



Can a workbook work? Examining whether a practitioner evaluation toolkit can promote instrumental use



Rebecca Campbell^{a,*}, Stephanie M. Townsend^{b,1}, Jessica Shaw^{c,2}, Nidal Karim^{d,3}, Jenifer Markowitz^{e,4}

^a Department of Psychology, Michigan State University, 127 C Psychology Building, East Lansing, MI 48824-1116, United States

^b Townsend Consulting & Evaluation, 8 Locke Drive, Pittsford, NY 14534, United States

^c National Institute of Justice

^d CARE USA, 151 Ellis Street, NE, Atlanta, GA 30303, United States

^e Forensic Nurse Consultant, 2308 Mt. Vernon Avenue, Suite 238, Alexandria, VA 22301, United States

ARTICLE INFO

Article history:

Received 14 June 2014

Received in revised form 24 April 2015

Accepted 29 April 2015

Available online 7 May 2015

Keywords:

Evaluation capacity building

Multi-site evaluations

Evaluation toolkits

Instrumental use

ABSTRACT

In large-scale, multi-site contexts, developing and disseminating practitioner-oriented evaluation toolkits are an increasingly common strategy for building evaluation capacity. Toolkits explain the evaluation process, present evaluation design choices, and offer step-by-step guidance to practitioners. To date, there has been limited research on whether such resources truly foster the successful design, implementation, and use of evaluation findings. In this paper, we describe a multi-site project in which we developed a practitioner evaluation toolkit and then studied the extent to which the toolkit and accompanying technical assistance was effective in promoting successful completion of local-level evaluations and fostering instrumental use of the findings (i.e., whether programs directly used their findings to improve practice, see Patton, 2008). Forensic nurse practitioners from six geographically dispersed service programs completed methodologically rigorous evaluations; furthermore, all six programs used the findings to create programmatic and community-level changes to improve local practice. Implications for evaluation capacity building are discussed.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

In large-scale, multi-site evaluations, developing and disseminating practitioner-oriented evaluation workbooks or toolkits are an increasingly common strategy for evaluation capacity building (see examples produced by the Kellogg Foundation, the United Way, the World Bank, the Pell Institute, among others).⁵ Toolkits demystify the evaluation process, present evaluation options, explain design choices, and offer step-by-step guidance to

practitioners. However, whether such resources can truly build evaluation capacity and foster the successful design, implementation, and use of evaluation findings merits empirical examination. Toolkits may be *efficient* in providing resources to many programs and practitioners, but they may or may not be *effective*. In this paper, we describe a multi-site project in which we developed a practitioner evaluation toolkit and then studied the extent to which the toolkit and accompanying technical assistance was effective in promoting successful completion of local-level evaluations and fostering instrumental use of the findings. In this project, we conceptualized ‘instrumental use’ consistent with Patton (2008) definition: “Instrumental use refer to evaluation findings directly informing a decision or contributing to solving a problem; the findings are linked to some subsequent action and in that sense become an *instrument* of action” (p. 102, emphasis in original).

1.1. Developing toolkits to develop evaluation capacity in local-level programs

The fundamental premise of evaluation toolkits is that the skills and methods of evaluation are teachable, and that with

* Corresponding author. Tel.: +1 517 432 8390; fax: +1 517 432 2945.

E-mail addresses: rmc@msu.edu (R. Campbell),

Stephanie.Townsend@earthlink.net (S.M. Townsend), Jessica.Shaw@ojp.doj.gov

(J. Shaw), nkarim@care.org (N. Karim), jenifer.markowitz@gmail.com

(J. Markowitz).

¹ Tel.: +1 585 690 9315.

² Tel.: +1 517 353 5015; fax: +1 517 432 7082.

³ Tel.: +1 404 979 9313; fax: +1 404 589 2632.

⁴ Tel.: +1 330 671 4470.

⁵ In this manuscript, we will be using the terms ‘workbook’ and ‘toolkit’ interchangeably to refer to comprehensive evaluation guides for practitioners. As noted below in Section 1.1, there is considerable variability in the literature in the extent to which these guides are content-specific and design-prescriptive.

appropriate guidance and support, practitioners can evaluate their work and use the findings to improve their practice (Cousins & Chouinard, 2012; Fetterman & Wandersman, 2005; Patton, 2008; Preskill & Boyle, 2008). Toolkits vary considerably in the extent to which they are “generalist” resources on evaluation or specific manuals for how to implement a specific design(s); furthermore, some workbooks are intended to span multiple substantive domains (e.g., “preventive health interventions”) whereas others are tailored to a specific content domain and service program. Typically, most toolkits provide “evaluation 101” instruction and explain evaluation design options, and then some go further to articulate step-by-step procedures for data collection and analysis, suggest strategies for utilization, and provide sample tools, measures, and other resources practitioners can use outright or adapt to their specific needs. Toolkits are a potentially cost-efficient way to build the evaluation capacity of organizations in that they can address multiple capacity needs, including: building evaluation knowledge, fostering positive dispositions toward evaluation, and developing evaluation skills (Wandersman, Imm, Chinman, & Kaftarian, 2000; Wandersman et al., 2008). In practice, toolkits can implement multiple capacity-building strategies simultaneously: they are a written resource that can be shared among multiple staff and stakeholders, their educational aspects can act as a substitute for more costly in-person training, and when accompanied by data analysis tools, they enhance technological capacity (Campbell et al., 2004; Dyckhoff, Zielke, Bultmann, Chatti, & Schroeder, 2012).

1.2. Evaluating whether toolkits develop capacity in local-level programs

To date, only a handful of projects have formally studied whether evaluation toolkits can build evaluation capacity and foster evaluation use. In one of the first studies to examine this issue, Oliver, MacBean, Conole, and Harvey (2002) developed a toolkit to help educators evaluate learning outcomes among students, and then they had both novice and experienced evaluators use the toolkit to design an evaluation. Although this project did not study actual implementation of the planned evaluations, Oliver et al. (2002) found that the toolkit was effective in helping both experienced and less-experienced evaluators develop good quality, tailored evaluation designs. As to whether this toolkit would support actual instrumental use of the design and the findings resulting from the design is not known (as it was not the focus of this project).

Expanding the scope to consider whether toolkits can support actual implementation, Campbell et al. (2004) conducted a state-wide project with human service organizations with the goal of using evaluation workbooks and technical assistance (e.g., in-person trainings, site-specific consultation) to guide all state-funded sexual assault services and prevention programs through the process of planning, implementing, and using evaluating findings to improve practice. Program staff reported high satisfaction with the resources provided, and when the research/evaluation team reviewed the programs' reports that they filed with the state funders, there was clear evidence that all sites had been able to successfully conduct their evaluations, with action plans for using the findings to improve practice (i.e., an intention for instrumental use). At a one-year follow-up interview, there had been considerable staff turnover in most agencies, but the new staff indicated that they were aware of and were continuing to use the evaluation workbooks (i.e., evaluation capacity building).

Taking implementation to an even larger scale, the *Getting to Outcomes* toolkit was designed to be a generalized resource for planning and evaluating preventive interventions' outcomes across multiple substantive domains (unlike the Oliver et al.,

2002 and Campbell et al., 2004 toolkits, which were tailored to specific content domains) (Chinman et al., 2008; Chinman, Tremain, Imm, & Wandersman, 2009; Hunter et al., 2009; Wandersman et al., 2000). *GTO* guides programs through a series of 10 questions that help staff analyze needs and resources, articulate goals and use of best practices, develop organizational capacity for evaluation, plan and implement evaluations, and develop approaches for continuous quality improvement and sustainability.

A multi-site evaluation of *GTO* and accompanying technical assistance found that these resources were effective at building both individual capacity and program performance, as well using outcome data for ongoing program improvement (i.e., instrumental use) (Chinman et al., 2008). An online, interactive format of *GTO* has also been developed and evaluated (*iGTO*) (Chinman et al., 2009). Fifty-six coalitions in two states used *iGTO* to develop experimental and quasi-experimental designs of their programs. *iGTO* was effective in supporting increased program performance, but participant perceptions of the online system were mixed and no participants reported intending to continue to use the system after the completion of the study (Chinman et al., 2009). More recently, Chinman, Hunter, and Ebener (2012) collaborated with a substance abuse service program that had participated in a *GTO* project to study how the evaluations contributed to instrumental use. Couched within a CQI (continuous quality improvement) model, the authors found that program staff were able to implement CQI approaches to create a “bridge between conducting self evaluation and making concrete changes to improve programming” (p. 613).

GTO, like most evaluation toolkits, focus on teaching program staff how to evaluate discrete, well-defined programmatic outcomes, but toolkits may also be effective in helping programs identify and evaluate more intangible aspects of their work related to programmatic values, such as equality or empowerment. For example, in a multi-national toolkit project, Burford et al. (2013) created and evaluated the *WeValue* toolkit, which helped community-based and non-governmental organizations assess values-based achievements, such as open and respectful communication, equal participation in decision making, and transparency in organizational mission and activities. The toolkit and accompanying training were successful in helping all sites develop at least one assessment tool and collect local data; in addition, the toolkit helped programs make sustained changes to their monitoring and evaluation strategies.

Taken together, the results of these projects suggest that toolkits can be effective in helping program staff learn about evaluation and implement evaluation projects. However, the literature has been less specific on exactly how programs have directly used their outcome evaluation findings to improve practice (i.e., instrumental use; Patton, 2008). Studies that have evaluated toolkits tend to report that instrumental use occurred, without demonstrating what exactly that use was and how it directly stemmed from the resources provided (i.e., the toolkit and technical assistance). To address this gap in the literature, we developed and evaluated a domain-specific toolkit (for sexual assault nurse practitioner programs) in a national-scale, multi-site project, with an emphasis on identifying evidence of instrumental use (i.e., using findings to create programmatic changes). To set the stage for this project, we will briefly describe the content domain of this service program and its intervention model.

1.3. Case example: developing & evaluating a toolkit for sexual assault nurse examiner (SANE) programs

Sexual violence is a pervasive social problem: national epidemiological data indicate that 18–25% of women are raped

in their adult lifetimes (Black et al., 2011; Kilpatrick, Resnick, Ruggiero, Conoscenti, & McCauley, 2007). To respond to rape victims' post-assault needs, many communities throughout the United States have implemented sexual assault nurse examiner (SANE) programs, whereby specially trained nurses provide comprehensive psychological, medical, and forensic services for sexual assault victims in either hospital or community-based clinic settings (Department of Justice [DOJ], 2013). Sexual assault forensic nurses are trained to offer crisis intervention and emotional support, health care (e.g., sexually transmitted infection [STI] screening and prophylaxis, pregnancy testing and emergency contraception), injury detection and treatment, and state-of-the-art forensic medical evidence collection. In addition, SANEs work with police and prosecutors to provide on-going case consultation and may testify as expert witnesses should a case go to trial (DOJ, 2013). Thus, SANE programs are often characterized as promoting the dual goals of improving patient health care (primary goal) and increasing sexual assault prosecution rates (secondary goal) (Campbell, Patterson, & Lichty, 2005; DOJ, 2013).

SANE programs spread quickly throughout the United States, growing from a handful of programs in the 1970s and 1980s to over 600 programs currently in existence (International Association of Forensic Nurses [IAFN], 2014). However, this diffusion occurred despite minimal evaluative data on the effectiveness of SANE programs. Early case studies in the field suggested that forensic nurses are an important resource to the legal community (see Campbell, Townsend, et al., 2005 for a review), and to date, two longitudinal quasi-experimental studies have documented increased prosecution rates after the implementation of SANE programs (Campbell, Patterson, & Bybee, 2012; Crandall & Helitzer, 2003). However, two rigorous evaluations are hardly sufficient to determine the effectiveness of SANE programs as an intervention model for post-assault care and improving sexual assault prosecution rates. Given the number of SANE programs in existence (600+), their geographic dispersion, and the fact that these interventions operate under a reasonably standardized scope of service and practice (IAFN, 2009), this is an excellent opportunity for building evaluation capacity through a practitioner-oriented toolkit.

1.4. The current study: evaluating whether a toolkit can promote instrumental use

The purpose of this project was to create and implement a forensic nursing practitioner evaluation toolkit. Six programs (two rural programs, two programs serving mid-sized communities, and two urban programs) received the step-by-step evaluation toolkit and accompanying technical assistance package (e.g., webinars, group consultation calls, individualized phone and email consultation, and in-person site visits). Our goal was to guide these six programs through the process of conducting local-level evaluations of how their services affected adult sexual assault prosecution rates in their communities. Given the dearth of literature regarding whether evaluation toolkits can promote instrumental use, we assessed whether the materials and technical assistance we provided were in fact sufficient to create programmatic changes in all six sites.

Therefore, as the six selected SANE programs moved through the evaluation steps outlined in the toolkit to assess program impact on prosecution rates, we (the evaluation team) collected data regarding: (1) what specific assistance and dosage of assistance was provided to the sites and whether it was perceived as useful by program staff; (2) whether program staff actually used the toolkit and technical assistance to complete methodologically rigorous, local-level evaluations; and (3) whether each of the six sites directly used their evaluation findings for instrumental

change (as defined previously: the findings directly influenced decisions, problem solving, and/or other actions to improve practice; see Patton, 2008), and if so, what was the nature of that instrumental use.

2. Methods

2.1. Overview

In this project, there were two “tracks” of data collection: the first was selecting the six SANE programs that would participate in the evaluation training and completing local-level evaluations of their programs; the second was studying how the programs used the resources we provided (i.e., the toolkit and accompanying technical assistance) to conduct their evaluations and how the sites used their findings to create programmatic changes (i.e., instrumental use).

2.2. Site selection

To recruit SANE programs to participate in this multi-site evaluation project, we collaborated with the International Association of Forensic Nurses (IAFN) and the National Sexual Violence Resource Center (NSVRC) to advertise this project via multiple list serve email announcements to all SANE programs within the United States, which at the time this study was conducted numbered approximately 600 programs. Interested SANE programs were instructed to contact the evaluation team for an application packet, which designed to screen programs for evaluation *readiness*, as opposed to evaluation *capacity*. Consistent with Preskill's model of evaluation capacity building, “readiness” means that the program had the organizational resources to participate in evaluation learning activities without compromising program operations; “capacity” means program staff have already learned evaluation skills and routinized evaluation within their organization (Preskill & Boyle, 2008; Russ-Eft & Preskill, 2001). Sampling programs based on evaluation *readiness* was appropriate because we wanted to target more “typical” program conditions, which would be a state of readiness rather than capacity. To that end, the application included questions regarding: (1) administrative staffing of the program (e.g., a paid program coordinator); (2) number of patients served per year and staffing levels to support those services; (3) record keeping/information documentation used in the program; and (4) selected items from Preskill and Torres' (2000) Readiness for Organizational Learning and Evaluation Instrument (ROLE) scale, which is a standardized measure of evaluation readiness (additional questions regarding the prior experience in research/evaluation were also included).

Programs would be deemed eligible to participate if they had: (1) a full-time SANE program coordinator (to serve as a stable point-of-contract with the evaluation team); (2) nursing staffing levels appropriate for their current patient caseloads so that staff would be able to devote time to participating in an evaluation project without adversely affecting program services; (3) secured access to the different data sources needed in order to have the potential to complete a pre-SANE/post-SANE evaluation of legal case outcomes; and (4) a mean score above the 25th percentile on the modified/shortened ROLE scale, which indicates good organizational readiness for evaluation activities. Given our intent to sample for organizational readiness, sites did not have to have prior experience with research and/or evaluation to be eligible to participate.

Seventy-three (73) SANE programs expressed interest in becoming an evaluation site: 30 programs completed the entire online application process; 7 programs had incomplete applications; and 36 programs did not activate their online application (i.e., they elected not to apply after expressing initial interest). The

Table 1
Overview of the SANE practitioner evaluation toolkit.

Toolkit step	Content focus
1. Evaluation questions	Reviews literature on different ways to conceptualize SANE effectiveness Summarizes evaluation 101 terminology, logic models, and theories of change Offers completed logic model for assessing SANE effectiveness re: <i>legal outcomes</i> Provides additional resources for assessing effectiveness re: <i>health outcomes</i>
2. Evaluation design	Explains pre–post design, post-only design, and ongoing evaluation design Presents the application of each evaluation design to assessing SANE legal outcomes Creates a decision tree for choosing evaluation design
3. Establish cooperative agreements	Outlines kinds of data needed for pre–post, post-only, and ongoing evaluation designs Explains the purpose of institutional review boards (IRBs) and their review process Presents strategies for working with police and prosecutors for evaluation
4. Sampling	Explains sampling designs, sampling frames, need for comparability of cases Presents example sampling criteria and exercises to practice sampling Outlines steps for selecting cases for pre–post, post-only, and ongoing evaluation designs
5. Data collection	Guides data collection for legal case outcome information (step-by-step directions) Sample data collection sheets (with instructions) provided for practitioner use
6. Data analyses	Teaches methods for data tabulation in Excel Explains methods for data checking/cleaning/verification Explains concept of statistical significance and relevance to pre–post design Guides process of interpreting findings (with examples)
7. Utilization	Discusses strategies for sharing findings with community partners Suggests next steps for positive and negative evaluation findings Provides resources for institutionalizing community responses to sexual assault Outlines process for creating action plan for change

30 complete applications were independently reviewed by each member of the evaluation team (which included a highly experienced forensic nurse consultant) to assess eligibility. Of the ten eligible programs, a stratified random sample was drawn consisting of two rural sites (Sites A and B), two mid-sized sites (Sites C and D), and two urban sites (Sites E and F).

2.3. Site evaluation training and technical assistance

Table 1 summarizes the content of the toolkit, which was the primary evaluation resource we provided to the sites, in addition to pre-programmed Excel data entry and data analysis files. Table 2 describes the technical assistance methods/modalities we used to support the sites' use of the toolkit, highlighting the dosage of evaluation assistance provided to the sites. To ensure fidelity to the toolkit's evaluation design and data collection procedures, all sites participated in a series of web-based trainings following their enrollment in the project. The trainings included live audio presentation, visuals, and discussion transmitted via an online conferencing platform. The trainings consisted of a series of 3 h-long sessions: (1) an overview of the project and evaluation process, (2) an explanation of the first steps of the evaluation design (understanding the design, identifying evaluation questions, establishing cooperative agreements, sampling cases, and collecting data), (3) instruction in analyzing data and interpreting results. Additionally, all-site conference calls were held periodically throughout the project to check on progress, share successes, and troubleshoot challenges. These calls were augmented by individual communication (via phone and email) between sites and the evaluation team (approximately 450 email threads throughout project). In addition, in-person site visits were conducted near the end of the project at each program to review data, assist with data analysis and interpretation, and develop action plans based on the findings.

2.4. Site-level evaluation designs

As part of the site-selection process, we screened programs for whether they had the potential to complete a pre–post evaluation

design to examine how sexual assault prosecution rates changed in their community after the implementation of the SANE program. After site selection and during the evaluation webinars (described above), we discovered that three sites would not be able to conduct a pre–post evaluation design due to unforeseen challenges in accessing pre-SANE data (e.g., a fire had destroyed records). Therefore, three sites needed to do a post-only evaluation design. These sites were evenly distributed across the sampling strata (i.e., one rural, one mid-sized, and one urban SANE program completed post-only designs).

2.5. Site-level data collection procedures and measures

Prior to the start of data collection, each SANE program obtained institutional review board (IRB) approval from their parent hospitals or host organizations given that they would be accessing sexual assault patients' medical records (to obtain a sampling frame for their evaluation).⁶ In addition, each site established a memorandum of understanding (MOU) with their local prosecutor's office to access legal case outcome records. As noted previously, the evaluation team provided training on evaluation ethics and IRB procedures, as well as on-going technical assistance during the process of securing proper approvals for this project. Program staff had never completed IRB applications before and they noted that it was particularly helpful to have the assistance of the evaluation team throughout this process.

All sites were provided with standardized instructions and data collection forms for sampling cases and obtaining legal case outcome data from the prosecutor's office. Briefly, pre-SANE and post-SANE cases needed to meet the following criteria to be eligible for inclusion: (1) the patient was at least 18 years old at the time of the sexual assault; (2) the patient had a complete medical forensic exam and evidence collection; (3) the exam and evidence

⁶ Each site also needed IRB approval because de-identified data would be later shared with the research team. We also had IRB approval from Michigan State University for the multi-site project, which provided oversight for our training and technical assistance work with the sites and for our own data collection (described in Section 2.6).

collection was not anonymous/de-identified; (4) the patient made a police report; (5) the police report was not anonymous or “Jane Doe;” and (6) the assault occurred within the focal county (counties) that was the subject of the evaluation.

Each site also had to determine the time period from which cases would be selected. Sites completing a post-only evaluation design selected cases that received services from their SANE program beginning one year after the program start date through one year prior to the evaluation (i.e., the first year and most recent year of SANE program operation were excluded).⁷ Sites completing a pre-post evaluation design selected the same post-SANE time period; the pre-SANE period was defined as three (or two, depending on record availability) years prior to the launch of their program. Because collecting prosecution case outcomes is a resource-intensive process (see Campbell et al., 2012), it was necessary to limit the number of cases studied, particularly in the context of program evaluations conducted by practitioners. We conducted a power analysis to determine the minimum number of cases per year that could be sampled and still yield 0.80 power for within-site and cross-site analyses. This analysis revealed that 30 cases per year would be necessary, which was also a program-matically feasible number for program staff (see Campbell, Townsend, Bybee, Shaw, & Markowitz, 2013 for details on these computations). Sites that had over 50 eligible cases per year randomly sampled 30 cases per year for inclusion in this project. The evaluation team provided each site with standardized directions and step-by-step forms when reviewing case files to determine case eligibility and sample cases.

Once program staff had drawn their pre-SANE and/or post-SANE samples, they needed to determine what actions the legal system had taken in each of these cases (recall that to be eligible for inclusion in the project, the sexual assault patient had to have made a report of the rape to the police). Consistent with measures used by Campbell et al. (2012), legal case outcomes were classified into one of five mutually exclusive categories: (1) the case was not referred by police/not charged by prosecutors; (2) the case was charged by prosecutors, but later dropped; (3) the case was plea bargained; (4) the case went to trial and ended in an acquittal; or (5) the case went to trial and ended in an conviction. For simplicity, categories (1) and (2) can be combined to reflect the percentage of cases that were not prosecuted; this simplified metric will be presented in this manuscript, given that the substantive results of the evaluations are not the focus of this paper. As noted previously, the evaluation team monitored data collection very closely via group conference calls and individual technical assistance to ensure that program staff was coding legal case outcomes consistently.

2.6. Evaluation team data collection procedures and measures

As the six sites followed the toolkit steps to complete their local-level evaluations, the evaluation team collected multiple kinds of data to address our three primary research goals (documenting the assistance provided to the sites and whether it was perceived as useful by program staff; determining whether staff used these resource to complete local-level evaluations; and assessing whether sites used their evaluation findings for

instrumental change). Specifically, we collected five main types of data:

2.6.1. Site contact log

Every contact with each site was logged in an Excel workbook that recorded the nature of the interaction (e.g., phone call, email, site visit) as well as a brief summary of the content. The site contact log provided a quantitative count of the frequency of interaction with the sites and helped us ascertain what assistance had been provided to each site and the progress of each site on their local-level evaluations.

2.6.2. Qualitative field notes

Every contact with each site that had sustained, substantive discussion regarding their use of the toolkit and their local-level evaluations was also captured in qualitative field notes (i.e., the group conference calls, individual site phone consultation, and individual in-person site visits, see Table 2, technical assistance modalities used with the sites). In these field notes we recorded information that program staff relayed to us regarding: how they were using the toolkit and technical assistance to complete their evaluations, revisions needed to improve the toolkit and other materials, challenges they were encountering, solutions identified for those problems, questions and additional requests for technical assistance, plans for sharing findings with stakeholders, and steps they were taking to use their findings to improve practice. In accord with the methods outlined by Emerson, Fretz, and Shaw (1995), fieldnotes were written by the project director, co-investigator, or principal investigator (depending on who had primary contact with the site) within 48 h of a contact (24 h typically for phone contact; 48 h for site visit contact). These notes included direct quotes and rich descriptions of the interactions; the researchers' reflective notes on the interaction were bracketed and separated from the primary notes.

2.6.3. Quantitative post-webinar satisfaction surveys

We collected quantitative satisfaction surveys for each instructional webinar via an online, anonymous survey that was distributed immediately following each webinar (100% response rate). Participants rated the quality of the training on multiple dimensions (e.g., coverage, clarity, helpfulness, intended use of materials provided) on a 1–5 scale (1 = strongly disagree to 5 = strongly agree) (see Table 3 for specific items).

2.6.4. Qualitative in-person site visit interviews

At the conclusion of the site visit, the two members of the evaluation team who conducted the visit (the project director, co-investigator, or principal investigator) interviewed the local program staff members who were directly involved in conducting the evaluation (typically 1–2 SANE program personnel per site, 100% participation rate). Consistent with qualitative interviewing methods outlined by Patton (2002), the interview protocol was semi-structured, meaning that we had designated topics to discuss, but question order and wording were flexible, given the issues raised by the participants; the interview probes were “open-ended or specific to the participants' comments, rather than to a preexisting theory” (Hsieh & Shannon, 2005, p. 1279). The interview questions covered five broad domains: (1) satisfaction with the toolkit and technical assistance received; (2) which resources were most helpful to the successful completion of their evaluation; (3) recommendations for improving the toolkit and technical assistance; (4) discussion of key substantive findings of their local findings; and (5) intentions for instrumental use of the results. Interviews were audio recorded with the participants' permission and transcribed verbatim.

⁷ The first year of program operation was excluded from the evaluation because the literature on the development of SANE programs has documented that there are often multiple changes in staffing, services, and community relationships during a program's launch (DOJ, 2013). Campbell et al. (2012) found that Year 2 cases were most appropriately modeled as the “start” of the program. The most recent year of program services was excluded because it typically takes one year (or more) for a case to move through the criminal justice system. Very recent cases would still have pending legal case outcomes, which could skew the evaluation findings.

Table 2
Technical assistance modalities used with the sites.

Activity	Planned or as needed	All site or site specific	Frequency	Dosage	Purpose
Webinars	Planned	All site	3 times	60–90 min	Webinars provided instruction on implementing specific toolkit evaluation steps.
Group conference calls	Planned	All site	3 times	60 min	Conference calls provided an opportunity for sites to update the evaluation team and each other on their progress.
Individual site email consultation	Both	Site specific	Weekly	Varied (450 email threads throughout project)	Emails contact provided information on: relationship building, data collection strategies, gaining access to hospital and prosecutor records, establishing sampling frame, IRB processes, and sharing findings.
Individual site phone consultation	As needed	Site specific	Monthly–bimonthly (each site had 5–12 calls throughout project)	10–60 min	Phone consultation provided sites with information on: evaluation design, data access, sampling, problem solving, and analysis assistance.
Individual site visits	Both	Site specific	1–2 times	1–2 days	The evaluation team provided on-site assistance with data collection, analysis, interpretation, presenting findings to community partners, and action planning.

2.6.5. Qualitative phone follow-up interviews

Six-to-eight weeks after the final project close-out group conference call, the project director conducted qualitative follow-up phone interviews with our primary contact at each of the six

sites (100% participation rate). The focus of these interviews was assessing intended and actual instrumental use of the findings. Following the same interviewing methodology outlined above, the questions covered four broad domains: (1) utilization activities

Table 3
Webinar post-training survey results.

Webinar #1: "Introduction to Evaluation" post-training survey results (n = 6)	
<i>Information covered</i>	
1 = Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
The information covered is relevant for our SANE program.	4.67 (SD = 0.516)
The information was provided in an user-friendly manner	4.83 (SD = 0.408)
I plan to use the information from the training in our evaluation.	4.83 (SD = 0.408)
The information provided was clear and understandable.	4.83 (SD = 0.408)
The information provided covered the areas needed for our program to complete an evaluation.	4.83 (SD = 0.408)
<i>Preparation for evaluation</i>	
1 = Strongly disagree; 2 = disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
The training helped me understand the purpose of evaluation.	5.00 (SD = 0)
The training helped me understand the key concepts in evaluation.	5.00 (SD = 0)
The training helped me understand the two main types of evaluation.	5.00 (SD = 0)
Webinar #2: "Moving From Design Through Data Collection" post-training survey results (n = 6)	
<i>Information covered</i>	
1 = Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
The information covered is relevant for our SANE program.	4.50 (SD = 0.548)
The information was provided in an user-friendly manner	4.50 (SD = 0.548)
I plan to use the information from the training in our evaluation.	4.83 (SD = 0.408)
The information provided was clear and understandable.	4.83 (SD = 0.408)
The information provided covered the areas needed for our program to complete an evaluation.	4.67 (SD = 0.516)
<i>Preparation for evaluation</i>	
1 = Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
The training helped me understand my role in designing and carrying out an evaluation.	4.67 (SD = 0.516)
The training helped me understand the types of evaluation questions we can investigate.	4.33 (SD = 0.516)
The training helped me understand how to design an evaluation.	4.67 (SD = 0.516)
The training helped me plan for how to build the relationships I need in order to carry out the evaluation (for example, with the prosecutor's office).	4.83 (SD = 0.408)
The training helped me understand how to sample cases for the evaluation.	4.67 (SD = 0.516)
Webinar #3: "Analyzing Your Data" post-training survey results (n = 7)	
<i>Preparation for using the data entry program</i>	
1 = Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
The training helped me understand how rows and columns are used in a spreadsheet.	4.00 (SD = 0.817)
The training helped me understand how to transfer the information I am collecting at the prosecutor's office into the spreadsheet.	4.43 (SD = 0.535)
The training helped me understand what I need to do to calculate the results after I have entered all the information from the prosecutor's office.	4.43 (SD = 0.787)
The training helped me understand the results I will get from the prosecutor's office.	4.57 (SD = 0.535)
<i>Expectations for the data entry program</i>	
1 = Strongly disagree; 2 = Disagree; 3 = Neutral; 4 = Agree; 5 = Strongly agree	
It will be easy to enter the information I collected at the prosecutor's office into the spreadsheet.	4.43 (SD = 0.535)
I will need to get additional help with entering the information from the prosecutor's office into the spreadsheet.	1.86 (SD = 0.690)
The way to enter information into the spreadsheet (i.e., type it in or choose from a dropdown list) will help to avoid mistakes in data entry.	4.29 (SD = 0.488)
It will be helpful to have an error message come up when there is missing data.	4.28 (SD = 0.488)
The graphs on the results page of the program will be easy to understand.	4.42 (SD = 0.535)

since site visit interview; (2) reactions from community stakeholders to the evaluation findings; (3) programmatic and/or community changes that stemmed from the evaluation findings; and (4) future plans for utilization. The phone interviews were also recorded and transcribed.

2.7. Evaluation team data analytic procedures

Quantitative data were analyzed with simple descriptive statistics (percentages, means, standard deviations). For the qualitative data, we compiled our field notes and interview transcripts to conduct a conventional content analysis of the key themes that emerged in the data (Hsieh & Shannon, 2005). Consistent with the procedures outlined by Hsieh and Shannon (2005), the narrative data were reviewed multiple times to immerse the coders in the data; coders recorded their initial impressions and developed preliminary codes to capture key themes related to the phenomena of interest (i.e., instrumental use). Codes were not developed a priori, but were emergent from the data themselves. The coding was conducted independently by two members of the evaluation team (the project director and the principal investigator), and then findings were reviewed and verified by a third member of the team (the co-investigator). Disagreements among analysts were noted, the data were re-checked, and a group consensus approach was used to reach final coding/interpretation decisions (see MacQueen, McLellan-Lemal, Bartholow, & Milstein, 2008).

Three primary themes emerged in the data regarding the nature of instrumental use exhibited across the six sites. First, we noted instances of when the sites *shared their findings with community partners* (i.e., the sites' plans to share their findings and their experiences of actually sharing their evaluation findings). For example, when one SANE noted, "*We [our multi-disciplinary team] met, we went over statistics [evaluation findings],*" this text was coded as 'instrumental use – sharing findings.' Second, we coded the *sites' plans to use their findings* (i.e., their intended next steps to make changes in their communities, based on their evaluation findings). For example, when one program staff member stated in a site visit interview, "*[our multi-disciplinary team] has some work to do to...understand what the data means,*" this text was coded as 'instrumental use – plans to use findings.' Finally, we coded descriptions of *actual use of findings* (as opposed to intentions to use findings) (i.e., actual changes that have been implemented in their programs/communities that are directly tied to the evaluation findings). For instance, when a SANE noted that, "*[our multi-disciplinary team] is now...meeting quarterly [as a result of the findings],*" this text was coded as 'instrumental use – actual use of findings.' The quotes selected for inclusion in this manuscript reflected sentiments expressed across program sites (i.e., the quote could have come from one of many SANE program personnel, as multiple people stated that same idea).

3. Results

3.1. Sites' perceptions of the utility of the resources provided

Given that our primary goal in this project was to facilitate successful completion of methodologically rigorous evaluations in all six sites and that the sites used the findings to create program and/or community-level improvement (i.e., instrumental change), we assessed whether the five main forms of technical assistance we provided (i.e., the toolkit and pre-programmed Excel files; webinars; group conference calls; individual site consultation; and individual site visits) were in fact helpful to program staff. In the qualitative site visit interviews, program staff reported that the tangible resources provided to the sites were construed as critical

resources to their efforts. Narrative data supporting this conclusion are presented below, organized by the feedback we received about each of the five technical assistance modalities.

With respect to the *toolkit* itself, program staff particularly appreciated the step-by-step directions, screenshots of the MS Excel analysis program, and sample data collection tables. For example, one SANE coordinator said:

"I like that it [the toolkit] was broken down into the steps. I liked how the circles [showed the steps], and then you had the one [identifying the current step] it was nice to see where [you were], what your reading was in context with, the bigger picture at all times" – MID-SIZED PROGRAM

Program staff liked that the toolkit was an all-inclusive reference book. In the context of this project, they could (and often did) call the evaluation team when they had questions, but they also realized that the answers they needed were in fact in the toolkit itself:

"The book is very user friendly...I'm a see-touch-do type person so that was helpful. I have referred back to it...if I had questions I could find the answers in there pretty much." – RURAL PROGRAM

The *accompanying pre-programmed MS Excel data analysis file* was undoubtedly the most appreciated resource provided to the sites. Program staff were concerned the data analysis and interpretation step would be time-consuming and challenging, especially if they did not have prior training in statistics. As one site noted:

"But I like that I didn't have to do the statistics 'cause I've never taken that class; doesn't sound like I want to either. Yeah, that's really cool how you guys just created that spreadsheet and you click when you're done, that's really clever." – RURAL PROGRAM

Feedback on each of the three training *webinars* was attained through quantitative post-training online surveys (see Table 3), which showed consistent positive feedback regarding the utility of this technical assistance modality in helping sites prepare and conduct their evaluations. In our site visit interviews, program staff also reported that they liked the webinars as they reinforced what was presented in the toolkit and helped bring the overall process into perspective. The webinars were prescheduled and sites reported that this helped to ensure they were making appropriate progress through the steps of the evaluation.

The *group conference calls* were another technical assistance strategy that was well-received by the sites. The conference calls helped to reinforce what had been presented in the toolkit and also helped ensure that sites were progressing through the steps in a timely way. Sites also found it helpful to interact with one another, to hear about others' experiences in implementing the evaluation, and challenges other sites were facing. This setting also allowed for group question and answer sessions. Sites reported that this was beneficial in that other sites may ask questions they had not yet thought of or did not feel comfortable asking. For instance, one SANE program coordinator said:

"It was really beneficial to listen to what was going on with the other sites...hearing (about others') challenges was really helpful because it made us all think outside the box" – RURAL PROGRAM

Individual technical assistance (provided by email or phone) was useful in addressing site-specific challenges. All sites reported the toolkit provided sufficient instruction to do the evaluation project independently, but they liked being able to contact a specific individual to answer questions, provide assistance, and encourage them throughout the implementation of the evaluation. As one program staff member reported:

“We’ve got 50 million other things to do to keep our program running and evaluation is, wow, I’d really love to do it, but I don’t have the time, I don’t have the money. . . So having a person there makes it more feasible for me to say, you know what, I’m going to bite this one off because there’s somebody there that I can call when I get stuck.” – URBAN PROGRAM

In the context of this project, we had the resources to conduct *site visits* with all six programs. We visited all six programs near the end of the project during the data analysis, interpretation, and utilization planning stages.⁸ Although the pre-programmed Excel file substantially reduced the labor for this task, program staff felt it was helpful to have a chance to talk through the results and plan next steps:

“[When you go through all the evaluation steps you want] the opportunity to have a conversation with somebody. . . to say, okay so that all looks great and they’re pretty charts but tell me what this means.” – MID-SIZED PROGRAM

Site visits may not be financially feasible in all projects, but program staff felt very strongly that the opportunity to have personal one-on-one contact with an experienced evaluator was essential, whether in person or by phone.

Overall, sites reported that the multiple modalities of technical assistance used to convey the evaluation information was helpful. Program staff liked that they were able to read the information in the toolkit, see it in the webinar, and talk about it during the conference calls, and that this reinforcement of material was crucial to their learning.

3.2. Sites’ completion of their local-level evaluations

As noted previously, our primary goal in this project was to guide these six programs through successful completion of methodologically rigorous evaluations. All sites followed the rigorous sampling and data collection procedures outlined in the toolkit, and as shown in [Table 4](#), three successfully conducted pre–post evaluations and three conducted post-only evaluations. All six sites had completed their evaluations within 12–14 months. Program staff expressed pride in their accomplishment and noted that they learned a tremendous amount through the experiences of being in this project, as two sites explained:

“I can come away from this project with that better understanding [of evaluation and research] and I also think if my board were to say, we need to do an evaluation of this or that or the other, I have a better sense of what that means and what it needs to look like depending on how formal of an evaluation they want done.” – URBAN PROGRAM

Although the substantive findings of these evaluations are not the focus on this paper, we summarize key results in [Table 4](#) for context, as these outcomes could certainly have impact on later instrumental use of the findings. The substantive results regarding sexual assault prosecution rates were remarkably consistent – and consistently negative – across all six sites. Overall, the vast majority of sexual assault cases in which victims had a medical forensic exam and reported to the police were never prosecuted (86% across the six sites), typically because the police never

forwarded the case to the prosecutors for consideration of charges.⁹ Within each individual site that completed a pre–post evaluation design, there were no significant increases in prosecution rates post-SANE (however, aggregate analyses did show a significant increase post-SANE, see [Campbell et al., 2014](#)). Though these results are consistent with the larger literature on the criminal justice system’s response to sexual violence ([Campbell et al., 2014](#)), these findings were a shock to program staff across all six sites. As one SANE explained:

“I was expecting that there would be many, many more referrals [for prosecution]. . . you know you think you have a strong program and it’s having positive outcomes on prosecutions and then when you see that it’s not having that impact, you’re like, *oh*” – URBAN PROGRAM

The successful completion of the local-level evaluations – with positive regard for the process – is particularly noteworthy, given that the substantive findings were not positive.

3.3. Sites’ instrumental use of their findings

Research on evaluation use has shown that negative findings can sometimes curb instrumental use, given stakeholders’ concerns that the program and its constituents could be perceived in a negative light ([Cousins, 2003](#); [Cousins & Chouinard, 2012](#); [Cousins & Whitmore, 1998](#); [Henry, 2000](#); [Patton, 2008](#); [Weiss, 1998](#)). However, the literature also suggests that with sufficient support, resources, and encouragement from evaluators throughout the process, instrumental use can occur, regardless of whether the findings are viewed as positive or negative (see [Cousins, 2003](#); [Patton, 2008](#)). In this project, we saw strong evidence of instrumental use in all six sites; narrative data supporting this conclusion are presented below, organized by three main types of use documented: sharing findings with community partners; developing plans to use their findings to enact community change; and creating changes in practice and policy as a direct result of the evaluation findings.

All programs made plans to *share their evaluation findings* with their home institutions and community partners, and most sites were able to accomplish this task within four months of obtaining their results. Programs developed either PowerPoint presentations summarizing their results, or informal “talking points” that highlighted key findings. Some programs shared their findings in their multidisciplinary team (MDT) meetings that included all partnering agencies:

“Well, we’re going to definitely share the results. . . in the 3½ years I’ve been here we’ve never really talked about data and so. . . let’s start slowly and sort of get them interested; understand the data and then in subsequent meeting begin to talk about – introduce the subject of, okay now we have this information about prosecution rates in [location], what do we want to do? Let’s think about what we want to do? Are there changes that we could implement?” – URBAN PROGRAM

Other sites met with stakeholder groups individually (e.g., prosecutor’s office, law enforcement, advocates, etc.) rather than as a collective group in order to address agency-specific problems revealed through the evaluation.

⁸ One of the urban sites had an additional site visit early in the project as they needed individualized on-site help with the early stages of sampling, given the large volume of cases treated in that program. There were no differences between this site and the other five regarding perceived utility of the technical assistance, the overall quality of the resulting evaluation conducted, the substantive findings of the evaluation, or the instrumental use of the findings.

⁹ When data were aggregated across the six sites, the results showed a significant, positive increase in prosecution pre-SANE post-SANE, but still, the overwhelming majority of adult sexual assault cases were *not* prosecuted. In other words, there was significant improvement, but the change was from *very* low prosecution rates to still quite low (just not *as* low) (see [Campbell et al., 2014](#)).

Table 4
Summary of the sites' evaluation designs and substantive findings.

	Design	Number of years pre	Number of years post	Sample size pre	Sample size post	Summary of site-specific key substantive findings
Site A (rural)	Pre/post	3	5	47	92	Pre: 89% were not prosecuted Post: 82% were not prosecuted Non-significant change
Site B (rural)	Post-only	(NA)	9	(NA)	253	Post (only): 91% were not prosecuted
Site C (mid-sized)	Pre/post	3	10	65	334	Pre: 96% were not prosecuted Post: 89% were not prosecuted Non-significant change
Site D (mid-sized)	Post-only	(NA)	7	(NA)	201 ^a	Post (only): 94% were not prosecuted
Site E (urban)	Pre/post	2	10	60 ^a	300 ^a	Pre: 95% were not prosecuted Post: 91% were not prosecuted Non-significant change
Site F (urban)	Post-only	(NA)	12	(NA)	344 ^a	Post (only): 86% were not prosecuted

^a Sampled 30 cases per calendar year.

After sharing the evaluation findings with community partners, many sites made *plans for how they could use the findings to enact change*. Some sites were not able to execute their plans before the evaluation project ended, but they had readily identified their next steps (e.g., reaching out to new community partners to address gaps in services revealed through the evaluation). Others were able to execute their plans for change rather quickly, which was quite remarkable as it indicates that programs were able to share, interpret, and utilize their evaluation findings in only a few months. For instance, in several sites, the evaluation findings illustrated the need for more frequent MDT meetings, regular sexual assault case reviews by the MDT, and the involvement of more law enforcement officers in the MDT. One site noted that the evaluation findings directly led to a change in MDT practice in their community:

“Before that we didn’t meet very often. Now we are meeting quarterly [as a result of sharing the findings] to try to put new protocols or education awareness.” – RURAL PROGRAM

Similarly, other sites used the evaluation findings to kick-start efforts to develop community-wide sexual assault response protocols. One site described how the evaluation findings were used for protocol development:

“I believe the [evaluation] findings really started the process. . . I know in the past that they had tried to. . . start a protocol and it never went through. . . And now I think with the statistics that it’s out there and all the cases that we are doing, now they are looking at [it] deeper.” – RURAL PROGRAM

Other sites used the evaluation findings to apply for grants to improve services for sexual assault victims in their communities. For instance, one site used the evaluation findings to apply for a federal grant to secure a sexual assault-specific law enforcement investigator:

“[We submitted] a grant to have one specific officer to respond to all of the domestics and sexual assaults versus whichever officer [is available]. . . It’s a federal grant. . . that will actually pay for his or her, the officer’s salary, benefits, education.” – RURAL PROGRAM

Another site used the findings to apply for a grant to fund sexual assault-related training for law enforcement personnel:

“A really great outcome is after sharing that [the evaluation findings]. . . we decided as a group that we needed to have some sexual assault investigative training for the department to make sure they were at least up to speed on current investigation techniques and tactics. . . now (they understand) that it is an actual issue [low prosecution rates and lack of training] and I

don’t think they recognized it before until we brought it to their attention, which is stunning” – MID-SIZED PROGRAM

For both of these communities, the grant applications were a direct result of the evaluation findings. As one program staff member explained:

“[Without the evaluation] we would not have had the data to support why it was so important. We would not have been able to say look, we’ve done a study, we’ve looked at our outcomes for X number of years. This is what’s actually happening, and clearly it’s a problem. . . you know nobody ever really likes to think that they’re doing the best job where they are. . . it’s just that they need to identify, maybe we have some weaknesses we need to work on. So that helped, it really helped.” – MID-SIZED PROGRAM

The fact that these sites were able to utilize the evaluation findings in such a short period of time suggests that additional change efforts may also be implemented over time.

4. Discussion

SANE programs are a national-scale intervention model for post-assault care of rape victims. These programs have disseminated rapidly throughout the United States (and other nations) and have been designated as “best practice,” despite limited evaluation data regarding their effectiveness (DOJ, 2004). To address this science-practice gap, we created and implemented a practitioner-oriented toolkit that could help local-level programs evaluate the impact of their services on sexual assault prosecution rates in their community, thereby building a larger knowledge base about this intervention model and improving practice.

Feedback from the six sites that participated in this study reported that the tangible resources provided (e.g., toolkit, MS Excel data analysis program) were not only helpful and user-friendly, but also alleviated many of their concerns about conducting an evaluation. The less tangible resources provided, including interactions with other program sites and the evaluation, were also well-received by sites. The technical assistance provided peer-to-peer support and encouragement, space for group question and answer sessions, and site-specific services. Furthermore, these resources were directly tied to the programs’ successes in developing, implementing, and completing methodologically rigorous evaluations. Program staff specifically stated that because of the resources provided (both the toolkit itself and the technical assistance), they were able to complete their planned evaluations. The observational data we collected throughout the project (e.g., field notes) confirmed the SANES’ interview accounts.

A primary goal for this project was to examine instrumental use and the extent to which the toolkit and technical assistance could help programs act upon their findings to create program-level and community-level changes. The extant literature on evaluation toolkits suggests that this approach to evaluation capacity building can promote instrumental use, but the specifics of how that happens and what it might look like have not been well articulated. In this project, we identified three specific manifestations of instrumental use. All sites shared their findings with their community partners, and given that the substantive results were generally negative, this alone is an important indicator of instrumental use, given that positive findings are typically more likely to be disseminated than negative results (Patton, 2008). After sharing the evaluation findings with their community partners, program staff made action plans to use the findings and then actually brought those plans to fruition to strengthen multidisciplinary collaborations (e.g., more regular MDT meetings), community-wide practice (e.g., develop community-wide response protocol), and apply for grants (e.g., to provide law enforcement training or personnel).

It is important to note, however, that because we did not use an experimental or quasi-experimental design in our study of whether the toolkit and technical assistance contributed to instrumental use, we cannot make a causal attribution that the resources we provided produced the observed changes in the six sites. We also cannot determine whether it was the toolkit itself or the toolkit plus the technical assistance package that was associated with the instrumental use documented in this project. Indeed, other toolkit evaluation projects have also included accompanying technical assistance to programs, so it is not clear whether a workbook alone would be sufficient to instruct program staff, motivate them through the evaluation process, and support instrumental use. Though we did not explicitly compare toolkit only vs. toolkit-plus-technical assistance, we suspect that the additional training and support was critical to programs (see Section 6, below). Future research on evaluation toolkits could explicitly compare the nature and dosage of assistance provided to sites, using experimental or quasi-experimental manipulations, which would provide useful information on evaluation capacity building and resulting instrumental use (see Hunter, Ober, Paddock, Hunt, & Levan, 2014 as an example).

We also acknowledge that the scale of this project is small relative to the number of SANE programs currently in operation in the United States (6–600) and that the six programs are not representative of the population of U.S. SANE programs. This study highlights that is possible for SANE programs with evaluation readiness to complete good quality evaluations, with technical assistance and support, but not all SANE programs have such readiness (and it was beyond the scope of this project to conduct a national assessment of program readiness). Although we identified and resolved many glitches in the toolkit, it is unlikely that we were able to identify all challenges SANE programs may face when attempting local-level evaluations, and we anticipate that future revisions of our materials will be necessary. The revised toolkit and accompanying resources were nationally disseminated through the National Criminal Justice Reference Service (www.ncjrs.gov), and as more SANE programs utilize this resource, we will learn more about the challenges facing practitioners when they embark upon evaluation activities.

5. Conclusion

Building evaluation capacity requires strengthening knowledge, skills, and attitudes about evaluation (Preskill & Boyle, 2008). The results of this study suggest that a practitioner-oriented toolkit, plus extensive technical assistance and support, can help

local programs conduct methodologically rigorous evaluations that results in instrumental use of the findings.

6. Lessons learned

We crafted the toolkit to guide SANE programs through the completion of a pre-SANE/post-SANE design, which is a common approach in outcome evaluations, but one that has been underutilized in the literature on the effectiveness of SANE programs. However, given that many SANE programs were established in the 1990s (Campbell, Townsend, et al., 2005; IAFN, 2014), a sizable proportion have been in existence for 10+ years – as were the programs that participated in this project. Given such longevity, a pre–post design poses some significant practical challenges. Program staff spent considerable time trying to locate records (often stored in remote locations) and then trying to understand and interpret changes in key paperwork over time. The three sites who completed post-only evaluation designs were also saddled with the task of finding old records, as these SANE program had been in operation for 7–12 years (i.e., the start of their “post” was nearly a decade ago). Furthermore, digging up old records, old practices, and old arguments among community stakeholders did not foster positive relationships in the here-and-now. When one of the SANES remarked in an all-site conference call that she wished the evaluation could just “start from right now” – meaning *not* tracking down old records and starting the evaluation from this point forward – we realized we needed to give programs more design options. We had designed the toolkit and technical assistance package around a methodologically rigorous design, which is not an unreasonable choice in terms of advancing the knowledge base about the effectiveness of this intervention, but it may not always be practical or useful to local-level stakeholders. Program staff wanted more design options for learning about their programs. Based on that insight, we revised the toolkit prior to its national dissemination in the National Criminal Justice Reference Service to include more flexible design options, including a new module called “From this Point Forward” for prospective program monitoring.

A second key lesson learned in this project is that fostering communication between sites in a multi-site project is critical for building an evaluation learning community (see also Toal et al., 2008). The six programs were geographically dispersed throughout the United States, so we had to form our learning community via conference calls, emails, and webinars (we had one in-person meeting at the discipline’s annual conference, but not all sites could attend). Although sites reported being satisfied with their communications with the evaluation team and their sister sites, they noted that even more program-to-program contact would have been welcomed. The learning community provided important social support and they would have liked more contact with each other throughout the project. For multi-site projects in particular, it might be useful for evaluators to develop multiple methods – above and beyond “standard” methods of conference calls and emails – for remote connection between sites. Online teaching platforms (such as Desire2Learn [D2L] and others) could be configured for an evaluation project, which offers multiple options for regular communication between learners.

Perhaps the most important lesson learned in this project is that the toolkit itself would likely not have been enough to create the level of instrumental use we documented in the six sites. The toolkit was designed to be a stand-alone resource with detailed instructions, examples, problem solving strategies, and ideas for post-evaluation instrumental use. Program staff noted that the toolkit did indeed contain the answers to their questions, but what they wanted and needed was being able to ask a person – a seasoned evaluator – their questions. As first-time evaluators, they

did not yet have a feel for what was “normal” and they did not yet trust their judgment to make on-the-ground decisions. Being able to call the evaluation team to talk through options was critically important to program staff – and this consultation did help prevent some methodological problems. As such, our experiences suggest that evaluation toolkits are useful resources for building evaluation capacity, but they are likely not sufficient for creating a learning community, for providing social support, and for guiding programs through instrumental use. Toolkits *plus* technical assistance contact with evaluators may be necessary to promote evaluation success and evaluation use.

Acknowledgements

This project was supported by Grant No. 2009-MU-MU-0002 awarded by the National Institute of Justice, Office of Justice Programs and U.S. Department of Justice. Opinions, findings, and conclusions or recommendations expressed in this publication/program/exhibition are those of the author(s) and do not necessarily reflect those of the Department of Justice.

References

- Black, M. C., Basile, K. C., Breiding, M. J., Smith, S. G., Walters, M. L., Merrick, M. T., et al. (2011). *The National Intimate Partner and Sexual Violence Survey (NISVS): 2010 Summary Report*. Atlanta, GA: National Center for Injury Prevention and Control Centers for Disease Control and Prevention.
- Burford, G., Velasco, I., Janouskova, S., Zahradnik, M., Hak, T., Podger, D., et al. (2013). Field trials of a novel toolkit for evaluating ‘intangible’ values-related dimensions of projects. *Evaluation and Program Planning*, 36, 1–14.
- Campbell, R., Bybee, D., Shaw, J. L., Townsend, S. M., Karim, N., & Markowitz, J. (2014). The impact of sexual assault nurse examiner (SANE) programs on criminal justice case outcomes: A multi-site replication study. *Violence Against Women*, 20, 607–625.
- Campbell, R., Dorey, H., Neageli, M., Grubstein, L. K., Bennett, K. K., Bonter, F., et al. (2004). An empowerment evaluation model for sexual assault programs: Empirical evidence of effectiveness. *American Journal of Community Psychology*, 34, 251–262.
- Campbell, R., Patterson, D., & Bybee, D. (2012). Prosecution of adult sexual assault cases: A longitudinal analysis of the impact of a Sexual Assault Nurse Examiner program. *Violence Against Women*, 18, 223–234.
- Campbell, R., Patterson, D., & Lichty, L. F. (2005). The effectiveness of sexual assault nurse examiner (SANE) program: A review of psychological, medical, legal, and community outcomes. *Trauma, Violence, & Abuse*, 6, 313–329.
- Campbell, R., Townsend, S. M., Bybee, D., Shaw, J. L., & Markowitz, J. (2013). *Implementation of a Sexual Assault Nurse Examiner (SANE) practitioner evaluation toolkit*. Washington, DC: National Institute of Justice.
- Campbell, R., Townsend, S. M., Long, S. M., Kinnison, K. E., Pulley, E. M., Adames, S. B., et al. (2005). Organizational characteristics of sexual assault nurse examiner programs: Results from the national survey of SANE programs. *Journal of Forensic Nursing*, 1, 57–64.
- Chinman, M., Hunter, S. B., & Ebener, P. (2012). Employing continuous quality improvement in community-based substance abuse programs. *International Journal of Health Care Quality Assurance*, 25, 604–617.
- Chinman, M., Hunter, S. B., Ebener, P., Paddock, S. M., Stillman, L., Imm, P., et al. (2008). The Getting to Outcomes demonstration and evaluation: An illustration of the prevention support system. *American Journal of Community Psychology*, 41, 206–224.
- Chinman, M., Tremain, B., Imm, P., & Wandersman, A. (2009). Strengthening prevention performance using technology: A formative evaluation of interactive Getting to Outcomes. *American Journal of Orthopsychiatry*, 79, 469–481.
- Cousins, J. B., & Chouinard, J. A. (2012). *Participatory evaluation up close: An integration of research-based knowledge*. Charlotte, NC: Information Age Publishing Inc.
- Cousins, J. B. (2003). Utilization effects of participatory evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (pp. 245–265). Boston: Kluwer.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, 80, 5–23.
- Crandall, C., & Helitzer, D. (2003). *Impact evaluation of a Sexual Assault Nurse Examiner (SANE) program. Award Number 98-WT-VX-0027*. Washington, DC: National Institute of Justice.
- Department of Justice (2013). *A national protocol for sexual assault medical forensic examinations: Adults/adolescents* (2nd ed.). Washington, DC: Author.
- Department of Justice (2004). *A national protocol for sexual assault medical forensic examinations: Adults/adolescents*. Washington, DC: Author.
- Dyckhoff, A. L., Zielke, D., Bultmann, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society*, 15, 58–76.
- Emerson, R. M., Fretz, R. L., & Shaw, L. L. (1995). *Writing ethnographic fieldnotes*. Chicago: University of Chicago Press.
- Fetterman, D. M., & Wandersman, A. (Eds.). (2005). *Empowerment evaluation principles in practice*. New York: Guilford Press.
- Henry, G. T. (2000). Why not use? *New Directions for Evaluation*, 88, 85–98.
- Hsieh, H., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15, 1277–1288.
- Hunter, S. B., Chinman, M., Ebener, P., Imm, P., Wandersman, A., & Ryan, G. W. (2009). Technical assistance as a prevention capacity-building tool: A demonstration using the Getting to Outcomes framework. *Health Education & Behavior*, 36, 810–828.
- Hunter, S. B., Ober, A. J., Paddock, S. M., Hunt, P. E., & Levan, D. (2014). Continuous quality improvement (CQI) in addiction treatment settings: Design and intervention protocol of a group randomized pilot study. *Addiction Science & Clinical Practice*, 9, 1–11.
- IAFN (2014). *Data base of the International Association of Forensic Nurses*. Retrieved from <http://www.forensicnurse.org>
- IAFN (2009). *Forensic nursing: Scope and standards of practice*. Silver Spring, MD: American Nurses Association.
- Kilpatrick, D. G., Resnick, H. S., Ruggiero, K. J., Conoscenti, L. M., & McCauley, J. (2007). *Drug-facilitated, incapacitated, and forcible rape: A national study*. Washington, DC: National Institute of Justice.
- MacQueen, K. M., McLellan-Lemal, E., Bartholow, K., & Milstein, B. (2008). Team-based codebook development: Structure, process, and agreement. In G. Guest & K. M. MacQueen (Eds.), *Handbook for team-based qualitative research* (pp. 119–136). Lanham, MD: Altamira/Rowman & Littlefield.
- Oliver, M., MacBean, J., Conole, G., & Harvey, J. (2002). Using a toolkit to support the evaluation of learning. *Journal of Computer Assisted Learning*, 18, 199–208.
- Patton, M. Q. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.
- Patton, M. Q. (2002). *Qualitative research & evaluation methods* (3rd ed.). Thousand Oaks, CA: Sage.
- Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29, 443–459.
- Preskill, H., & Torres, R. T. (2000). *Readiness for organizational learning and evaluation instrument (ROLE)*. Claremont, CA: Authors.
- Russ-Eft, D., & Preskill, H. (2001). *Evaluation in organizations: A systematic approach to enhancing learning, performance, and change*. New York: Basic Books.
- Toal, S. A., King, J. A., Johnson, K., & Lawrenz, F. (2008). The unique character of involvement in multi-site evaluation settings. *Evaluation and Program Planning*, 32, 91–98.
- Wandersman, A., Imm, P., Chinman, M., & Kaftarian, S. (2000). Getting to outcomes: A results-based approach to accountability. *Evaluation and Program Planning*, 23, 289–395.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., et al. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology*, 41, 171–181.
- Weiss, C. H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19, 21–33.
- WW Kellogg Foundation (2010). *Evaluation handbook*. Battle Creek, MI: Author.

Rebecca Campbell, Ph.D., is a Professor of Psychology at Michigan State University. Her research examines how the legal, medical, mental health systems and rape crisis centers respond to the needs of sexual assault victims. Her evaluation practice focuses on improving community services for victims of sexual assault.

Stephanie M. Townsend, Ph.D., is President of Townsend Consulting & Evaluation. As a community psychologist, she specializes in developing practice-based evidence in the areas of sexual violence prevention and community responses to sexual violence.

Jessica Shaw, M.A., is a doctoral candidate in the Ecological-Community Psychology program at Michigan State University. Her research interests include community responses to sexual assault within the medical and criminal justice systems and program and policy evaluation.

Nidal Karim, Ph.D., works at CARE as a Senior Advisor for Gender and Empowerment Impact Measurement. She has also worked as a Behavioral Scientist at the Division for Violence Prevention at CDC.

Janifer Markowitz, ND, RN, WHNP-BC, is a forensic nursing consultant who specializes in issues related to sexual assault and domestic violence, including medical-forensic examinations and professional education and curriculum development. In addition to teaching internationally, she provides expert testimony, case consultation, and technical assistance; and develops training materials, resources, and publications for a variety of military and civilian clients.