and above on the Maryland Scale, two of which were RCTs (Archer, Boittin, & Mo, 2016; Gausman, Chernoff, Duger, Bhabha, & Chu, 2016). Those that were post reviews of project documentation, interviews, and end-line surveys were explicit about their triangulation methods. While the updated systematic review is not complete, this does point to a promising development in terms of quality.

The anti-trafficking evaluations in the promising practices database were often limited to process evaluation, giving a skewed sense of success without really informing whether an intervention is effective at producing results. The US Government Accounting Office (2011) defines four types of evaluation: (1) Process or Implementation Evaluation; (2) Outcome Evaluation; (3) Impact Evaluation; (4) Cost-Benefit and Cost-Effectiveness Analyses. Process evaluations look at whether the program was implemented according to plan, whereas Outcome and Impact evaluations go deeper into whether the program met its objectives for change. Twenty of the evaluations either determined that the program was “Inconclusive” in the achievement of its objectives or it was too difficult for the research team to determine if the program was a success (tagged as “Unclear”). This is in part due to the quality of program design. Often missing was an explicit theory of change underpinning program logic. This can lead to program objectives that are not clearly defined, are overly broad and impossible to evaluate or too narrow to reflect meaningful impact (Bryant & Joudo, 2018; Davy, 2016). Program objectives which are clear, meaningful, achievable, and measurable frame an entire project. They guide the development of baseline data, indicators for monitoring, and act as a yardstick for evaluation.

Of the 90 anti-trafficking evaluations in the database, 73 were conducted by independent evaluators. However, often little or no details were given on the author of the study, their relationship to the organization or program being evaluated, or how they were funded, which may suggest that the number of independent evaluations is much lower. This is supported by Davy (2016), who found that many evaluations had been undertaken by program staff rather than an external, independent evaluator, especially in smaller NGOs. The relationship between the evaluator and the implementor of the project remains unclear in the 11 more recent evaluations. Internal

¹⁹Initial searches were conducted for truncated versions of “trafficking” and “evaluation” of academic, 3ie, UNODC, ILO, IOM, and DEC databases. These searches took place in August 2019 and will continue throughout latter part of 2019.

evaluations are more prone to bias, and not publishing the details of the evaluator points to a lack of transparency which affects reliability.

Barriers to Effective Monitoring and Evaluation and Impact Assessments

In many instances, the findings from the review of 90 anti-trafficking evaluations highlight the same barriers to conducting monitoring and evaluation and impact assessments across the international development sector. Limited resources, expertise and data, bias in methodology, ethical considerations regarding the use of RCTs, attribution challenges in conducting impact assessments, political constraints, and short-term project timescales are all barriers which prevent thorough evaluation of international development programming (Bamberger, Rugh, & Mabry, 2006, 2012), including anti-trafficking projects.

Evaluations can be costly and consume many staff hours, especially when designing and implementing impact assessments. If insufficient budget has been set aside for this work, it is difficult to achieve a consistent and rigorous approach (Bamberger et al., 2006, 2012). These same issues are reflected in evaluations conducted of anti-trafficking interventions. Cost can play a role in identifying who should be involved in the evaluation process (Gallagher & Surtees, 2012), and limit the methodologies that can be employed.

Furthermore, monitoring and evaluation conducted without expertise or training on data collection or analysis, or without proper care and planning, can impede results (Bamberger et al., 2006, 2012). This is often the case where evaluations are conducted in-house rather than by an external expert, and the weaknesses show in the methodology of international development and anti-trafficking evaluations alike. Aside from issues of methodology, independence in evaluations is essential for non-biased results. Bias is more likely in evaluations conducted where funding is dependent on a positive review rather than as part of a culture of learning and improving program planning based on real evidence and measurable experience. Competitiveness for scarce resources leads to inflation of results and suppression of lessons learned regarding initiatives that may be unsuccessful. This competitiveness affects the anti-trafficking sector. A 2018 report by United Nations University revealed that Official Development Assistance to tackle modern slavery, forced labor, human trafficking, and child labor, while an increase from US\$150 million in 2001, stood at approximately US\$433.7 million by 2013 (Gleason & Cockayne, 2018). By way of comparison, the ILO estimated in 2014 that the profits from forced labor alone equaled US\$150 billion (International Labour Office, 2014). Lessons that initiatives are unsuccessful are arguably as important as more positive learnings; it is therefore important that a safe space is created where these findings can be shared. Viewing evaluation as a separate activity to program design also hinders a culture of learning, by preventing the collection of baseline data, or incorporating findings into a cyclical approach to program design.

Running an RCT can significantly strengthen the findings of a study and allows an appreciation of the actual impact the intervention made, controlling for other factors (CIMA Global Academic Research Programme, 2017; Gausman et al., 2016). However, it poses great ethical challenges (Burrell, 2012) that should be carefully considered in any development program, including human trafficking. Even once conducted, measurement and attribution challenges due to the complex variables to assess (CIMA Global Academic Research Programme, 2017) can also be found in impact assessments of trafficking programming (Gallagher & Surtees, 2012). Finally, international development and anti-trafficking interventions usually depend on short-term funding from a variety of sources, which means programs can only be designed to fit short-term needs, thereby preventing longitudinal studies examining the long-term impacts of a program on beneficiaries and other stakeholders.

Barriers to Effective Monitoring and Evaluation and Impact Assessments in the Anti-trafficking Sector

Despite these commonalities, there are some barriers identified in the review of the evaluations in the promising practices database that are more specific to anti-trafficking interventions. For example, the concealed nature of human trafficking makes it difficult to estimate and establish baselines against which to measure. Estimates of human trafficking are conducted by using multiple systems estimation (MSE), a form of capture-tag-recapture using administrative datasets (Bales, 2017; Bales, Hesketh, & Silverman, 2015; Bales, Murphy, & Silverman, 2019, 2018b, 2018c; Van Dijk, Cruyff, & van der Heijden, 2018a; van Dijk, van der Heijden, & Heerdink, 2016), random-sample surveys (Joudo Larsen & Diego-Rosell, 2017; Pennington, Ball, Hampton, & Soulakova, 2009; Zhang, 2012), and/or other forms of reporting on trafficking incidents. MSE has not yet been used to track change over time, while random sample surveys are costly to implement. Any impact assessment, which would aim to show reduction in prevalence over time, is therefore challenging to implement. Accounting for displacement is a related issue, when, in response to an intervention, trafficking patterns move elsewhere. While surveys could identify displacement when conducted in adjoining areas, the location where trafficking is displaced to is often difficult to predict.

Issues of displacement are highlighted by the evaluation of International Justice Mission's program "Project Lantern" in Cebu, Philippines. The project provided capacity building for law enforcement, prosecutors, judges, court personnel, government officials and service providers; developed good practice training manuals, guidebooks, and curricula; procured resources, including equipment and supplies for law enforcement, prosecutorial officials and aftercare staff; and mobilized civil society to put pressure on the justice system to take targeted action to combat human trafficking. The evaluation concluded that there was an observed reduction in the availability of children for commercial sexual exploitation in Cebu, and that Project Lantern contributed to this change (Jones, Schlangen, & Bucoy, 2010). However, qualitative data collected during the evaluation confirmed that while children were less visible, this did not necessarily signify a reduction in trafficking. Instead, trafficking had potentially been driven underground or driven out of metro Cebu to other areas. Others pointed to the arrival of the "cottage porn industry" in private homes around the timeframe of the project, which may also explain the reduced visibility of children exploited for commercial sex (Jones et al., 2010).

A further barrier lies in the complexity of human trafficking crimes, which can limit the identification of long-term changes and impact. Human trafficking is a multifaceted crime, which encompasses many perpetrators, routes, sectors, victims, and forms of exploitation. However, nearly 50% of the evaluations ($n = 43$) did not specify which sector or form the program was focusing on, but instead stated that the interventions aimed to tackle human trafficking more generally. This highlights one of the weaknesses with the design of trafficking programs and subsequent evaluations – without clear understanding of the sectors or forms of exploitation that the program was aiming to address, it is difficult, if not impossible, to identify exactly which change is desired and therefore which intervention will or will not have an impact.

This complexity is also difficult to translate into survey instruments for the purposes of pre- and post-assessments. For instance, the Harvard FXB Center for Health and Human Rights conducted a comprehensive impact assessment of the Freedom Fund's project in Northern India aiming at eradicating forced labor and increasing socio-economic benefits in the targeted community (Gausman et al., 2016). The study, conducted 3 years after the beginning of the project, benefitted from a strong methodological framework, using both qualitative and quantitative indicators, a baseline study and a comparison group. The comparison group, with similar features, did not receive any direct intervention until later in the implementation of the program (Gausman et al., 2016). The end-line survey found no evidence of trafficking in any of the groups that received either full intervention, partial intervention, or no/limited intervention. Qualitative interviews conducted at the same time, however, revealed that trafficking was a problem, suggesting that the lack of evidence

at the end-line could be due to the hesitation in “providing frank answers or difficulty in understanding exactly what was being asked” (Gausman et al., 2016).

The level of trauma victims of human trafficking experience can make an effective evaluation more difficult. Several studies (Altun, 2017; Hossain, Zimmerman, Abas, Light, & Watts, 2010; Kiss et al., 2015) prove that those who have experienced trafficking show high levels of depression, anxiety, and post-traumatic stress disorder. This can preclude any ability or wish to engage in evaluation processes and can hinder longitudinal studies. One exception to this is a ten-year research project launched by Chab Dai in 2010 to better understand reintegration for survivors of trafficking for sexual purposes. The project releases reports one or two times a year, providing a balance of current, continued outputs with long-term investment in research on the process of reintegration (Tsai, Vanntheary, & Channtha, 2018). While the study boasts high retention rates – 76% in 2013²⁰ (Miles, Sophal, Vanntheary, Channtha, & Phally, 2014) – it attributes this to the high levels of trust built during the course of the project. This trust was built by conducting interviews three times per year in the initial year of study, maintaining a database of participants’ contacts, as well as “being available by phone for contact 24/7” (Miles et al., 2014). While commendable, such resource intensive approaches remain the exception rather than the rule.

A final barrier is the specific unintended consequences of anti-trafficking interventions. Much has been written of the linkages between human trafficking programing and anti-sex work policies (Kempadoo, Sanghera, & Pattanaik, 2005; Lerum & Brents, 2016), while recent improvements in governmental responses to modern slavery (Minderoo Foundation, 2019), have been undermined by the concurrent creation of hostile migration policies that exacerbate vulnerability (David, Bryant, & Joudo Larsen, 2019; Galos, Bartolini, Cook, & Grant, 2017; Minderoo Foundation, 2019). Accordingly, scholars and practitioners have highlighted the importance of conducting impact assessments on the unintended consequences of an anti-trafficking intervention (Gallagher, 2007; Global Alliance Against Trafficking in Women (GAATW), 2007), including adverse impact on the enjoyment of human rights. Conducting a human rights impact assessment is then crucial to be able to identify violations of the right to liberty after having freed enslaved or trafficked women (GAATW, 2007) or the adverse impact of anti-trafficking programmes on the rights to health of sex-workers (Ahmed & Seshu, 2012), as just two examples of potential human rights violations.

So, 20 Years On, Do We Know Nothing about What Works?

The ability to draw concrete conclusions regarding what works to combat human trafficking is hampered by the quality of evaluation of anti-trafficking programing and subsequent evaluations. While there appears to be an improvement in the quality of evaluations conducted since 2016, the number of quality evaluations remains limited.

However, it is not the intention of this paper to throw out all findings from the review of the anti-trafficking evaluations. To disregard all programing and evaluations is counter-productive and leaves us no further along in our understanding than we were in the year 2000. A thematic analysis of the evaluations in the promising practices database provides us with some potential lessons learned. While not strong enough to conclusively answer the question “what works,” it provides a snapshot of where there might be some promising practices for future interventions.

We have also learned much about the specificities of conducting monitoring and evaluation in the anti-trafficking sector. The difficulties encountered does not mean that we should not conduct monitoring and evaluation, preferring instead to focus limited resources on interventions and support for survivors. Monitoring, evaluation, and impact assessment are based on the careful specification of aims and objectives for ending human trafficking or addressing a core component of exploitative practices, measuring and monitoring the progress of a particular intervention, and providing a framework for understanding the overall impact of anti-trafficking work. Without these

²⁰Most recent data that could be found.

evaluations, it is not possible to target limited resources and provide the means through which to judge whether anti-trafficking efforts are having an impact on those affected, bringing about a strengthening of prevention mechanisms, and ultimately a reduction in prevalence. Based on the review of the methodologies underpinning these evaluations, this paper concludes by highlighting some monitoring and evaluation methods that require further investigation.

Conducting Raising Awareness Campaigns

The use of raising awareness campaigns in anti-trafficking programming is common. They tend to include educating the public on definitions or indicators of human trafficking and how to report incidences, as well as more targeted interventions highlighting risks for specific groups. Of the 90 in the database, 47 evaluations described programs that included some form of raising awareness campaign. Many of these evaluations conclude that raising awareness campaigns had mixed results, but that in order to be effective, campaigns need to be targeted to particular groups, have a clear message, and be adapted to local contexts.

For example, a 2006 evaluation of International Organization for Migration (IOM)'s information campaigns in Cambodia made limited use of pre- and post-assessment and a control group. It found that awareness levels increased as a result of a mass information campaign and village-based activities, in that participants were better able than control groups to repeat the main messages from the campaign and to recite the hotline number and the penalties associated with trafficking crimes. However, the evaluation also found that there was limited understanding of what trafficking constitutes and highlighted the need for a more user-friendly trafficking definition (Sainsbury, 2006). A 2014 evaluation of UN Women's anti-human trafficking program in India also concluded that the program's raising awareness campaign reported mixed results, with some target audiences reporting that they had received the information, and others reporting that they had not (UN Women, 2014). A review of MTV Exit, a large-scale multi-media campaign designed to raise awareness of human trafficking in the Asia-Pacific region, revealed that programs "need to have clear and targeted messaging with a call to action and that promotes behaviour change" (Skuse and Downman, 2012). A 2012 evaluation of an anti-trafficking campaign in Kosovo also determined that there was a slight increase in knowledge on human trafficking, and that more information should be targeted to the affected groups and communities in a clear and simple manner (Kuneviciute, Waeyenberge, Viaene, & Moens, 2012).

The literature and an evaluation conducted in 2016 support these findings. A 2017 review of 55 demand-side raising awareness campaigns (5% of which had been evaluated) found that these campaigns needed to go beyond awareness raising to target specific behaviors. It concluded that campaigns only have a positive impact on reducing exploitation and human trafficking in combination with other interventions, if at all (Vogel & Cyrus, 2017). The 2016 RCT of awareness raising campaigns in Nepal revealed more positive findings, concluding that human trafficking awareness campaigns increase the ability of respondents to self-identify as having been trafficked and to recognize the occurrence of human trafficking among family and friends (Archer et al., 2016). The campaigns that were subject to the RCT were highly adapted to the Nepalese context; before the campaign was implemented, the team researched the different media types and respondent access in Nepal, interviewed NGOs that had conducted similar campaigns in the past, gathered stories through source interviews, and piloted the materials (Archer et al., 2016).

Strengthening Legislation

Forty-six evaluations described programs that included elements of technical support to governments to strengthen legislation, among other outcomes. Many of these evaluations highlighted the importance of local ownership and embedding trafficking interventions in existing national structures. UNODC's global programme, GLOT55, aimed to promote the *UN Trafficking Protocol* in 55 countries by providing assistance to selected states to strengthen their legal frameworks; conducting

national and regional trainings; improving support for victims; and supporting raising awareness activities (Marshall & Berman, 2013). The evaluation commended the program's flexibility, which allowed countries to build national ownership of the response to trafficking and highlighted that the program did support many countries to begin the process of drafting national legislation. However, there was no concrete change in the number of ratifications or the passage of national legislation during the program's implementation. Further, the global design prevented the adaptation of activities by local stakeholders. The importance of local ownership and the necessity for broad consultation with key stakeholders was also highlighted in an evaluation of an UNODC project in South Africa, which was implemented between September 2009 and March 2013. Legislation was not passed during the project timeframe, while there remained uncertainty if the establishment of a Border Management Agency, one of the activities of the project, would be achieved. External factors, such as hosting of the World Cup in 2012, led to project delays (Plessis, 2013). Comprehensive anti-trafficking legislation was passed in South Africa just a few months later in July 2013 (Act No. 7 of 2013: Prevention and Combating of Trafficking in Persons Act, 2013).

What these evaluations reveal is not unique to these projects. Lack of national ownership, limited consultation with key stakeholders, and the need for longer program timeframes to ensure the sustainability of results was common to a number of evaluation findings in the database.

Supporting Survivors

Central to most anti-human trafficking programming are interventions to identify and provide support to trafficking victims. Fifty-three evaluations in the database include initiatives such as training for front line service providers, case management, medical support, shelters, support groups, a hotline, and longer-term reintegration services such as education and job placement programs.

Fourteen evaluations in the database were tagged as assessing programs that included some form of case management. A significant number of these were implemented in the U.S. ($n = 9$), followed by the Philippines ($n = 2$) and the UK ($n = 2$). Case management is defined as the provision of coordinated and comprehensive services to victims of modern slavery, where that assistance is tailored to an individual's needs, as opposed to a more general, "one size fits all" approach to victim assistance (Bryant & Joudo, 2019). A 2014 evaluation of services provided to victims of human trafficking in New York used participant observation, review of documents, and interviews with key staff to determine the effectiveness of a trauma-focused approach, concluding that, "being trafficked is in itself a traumatic experience, then a trauma-informed care framework for providing services to traumatised individuals makes sense" (Heffernan & Blythe, 2014). The importance of trauma-focused care was also central to the *Oshkiniigikwe* program, which provides support services to American Indian and Alaska Native women, adolescent girls and their families. The evaluators conducted interviews with clients at entry point into the program and again six months later to determine that delivering services in culturally sensitive ways and supporting victims to strengthen their sense of identity were an integral part of case management (Pierce, 2012). Finally, a 2014 evaluation of NGO programming in Albania, Kosovo, Serbia, Romania, and Bulgaria, highlighted the importance of longer term support for victim assistance initiatives to allow sufficient time for recovery (Surtees, 2015).

Improving Monitoring and Evaluation of Anti-trafficking Interventions

Reflecting on barriers to monitoring and evaluation and the specific difficulties regarding evaluations of anti-trafficking interventions highlights some interesting approaches to be taken into account when conducting future evaluations of human-trafficking programming.

The ultimate goal of many trafficking programs in the promising practices database is a reduction of prevalence. However, given the difficulties in measuring impact, and the ethical issues surrounding the use of RCTs, this should not be the focus of every anti-trafficking

intervention. Nor should we raise the expectation that every project should necessarily reduce prevalence. Tracking progress toward proxy indicators, such as change in behaviors or a reduction in risk factors and strengthening in protective factors, may be more achievable for many projects. Measuring prevalence would then become the remit of impact assessments of programs targeting whole sectors or geographic regions. Alongside this, and for prevention projects, in particular, there is also a need to expand understanding of human trafficking to assess the myriad of contextual factors and underlying systems that combine to cause human trafficking. For example, any impact assessments of labor trafficking projects should look at labor migration policies, social protections, and transparent labor recruitment methods (Kiss & Zimmerman, 2019). This also requires a change in donor behavior to prioritize robust evaluation and see impact evaluation as a tool to assess these broader interventions.

RCTs are now being conducted, as shown by the RCT of raising awareness campaigns in Nepal (Archer et al., 2016), community empowerment interventions in India (Gausman et al., 2016), and forthcoming studies in northern and southern India (The Freedom Fund, 2019). In India, ethical issues were resolved by comparing impact on three cohorts: (1) hamlets where the intervention was considered “mature” (full intervention), (2) where the intervention was not considered “mature”, but where work was ongoing (partial intervention), and (3) those where the intervention was limited and the intervention took place toward the end of the study period (comparison group) (Gausman et al., 2016). In other policy areas, RCTs are used in a lagged fashion, so that interventions are implemented at a later time in those areas where one was not completed in the first instance. This is one interesting method which should be applied more rigorously in the anti-trafficking field.

Despite donor preference for quantitative methods in anti-trafficking evaluations (Gallagher & Surtees, 2012), qualitative methods to evaluation can lead to improved understanding of the effectiveness of projects. Techniques such as observation, participant observation, key informant interviews, and focus groups can give us insight into changes or lack thereof. These techniques can be successful as long as these are conducted critically and by triangulating with other data sources (Patton, 2002). Mixed method approaches, however, remain the ideal for conducting impact assessments.

Survivors should play a central role in monitoring and evaluation of anti-trafficking programing. Participatory approaches can be a promising approach to evaluation that improves retention rates and acts as a vehicle for rehabilitation. Studies on Freedom Fund’s programme in Bihar and Uttar Pradesh, for example, utilized life stories and participatory statistics – a participatory method to collect and analyze data to assess whether the program had made a real contribution to the eradication of modern slavery and bonded labor in the targeted communities (Burns, Oosterhoff, Raj, & Nanda, 2015; Oosterhoff et al., 2016). The participatory model to impact assessment has been generally considered effective in the anti-modern slavery field since it engages with local communities and survivors to better understand the long-term impact of the intervention on their lives (International Council on Human Rights Policy, 2011). A participatory model can “give people affected by slavery a voice about what should be counted and gives them a chance to input into how the survey results could be used for locally relevant action” (Oosterhoff et al., 2016). While promising, it is still important to assess which stakeholders to include in this process, the role of psycho-social support during the evaluation process, and how to provide adequate and uniform training for those who participate.

Conclusion

Drawing this together, nearly 20 years on from the ratification of the *UN Trafficking Protocol*, we do not have concrete answers to the question “what works” to combat human trafficking. However, there are interesting lessons learned and suggestions on how to strengthen our use of monitoring and evaluation. The quality of anti-trafficking evaluations appears to be improving, with increased use of impact assessment, as part of a toolkit of monitoring and evaluation techniques. Its use,

however, only makes sense when evaluating the impact of a significant program or an organizational theory of change. Use of participatory statistics and other participatory models are an important development to tracking impact of anti-trafficking programming. Qualitative data collection techniques should not be dismissed and can highlight findings when evaluators triangulate data and critically engage with program documents and interviewees. Donors should reflect the importance of monitoring and evaluation by providing for longer timeframes for calls for proposals and program implementation as well as resources to conduct monitoring and evaluation.

Despite the limitations of the methodologies used in anti-trafficking interventions, there are some common findings from existing evaluations that allow us to draw some conclusions regarding what works. The impact of raising awareness campaigns is limited when these are not targeted to specific communities with a clear message. Support for governments to pass legislation can be deemed to have had an impact, but only when there is national ownership and sufficient time allocated to reflect the length of time it takes to implement legislative and policy change. Support for victims is effective when it is victim-centered, applies a trauma – focused lens and prioritizes the sense of identity of the victim.

With an estimated 40.3 million people in some form of modern slavery in the world and a global commitment to end human trafficking and modern slavery by 2030, it is important to understand “what works” in the fields of anti-trafficking. Across anti-trafficking work, real positive change needs to be demonstrated through effective use of baseline data, the setting of real and deliverable outcomes that can be measured in some way, and a framework for analyzing the contribution and direct impact of purposive interventions designed to combat human trafficking and modern slavery. In the absence of tangible evidence and well thought out programs that can be evaluated, anti-trafficking and anti-modern slavery interventions will struggle in their efforts to end modern slavery in all its forms and to maintain crucial support from a variety of donors and funders within a highly competitive environment.

Acknowledgments

We are particularly grateful to Bernadette Joudo who completed many of the searches for the promising practices database. The Minderoo Foundation funded the development of the promising practices database, with support from the Freedom Fund.

We are also grateful to Elena Abrusci and Kathryn Walker who carried out the background research on monitoring, evaluation and impact assessment for this article. Funding has come from the Rights Lab, a Beacon of Excellence at the University of Nottingham.

Todd Landman receives funding from the UK’s Economic and Social Research Council’s Impact Accelerator Award (IAA) and is the host of The Rights Track (www.rightstrack.org), a human rights podcast series funded by the Nuffield Foundation, the ESRC, and the Rights Lab. Series 3 and 4 of The Rights Track are dedicated to the issue of modern slavery. This work was supported by the Economic and Social Council [grant number ES/K00803X/1]; the University of Nottingham; and the Nuffield Foundation.

References

- Act No. 7 of 2013: Prevention and Combating of Trafficking in Persons Act, (2013).
- Ahmed, A., & Seshu, M. (2012). “We have the right not to be ‘rescued.’”: When anti-trafficking programmes undermine the health and well-being of sex workers. *Anti Trafficking Review*, 1, 149–168.
- Altun, S., Abas, M., Zimmerman, C., Howard, L. M., & Oram, S. (2017). Mental health and human trafficking: Responding to survivors’ needs. *BJPsych International*, 14(1), 21–23.
- Archer, D., Boittin, M., & Mo, C. H. (2016). Reducing vulnerability to human trafficking: An experimental intervention using anti-trafficking campaigns to change knowledge, attitudes, beliefs, and practices in Nepal. Retrieved from <https://www.iie.org/Research-and-Insights/Publications/DFG-Vanderbilt-Publication>
- Bales, K. (2017). Unlocking the statistics of slavery. *Chance*, 30(3), 4–12. doi:10.1080/09332480.2017.1383105
- Bales, K., Hesketh, O., & Silverman, B. (2015). Modern slavery in the UK: How many victims? *Significance*, 12(June), 16–21. doi:10.1111/j.1740-9713.2015.00824.x
- Bales, K., Murphy, L., & Silverman, B. (2019). How many trafficked people are there in greater New Orleans? Lessons in measurement. *Journal of Human Trafficking*, Forthcoming, 1–13. doi:10.1080/23322705.2019.1634936
- Bamberger, M., Rugh, J., & Mabry, L. (2006). *Real world evaluation: Working under budget, time, data, and political constraints*. Thousand Oaks, California, USA: Sage Publications.

- Bamberger, M., Rugh, J., & Mabry, L. (2012). *Real world evaluation: Working under budget, time, data, and political constraints* (2nd ed.). Thousand Oaks, California, USA: Sage Publications.
- Blankenberg, F. (1995). *Methods of impact assessment research programme, resource pack and discussion*. The Hague, Netherlands: Oxfam UK/ I and Novib.
- Bryant, K., & Joudo, B. (2018). What works? A review of interventions to combat modern day slavery. Retrieved from <https://www.walkfreefoundation.org/news/resource/works-review-interventions-combat-modern-day-slavery/>
- Bryant, K., & Joudo, B. F. (2019). *What works? Lessons learned in case management*. Perth: Walk Free Foundation.
- Burns, D., Oosterhoff, P., Raj, A. M., & Nanda, R. B. (2015). Patterns and dynamics of slavery and bonded labour in Bihar and Uttar Pradesh: Findings from life story analysis. Retrieved from <https://d1r4g0yjc7lx.cloudfront.net/wp-content/uploads/IDS-Dynamics-of-slavery-in-UP-and-Bihar-11th-Jan-2016-FINAL-W-NAMES-CHANGED-.pdf>
- Burrell, J. (2012). On the overuse of randomised control trials in the aid sector. *Global Policy Journal* [Online]. <http://www.globalpolicyjournal.com/blog/30/10/2012/overuse-randomized-controlled-trials-aid-sector>
- CIMA Global Academic Research Programme. (2017). *Impact assessment in a non-government organisation*. United Kingdom.
- David, F., Bryant, K., & Joudo Larsen, J. (2019). Migrants and their vulnerability to human trafficking, modern slavery and forced labour. Retrieved from https://publications.iom.int/system/files/pdf/migrants_and_their_vulnerability.pdf
- Davy, D. (2015). Understanding the support needs of human-trafficking victims: A review of three human-trafficking program evaluations. *Journal of Human Trafficking*, 1(4), 318–337. doi:10.1080/23322705.2015.1090865
- Davy, D. (2016). Anti-human trafficking interventions: How do we know if they are working? *American Journal of Evaluation*, 37(4), 486–504. doi:10.1177/1098214016630615
- Farrington, D. P., Gottfredson, D. C., Sherman, L. W., Welsh, B. C. (2002). Maryland scientific methods scale. In L. W. Sherman & D. P. Farrington (Eds.), *Evidence-based crime prevention* (pp. 13–21). London: Routledge.
- Gallagher, A. T. (2007). *The law and politics of human trafficking: Presentation at the Centre for International Justice and Governance* Canberra: Australian National University.
- Gallagher, A. T., & Surtees, R. (2012). Measuring the success of counter-trafficking interventions in the criminal justice sector: Who decides- and how? *Anti-Trafficking Review*, 1, 10–30.
- Galos, E., Bartolini, L., Cook, H., & Grant, N. (2017). Migrant vulnerability to human trafficking and exploitation: Evidence from the Central and Eastern Mediterranean migration routes. Retrieved from https://publications.iom.int/system/files/pdf/migrant_vulnerability_to_human_trafficking_and_exploitation.pdf
- Gausman, J., Chernoff, M., Duger, A., Bhabha, J., & Chu, H. (2016). When we raise our voice: The challenges of eradicating labor exploitation. Retrieved from <https://d1r4g0yjc7lx.cloudfront.net/wp-content/uploads/FINAL-When-We-Raise-Our-Voice-.pdf>
- Gleason, K. A., & Cockayne, J. (2018). Official development assistance and SDG target 8.7: Measuring aid to address forced labour, modern slavery, human trafficking and child labour. Retrieved from http://collections.unu.edu/eserv/UNU:6612/Development_Assistance_and_SDG_Target_8.7_FINAL_WEB_7_.pdf
- Global Alliance Against Trafficking in Women (GAATW). (2007). *Collateral damage. The impact of anti-trafficking measures on human rights around the world*. Bangkok, Thailand: Global Alliance Against Trafficking in Women.
- Götzmann, N., Vanclay, F., & Seier, F. (2016). *Social and Human Rights Impact Assessments: What Can They Learn from Each Other? Impact Assessment and Project Appraisal*, 34(1), 14–23. doi:10.1080/14615517.2015.1096036
- Hames, C., Dewar, F., & Napier-Moor, R. (2010). *Feeling good about feeling bad ... a global review of evaluation in anti-trafficking initiatives*. Bangkok, Thailand: Global Alliance Against Trafficking in Women.
- Harkins, B. (2017). Constraints to a robust evidence base for anti-trafficking interventions. *Anti Trafficking Review*, 8, 113–130.
- Harrison, J., & Stephenson, M-A. (2010). *Human Rights Impact Assessment: Review of Practice and Guidance for Future Assessments*, Scottish Human Rights Commission. Retrieved from <https://fian-ch.org/content/uploads/HRIA-Review-of-Practice-and-Guidance-for-Future-Assessments.pdf>
- Heffernan, K., & Blythe, B. (2014). Evidence-based practice: Developing a trauma-informed lens to case management for victims of human trafficking. [Human trafficking]. *Global Social Welfare*, 1, 169–177. doi:10.1007/s40609-014-0007-8
- Hossain, M., Zimmerman, C., Abas, M., Light, M., & Watts, C. (2010). The relationship of trauma to mental disorders among trafficked and sexually exploited girls and women. *American Journal of Public Health*, 100(12), 2442–2449. doi:10.2105/AJPH.2009.173229
- Inter-agency Coordination Group Against Trafficking in Persons (ICAT). (2016). Issue paper: Pivoting towards the evidence: Building effective counter-trafficking responses using accumulated knowledge and a shared approach to monitoring, evaluation and learning. Retrieved from http://icat.network/sites/default/files/publications/documents/16-10259_Ebook.pdf
- International Council on Human Rights Policy. (2011). *Assessing the impact of human rights work: Challenges and choice*. Retrieved from: http://www.ichrp.org/files/papers/186/impact_assessment_human_rights_approach_paper.pdf
- International Labour Office. (2014). *Profits and poverty: The economics of forced labour*. Geneva, Switzerland. Retrieved from https://www.ilo.org/wcmsp5/groups/public/—ed_norm/—declaration/documents/publication/wcms_243391.pdf

- International Labour Organization, Walk Free Foundation, & International Organization for Migration. (2017). Global estimates of modern slavery: Forced labour and forced marriage. Retrieved from https://www.ilo.org/global/publications/books/WCMS_575479/lang-en/index.htm
- Jones, A., Schlangen, R., & Bucoy, R. (2010). An evaluation of the international justice mission's "Project lantern" assessment of five-year impact and change in the public justice system. Retrieved from <https://www.ijm.org/sites/default/files/download/resources/120610-Project-Lantern-Impact-Assessment-AJ.pdf>
- Joudo Larsen, J. L., & Diego-Rosell, P. (2017). Using surveys to estimate the national prevalences of modern slavery: Experience and lessons learned. *CHANCE*, 30(3), 30–35. doi:10.1080/09332480.2017.1383110
- Kempadoo, K., Sanghera, J., & Pattanaik, B. (2005). *Trafficking and prostitution reconsidered: New perspectives on migration, sex work, and human rights*. Boulder, Colorado, US: Paradigm Publishers.
- Kiss, L., Pocock, N. S., Naisangansri, V., Suos, S., Dickson, B., Thuy, D., ... Zimmerman, C. (2015). Health of men, women, and children in post-trafficking services in Cambodia, Thailand, and Vietnam: An observational cross-sectional study. *The Lancet Global Health*, 3, e154–e161. doi:10.1016/S2214-109X(15)70016-1
- Kiss, L., & Zimmerman, C. (2019). Human trafficking and labor exploitation: Toward identifying, implementing, and evaluating effective responses. *PLoS Medicine*, 16(1), e1002740. doi:10.1371/journal.pmed.1002740
- Kuneviciute, I., Waeyenberge, L., Viaene, N., & Moens, M. (2012). Anti trafficking campaign in Kosovo 2012: Final campaign evaluation report. Retrieved from https://childhub.org/en/system/tdf/library/attachments/1589_Final_Anti-Trafficking_Campaign_Report_%2829_January%29_original.pdf?file=1&type=node&id=6297
- Landman, T. (2006). *Studying human rights*. London: Routledge.
- Lerum, K., & Brents, B. G. (2016). Sociological Perspectives on Sex Work and Human Trafficking, *Sociological Perspectives*, 59(1), 17. doi: 10.1177/0731121416628550
- Marshall, P., & Berman, J. (2013). Promoting the implementation of the protocol to prevent, suppress and punish trafficking in persons, especially women and children, and the protocol against smuggling of migrants by land, sea and air, both supplementing the United Nations convention against transnational organized crime. Retrieved from http://www.unodc.org/documents/evaluation/Independent_Project_Evaluations/2014/GLOT55_Independent_Project_Evaluation_Report_2014.pdf
- Miles, S., Sophal, H., Vannthear, L., Channtha, N., & Phally, S. (2014). Butterfly methodology change: A reflection paper. Retrieved from <https://static1.squarespace.com/static/55a81f9be4b01a30079bb9d3/t/55b9519ce4b07fef1f59c9fd/1438208412763/Butterfly+2014+Methods.pdf>
- Minderoo Foundation. (2019). Measurement, action, freedom: An independent assessment of government progress towards achieving UN sustainable development goal 8.7. Retrieved from <https://www.minderoo.com.au/measurement-action-freedom/>
- OECD/Development Assistance Committee. (1991). *DAC Principles for Evaluation of Development Assistance*. Paris, France: OECD Publishing. Retrieved from: <https://www.oecd.org/dac/evaluation/2754804.pdf>
- Office of the High Commissioner for Human Rights. (2011). Guiding principles on business and human rights: implementing the united nations "Protect, Respect and Remedy" framework. Retrieved from: https://www.ohchr.org/documents/publications/GuidingprinciplesBusinesshr_eN.pdf
- Oosterhoff, P., Bharadwaj, S., Burns, D., Raj, A. M., Nanda, R. B., and Narayanan, P. (2016). *Using Participatory Statistics to Examine the Impact of Interventions to Eradicate Slavery: Lessons from the Field*, Centre for Development Impact. Retrieved from <https://www.cdimpact.org/publications/using-participatory-statistics-examine-impact-interventions-eradicate-slavery-lessons>
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). London, UK: Sage.
- Pennington, J. R., Ball, W. A., Hampton, R. D., & Soulakova, J. N. (2009). The cross-nation market in human beings. *Journal of Macromarketing*, 29(2), 119–134. doi:10.1177/0276146708327630
- Pierce, A. (2012). American Indian adolescent girls: Vulnerability to sex trafficking, intervention strategies. [Human trafficking; sale or exploitation of children]. *American Indian and Alaska Native Mental Health Research: the Journal of the National Center*, 19(1), 37–56. doi:10.5820/aian.1901.2012.37
- Plessis, P. D. (2013). Strengthening law enforcement capacity (Border control operations) and criminal justice response to smuggling of migrants and trafficking in person. Retrieved from http://www.unodc.org/documents/evaluation/Independent_Project_Evaluations/2013/ZAF_T54_Independent_Project_Evaluation_Report_01Aug2013.pdf
- Sainsbury, C. (2006). *Information campaign to combat trafficking in women and children in Cambodia*. Retrieved from If you are interested in the following evaluation report, please contact OIG/Evaluation at Headquarters (eva@iom.int).
- Skuse, A., & Downman, S. (2012). MTV EXIT ASIA III: A Campaign to Increase Awareness and Prevention of Trafficking in Persons. An independent review. Retrieved from: <https://dfat.gov.au/about-us/publications/Documents/mtv-exit-iii-review-report.pdf>
- Surtees, R. (2015). *Beyond trafficking. The re/integration of trafficking victims in the Balkans 2007 to 2014*. Brussels, Belgium: King Baudouin Foundation
- The Freedom Fund. (2019). Planning and evaluation of the northern India hotspot. Retrieved from <https://freedomfund.org/programs/monitoring-evaluation/institute-of-development-studies-university-of-sussex/>

- Tsai, L. C., Vanntheory, L., & Channtha, N. (2018). Experiences in shelter care: Perspectives from participants in the butterfly longitudinal study. Retrieved from <https://chabdai.org/butterfly/>
- UN General Assembly. (2015). Transforming our world: The 2030 agenda for sustainable development, A/RES/70/1. Retrieved from <http://www.refworld.org/docid/57b6e3e44.html>.
- UN Women. (2014). End-line evaluation: UN women's anti-human trafficking program. Retrieved from <http://gate.unwomen.org/EvaluationDocument/Download?evaluationDocumentID=3941>
- United Nations Convention against Transnational Organized Crime and the Protocols Thereto (CTOC), (2004).
- UNODC. (2008). Toolkit to combat trafficking in persons: Global programme against trafficking in human beings. Retrieved from https://sherloc.unodc.org/res/cld/bibliography/toolkit-to-combat-trafficking-in-persons_html/07-89375_Ebook1.pdf
- UNODC. (2009). International framework for action to implement the trafficking in persons protocol. Retrieved from https://www.unodc.org/documents/human-trafficking/Framework_for_Action_TIP.pdf
- US Department of State (2018). Trafficking in persons report 2018. Retrieved from <https://www.state.gov/documents/organization/282799.pdf>
- US Government Accountability Office. (2011). *Performance measurement and evaluation*. (GAO-11-646SP). Washington D.C., USA Retrieved from <https://www.gao.gov/assets/80/77277.pdf>.
- van der Laan, P. H., Smit, M., Busschers, I., & Aarten, P. (2011). Cross-border trafficking in human beings: Prevention and intervention strategies for reducing sexual exploitation. *Campbell Systematic Reviews*, 7, 1–50. doi:10.4073/csr.2011.9
- van Dijk, J., Cruyff, M., & van der Heijden, P. G. M. (2018a). Monitoring target 16.2 of the United Nations sustainable development goals: Multiple systems estimation of the numbers of presumed victims of trafficking in persons. Ireland. United Nations Office on Drugs and Crime. Vienna, Austria. Retrieved from https://www.unodc.org/documents/data-and-analysis/glotip/MSE_Research_Brief_Ireland.pdf
- van Dijk, J., Cruyff, M., & van der Heijden, P. G. M. (2018b). Monitoring target 16.2 of the United Nations sustainable development goals: Multiple systems estimation of the numbers of presumed victims of trafficking in persons. Romania. United Nations Office on Drugs and Crime. Vienna, Austria. Retrieved from https://www.unodc.org/documents/data-and-analysis/glotip/MSE_Research_Brief_Romania.pdf
- van Dijk, J., Cruyff, M., & van der Heijden, P. G. M. (2018c). Monitoring target 16.2 of the United Nations sustainable development goals: Multiple systems estimation of the numbers of presumed victims of trafficking in persons. Serbia. United Nations Office on Drugs and Crime. Vienna, Austria. Retrieved from: https://www.unodc.org/documents/data-and-analysis/glotip/MSE_Research_Brief_Serbia.pdf
- van Dijk, J., van der Heijden, P. G. M., & Heerdink, S. (2016). *Multiple systems estimation for estimating the number of victims of human trafficking across the world*. Southampton, UK: University of Southampton Institutional Repository.
- Vanclay, F. (2003). International principles for social impact assessment. *Impact Assessment and Project Appraisal*, 21 (1), 5–12. doi: 10.3152/147154603781766491
- Vogel, D., & Cyrus, N. (2017). How successful are campaigns addressing the demand-side of trafficking in human beings? Retrieved from [https://www.demandat.eu/sites/default/files/D12.4%20DemandAT-PB-%20Campaigns_web version.pdf](https://www.demandat.eu/sites/default/files/D12.4%20DemandAT-PB-%20Campaigns_web%20version.pdf)
- Voyages, S. (2019). Trans-Atlantic slave trade - estimates. Retrieved from <https://www.slavevoyages.org/assessment/estimates>
- Walk Free Foundation. (2018). The global slavery index 2018. Retrieved from www.globalslaveryindex.org
- Zhang, S. X. (2012). Trafficking of migrant laborers in San Diego county: Looking for a hidden population. Final REport submitted to the national institute of justice. Retrieved from <https://www.ncjrs.gov/pdffiles1/nij/grants/240223.pdf>

Appendix A: Search terms

Evaluation (if necessary) OR Assessment (if necessary)

AND

-Modern slavery OR

-Human trafficking OR

-Worst forms of child labor OR

-Slavery OR

-Forced labor OR

-Domestic servitude OR

-Debt bondage OR

-Forced marriage OR

-Servile marriage OR

-Early marriage OR

- Child marriage OR
- Sale or exploitation of children OR
- Use of child soldiers OR
- Child labor OR
- Prostitution OR
- Refugees OR
- Internally displaced persons OR
- Female genital mutilation OR
- Safe migration OR
- Labor migration

A total of 1,787,748 sources were identified from these searches, largely due to the number of sources identified through internet searches. For internet searches that revealed a large number of evaluations, the team reviewed the first 10 pages for relevant evaluations. From all of our searches, a total of 410 evaluations were included in Endnote. These were then reviewed for duplications, leaving 344 evaluations in the library.

After further review, another 165 were then removed, leaving a total of 179. These were removed because they were descriptions of programs, were a literature review, were lists of good practice determined by other organizations, were mid-term evaluations, were formative (or pre-assessment) evaluations, were not in English, were a summary of a larger document which was not available, were annual reports rather than evaluations, or did not include an explicit methodology of how the evaluation was conducted. Explicit methodologies were defined as the inclusion of a methodology section or a description of the actions taken by the evaluation team; those evaluations missing this were also excluded. Lists of good practice determined by other organizations were removed as these had inconsistent or contradictory criteria from the work being conducted of what constituted a good practice. Systematic reviews of evaluations, or reviews of reviews were also excluded from the final database. However, systematic reviews, literature reviews, and annual reports were used to inform the drafting of individual policy papers where interesting lessons could be drawn.

Appendix B: Inclusion criteria

(1) Is the report or website an evaluation of a program or activity?

An evaluation measures progress toward outputs, or change in outcomes, or an assessment of impact, of a development program, policy or intervention.

An international development program or project is a set of activities which deliver outputs and outcomes that promote social change and behavior change related to modern slavery or related social issues.

An activity is a singular action taken to address modern slavery, which could be part of a program, project or a standalone activity. Examples include giving training, conducting inspections, distributing leaflets, providing technical support to the government.

YES- Go to Q2

NO- Exclude from Endnote

(2) Does the evaluation reference a form of modern slavery?

Modern slavery defined as:

-Human trafficking;

-Worst forms of child labor;

-Slavery;

-Forced labor;

-Domestic servitude;

-Debt bondage;

-Forced, servile or early marriage;

-Sale or exploitation of children; and

-Use of child soldiers.

YES- Include in Endnote

NO- Go to Q3

(3) Does the evaluation reference a related social issue?

Similar social issue includes child labor, prostitution, refugees, internally displaced persons, female genital mutilation, safe migration, and labor migration. These are included as individuals who experience these issues may become vulnerable to being enslaved, thus successful programs may become relevant as prevention activities.

YES- Include in Endnote

NO- Exclude from Endnote