

Article

# Developing Evaluation Approaches for an Anti-Human Trafficking Housing Program

Rebecca J. Macy<sup>1</sup> , Amanda Eckhardt<sup>2</sup>,  
Christopher J. Wretman<sup>1,3</sup>, Ran Hu<sup>2,4</sup>,  
Jeongsuk Kim<sup>1</sup>, Xinyi Wang<sup>2</sup>,  
and Cindy Bombeck<sup>2</sup>

American Journal of Evaluation

2022, Vol. 43(4) 539-558

© The Author(s) 2021



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)

DOI: 10.1177/10982140211056913

[journals.sagepub.com/home/aje](https://journals.sagepub.com/home/aje)



## Abstract

The increasing number of anti-trafficking organizations and funding for anti-trafficking services have greatly out-paced evaluative efforts resulting in critical knowledge gaps, which have been underscored by recent recommendations for the development of greater evaluation capacity in the anti-trafficking field. In response to these calls, this paper reports on the development and feasibility testing of an evaluation protocol to generate practice-based evidence for an anti-trafficking transitional housing program. Guided by formative evaluation and evaluability frameworks, our practitioner-researcher team had two aims: (1) develop an evaluation protocol, and (2) test the protocol with a feasibility trial. To the best of our knowledge, this is one of only a few reports concerning anti-trafficking housing program evaluations, particularly one with many foreign-national survivors as evaluation participants. In addition to presenting evaluation findings, the team herein documented decisions and strategies related to conceptualizing, designing, and conducting the evaluation to offer approaches for future evaluations.

## Keywords

anti-trafficking, evaluability, evaluation protocol, formative evaluation, housing, human trafficking, participatory evaluation, practice-based evidence

Human trafficking (HT) for labor and/or sexual exploitation is associated with an array of detrimental consequences for trafficked persons (Dell et al., 2019; Muftić & Finn, 2013). Accordingly, trafficking survivors often benefit from comprehensive services to address legal and immigration needs, ensure

<sup>1</sup> School of Social Work, University of North Carolina at Chapel Hill

<sup>2</sup> Restore NYC, New York, New York

<sup>3</sup> The Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill

<sup>4</sup> Factor-Inwentash Faculty of Social Work, University of Toronto

## Corresponding Author:

Rebecca J. Macy, University of North Carolina at Chapel Hill, School of Social Work, 325 Pittsboro Street, CB #3550, Chapel Hill, NC 27599, USA.

Email: [rjmacy@email.unc.edu](mailto:rjmacy@email.unc.edu)

their safe and permanent housing, establish fair and living-wage employment, and promote their physical and mental health (Kaufman & Crawford, 2011; Macy & Johns, 2011). To address survivors' needs, growing numbers of non-governmental organizations (NGO) are providing services to promote survivors' safety and well-being, as well as prevent their re-trafficking. Governmental and private funders have made substantial expenditures to support such anti-trafficking services. For example, in 2020, the United States (US) Department of Justice (DOJ) awarded more than \$100 million to initiatives and programs focused on HT (DOJ Office of Public Affairs, 2020). Nonetheless, the increasing number of such organizations and related funding have greatly out-paced evaluative efforts resulting in critical knowledge gaps concerning anti-trafficking services, which have been underscored by recent recommendations for the development of greater evaluation capacity in the anti-trafficking field. (Davy, 2015; Dell et al., 2019; Krieger et al., 2020; Okech et al., 2018). In response to these calls, this paper seeks to help address gaps by reporting on the development and feasibility testing of an evaluation protocol to generate practice-based evidence concerning a transitional housing program for trafficking survivors.

## **Evaluation of Anti-Trafficking Services**

Despite the gaps noted above, evaluations of anti-trafficking services have been undertaken and presented in both reports and in the peer-reviewed literature. Although the types of programs evaluated, methods and data used, and targeted outcomes have varied considerably, findings from this prior work provide a valuable context in which the current effort was conducted.

HT program evaluations have included, as a few examples, those related to arrest diversion (Roe-Sepowitz et al., 2014), comprehensive case management (Goździak & Lowell, 2016; Potocky, 2010), comprehensive initiatives (Caliber, 2007), shelter-based interventions (Crawford & Kaufman, 2008; Munsey et al., 2018), mentoring (Criswell, 2014; Rothman et al., 2020), rehabilitation (Aborisade & Aderinto, 2008; Honeycutt, 2012), trauma-focused cognitive behavioral therapy (Bass et al., 2011; O'Callaghan et al., 2013), and services tailored to trafficked minors (Gibbs et al., 2014).

Methods have included analyses of administrative data (Goździak & Lowell, 2016; Roe-Sepowitz et al., 2014), case and chart reviews (Aborisade & Aderinto, 2008; Aderinto & Aborisade, 2008; Caliber, 2007; Crawford & Kaufman, 2008; Potocky, 2010), evaluability assessments (Caliber, 2007), field work (Goździak & Lowell, 2016), qualitative interviews with key informants and survivors (Aborisade & Aderinto, 2008; Aderinto & Aborisade, 2008; Potocky, 2010; Roe-Sepowitz et al., 2014), participatory process evaluation (Gibbs et al., 2014), pre-posttest surveys (Bass et al., 2011; Criswell, 2014), longitudinal data tracking over the course of service provision (Munsey et al., 2018; Rothman et al., 2020), and randomized studies (O'Callaghan et al., 2013).

Outcomes targeted for change have included survivors' adjustment patterns (Aderinto & Aborisade, 2008), re-victimization (Rothman et al., 2020), arrest and re-arrest (Roe-Sepowitz et al., 2014; Rothman et al., 2020), community re-integration (Honeycutt, 2012), service goal attainment (Potocky, 2010), physical and mental health (Bass et al., 2011; Crawford & Kaufman, 2008; Honeycutt, 2012; Munsey et al., 2018; O'Callaghan et al., 2013), social support and coping skills (Rothman et al., 2020), social networks (Criswell, 2014), and vocational changes (Criswell, 2014).

Although these prior evaluation efforts constitute a beneficial foundation, critical evaluation evidence gaps persist, with many of the HT services currently being delivered remaining uninvestigated (Davy, 2015; Krieger et al., 2020; Okech et al., 2018). The consequences of delivering untested services can be benign but can also have negative consequences if well-meaning service providers unintentionally do more harm than good in their efforts to support survivors (Kaufman & Crawford, 2011). Even benign services can cause harms when scarce resources are spent on ineffective programs. For these reasons, increased efforts to evaluate services for survivors of all ages, genders,

cultural backgrounds, social circumstances, and trafficking types are needed (Davy, 2015; Dell et al., 2019; Felner & DuBois, 2017).

Nonetheless, evaluations of anti-trafficking services are hindered by several related factors (Krieger et al., 2020). First, service *settings* that are frequently complex and dynamic. Second, service *providers* who are often overworked, lack financial resources and/or institutional support for evaluation, and have little guidance regarding the collection and treatment of data. Third, service *users* who may be reasonably distrustful of macro-level institutions, as well as in urgent need of resources to promote their safety and well-being. Fourth, service *funders* who are heterogeneous, and only recently expressing an interest in evaluation-derived outcomes. For all these reasons, a key part of the evaluation challenge is in the nature of the evaluation work itself.

One important approach toward helping to address this challenge, especially as HT services proliferate, is an emphasis on *practice-based evidence*, which is developed through research with NGO's (Green, 2006; Nnawulezi et al., 2018a) and often generated using community-based and participatory research strategies (Ammerman et al., 2014). While practice-based evidence is developed systematically and rigorously with attention to internal validity, it also emphasizes ecological validity by attending to the complex, dynamic contexts in which service provision occurs. Thus, informed by practice-based evidence approaches, evaluators can develop methods for investigating anti-trafficking programs.

## Restore NYC and Transitional Housing Services

Restore NYC, a 501(c)(3) NGO organization in New York, New York, is a service provider for HT survivors in the United States (US). Conducting service provision since 2009, Restore NYC focuses on serving adult, self-identified women who have been trafficked into the greater New York city area. Since its inception, Restore NYC has served over 2,000 HT survivors—and those at risk of trafficking—from over 80 countries, with approximately 200 HT survivors served annually in its programs of housing, economic empowerment, counseling, and case management (Restore NYC, 2019). To help ensure that programs work as intended, Restore NYC has developed an ongoing commitment to evaluation.

One of Restore NYC's primary targets for evaluation is its transitional housing program. Residents meet the federal criteria for trafficking and are referred by self, attorneys who represent HT survivors, law enforcement (e.g., Homeland Security, the Federal Bureau of Investigations), and other not-for-profit organizations. Once referred, potential residents are screened for eligibility and assessed via a comprehensive intake process led by the Restore NYC housing manager using a low-barrier framework. That is, survivors who might be excluded from traditional housing services due to challenges such as substance abuse are not turned away (Nnawulezi et al., 2018b). At program entry, program staff work with each resident to create an individualized service plan to address immediate, ongoing, and long-term needs based on recommended practices (e.g., Macy & Johns, 2011). Plan development is guided by a detailed program theory of change, logic model, and extensive training for staff. In addition to housing, residents receive comprehensive and coordinated case management that includes connecting survivors to services outside of the housing program, both at Restore NYC and in the greater community, based on their individualized goals and needs. Examples of such services include (a) job readiness training and placement, (b) English as a second language instruction, (c) immigration legal services, (d) medical services, and (e) mental wellbeing services. Residents typically live in the housing program and receive services for 12 months, with stays of up to 18 months as needed on an individual basis.

Evaluation of the housing program has been an organizational-level priority for three reasons. First, transitional housing is a cornerstone of the organization, both philosophically and fiscally,

comprising upward of 20% of total annual expenses in recent years (Restore NYC, 2019). Second, the resources required to support the delivery of transitional housing, given its intensive nature, are significant and thus potentially wasteful without demonstrated program helpfulness. Third, despite laudable efforts to investigate housing programs for survivors of trauma, violence, and/or trafficking, there is a documented dearth of rigorous evidence (i.e., experimental, quasi-experimental) concerning the outcomes of such programs (Klein et al., 2021).

Despite a strong desire to evaluate this program, Restore NYC was limited in its evaluation capacity. This changed in 2014 when Restore NYC connected with a team at the School of Social Work at the University of North Carolina at Chapel Hill (UNC-CH). This partnership combined Restore NYC's significant on-the-ground practice expertise with a university-based research team experienced in conducting community-engaged and trauma-informed evaluations with violence survivors, which led to the development of the practitioner-research team described herein.

## Aims and Research Questions

Guided by formative evaluation and evaluability frameworks (e.g., Bowen et al., 2009; Dehar, Casswell, & Duignan, 1993; Leviton et al., 2010; Orsmond & Cohn, 2015; O' Cathain et al., 2019; Trevisan & Walsler, 2014) the practitioner-researcher team held two broad aims. First, the team sought to develop a comprehensive *evaluation protocol* with guiding principles for evaluation, outcome measures, as well as data collection, management, and analysis procedures. Second, the team sought to test the practicability of the evaluation protocol with a *feasibility trial* as well as to generate formative, practice-based evidence about the housing program.

In addition to these aims, the project was guided by three research questions (RQ). RQ1: Can an evaluation protocol be developed for this type of HT service provision, particularly considering the complexity inherent in the service setting, service delivery, and service users' needs? RQ2: Can a feasibility evaluation trial be practically and successfully implemented among a sample of survivor service users? RQ3: Can results inform practice with HT survivors in this program and setting?

The development and feasibility testing of a formal protocol was viewed by the entire team as a necessary step towards codifying the team's ongoing evaluative efforts overall. Moreover, considering the challenges the anti-trafficking field faces, documenting the decisions and strategies related to conceptualizing, designing, and conducting the evaluation was theorized as potentially helpful for the anti-trafficking field broadly, particularly considering current recommendations from the field to develop greater evaluation capacity (Davy, 2015; Dell et al., 2019; Krieger et al., 2020; Okech et al., 2018).

## Aim One Methods—Develop an Evaluation Protocol

For the first project aim, two activities were initially undertaken to develop the evaluation protocol. First, the team conducted a literature review of relevant research. Second, the project sought to solicit feedback from Restore NYC's administrators, staff, and survivors. These development activities occurred from 2014 to 2018, with the primary work being conducted from 2016 onwards. This latter stage was funded by an award through the Office for Victims of Crime, Office of Justice Programs, U.S. Department of Justice. All research procedures were reviewed by the Institutional Review Board (IRB) of the Office of Research Ethics at the UNC-CH (IRB #17-422; IRB #19-2744).

### Literature Reviews

The literature reviews were comprised of (a) a narrative review of substantive and methodological research related to HT evaluation, and (b) a systematic review of substantive research concerning

measures for evaluating anti-trafficking services. The narrative review focused on determining key evaluation practices, while the systematic review focused on key targets (i.e., evaluation outcomes).

The narrative review included two general domains. First, substantive research was surveyed related to HT evaluation research including those such as (a) Davy's review of existing HT program evaluations (2015), (b) Macy's and John's overview of trafficking services and programs (2011), and (c) Zimmerman and colleagues' conceptual model on HT (2011). Second, methodological research was surveyed related to social services evaluation including articles by (a) Shulha and colleagues' principles to guide collaborative evaluations (2016), (b) Bowen and colleagues' suggestions for designing feasibility studies (2009), and (c) Shadish and colleagues' technical information regarding quasi-experimental designs (2002). Research in this review was selected based on the expertise of all team members. All articles were scrutinized by the team, with key findings noted to inform the evaluation protocol.

Details concerning the systematic substantive review component can be found in the article summarizing the review itself (Graham et al., 2019). Briefly here, the search involved two strategies: (a) a systematic search of electronic databases housing peer-reviewed journal articles and (b) reference harvesting of identified studies. Review inclusion criteria were having (a) been published in English in a peer-reviewed journal, (b) included data collection pertinent to needs and/or outcomes of trafficking survivors, (c) acknowledged that some portion of their sample included trafficking survivors, and (d) included specific details regarding constructs (i.e., variables) and/or measures (e.g., instruments). Articles were systematically searched, tracked, and reviewed in accordance with the preferred reporting items for systematic reviews and meta-analyses statement (Moher et al., 2009). In all, 22 (41.5%) articles identified a total of 34 specific published measures relevant for outcomes evaluations of anti-trafficking programs.

### *Administrative Leadership, Staff, and Survivor Feedback*

The second aim of the evaluation protocol was to gather feedback from key groups, including Restore NYC administrative leadership, as well as staff and survivor residents in the housing program. The team conducted guided feedback sessions facilitated by the research team lead (first author) with coordination by the practice team lead (second author). Prior to each feedback session, the team worked collaboratively to determine specific questions of interest, including questions about appropriateness and feasibility of data collection, potential outcomes and measures, and means to ensure survivor input, confidentiality, and safety in the evaluation.

During the evaluation protocol development, seven such sessions in total were conducted. As noted above, feedback sessions included three types of personnel individually and collectively: (a) survivors who were current residents in the housing program at the time the sessions were held ( $N \approx 15$ ), (b) administrative leadership (e.g., Executive Director;  $N \approx 5$ ), (b) housing program staff (e.g., Housing Director;  $N \approx 12$ ). In addition to the two feedback sessions with survivor residents, due to Restore NYC's purposive employment and survivor empowerment strategies, members of the staff included individuals with lived experience of human trafficking, though the identity of these staff was blinded to university researchers. Detailed notes were taken at each session, which were then used to guide evaluation protocol development.

In addition to these efforts to ensure that survivors had input into the design of the evaluation, in the implementing the feasibility trial, which is detailed subsequently, a post-assessment feedback form was implemented for survivor residents to share their experiences with the process of completing data collection assessments as well as to make suggestions for improvement. To ensure that the feasibility trial, as it was implemented, was conducted in survivor-centered ways, questions on the feedback form included "How do you feel about the length of the time taken to complete the assessments," "How did you feel at the end of the assessment," and "Did you feel all the questions

were clear?" among others. Trial procedures were refined in minor ways based on this regularly collected feedback from survivors.

## **Aim One Results—Evaluation Protocol**

Findings from the aim one activities were compiled into three distinct yet overlapping domains including (a) characteristics of the evaluation, (b) methodological best practices to guide the implementation of the protocol, and (c) targets for the evaluation. Together, the team conceptualized these results as comprising the evaluation protocol itself.

### *Evaluation Characteristics*

Four characteristics were determined to define the *internal operations* of the evaluation protocol, including (a) relationship-focused (i.e., promoting constructive working relationships in the team), (b) mutually-understanding (i.e., bridging differences between practitioners and researchers on the team), (c) participation-promoting (i.e., ensuring active involvement by all team members), and (d) progress-monitoring (i.e., ensuring that planned activities were carried out in timely ways and to completion). Likewise, four characteristics were determined to define the *external implementation* of the evaluation protocol: (a) strengths-based (i.e., with an emphasis on survivors' capacities and resources), (b) trauma-informed (i.e., with an understanding of how violence and victimization might impact survivors' lives, including their research participation; Elliot et al., 2005), (c) culturally-sensitive (i.e., research approaches that are responsive to all survivors no matter their background, languages, and situations), and (d) survivor-centered (i.e., research approaches that are carried out with survivors' perspectives and recommendations at the forefront). All eight characteristics were included in the protocol and woven into all subsequent evaluation activities and products.

### *Evaluation Practices*

Five sets of evaluation practices were determined to define the evaluation protocol. As Table 1 fully presents both the overall domains, as well as related recommendations, key examples findings are briefly presented here, including (a) measurement selection (e.g., whenever possible, select measures that have been cross-culturally validated); (b) data collection (e.g., build data collection into regular service provision); (c) data management (e.g., regularly consult among practice and research team members regarding data quality); (d) data analysis (e.g., use non-parametric tests with small or non-normally distributed data); and (e) interpretation of results (e.g., link results to specific research questions at the program level to inform service evaluation and refinement and at the survivor level to support ongoing individualized service planning). Like the evaluation characteristics noted above, all five evaluation practice domains were included in the protocol and woven into all subsequent evaluation activities and products.

### *Evaluation Targets*

Overall, the team identified six evaluation target domains, which were deemed to be potentially modifiable by the housing program given its existing theory of change, logic model, and service strategies, (a) physical and mental well-being, (b) stress and post-traumatic stress disorder (PTSD), (c) depression, (d) coping, (e) post-traumatic growth, and (f) employment and housing. Likewise, these key constructs have been previously used in both typical anti-trafficking service provision and in prior evaluations (Graham et al., 2019). In addition, as described above, the review determined that 34 identified measures have been used in extant evaluations of anti-trafficking services (Graham et al., 2019). Accordingly, while considering measures that had been previously used in HT research

**Table 1.** Anti-Trafficking Program Evaluation Recommendations—Activities and Practices.

Activities	Recommended Practices
<b>Measurement</b>	<p>Assess constructs with multi-item measures (i.e., scales) capturing holistic conceptualizations of constructs (i.e., latent variables)</p> <p>Select measures (1) previously used in trafficking and/or violence services evaluation research and (2) validated by via psychometric analyses</p> <p>Select measures with (1) total and (2) sub-scale internal consistency reliability values (e.g., Cronbach's <math>\alpha</math>) greater than 0.60</p> <p>Select measures with demonstrated construct validity</p> <p>Select measures that have been cross-culturally validated</p> <p>Match measures to pre-specified evaluation targets</p> <p>Include variables related to (1) demographic characteristics, (2) trafficking experiences, and (3) resident status</p>
<b>Data collection</b>	<p>Build data collection into regular service provision</p> <p>Collect data at (1) baseline and at (2) multiple points post-baseline (i.e., time series design)</p> <p>Ensure that survivors are given appropriate, clear, and meaningful informed consent for data collection and use</p> <p>Ensure that service receipt is not dependent on survivors' evaluation participation</p> <p>Use proximal service providers to collect data</p> <p>Address discomfort or emotional distress caused by completing assessment questions</p> <p>Translate relevant materials into appropriate languages</p>
<b>Data management</b>	<p>Regularly consult among practice and research team members regarding data quality</p> <p>Include appropriate (1) time, (2) individual, and (3) group identifiers</p> <p>Include relevant dates (e.g., program intake)</p> <p>Pre-specify (1) missingness, (2) intra-variable (i.e., out-of-range), and (3) inter-variable (i.e., non-compliance) errors</p>
<b>Data analysis</b>	<p>Build analyses from (1) univariate statistics to (2) bivariate statistics and tests to (3) multivariate tests and models</p> <p>Analyze data using both (1) paired hypotheses (e.g., longitudinal by time) and (2) unpaired hypotheses (e.g., cross-sectional by group)</p> <p>Use non-parametric tests with (1) small or (2) non-normally distributed data</p> <p>Calculate missing data by (1) variable/measure, (2) survivor, and (3) group</p> <p>Check for outlier (1) values and (2) observations</p> <p>Conduct (1) sensitivity analyses and (2) sub-group analyses</p>
<b>Interpretation</b>	<p>Report missing data characteristics</p> <p>Link results to specific research questions (1) at the program level to inform service evaluation and refinement and (2) at the survivor level to support ongoing individualized service planning</p> <p>Interpret numerical estimates in terms of (1) effect sizes, (2) relative change, and/or (3) standard deviation change</p> <p>Avoid claims of causality in absence of a randomized design</p>

as much as possible, the team selected measures for the evaluation target domains that had shown reliability and validity, had been cross-culturally validated, and that capturing holistic conceptualizations of the evaluation outcomes, whenever possible.

## Aim Two Methods—Feasibility Trial

Guided by the protocol, the team conducted the feasibility trial, which occurred from 2018 to 2019 and was conducted under the same IRB auspices as the evaluation development.

## Participants

The feasibility trial was pre-specified to be all 33 residents enrolled in Restore NYC's transitional housing program comprising survivors who entered the program from March of 2014 to October of 2019. Residents were adult, self-identifying women, and primarily those with either pre-certification or certification in accordance with the Trafficking Victims Protection Act (TVPA; 2000), which allows adult HT survivors who are not US citizens or Lawful Permanent Residents to be eligible to receive federal and state benefits to the same extent as refugees. Importantly, Restore NYC's program features (a) rolling admissions such that participants begin the program whenever they are referred and (b) varying lengths of program completion time as participants are allowed to exit at their choosing.

## Data Collection

To ensure that the program tailors services to residents' needs, transitional home staff have collected data on residents on a quarterly basis since the program began. Consequently, data collection for the feasibility trial occurred by adapting Restore NYC's existing assessments and practices. Consistent with typical program practices, assessments included (a) a comprehensive intake assessment at program entry and (b) regular 3-month follow-up assessments throughout the entirety of a survivor's time in the program, which may be up to 18 months.

All assessments were conducted by trained Restore NYC staff in private settings, with attention to confidentiality, and in residents' first language. Restore NYC has translated the assessments into multiple languages to ensure that the assessments can be conducted in survivors' preferred languages. Restore NYC also has bicultural and bilingual staff who can address most survivors' language preferences. In addition, when necessary, Restore NYC engages interpreters to assist in the administration of the assessments. Restore NYC staff report that typically the assessments were conducted with survivors in 15 to 30 min.

Prior to any assessments and data collection, Restore NYC staff review with new residents their rights via a standardized Notice of Clients' Rights document, as well discusses informed consent via a standardized Informed Consent document. Survivors are informed that they may withdraw from participating in any assessment and/or decline to complete any specific assessment questions without any consequences for their services and/or program participation.

After conducting the assessments, staff entered data into a secure, cloud-based case-management database using Apricot (Social Solutions, Austin, TX). Also, as noted earlier, a post-assessment feedback form was implemented for survivor residents to share their experiences with the assessments.

## Measures

Guided by the evaluation protocol (i.e., aim one), the trial assessments were comprised of standardized instruments to enable a holistic evaluation of residents' needs. Moreover, the assessments instruments were developed and selected to provide meaningful information concerning changes in survivors' needs, improvements, and outcomes over time at both the (a) survivor-level to inform ongoing service planning with individual survivors and (b) program-level to inform program evaluation in the six evaluation target domains (i.e., physical and mental well-being, stress and PTSD, depression, coping, post-traumatic growth, employment, and housing) that were determined in the development of the evaluation protocol.

*Resident characteristics.* The feasibility trial collected information related to residents' (a) demographic (e.g., age, gender identity), (b) immigration (i.e., documentation status), and (c) employment (i.e., current employment status) characteristics at intake/entry to the program to characterize the sample.



**Well-being.** Six validated instruments were used to collect information on well-being. Generalized well-being was assessed using the Short-Form Health Survey (SF-12; Ware et al., 1996). The SF-12 is a 12-item scale measuring well-being via two primary subscales for (a) *physical well-being* and (b) *mental well-being*, ranging from 0 to 100 for both. Within these sub-scales are eight domains comprising (a) “Physical functioning,” (b) “Role-physical,” (c) “Bodily pain,” (d) “General health,” (e) “Mental health,” (f) “Role-emotional,” (g) “Social functioning,” and (h) “Vitality.” The SF-12 Short-Form has two versions using two different recall periods—asking respondents to consider either (1) past week or (2) past month when answering questions (Ware et al., 2007). The one-month recall version is used in this evaluation.

**Stress** was measured with the Perceived Stress Scale (PSS; Cohen et al., 1983), a 14-item scale of individuals’ perception of the degree to which situations in their life are stressful. Scores range from 0 to 40, with scores  $\geq 20$  indicating a “high” level of stress.

**Post-traumatic stress** was assessed with a measure pegged to the Diagnostic and Statistical Manual of Mental Disorders (DSM) titled the PTSD Checklist for DSM-5 (PCL-5; Blevins et al., 2015). This is a 20-item scale of post-traumatic stress with a score range of 0 to 80. A score  $\geq 38$  indicates the potential for a formal diagnosis of PTSD.

**Depression** was assessed using the Center for Epidemiologic Studies Depression Scale (CES-D; Eaton et al., 1977; Radloff, 1977). This is a 20-item scale of items pegged to the DSM-5, which measures depression in the past week with scores ranging from 0 to 60. A score  $\geq 16$  indicates a potential clinical level of depression.

**Coping** was assessed using the Brief COPE (Carver, 1997), a 28-item scale of coping styles comprising two broad subscales for adaptive (e.g., reframing) and maladaptive (e.g., denial) coping strategies. Within these subscales are 14 domains comprising (a) “Active coping,” (b) “Emotional support,” (c) “Instrumental support,” (d) “Positive reframing,” (e) “Planning,” (f) “Humor,” (g) “Acceptance,” (h) “Religion,” (i) “Self-distraction,” (j) “Denial,” (k) “Disengagement,” (l) “Venting,” and (m) “Self-blame.”

**Post-traumatic growth** was assessed using the Post-Traumatic Growth Inventory (PTGI; Tedeschi & Calhoun, 1996) to assess survivors’ growth in response to experienced trauma. The PTGI is a 21-item scale measuring positive changes experienced in the aftermath of a traumatic event, such as trafficking, with a total range from 0 to 105. Within the total measure are five subscales comprising (a) “Appreciation of life,” (b) “New possibilities,” (c) “Personal strength,” (d) “Spiritual change,” and (e) “Relating to others.”

**Employment and housing.** Two categorical measures were used assess changes in survivors’ employment and housing, including (a) a measure for *employment* (e.g., “Are you currently employed?” (No = 0, Yes = 1) and (b) a measure for *housing* (e.g., “What is your current housing status?” Homeless = 0, Supportive housing = 1, Community dwelling = 3). Both measures were identified by Restore NYC as being feasible to collect and potentially subject to change by the housing program.

## Data Analysis

The outcome analysis was focused on descriptive univariate characterization of residents’ outcomes at intake and at the 3-, 6-, 9-, 12-month intervals and at case closing when a resident exits the home (i.e., at about 18 months), as well as calculation of outcome change over time. Change was calculated as relative change percentage ( $\Delta\%$ ) using the formula of  $(\text{post-intake score} - \text{intake score}) / \text{intake score} \times 100$  for each post-intake interval. These comparisons were conducted on pooled resident data for 3-, 6-, 9- and 12-months compared with pooled intake data, and for case closing only on the subset of residents with both intake and closing data only. Relative change, which assessed by

what percentage a score changed from an original score, was used in the analysis so that resident change across the 35 scale and subscale measures could be standardized given the variability in the measures' ranges.

Also, as the above scales have not been widely used with HT survivors, period-based internal consistency reliability estimates were calculated using Guttman's  $\lambda_2$ . Estimates of  $\lambda_2 \geq 0.70$  were sought to investigate reliability of the scales within this population. Due to the small size of the sample of residents, bivariate and/or multivariate tests (e.g., McNemar's tests, multilevel growth curve models) comparing residents' outcomes across time were not conducted in this feasibility trial, resulting in no tests of statistical significance. All analyses were conducted by the research team using Stata 16.1 (StataCorp, College Station, TX).

## Aim Two Results—Feasibility Trial

### *Participant Characteristics*

Table 2 shows characteristics related to survivor characteristics at program intake. The residents were all self-identifying women, as was consistent with the program model, with a mean age of almost 40 years old. There were 23 different countries of origin, mostly east Asia and Latin America, with 18 different languages. In terms of English language skills, almost two-thirds (64%) had limited or intermediate ability. At program entry, only about 19% had an immigration status that enabled them to work and/or receive federal benefits (e.g., human trafficking visa, lawful permanent resident status). In addition, over 90% were working less than full-time with 75% being unemployed. In terms of Victims of Trafficking and Violence Protection Act (TVPA) certification<sup>1</sup>, almost were certified, pre-certified, or provisionally certified (79%).

### *Well-Being*

Extant counts of the well-being outcomes varied at intake as Restore NYC does not require resident data to be collected to deliver services. Among all 32 residents with any intake data, counts of well-being intake data included 22 (69%) for the PCL-5, 23 (72%) for the SF-12, 29 (91%) for the PTGI, 30 (94%) for the CES-D and PSS, and 32 (100%) for the Brief COPE. Well-being data counts at post-intake intervals tended higher, with counts of 22–24 (88–96%) among the 25 residents with 3-month data, 23–24 (96–100%) among the 24 residents with 6-month data, 22–24 (92–100%) among the 24 residents with 9-month data, and 21 (100%) among the 21 residents with 12-month data. In total, 24 residents had any closing data, with actual well-being observations ranging from 16 to 18 counts across all variables.

Residents' mean well-being scores and relative change percentages are presented in Table 3. The primary comparisons are at the 3-, 6-, 9-, and 12-month periods where pooled resident data is compared, with a secondary comparison at case closing on matched data only.

For the SF-12, residents demonstrated consistent and desired improvements on their "General health" ( $\geq +6\%$ ), "Mental total" ( $\geq +6\%$ ), "Mental health" ( $\geq 5\%$ ), and "Social functioning" ( $\geq +2\%$ ) scores at 3-, 6-, 9-, and 12-months. Internal consistency for nine of the 10 SF-12 measures demonstrated acceptability across intake and the four primary time periods ( $\lambda_2$  range: 0.78–0.91), with only "Bodily pain" being low ( $\lambda_2 = 0.59$ ).

For the PSS, all post-intake scores were consistently in the desired direction ( $\leq -6\%$ ), and the internal consistency of the measure across 3-, 6-, 9-, and 12-months was good ( $\lambda_2 = 0.84$ ).

For the PCL-5, all post-intake scores were consistently negative ( $\leq -21\%$ ) with good internal consistency across the same period ( $\lambda_2 = 0.82$ ).

The CES-D did not demonstrate consistently desired change, but residents' mean scores at 6-months ( $-17\%$ ) and 12-months ( $-14\%$ ) were notably highly negative relative to baseline,

**Table 2.** Anti-Human Trafficking Housing Program Evaluation—Survivor Intake Demographic Characteristics (N = 33).

Characteristics	Mean (SD) or N (%)
<b>Age at intake</b>	39.5 (14.3)
<b>First language</b>	
Spanish	10 (30.3)
English	4 (12.1)
Russian	3 (9.1)
Others	16 (48.5)
<b>English language proficiency</b>	
Limited	12 (36.4)
Intermediate	9 (27.3)
Advanced	12 (36.4)
<b>Immigration status</b>	
Undocumented	9 (28.1)
Unknown	6 (18.8)
Out of status	5 (15.6)
Human trafficking visa	4 (12.5)
Continued presence	2 (6.3)
Other	6 (18.8)
<b>Employment status</b>	
Unemployed	24 (75.0)
Part-time	3 (9.4)
Hourly	2 (6.3)
Full-time	2 (6.3)
Unknown	1 (3.1)
<b>Trafficking identification status</b>	
Pre-certified TVPA	15 (51.7)
Certified TVPA	5 (17.2)
Provisional TVPA	3 (10.3)
Other	6 (20.7)

Note. TVPA = Trafficking Victims Protection Act. Mean and SD calculated for continuous age variable only. Resident sample size ranges from 31–33 due to missing data.

indicative of strong decreases in depression. In addition, the CES-D change at closing was notably strong ( $-29\%$ ). The internal consistency of the CES-D across the four primary time periods was acceptable ( $\lambda_2 = 0.76$ ). For the Brief COPE, residents demonstrated consistent and desired improvements on their “Emotional support” ( $\geq +6\%$ ), “Self-distraction” ( $\geq -6\%$ ), “Denial” ( $\geq -7\%$ ), and “Disengagement” ( $\geq -2\%$ ) scores.

Internal consistency for 13 of the 16 Brief COPE measures demonstrated acceptability across time ( $\lambda_2$  range: 0.70–0.92), with only “Planning” ( $\lambda_2 = 0.66$ ), “Self-distraction” ( $\lambda_2 = 0.65$ ), and “Venting” ( $\lambda_2 = 0.56$ ) being low.

Finally, for the PTGI, residents demonstrated consistent and desired improvements on their “Personal strength” ( $\geq +3\%$ ) and “Spiritual change” ( $\geq +3\%$ ) scores. Internal consistency for all six PTGI measures demonstrated acceptability across time periods ( $\lambda_2$  range: 0.83–0.93).

Also, in Table 3, any change at any time post-intake that was found to be  $\geq 5.0\%$  in a desirable direction has been shaded in grey. These changes represent potential small-to-large increases to residents’ outcomes over the course of their participation in the program. Across all 35 measures and 4 post-intake intervals, 63 (45%) of the  $35 \times 4 = 140$  possible primary post-intake scores were found to be in the desired direction and of a  $\geq 5.0\%$  relative change. Changes that were  $\geq 5.0\%$  and

**Table 3.** Anti-Human Trafficking Housing Program Evaluation—Preliminary Survivor Well-Being Outcomes, by Measure and Period.

Measures (Desired direction, Range)	Intake	3 months	6 months	9 months	12 months	Closing <sup>a</sup>
	N = 32 Mean	N = 25 Mean, Δ	N = 24 Mean, Δ	N = 24 Mean, Δ	N = 21 Mean, Δ	N = 18 Mean, Δ
<b>SF-12 (+, 0–100)</b>						
Physical total	53.5	52.4, – 2.0%	51.4, – 4.0%	54.4, + 1.7%	52.2, – 2.4%	52.7, – 0.6%
Physical functioning	50.7	50.0, – 1.4%	52.5, + 3.6%	51.3, + 1.3%	51.6, + 1.8%	52.5, – 0.1%
Role-physical	49.2	50.4, + 2.4%	50.5, + 2.6%	49.5, + 0.6%	47.2, – 4.2%	49.3, – 1.1%
Bodily pain	48.1	47.0, – 2.2%	45.4, – 5.7%	50.9, + 5.8%	46.9, – 2.5%	50.0, + 4.6%
General health	50.2	53.6, + 6.7%	53.0, + 5.6%	53.6, + 6.8%	53.6, + 6.7%	54.1, + 9.1%
Mental total	40.1	42.9, + 6.9%	42.8, + 6.7%	42.6, + 6.1%	43.4, + 8.2%	48.4, + 20.6%
Mental health	40.4	42.7, + 5.7%	42.6, + 5.4%	44.5, + 10.2%	45.0, + 11.6%	48.8, + 20.2%
Role-emotional	41.3	43.5, + 5.3%	41.7, + 0.9%	41.3, – 0.2%	41.9, + 1.4%	45.2, + 5.2%
Social functioning	41.8	46.6, + 11.3%	46.3, + 10.6%	43.6, + 4.1%	42.6, + 1.9%	50.0, + 7.6%
Vitality	51.1	55.4, + 8.4%	49.7, – 2.8%	54.6, + 6.8%	53.0, + 3.7%	55.6, + 10.8%
<b>PCL-5 total (–, 0–40)</b>	18.6	18.2, – 2.1%	17.5, – 6.1%	17.7, – 4.8%	17.4, – 6.3%	14.2, – 25.1%
<b>PCL-5 total (–, 0–80)</b>	30.2	26.2, – 13.3%	27.8, – 7.8%	29.1, – 3.6%	23.7, – 21.4%	16.9, – 44.7%
<b>CES-D total (–, 0–60)</b>	23.1	22.7, – 1.6%	19.1, – 17.1%	23.7, + 2.9%	19.8, – 14.1%	17.4, – 28.9%
<b>Brief COPE</b>						
Adaptive total (+, 16–64)	43.5	45.2, + 3.9%	46.5, + 6.8%	44.3, + 1.7%	42.2, – 3.0%	46.3, + 8.1%
Active coping (2–8)	6.2	6.0, – 3.5%	6.2, – 0.2%	6.0, – 4.2%	5.9, – 5.0%	6.3, + 9.2%
Emotional support (2–8)	4.8	5.5, + 15.5%	5.5, + 15.9%	5.1, + 6.4%	5.1, + 6.6%	5.7, + 13.3%
Instrumental support (2–8)	5.6	5.6, + 0.0%	6.2, + 11.6%	6.0, + 7.9%	5.3, – 4.1%	6.0, + 16.9%
Positive reframing (2–8)	5.8	5.8, – 0.3%	6.0, + 2.0%	5.8, – 0.3%	5.3, – 9.5%	6.1, + 8.2%
Planning (2–8)	6.1	6.4, + 4.3%	6.8, + 10.2%	6.1, – 0.6%	5.9, – 3.6%	6.9, + 15.6%
Humor (2–8)	3.4	3.5, + 4.0%	3.2, – 4.0%	3.1, – 7.7%	3.0, – 10.3%	2.9, – 18.5%
Acceptance (2–8)	5.8	6.3, + 8.3%	6.3, + 8.8%	5.9, + 2.3%	5.9, + 1.3%	6.2, + 13.4%
Religion (2–8)	5.9	6.2, + 5.1%	6.3, + 7.8%	6.3, + 7.3%	5.9, – 0.3%	6.1, – 1.1%
<i>Maladaptive total (–, 12–48)</i>	27.5	28.1, + 2.3%	25.8, – 6.2%	24.9, – 9.6%	23.3, – 15.2%	21.9, – 22.2%
Self-distraction (2–8)	5.8	5.3, – 9.0%	5.2, – 10.6%	5.2, – 10.5%	4.9, – 15.2%	5.7, – 5.6%
Denial (2–8)	4.6	4.3, – 6.6%	3.8, – 17.8%	3.7, – 18.0%	3.4, – 24.9%	2.7, – 40.3%
Substance Use (2–8)	3.8	3.9, + 1.8%	3.3, – 13.3%	3.0, – 20.8%	3.1, – 19.5%	2.7, – 36.9%
Disengagement (2–8)	4.4	4.3, – 3.0%	4.3, – 2.3%	3.9, – 11.8%	3.7, – 17.4%	3.1, – 30.9%
Venting (2–8)	4.5	5.4, + 20.6%	4.5, + 0.7%	4.7, + 6.1%	4.2, – 5.2%	4.3, – 7.1%
Self-blame (2–8)	4.4	5.0, + 13.5%	4.7, + 6.9%	4.3, – 3.3%	4.0, – 9.2%	3.4, – 19.0%

(continued)

**Table 3.** Continued.

Measures (Desired direction, Range)	Intake	3 months		6 months		9 months		12 months		Closing <sup>a</sup>
	N = 32 Mean	N = 25 Mean, Δ	N = 24 Mean, Δ	N = 24 Mean, Δ	N = 24 Mean, Δ	N = 21 Mean, Δ	N = 18 Mean, Δ			
<b>PTGI (+)</b>										
Total (0–105)	74.8	75.4, + 0.8%	76.9, + 2.8%	74.3, – 0.5%	<b>79.4, + 6.2%</b>	71.3, + 0.5%				
Appreciation of life (0–15)	10.7	11.0, + 3.0%	11.0, + 2.6%	10.4, – 3.1%	<b>11.6, + 7.9%</b>	10.2, – 0.6%				
New possibilities (0–25)	18.3	19.1, + 4.0%	18.9, + 3.1%	18.4, + 0.5%	<b>19.3, + 5.4%</b>	17.5, 0.0%				
Personal strength (0–20)	14.0	14.5, + 3.6%	<b>15.0, + 7.2%</b>	14.4, + 2.5%	<b>15.9, + 13.3%</b>	<b>14.7, + 12.8%</b>				
Spiritual change (0–10)	7.1	7.3, + 2.6%	<b>8.1, + 14.4%</b>	7.3, + 2.7%	<b>7.5, + 6.4%</b>	7.3, + 2.8%				
Relating to others (0–35)	24.6	23.5, – 4.6%	23.8, – 3.1%	23.9, – 2.9%	25.0, + 1.9%	21.7, – 6.3%				

Note. Change (Δ) represents relative change from intake; Values ≥ 5.0% in desired direction bolded. SF-12 = The Short-Form Health Survey; PSS = Perceived Stress Scale; PCL-5 = PTSD Checklist for DSM-5; CES-D = Center for Epidemiologic Studies Depression Scale; PTGI = Post-Traumatic Growth Inventory. <sup>a</sup> Closing-based relative change compares only the subset of residents with both intake and closing data; Closing time varied from 3 to 18 months depending on the individual resident.

**Table 4.** Anti-Human Trafficking Housing Program Evaluation—Preliminary Survivor Employment and Housing Outcomes, by Measure and Period.

Measures	Intake N = 32 n (%)	3 months N = 28 n (%)	6 months N = 25 n (%)	9 months N = 24 n (%)	12 months N = 21 n (%)	Closing N = 24 n (%)
<b>Employed</b>	10 (31.3)	9 (32.1)	17 (68.0)	17 (70.8)	14 (66.7)	20 (83.3)
<b>Housing</b>						7 (31.8)
Homeless	2 (11.1)					
Supportive	14 (77.8)	12 (85.7)	7 (70.0)	12 (100.0)	10 (58.8)	9 (39.1)
Community dwelling	2 (11.8)	2 (14.3)	3 (30.0)		7 (41.2)	14 (60.9)

in an undesirable direction (not pictured), were found to represent just 9 (6%) of all scores. Secondly, the case closing comparison among the reduced subset of residents with paired intake and closing data found 24 (68.6%) of the 35 outcomes to be in the desired direction with a  $\geq 5.0\%$  relative change.

### *Economic Empowerment*

Table 4 shows that residents' employment status improved over time increasing from 31% of residents having employment at intake to 67% having employment after 12 months and 83% at closing. In terms of housing, while nearly 59% were in some form of supportive housing at 12 months, 41% were in a community dwelling, with none being homeless. At closing, 61% were in a community dwelling, representing a 49% absolute increase from intake.

### **Discussion**

Despite the growth of anti-trafficking services, increasing funding for such services, and growing evaluation efforts, the anti-trafficking field is challenged with a dearth of high-quality program evaluations (Davy, 2015; Dell et al., 2019; Krieger et al., 2020; Okech et al., 2018). Consequently, the field faces significant knowledge gaps, including whether services are working as intended and whether services are making positive differences for survivors. Such a situation is worrisome given the potential for even well-intentioned programs to do harm, and the potential of wasted resources if programs do not work as intended. Nonetheless, the lack of evaluations on anti-trafficking programs is understandable given the considerable challenges such efforts face, including the overall lack of guidance regarding how programs can best be evaluated (Krieger et al., 2020). With the goal of helping to respond to calls for developing evaluation capacity in the anti-trafficking field, this paper presented our team's efforts to develop (aim one) and feasibility test (aim two) a program evaluation protocol, including our decisions and strategies in conceptualizing, designing, and conducting the evaluation.

### *Limitations*

Before reflecting on the research questions that guided our efforts, we encourage readers to be mindful of the study's limitations. Like many other anti-trafficking programs, Restore NYC's aims to connect survivors residing in the housing program to key services (e.g., English instruction, job readiness training, legal services, physical and mental health services) via comprehensive case management. Nevertheless, this effort represents one evaluation of one program. Given the heterogeneity of anti-trafficking programs, the protocol, evaluation strategies, and recommendations

generated here may have limited implications for other programs, communities, or contexts. Also, the data collected herein is unique to this program and has important limitations due to the varying enrollment and participation periods of participants. Likewise, the findings may have been shaped by the practitioner-researcher team that came together for this project. Even if the findings presented here do not translate readily to other settings, it is our hope that the protocol and recommendations presented here, including those in Table 1, offer helpful strategies to future evaluations. We also note here that this evaluation protocol did not seek to conduct longitudinal follow-up after survivors exited the housing program, which is a limit of our current effort and a recommendation for future ones.

### *Evaluation Protocol—Discussion and Implications*

In response to our first research question (RQ1: Can an evaluation protocol be developed for an anti-trafficking housing program?), our team found that we were able to develop a meaningful protocol (i.e., aim one), which was guided by the activities and recommendations summarized in Table 1.

Here we would like to note that the literature review efforts greatly informed our protocol. The extant evaluation literature—particularly the guidance on community-engaged, participatory, and formative evaluations—helpfully guided the protocol's development (e.g., Bowen et al., 2009; Leviton et al., 2010; Orsmond & Cohn, 2015). Likewise, the emerging evaluation literature on anti-trafficking programs to-date was helpful (e.g., Davy, 2015; Gibbs et al., 2014; O'Callaghan et al., 2013). In particular, the review findings underscored that practitioner-researcher evaluation partnership should be relationship-focused, mutually-understanding, participation-promoting, and progress-monitoring. In addition, evaluations of anti-trafficking programs should be strengths-based, trauma-informed, culturally sensitive, and survivor-centered. Moreover, evaluation teams must ensure the following: (a) measures capture holistic conceptualizations of evaluation outcomes; (b) data collection strategies are practical and respect regular service provision; (c) data management is active, regular, and strives to enhance data quality; (d) data analysis is robust to the types of data that will be typically collected in such evaluations (e.g., small samples, non-normally distributed data, and missingness); (e) interpretation of results are meaningful for programs, practitioners, and survivors.

In addition to the value of the literature reviews for developing the protocol, we would also like to underscore the importance of gathering feedback from other practice leaders and providers, as well as survivors. The feedback sessions with survivors, administrative leadership, and program staff helpfully enabled our team to receive pragmatic guidance from those who would both be most impacted by the data collection activities, as well as potentially benefit from the evaluation findings. Notably, survivor input at key points in the protocol's development and implementation (i.e., in both the feedback sessions and in the post assessment feedback forms) especially shaped data collection activities and evaluation outcomes.

### *Feasibility Trial—Discussion and Implications*

Overall, the feasibility trial findings (i.e., aim two) show that the second research question (RQ2: Can a feasibility evaluation trial be practically and successfully implemented among a sample of survivor service users?) and the third research question (RQ3: Can results inform practice with HT survivors in this program and setting?) were also answered affirmatively.

The trial showed that comprehensive data across key outcomes (i.e., physical and mental well-being, stress and PTSD, depression, coping, post-traumatic growth, employment, and housing) could be collected not only at program entry and program completion but also over time throughout service delivery. Notably, these data came from survivors who represented diversity in their national origins and primary languages. Many survivors included in the evaluation had limited English proficiency and uncertain immigration status.

Although the trial did not result in complete data for all survivors, enough data was collected for a meaningful, albeit preliminary analysis. Likewise, to the best of our knowledge, this is one of only a few reports concerning anti-trafficking housing program evaluations particularly one with many foreign-national survivors as participants. Accordingly, this trial adds to the preliminary, though growing, evaluation evidence base showing that—even in the context of a complex, dynamic service environment—comprehensive evaluation data can be collected from survivors with diverse backgrounds, experiences, and situations using standard survey instruments. The evaluation findings also suggest that, not only can these data be collected, but that such data can be useful in indicating changes in survivors' outcomes over time. In turn, such findings hold implications for informing practice both at the individual level (i.e., informing survivors' unique services plans), and the program level (i.e., guiding changes to service content, delivery, and strategies for the program overall).

In considering recommendations for future evaluations, we reflect here that the feasibility of the evaluation might have been helped by: (a) an emphasis on survivors' capacities and resources, as well as their needs; (b) responsiveness to survivors regardless of background, languages, and situations; and (c) efforts to forefront survivors' preferences and recommendations for data collection.

In addition, as our team worked together over the course of developing and implementing the evaluation, we realized the importance of our practitioner-researcher collaboration in that both sides brought critical expertise and capacities to the partnership (Ammerman et al., 2014). The practitioners brought deep knowledge of survivors' service needs and research preferences, as well as expertise concerning data collection strategies that could be successfully implemented in the context of typical service delivery. The researchers brought expertise concerning data organization and management, data analysis strategies, as well as in the interpretation of statistical findings. Likewise, both side of the partnership brought organizational resources to the collaboration that augmented the project. As examples, the practitioners ensured that the evaluation protocol was survivor-centered through the inclusion of those with lived experience in the feedback groups, including both program staff and program participants. The researchers were able to augment the project through the inclusion of students on the team to support the evaluation at no cost. Also, project protocols were independently reviewed for human research protections through the research team's university-based IRB.

### *Recommendations for the Future and Conclusion*

Although some robust evaluations have been carried out in anti-trafficking programs, including longitudinal data collection over the course of service provision (e.g., Munsey et al., 2018; Rothman et al., 2020), and randomized studies (e.g., O'Callaghan et al., 2013), the use of such research methods is still rare in the anti-trafficking field. Accordingly, and considering calls to develop greater evaluation capacity (Davy, 2015; Dell et al., 2019; Krieger et al., 2020; Okech et al., 2018), our team sought to document the decisions and strategies related to conceptualizing, designing, and conducting a longitudinal program evaluation of anti-trafficking housing program. Although this paper offers only one protocol concerned with one program, it is our hope that other evaluation teams will develop and share their protocols so that the entire anti-trafficking field can benefit from the lessons learned by a growing number of projects.

Also, for the future, we recommend that practice, policy, and/or research leaders in the anti-trafficking field, promote practice and research partnerships in evaluation, as well as promote evaluation and data capacities within organizations that serve HT survivors. If growing numbers of anti-trafficking programs were able to undertake evaluations in coordinated ways, the field could develop multi-site data and evaluation efforts, which in turn and in combination, could lead to rigorous, quasi-experimental studies with large samples that are fully powered to detect statistical



effects on key outcomes. Moreover, such multi-site data and evaluation efforts would be community-based, ecologically valid, and strongly based in practice.

While a multi-site data and evaluation project may not be feasible in the short-term, this study adds to others in showing that comprehensive, longitudinal program evaluation can successfully be carried out in the context of dynamic anti-trafficking service provision. Likewise, this study suggests that the current situation, of limited program evaluation, is no longer adequate. Accordingly, we hope that the findings and recommendations determined from this study help to offer a long-term vision for the potential of anti-trafficking program evaluation in the future. In the short-term, we also hope that this paper will encourage practitioners, researchers, and survivor-leaders to partner together to develop and carry out their own evaluations in their own communities and programs.

### Acknowledgments

The authors wish to thank all the survivors who gave their feedback and data to facilitate this project. The authors also acknowledge and thank Kristin Weschler, Victim Justice Program Specialist at the Office for Victims of Crime, U.S. Department of Justice, for her review of and feedback on an earlier version of this manuscript.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Office for Victims of Crime, Office of Justice Programs, U.S. Department of Justice (grant number #2016-VT-BX-K057). The opinions, findings, and conclusions or recommendations expressed in this product are those of the contributors and do not necessarily represent the official position or policies of the U.S. Department of Justice. Additional support for the development of this manuscript was provided by the L. Richardson Preyer Distinguished Chair for Strengthening Families fund of the School of Social Work at the University of North Carolina at Chapel Hill.

### ORCID iD

Rebecca J. Macy  <https://orcid.org/0000-0003-4212-9753>

### Note

1. TVPA certification captures three main stages of Trafficking identification status: (1) Pre-certified TVPA refers to an identification stage when a client has been screened by an immigration attorney and meeting TVPA criteria but has not received their official certification letter from the US Department of Health and Human Services (HHS) Office of Refugee Resettlement (ORR) that they are a certified victim of trafficking according to TVPA Criteria; (2) Certified TVPA means a client has received their official certification letter; (3) Provisional TVPA describes a stage where a client has been screened by a service provider to show enough evidence for trafficking but has not yet been screened by an attorney.

### References

- Aborisade, R. A., & Aderinto, A. A. (2008). Adjustment patterns and obstacles against social rehabilitation of sex workers in Nigeria. *African Sociological Review/Revue Africaine de Sociologie*, *12*, 128–143.
- Aderinto, A. A., & Aborisade, R. A. (2008). Pathways to sex workers' social rehabilitation and assessment of rehabilitation approaches in Nigeria: Evidence from a qualitative study. *Gender and Behaviour*, *6*, 1925–1959. <https://doi.org/10.4314/gab.v6i2.23428>

- Ammerman, A., Smith, T. W., & Calancie, L. (2014). Practice-based evidence in public health: Improving reach, relevance, and results. *Annual Review of Public Health, 35*, 47–63. <https://doi.org/10.1146/annurev-publhealth-032013-182458>
- Bass, J., Bearup, L., Bolton, P., Murray, L., & Skavenski, S. (2011). *Implementing trauma focused cognitive behavioral therapy (TF-CBT) among formerly trafficked-sexually exploited and sexually abused girls in Cambodia: A feasibility study*. World Vision.
- Blevins, C. A., Weathers, F. W., Davis, M. T., Witte, T. K., & Domino, J. L. (2015). The posttraumatic stress disorder checklist for DSM-5 (PCL-5): Development and initial psychometric evaluation. *Journal of Traumatic Stress, 28*, 489–498. <https://doi.org/10.1002/jts.22059>
- Bowen, D. J., Kreuter, M., Spring, B., Cofta-Woerpel, L., Linnan, L., & Weiner, D., ... M. Fernandez (2009). How we design feasibility studies. *American Journal of Preventive Medicine, 36*, 452–457. <https://doi.org/10.1016/j.amepre.2009.02.002>
- Caliber (2007). *Evaluation of comprehensive services for victims of human trafficking: Key findings and lessons learned*. US Department of Justice, <https://www.ncjrs.gov/pdffiles1/nij/grants/218777.pdf>.
- Carver, C. S. (1997). You want to measure coping but your protocol's Too long: Consider the brief COPE. *International Journal of Behavioral Medicine, 4*, 92–100. [https://doi.org/10.1207/s15327558ijbm0401\\_6](https://doi.org/10.1207/s15327558ijbm0401_6)
- Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*, 385–396. <https://doi.org/10.2307/2136404>
- Crawford, M., & Kaufman, M. R. (2008). Sex trafficking in Nepal: Survivor characteristics and long-term outcomes. *Violence Against Women, 14*, 905–916. <https://doi.org/10.1177/1077801208320906>
- Criswell, L. K. (2014). *The Thriving Conversation Project: A pilot assessment of the use of thriving principles with sexually exploited youth*. (Publication No. 3702701) [Doctoral dissertation, Fuller Theological Seminary], ProQuest Dissertations Publishing.
- Davy, D. (2015). Understanding the support needs of human-trafficking victims: A review of three human-trafficking program evaluations. *Journal of Human Trafficking, 1*, 318–337. <https://doi.org/10.1080/23322705.2015.1090865>
- Dehar, M. A., Casswell, S., & Duignan, P. (1993). Formative and process evaluation of health promotion and disease prevention programs. *Evaluation Review, 17*, 204–220. <https://doi.org/10.1177/0193841X9301700205>
- Dell, N. A., Maynard, B. R., Born, K. R., Wagner, E., Atkins, B., & House, W. (2019). Helping survivors of human trafficking: A systematic review of exit and postexit interventions. *Trauma, Violence, & Abuse, 20*, 183–196. <https://doi.org/10.1177/1524838017692553>
- Department of Justice (2020, September 21). *Justice News: Justice Department Awards Nearly \$101 Million to Combat Human Trafficking*. <https://www.justice.gov/opa/pr/justice-department-awards-nearly-101-million-combat-human-trafficking>.
- Eaton, W. W., Smith, C., Ybarra, M., Muntaner, C., & Tien, A. (2004). Center for epidemiologic studies depression scale: Review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcomes assessment: Instruments for adults* (pp. 363–377). Lawrence Erlbaum Associates Publishers.
- Elliott, D. E., Bjelajac, P., Fallot, R. D., Markoff, L. S., & Reed, B. G. (2005). Trauma-informed or trauma-denied: Principles and implementation of trauma-informed services for women. *Journal of Community Psychology, 33*(4), 461–477. <https://doi.org/10.1002/jcop.20063>
- Felner, J. K., & DuBois, D. L. (2017). Addressing the commercial sexual exploitation of children and youth: A systematic review of program and policy evaluations. *Journal of Child & Adolescent Trauma, 10*, 187–201. <https://doi.org/10.1007/s40653-016-0103-2>
- Gibbs, D., Walters, J. H., Lutnick, A., Miller, S., & Kluckman, M. (2014). *Evaluation of services for domestic minor victims of human trafficking*. US Department of Justice, <https://www.ncjrs.gov/pdffiles1/nij/grants/248578.pdf>.

- Goździak, E. M., & Lowell, B. L. (2016). *After rescue: Evaluation of strategies to stabilize and integrate adult survivors of human trafficking to the United States*. Institute for the Study of International Migration, <https://www.ncjrs.gov/pdffiles1/nij/grants/249672.pdf>.
- Graham, L. M., Macy, R. J., Eckhardt, A., Rizo, C. F., & Jordan, B. L. (2019). Measures for evaluating sex trafficking aftercare and support services: A systematic review and resource compilation. *Aggression and Violent Behavior, 47*, 117–136. <https://doi.org/10.1016/j.avb.2019.04.001>
- Green, L. W. (2006). Public health asks of systems science: To advance our evidence-based practice, can you help us get more practice-based evidence? *American Journal of Public Health, 96*, 406–409. <https://doi.org/10.2105/AJPH.2005.066035>
- Honeycutt, R. J. (2012). *Sex trafficked survivors' recovery: program evaluation of Transitions Global Cambodian safe house*. (Publication No. 3506888) [Doctoral dissertation, George Fox University], ProQuest Dissertations Publishing.
- Kaufman, M. R., & Crawford, M. (2011). Research and activism review: Sex trafficking in Nepal: A review of intervention and prevention programs. *Violence Against Women, 17*, 651–665. <https://doi.org/10.1177/1077801211407431>
- Klein, L. B., Chesworth, B. R., Howland-Myers, J. R., Rizo, C. F., & Macy, R. J. (2021). Housing interventions for intimate partner violence survivors: A systematic review. *Trauma, Violence, & Abuse, 22*(2), 249–264. <https://doi.org/10.1177/1524838019836284>
- Krieger, K. E., Gibbs, D. A., & Cutbush, S. L. (2020). Evaluating human trafficking service programs: What can be learned from domestic violence service program evaluations. *Journal of Human Trafficking, 1*–27. <https://doi.org/10.1080/23322705.2020.1784686>
- Leviton, L. C., Khan, L. K., Rog, D., Dawkins, N., & Cotton, D. (2010). Evaluability assessment to improve public health policies, programs, and practices. *Annual Review of Public Health, 31*, 213–233. <https://doi.org/10.1146/annurev.publhealth.012809.103625>
- Macy, R. J., & Johns, N. (2011). Aftercare services for international sex trafficking survivors: Informing US service and program development in an emerging practice area. *Trauma, Violence, & Abuse, 12*, 87–98. <https://doi.org/10.1177/1524838010390709>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Plos Medicine, 6*, e1000097. <https://doi.org/10.1371/journal.pmed1000097>
- Muftić, L. R., & Finn, M. A. (2013). Health outcomes among women trafficked for sex in the United States: A closer look. *Journal of Interpersonal Violence, 28*, 1859–1885. <https://doi.org/10.1177/0886260512469102>
- Munsey, S., Miller, H. E., & Rugg, T. (2018). Generatehope: A comprehensive treatment model for sex-trafficked women. *Journal of Evidence-Informed Social Work, 15*(4), 420–431. <https://doi.org/10.1080/23761407.2018.1467809>
- Nnawulezi, N., Godsay, S., Sullivan, C. M., Marcus, S., & HacsKaylo, M. (2018a). The influence of low-barrier and voluntary service policies on survivor empowerment in a domestic violence housing organization. *American Journal of Orthopsychiatry, 88*(6), 670–680. <https://doi.org/10.1037/ort0000291>
- Nnawulezi, N., Lippy, C., Serrata, J., & Rodriguez, R. (2018b). Doing equitable work in inequitable conditions: An introduction to a special issue on transformative research methods in gender-based violence. *Journal of Family Violence, 33*, 507–513. <https://doi.org/10.1007/s10896-018-9998-8>
- O'Callaghan, P., McMullen, J., Shannon, C., Rafferty, H., & Black, A. (2013). A randomized controlled trial of trauma-focused cognitive behavioral therapy for sexually exploited, war-affected Congolese girls. *Journal of the American Academy of Child & Adolescent Psychiatry, 52*, 359–369. <https://doi.org/10.1016/j.jaac.2013.01.013>
- O'Cathain, A., Croot, L., Sworn, K., Duncan, E., Rousseau, N., & Turner, K., ... P. Hoddinott (2019). Taxonomy of approaches to developing interventions to improve health: A systematic methods overview. *Pilot and Feasibility Studies, 5*, 41. <https://doi.org/10.1186/s40814-019-0425-6>
- Okech, D., Choi, Y. J., Elkins, J., & Burns, A. C. (2018). Seventeen years of human trafficking research in social work: A review of the literature. *Journal of Evidence-Informed Social Work, 15*(2), 103–122. <https://doi.org/10.1080/23761407.2017.1415177>

- Orsmond, G. I., & Cohn, E. S. (2015). The distinctive features of a feasibility study: Objectives and guiding questions. *OTJR: Occupation, Participation and Health, 35*, 169–177. <https://doi.org/10.1177/1539449215578649>
- Potocky, M. (2010). Effectiveness of services for victims of international human trafficking: An exploratory evaluation. *Journal of Immigrant & Refugee Studies, 8*, 359–385. <https://doi.org/10.1080/15562948.2010.522462>
- Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401. <https://doi.org/10.1177/014662167700100306>
- Restore NYC (2019). *Fiscal year 2019 gratitude report*. [https://issuu.com/restorenyc/docs/restore\\_gratitudereport2019\\_web\\_single\\_\\_1\\_](https://issuu.com/restorenyc/docs/restore_gratitudereport2019_web_single__1_)
- Roe-Sepowitz, D. E., Gallagher, J., Hickle, K. E., Pérez Loubert, M., & Tutelman, J. (2014). Project ROSE: An arrest alternative for victims of sex trafficking and prostitution. *Journal of Offender Rehabilitation, 53*, 57–74. <https://doi.org/10.1080/10509674.2013.861323>
- Rothman, E. F., Preis, S. R., Bright, K., Paruk, J., Bair-Merritt, M., & Farrell, A. (2020). A longitudinal evaluation of a survivor-mentor program for child survivors of sex trafficking in the United States. *Child Abuse & Neglect, 100*, 104083. doi:10.1016/j.chiabu.2019.104083
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2016). Introducing evidence-based principles to guide collaborative approaches to evaluation: Results of an empirical process. *American Journal of Evaluation, 37*, 193–215. <https://doi.org/10.1177/1098214015615230>
- Tedeschi, R. G., & Calhoun, L. G. (1996). The posttraumatic growth inventory: Measuring the positive legacy of trauma. *Journal of Traumatic Stress, 9*, 455–471. <https://doi.org/10.1007/BF02103658>
- Trafficking Victims Protection Act of 2000, Pub. L. No. 106–386, § 102(b) (14–15) (2000).
- Trevisan, M. S., & Walser, T. M. (2014). *Evaluability assessment: Improving evaluation quality and use*. Sage Publications.
- Ware, J.Jr., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey: Construction of scales and preliminary tests of reliability and validity. *Medical Care, 34*, 220–233. <https://doi.org/10.1097/00005650-199603000-00003>
- Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2007). *User's manual for the SF-12v2TM health survey:(with a supplement documenting SF-12 health survey)*. Lincoln, RI: QualityMetric Incorporated.
- Zimmerman, C., Hossain, M., & Watts, C. (2011). Human trafficking and health: A conceptual model to inform policy, intervention and research. *Social Science & Medicine, 73*, 327–335. <https://doi.org/10.1016/j.socscimed.2011.05.028>