



Methodological Challenges of Collecting Evaluation Data From Traumatized Clients/Consumers

A Comparison of Three Methods

Rebecca Campbell

Adrienne E. Adams

Michigan State University

Debra Patterson

Wayne State University

This project integrates elements of responsive evaluation and participatory evaluation to compare three evaluation data collection methods for use with a hard-to-find (HTF), traumatized, vulnerable population: rape victims seeking postassault medical forensic care. The first method involves on-site, in-person data collection, immediately postservices; the second, telephone follow-up assessments, 1 week postservices; and the third, private, self-administered surveys completed immediately postservices. There are significant differences in response rates across methods: 88% in-person, 17% telephone, and 41% self-administered. Across all phases, clients gave positive feedback about the services they received and about all three methods of data collection. Follow-up analyses suggested that nonresponders did not differ with respect to client characteristics, assault characteristics, or nursing care provided. These findings suggest that evaluations with HTF service clients may need to be integrated into on-site services because other methods may not yield sufficient response rates.

Keywords: *hard-to-find populations; responsive evaluation; participatory evaluation; rape; sexual assault*

Research and evaluation with hard-to-find (HTF), traumatized, and vulnerable populations involves multiple methodological challenges (Benoit, Jansson, Millar, & Phillips, 2005; Campbell, Sefl, Wasco, & Ahrens, 2004; Liamputtong, 2007; Miller & Campbell, 2008; Rosenbaum & Langhinrichsen-Rohling, 2006). These groups are often HTF because they are socially or physically disenfranchised (e.g., urban poor, homeless, undocumented immigrants), stigmatized (e.g., drug users, LGBT, sex workers), and/or traumatized (e.g., battered women, rape survivors; Birman, 2005; Miller, Forte, Wilson, & Greene, 2006; Sullivan & Cain, 2004). Individuals from such groups may at some point come into contact with community-based organizations (CBOs), such as public aid programs, shelters, crisis centers, and treatment facilities. When HTF individuals surface at social service agencies, evaluators may be able to collect data from them to learn about the service needs of these groups.

Sampling HTF populations can be frustrating for researchers/evaluators because it can be quite difficult to obtain nonbiased samples at CBOs (Fleishman, Mor, Cwi, & Piette, 1992; Heckathorn, Broadhead, Anthony, & Weakliem, 1999; Miller & Campbell, 2008; O'Connell,

Authors' Note: Correspondence concerning this article should be addressed to Rebecca Campbell, PhD, Department of Psychology, Michigan State University, 127C Psychology Building, East Lansing, MI 48824-1116; e-mail: rnc@msu.edu.

2000). Although there has been promising development in sampling techniques for HTF populations (e.g., respondent-driven sampling, Heckathorn, 1997, 2002; time-space sampling, Muhib et al., 2001), for many CBOs, the more pressing issue is trying to collect data from their current clients for program improvement. Program staff often need information directly from clients regarding the utility of services and unmet needs (Garland, Saltzman, & Aarons, 2000; Molnar & Stup, 1994; Sanders, Trinh, Sherman, & Banks, 1998). However, there is real concern that collecting data could interfere with service delivery and strain the tentative relationship between the client and CBO (Riger et al., 2002; Sullivan & Cain, 2004). These clients are already sensitive to being questioned, blamed, and traumatized—will they tolerate more questions? Not to suggest that sampling is unimportant, but in some situations, such as evaluating new services for HTF groups, collecting any data could be daunting.

This raises a different set of methodological challenges for researchers/evaluators working with vulnerable groups—we have access at the CBO, but what is the best method for collecting data? In this situation, researchers/evaluators would likely consider archival, nonintrusive observational methods. But CBOs often want feedback directly from their clients, and certainly there is a long-standing value for consumers' input in evaluation scholarship and practice (Lebow, 1982, 1983; Neigher & Schulberg, 1982; Ware & Davies, 1983). In that context of demand for data directly from HTF, traumatized, and/or vulnerable clients, what method will yield the highest response rate, and at the same time, be well received by clients? This is the methodological challenge we examined in this project, and our setting was a CBO that serves rape victims immediately after they have been raped.

The Context of Our Project

The CBO for this evaluation was a sexual assault nurse examiner (SANE) program, which provides services to victims of sexual assault/rape including crisis intervention, medical care, and state-of-the-art forensic exams for evidence collection. Traditionally, this care had been provided in hospital emergency departments, but given the frenetic pace of these settings, many victim advocates and health care providers felt that alternative programs were needed so that victims could have individualized care in a calm, quiet place. As a result, the nursing profession created SANE programs whereby specially trained forensic nurses work with victims in either hospital clinic-type settings or community-based centers to provide comprehensive medical-forensic care. In addition, many SANE programs also have volunteer victim advocates who focus on responding to the emotional needs of victims and their families during the exam process (see Campbell, Patterson, & Lichty, 2005, for a review). Throughout the 1990s, SANE programs spread quickly and there are now approximately 475 in existence in the United States (International Association of Forensic Nurses, 2008). Although SANE programs are widely considered to be the best-practice model of care (see Littell, 2001), there have been very few evaluations of their services (Campbell et al., 2005).

Evaluating the work of SANE programs presents several challenges, which may explain why there are so few studies. First and foremost, the programs' clients are highly vulnerable and traumatized as most have been raped within the past 72 hr and the majority within the past 12 hr (Campbell, Patterson, Bybee, & Diegel, 2007). Immediately postassault, many victims exhibit acute trauma symptoms including fear, anxiety, anger, dissociation, and difficulty concentrating. Therefore, it is necessary to consider carefully if and how an evaluation can be conducted in ways that are not burdensome and yet yield information of sufficient quality.

Second, the SANE nurses obtain a medical history prior to the exam, which includes detailed questions about nature of the assault, the parts of the victims' bodies that were

violated (and exactly how), and the victims' health needs. In addition, many victims also have contact with the police before their SANE exam. Police officers also ask victims detailed questions about the assault, and it is not uncommon that they ask some questions over and over again to check for consistency in the victims' accounts. As a result, during the process of the police report and forensic exam, victims must endure hours of questioning, and when this is finally over, most want nothing more than to go home immediately and be left alone. Quite simply, victims may not want to answer even more questions for the purpose of evaluation.

Third, many program clients may not come back for follow-up services. Previous research on traditional hospital emergency department care suggests that only 23% to 31% of sexual assault victims complete follow-up medical care (Boykins & Mynatt, 2007; Holmes, Resnick, & Frampton, 1998), which is not entirely surprising given the nature of what happens during their first visit. Although SANE programs appear to have addressed many of the shortcomings of traditional hospital care, it is still quite likely that many victims just want to put this experience behind them and move on (Herman, 1992). Therefore, it is reasonable to assume that many clients will not have return follow-up contact with SANE programs. As a result, many clients may be one-time users of program services, and there may not be other opportunities to capture program clients for the purpose of evaluation.

In view of these challenges, our approach integrated elements of responsive evaluation (Abma & Stake, 2001; Stake, 2004) and participatory evaluation (Cousins & Earl, 1992, 1995; Cousins & Whitmore, 1998).^{1,2} Stake and Abma (2005) noted that responsive evaluation embodies "responsiveness to key issues or problems, especially those recognized by people at the sites" (p. 376). Such an approach was appropriate, and in our estimation, fully necessary, because service provision is so delicate in this context, the clients are highly traumatized, and program staff have unique expertise about working with this population. Consistent with this approach, the researchers/evaluators spent approximately 1 year becoming acquainted with the history and operations of this program, which reaffirmed that the nurses and advocates' insider perspectives were essential for figuring out how the evaluation could be incorporated into this complex process so that it would not interfere with program services.

Stake and Abma (2005) noted, "There is a common misunderstanding that responsive evaluation features collaborative methods. It sometimes fits well with them, but the two are not the same" (p. 376). For this project, we felt collaborative, participatory methods would work well because there was a strong organizational commitment at this CBO to develop capacity in program evaluation. Consistent with practical participatory evaluation (P-PE; Cousins & Earl, 1992, 1995), the primary users, the nurses and advocates, were involved in designing and conducting the evaluation. Decision-making control was balanced between the evaluators and program staff.

The nurses and advocates wanted to know how their clients felt about the services they received and how they were treated by program staff (e.g., did they feel they were given options, did they feel they were treated with dignity and respect, did they feel supported, did they get all of the resources they needed?). We ruled out observational methods (too invasive) and archival methods (incapable of capturing desired information), so we had to figure out how to make survey/interview methods work in this context. Therefore, the questions posed to the clients needed to be brief, limited in number, and absolutely on target and essential to the evaluation ("no pork," as one nurse noted). We codeveloped a program theory logic model to guide this process (see Campbell, Patterson, Adams, Diegel, & Coats, 2008) and wrote questions linked to the final model. From this process, we finalized 15 closed-ended and two open-ended questions that were specific to the services and care provided by the nurses (rather than the advocates or the program overall).

Once we knew what we wanted to ask, we considered who should be collecting the data. Because the program provides 24/7 services, it would be practically difficult to get an evaluator on-site for every case; however, advocates respond to each call and because the evaluation questions were specific to the nurses, having the advocates collect evaluation data would be possible. In addition, to be responsive to the needs of the clients in this highly traumatizing period of time, bringing an outside party (an evaluator) into the dynamics of the exam was inadvisable. Because there is precedent for having program staff collect evaluation data within various participatory approaches (see Cousins & Whitmore, 1998, for a review), it seemed both necessary and reasonable to train the advocates in data collection (described below).

Our final decisions were when to ask the questions and how. We did not want the evaluation to interrupt services, so we looked for “windows” in the exam process when data collection could be woven into normal program operations. The researchers/evaluators proposed three methods, we debated the pros and cons of each with program staff, and the collective group could not reach consensus. Stake (2004; Abma & Stake, 2001) reminded us that as evaluators we do not need to create consensus when it isn’t there and that different points of view can coexist. We proposed approaching this dilemma empirically to test which method of administration would work best. Empirical comparisons of different evaluation methods have been a useful strategy in client-based evaluations in general (Bausell & Rinkus, 1979; Warner, Berman, Weyant, & Ciarlo, 1983) and with HTF populations specifically (Fleishman et al., 1992).

Although program staff had concerns about some of the proposed methods (described below), we had developed enough trust with them during the planning process that they agreed to this plan. We all recognized that if we could identify a data collection method that would yield a high response rate without upsetting clients or interfering with services, this would be helpful information for others who work with HTF, traumatized, and/or vulnerable clients. Specifically, we had three methodological evaluation questions: (a) Would there be differences in the response/participation rates across the three methods? (b) Would there be differences in how the clients liked or disliked the three methods? and (c) Would the clients’ responses vary by data collection method (i.e., more positive or more negative by method)?

Methods and Results

The first data collection method was, “on-site, in-person administration by advocate.” After the clients had completed the medical forensic exam, they typically spent approximately 20 min with the advocates for questions and referrals. Near the end of that time, the advocates asked clients whether they could ask them a few brief questions about their experiences with the nurses. If so, the advocate completed an oral consent process that highlighted the purpose of the evaluation, the procedures, confidentiality, and risks/benefits. For the clients who agreed to participate in the evaluation, the advocates read the questions to them and recorded their answers (see the following sections for details on advocate training). This method most concerned program staff—they were worried it would be too “in their [the clients] face” and that because the data were collected so quickly after services, clients would not have had time yet to absorb and evaluate the care they had received.

The second method was, “telephone follow-up by advocate.” As part of their normal program practice, advocates asked all clients whether they would like a phone check-in a couple of days after the exam to see how they were doing. Some clients do not agree to be contacted, but for those who did and were reachable by the staff, the advocates asked them near the end of their phone call whether they had a few more minutes to answer questions about their experiences with the nurses. The same procedures for oral consent and evaluation

administration, as described above, were then repeated on the phone for those who agreed to participate. This was the method most favored by program staff in the planning process because they viewed it as the least invasive, and because there was a delay between services and data collection, clients would have more perspective on their experiences with the nurses.

The third method was, "on-site, client self-administered": At the end of their time with the advocates, clients were asked whether they would be willing to fill out a brief questionnaire about their experiences with the nurses. The questionnaire contained the same items and response categories as the prior two methods; new formatting and introductory material were added so that the instrument functioned as a self-administered survey. Advocates gave a written consent form to clients who expressed a willingness to participate, and clients were given time in private to complete the questionnaire and place it in a locked collection box. Although the program staff liked that this method gave clients more privacy, which they felt would minimize social desirability bias, they were still concerned about further questioning so soon after services.

The three methods were implemented consecutively in phases, each lasting 5 weeks (15 weeks total). We selected 5 weeks as our time frame because it is common for violence-against-women CBOs to do evaluations for less than a full year due to resource constraints (e.g., collect data from all clients in just one month, one quarter, etc.; Campbell et al., 2000; Riger et al., 2002; Sullivan & Alexy, 2001). It was necessary to implement the three phases sequentially because there are well-documented seasonal variations in rates of sexual assault (Michael & Zumpe, 1983) and all evaluation data needed to be collected within the same season.

The SANE program advocates played a central role in each of the three data collection methods. The advocates were mostly White (88%), female volunteers whose age ranged from 25 to 62 ($M = 39$). All of the advocates had participated in 70 hrs of intensive sexual assault advocacy training prior to the evaluation. This training included information on multiple forms of oppression and diversity, sexual assault and domestic violence, trauma effects of victimization, crisis intervention, and systems advocacy. Only those advocates who demonstrated strong advocacy skills at the end of training, as assessed by the advocacy program coordinator, went on to provide direct services to survivors. In addition to this training, the advocates had to successfully complete evaluation methodological training conducted by the researchers/evaluators. The training provided background information about the purpose of the project, how the methods and questions were developed, how the consent process should be conducted, and how the questions should be asked and responses recorded. The researchers/evaluators demonstrated two mock administrations, and then the advocates practiced via role plays. The researchers/evaluators observed the role plays and provided feedback. A final mock administration was conducted to check reliability of coding, and all advocates coded the information accurately. The researchers/evaluators met with the advocates and nurses throughout data collection to ensure that the methods were being implemented correctly.

During the three phases of the evaluation, the SANE program provided services to 52 adult sexual assault survivors, ranging in age from 18 to 59 ($M = 29$), and the majority were White (68%). The advocates were demographically quite similar to the clients they were serving, and indeed similar to county-level data regarding race/ethnicity. Of the rape survivors studied, 32% had some type of disability, which included mobility/visual/hearing impairments, mental/emotional disabilities, cognitive/developmental disabilities, and/or chemical dependency. Previous studies have documented that women with disabilities are at disproportionately higher risk for sexual assault (e.g., 40%-83% of women with disabilities have been sexually assaulted; see Elman, 2005, for a review), so the rate obtained in this project is, unfortunately, not surprising. Most victims were sexually assaulted by someone they knew (66%), which is consistent with other studies of sexual assault survivors who seek community services (e.g., Campbell

et al., 1999). Most victims had sustained some degree of physical injury (78%) and/or anogenital injury (56%) as a result of the assault, and these rates are somewhat higher than what has been reported in previous studies (e.g., Campbell et al., 1999). We suspect these differences are due to the fact that the victims in the current study were all examined by SANEs who conduct significantly more thorough medical exams and injury documentation than non-SANE medical personnel (Derhammer, Lucente, Reed, & Young, 2000). All clients aged 18 or older treated in the SANE program during these 15 weeks were asked to participate in the evaluation (i.e., there was no sampling; all clients were recruited).

Clients' Response/Participation Rates

As can be seen in Table 1, the number of eligible clients in each evaluation phase was consistent (suggesting that we were successful in avoiding seasonal-rate variations), but we had strikingly different participation rates across the three methods: $\chi^2(10, N = 52) = 48.10, p < .0001$. The participation rate for on-site, in-person administration by advocate was significantly higher than the other two methods at 88% consent/completion. The telephone follow-up method yielded the lowest participation rate (17%). Although some clients ($n = 4$) did not want any follow-up contact from program staff, most did ($n = 14$), but the advocates were unable to reach many of them ($n = 9$). As a result, most clients in this phase did not receive the program's usual postexam follow-up or the evaluation. The advocates made repeated attempts to reach the clients (usually three calls), which is a reasonable effort in the context of nonprofit, social service agencies, but it is of course much lower than the seven (or more) attempts that are often necessary in research studies with HTF populations (e.g., Gondolf & Deemer, 2004). Almost all clients in the on-site, client self-administered method initially agreed to participate in the evaluation (94%), but only 41% actually completed a questionnaire; the others returned a blank or nearly all-blank questionnaire in the lock box.

Clients' Perceptions of the Data Collection Methods

In all three phases, we collected qualitative data about how each method was perceived by the clients. Such metaevaluation data may be able to shed some light on why there were differential response rates: Perhaps one method (or methods) was particularly liked or disliked by these clients. In Phase 1 (in-person, on-site administration by advocate) and Phase 2 (telephone follow-up), the advocates recorded open-ended notes regarding how the clients' reacted to the evaluation procedures. In addition, they were asked to comment on their own comfort with administration of the method. For Phase 1, the overwhelming majority of advocates characterized the clients as *receptive, willing, and/or happy to do it*. When clients refused to participate, it was because they were already very tired, they did not want to answer any more questions, and they simply wanted to go home. With respect to their own experiences administering the evaluation, advocates noted that they did not encounter difficulties, and although their first administration was sometimes awkward, overall the evaluation was easier to conduct than they had anticipated (somewhat to their surprise). For the few clients who completed the Phase 2 telephone follow-up version of the evaluation, the advocates noted that they had a similarly positive reaction as they did in the in-person clients. The primary difficulty with this method noted by the advocates was, as expected, the difficulty of reaching clients by phone.

For Phase 3 (on-site, client self-administered), many clients in this phase initially agreed to participate, but they did not complete the questionnaire; it was not feasible or ethical for the advocates to ask clients about why they changed their minds. However, most clients who refused questioning in Phase 1 did so because of fatigue, so it seems probable that this may

Table 1
Participation Rates by Data Collection Method

	Phase 1: On-Site, In-Person Administration by Advocate	Phase 2: Telephone Follow-Up by Advocate	Phase 3: On-Site, Client Self- Administered	All Phases Combined
Number of eligible clients served by program ^a	17	18	17	52
Number who completed evaluation	15 (88%)	3 (17%)	7 (41%)	25 (48%)
Number who did not participate in evaluation	2	15	10	27
Client was asked to participate but refused	1	0	0	1
Client started evaluation, but changed mind and withdrew	1	2	9	12
Client did not want any follow-up from program	n/a	4	n/a	4
Client agreed to follow-up but staff could not reach them	n/a	9	n/a	9
Staff did not have an evaluation survey with them at time of services	0	n/a	1	1

a. Clients who were under 18 years of age were not eligible for the evaluation.

have been an issue here as well. For those who did participate, there was an additional question in the survey: "Please tell us what it was like to fill [this survey] out. Was there anything that was difficult or that you didn't like?" Only half of those who participated in this phase answered this question, and all stated it was easy to complete and they had no concerns about participating. Across all phases, clients gave positive feedback about all three methods of data collection. Therefore, the differential response rates do not appear to be a function of client preferences (or dislikes) for a particular method.

Clients' Responses Across Data Collection Methods

The final question we needed to explore was whether there were substantive differences in clients' responses across the three methods of data collection—did the clients rate the quality of the care they received from the nurses differentially as a function of evaluation method? For example, if the clients rated services significantly higher in Phase 1, which also had the highest response rate, this would suggest that the evaluation was affected by both demand effects and social desirability bias—clients may have felt they had to answer and in so doing, gave the right answers. The participants' responses in each method were compared using the two criteria established by Bausell and Rinkus (1979): Methods of administration were deemed comparable if the items were highly correlated across method of administration (.60 or higher) and if the mean response for the item did not differ significantly across methods. Because of the small sample size of this project, we used nonparametric tests for these analyses (Pett, 1997). All Kendall's Tau correlations were above .90 and all Kruskal-Wallis one-way ANOVA tests comparing means were nonsignificant (see Table 2). These findings suggest that the substance of clients' answers were consistent across all three methods of administration. Furthermore, the quantitative ratings were consistently very positive across all phases, as were the qualitative comments (with the exception of one Phase 2 client), so either the clients gave socially desirable answers across all data collection methods (even the ones completed in private, which presumably would be more likely to capture negative feedback), or maybe, the clients did have positive experiences with the nurses, no matter how you asked them.

Table 2
Client Responses by Data Collection Method

	Phase 1 <i>M</i>	Phase 2 <i>M</i>	Phase 3 <i>M</i>	Kruskal- Wallis ANOVA
Survey part I: In your interaction with the nurse, how often did the nurse ^a				
1. explain what was going to happen next in the exam?	4.0	4.0	3.86	2.71, <i>ns</i>
2. ask if you had questions?	3.63	4.0	3.57	1.18, <i>ns</i>
3. tell you had a chance to help with the exam if you wanted to (e.g., pull or comb own hair)?	3.47	3.67	3.14	0.79, <i>ns</i>
4. tell you how parts of the exam might feel before she did them?	3.94	4.0	3.71	0.78, <i>ns</i>
5. take your needs and concerns seriously?	4.0	4.0	4.0	0.00, <i>ns</i>
6. listen to you?	4.0	4.0	4.0	0.00, <i>ns</i>
7. ask you whether you felt you could take a break during the exam or say, "no" to any part of the exam?	3.93	3.67	4.0	3.13, <i>ns</i>
8. explain why each part of the exam was important?	3.88	3.00	3.79	1.49, <i>ns</i>
Survey part II: During and after your interaction with the nurse, ^b				
9. how much care and compassion did you feel that the nurse showed?	4.0	4.0	4.0	0.00, <i>ns</i>
10. how much control did you feel that you had during the exam?	3.57	4.0	3.71	1.50, <i>ns</i>
11. how informed did you feel about what was happening in the exam?	4.0	4.0	4.0	0.00, <i>ns</i>
12. how clear were the nurse's instructions about the medications?	4.0	4.0	4.0	0.00, <i>ns</i>
13. how informed did you feel about where to go for follow-up medical care?	3.86	3.33	3.86	1.61, <i>ns</i>
14. how much pressure did you feel from the nurse to go through with prosecution? (reverse coded)	3.43	3.00	3.17	0.81, <i>ns</i>
15. how likely will you be to contact the nurse if you have a problem?	4	4	4	0.00, <i>ns</i>

a. Response categories: 1 = *none of the time*; 2 = *some of the time*; 3 = *most of the time*; 4 = *all of the time*.

b. Response categories: 1 = *not at all*; 2 = *a little*; 3 = *somewhat*; 4 = *a lot or completely*.

But, not all clients participated in the evaluation and provided feedback about the program (as evidenced in the significantly different response rates). It is possible that those who completed the evaluation found the services to be helpful—that's why they completed the evaluation and why their answers are consistently positive across evaluation method. Those who did not find the services to be helpful, did not participate; therefore, their negative feedback goes undocumented, and we end up with a biased assessment of services. There were more clients in Phases 2 and 3 who did not participate in the evaluation, so it is worth checking if there was something different about the clients (e.g., demographics, nature of assaults experienced), program operation, and/or patient care across each phase.

Data collection was anonymous (to the evaluators) in all phases, so we cannot compare characteristics of those who participated in the evaluation versus those who declined, on the basis of information from the SANE program records. However, we were allowed to examine deidentified SANE records from clients who would have been eligible for the evaluation and who were treated during the time in which the evaluation was taking place to explore whether client characteristics differed across the three phases. We compared demographic (age, race) and assault (type of rape, physical injury, anogenital injury) characteristics of the eligible clients treated in the program during Phase 1 versus Phase 2 versus Phase 3. Again, given our limited sample size, nonparametric tests were appropriate (Fisher exact test for categorical variables and Mann-Whitney *U* tests for continuous variable) and all were nonsignificant. In addition, there were no changes in program operation across the three phases. We asked the program director to provide patient care performance ratings for each of the nurses in the program, and there were no

significant differences in nurse ratings by phase. Taken together, these data suggest that the clients served (and how they were served) in the lower response-rate phases were not significantly different from those in the phase with the highest response rate.

Discussion

Program staff and researchers/evaluators share a long-standing concern regarding if and how it is possible to collect evaluation data directly from HTF, traumatized, and/or vulnerable populations when they surface at CBOs for services (see Miller & Campbell, 2008, for a review). Program staff may be worried that data collection could have an inadvertent negative impact on the burgeoning relationship between clients and staff; in addition, researchers/evaluators know that sampling, nonresponse, and social desirability biases are particularly challenging to resolve with these populations.

In this project, we put these concerns to an empirical test, comparing the data yielded on the same survey items administered by three different data collection methods. The first method involved direct, oral questions posed to rape victims after they had received medical forensic exam services, but before they left the SANE program. Program staff were most concerned about this approach, and they were more favorably inclined to the other two methods: phone follow-up a few days after the clients had left the program and a self-administered survey completed privately before leaving the program.

Interestingly, there were no differences in the metaevaluation data regarding how clients perceived each method. Neither staff nor clients expressed problems with or preferences for a particular data collection method. In addition, the responses given by the clients were comparable across method: they gave similarly positive feedback about the care they received from the nurses whether asked face-to-face, after delay of time over the phone, or in a private anonymous survey, when presumably more negative feedback could surface. Whereas we cannot rule out possibility of social desirability bias, we believe the pattern of results suggests that the clients indeed felt that they received helpful, supportive care from their nurses, which is consistent with other studies (e.g., Ericksen et al., 2002) and subsequent follow-up studies of this same SANE program (e.g., Patterson, Campbell, & Cabral, 2007).

We also compared client and assault characteristics for the individuals treated in the program during each of three phases of data collection, and again, it appeared that those served in the program were consistent across phases. Program operations and quality of patient care were also comparable in all phases of data collection. We cannot discount the possibility of nonresponse bias; however, on the basis of the information we do have, we find that the pattern of findings does not indicate that there were interactions between data collection phase and client or program operation/care.

What was different, and significantly so, was the response/participation rate across the phases: 88% for in-person, 17% for telephone follow-up, and 41% for self-administered. If directly asked to answer a few more questions about the care they received from their nurses, clients did answer. However, once they left the program, clients were very hard to find again for either programmatic follow-up or evaluation. And, if you give tired, traumatized individuals the option privately to decline answering more questions after they had already answered hundreds of others, most will. For other researchers/evaluators and CBOs who want to collect evaluation data from HTF, vulnerable clients, our data suggest that it is necessary to ask them directly, while they are still physically at the service program. This method had the highest response rate and was not distinguishable from the other methods with respect to implementation comfort/discomfort or substantive response.

To some extent, the success of the in-person administration could be attributed to the role of the advocates, who work hard to establish rapport with clients to help them feel safe and comfortable. It is possible that this connection positively affected clients' willingness to participate in the evaluation. The advocates were also, in general, demographically similar to their clients, which may have also affected rapport and participation. Whether the results would be different if there were more dissimilarities between advocates and clients is unknown, but the findings regarding disability status shed some light on this issue. Nearly one-third of the victims who participated in the evaluation had some type of disability, which suggests that the advocates were successful in working with dissimilar clients, and as such, we were able to collect needed information from an understudied group within this HTF, traumatized, and vulnerable population.

It may have been methodologically interesting to compare response rates when advocates ask clients to participate versus an independent evaluator to disentangle whether there were demand effects. However, in the context of this project, being responsive to the needs of the rape survivors was paramount and introducing an unknown person into the exam process, solely for the purposes of evaluation, was inadvisable. We suspect this situation may arise in other evaluations with other HTF, traumatized, and vulnerable populations where program staff are best positioned to collect data. In our project, we tried to limit demand effects by using an informed consent process, so clients knew they could refuse participation without consequence, and by ensuring that the staff member collecting evaluation data was not the subject of the evaluation itself (i.e., questions focused on the care provided by the nurses, not the advocates).

Our findings cannot be disentangled from the evaluation approaches that guided their generation. We suspect that the reason why the evaluations were well tolerated was due in part to the responsive, participatory approach we used in this project. The methods were designed with the needs of the clients and program staff in mind. We spent considerable time learning about this program and its clients, and we styled our evaluation such that it could flow into regular program operation with minimal disruption. We worked with the program staff to design a survey instrument that was brief and contained questions that would not invoke further trauma. Furthermore, to respect the client's privacy, the advocates administered the survey rather than a member of the evaluation team. We worked closely with the advocates during data collection training and follow-up monitoring during which we communicated our high degree of trust and respect for them and their role in the program. This likely also benefited the evaluation as the advocates may have felt positively inclined toward the project and invested in its purpose.

Therefore, we do not offer an unqualified recommendation that it is reasonable to have program staff directly ask questions of HTF, vulnerable clients during service delivery—and expect good response rates and a well-received evaluation. Rather, our experiences suggest that if the evaluation is responsive to the unique circumstances of the program, its staff, and clients, it is possible to collect useful data from HTF service populations. We are not arguing that responsive or participatory evaluation approaches are uniquely suited for these populations. We concur with Miller and Campbell (2006) that metaevaluation research is necessary for understanding how multiple theories and approaches are carried out in practice and how those models shape what is learned in any given evaluation. It remains an unanswered empirical question as to what other evaluation approaches (e.g., empowerment, realist, deliberative democratic, or transformative evaluation) or methods (e.g., diary methods, visual techniques such as Photovoice) could also be effective for working with HTF, traumatized, and/or vulnerable populations.

We acknowledge two key methodological limitations of this study that temper the strength of the conclusions that can be drawn from this work. First, the sample sizes in all three phases are small. Therefore, we used nonparametric statistics, but the power to detect significant effects is low. Nevertheless, it is also important to remember that we specifically set the evaluation phases to be at a length consistent with evaluation practice in many victim-services CBOs (Campbell et al., 2000; Riger et al., 2002; Sullivan & Alexy, 2001). Our data provide a realistic picture of evaluation practice in this field, given the resources these CBOs typically have to devote to program evaluation (Sullivan & Alexy, 2001).

Second, in Phases 1 and 2, the data regarding client perceptions of the data collection methods came from the advocates who administered the measures. It is possible that clients had a more negative reaction to the evaluations than was noticed or interpreted correctly by the advocates; however, in Phase 3, we did collect data directly from clients, which indicated that the questions were not troublesome. The advocates' descriptive accounts of the data collection process and the clients' visible responses were far more detailed than what we received directly from the clients in the final phase.

In spite of these limitations, this project provided useful information for this particular CBO and for other programs that work with HTF, traumatized, and/or vulnerable populations. By comparing response rates across evaluation methods, we discovered that telephone follow-up was not effective in reaching a substantial proportion of clients—for either evaluation or for program services. These findings prompted the administrative staff to reevaluate what information is given to clients when they are on-site versus what may be planned for follow-up contact because there may not be an opportunity for more contact with clients. This also means that CBOs that work with HTF populations will likely need to invest more staff time and effort if they want to have postservices contact with their clients. The SANE program staff also found it helpful to receive feedback directly from their clients, and they were relieved it did not appear to cause any harm in the process. The clients' responses were quite positive, which provided some much-needed affirmation and support to the nurses. Professional burnout is common in human services professions, and those who work with HTF and traumatized populations are at particularly high risk (Townsend & Campbell, *in press*). Evaluation data can help mitigate such stress by providing positive feedback (if indeed it is warranted) or by suggesting targeted areas for improvement, which can also be motivating.

In addition to the evaluation findings, the survey instrument itself could prove useful to the CBO. The survey questions were collaboratively developed to capture the specific activities nurses are to perform and the corresponding intended outcomes for clients. This instrument could be used to train new program staff or as a performance measurement tool to gauge the extent to which staff are engaging in the desired activities. Although there are critical issues to consider when collecting data from vulnerable clients, it can be done in ways that provide useful information for program improvement and give these clients, who are often so marginalized, an opportunity to provide feedback on the services they have received.

Notes

1. We do not claim that this project is a full-scale implementation of either responsive evaluation or participatory evaluation. This is an exploratory methodological study, but we recognized the value of responsiveness to program staff and clients, and used participatory approaches during project planning (see above).

2. We were in a fortunate situation in that the evaluation was not mandated by any external party but rather was undertaken at the joint initiative of the CBO and the evaluation team for the purposes of program improvement. Consequently, we were not required to use any particular theory, approach, or design.

References

- Abma, T. A., & Stake, R. E. (2001). Stake's responsive evaluation: Core ideas and evolution. *New Directions for Evaluation, 92*, 7-21.
- Bausell, R. B., & Rinkus, A. J. (1979). A comparison of written versus oral interviews. *Evaluation & the Health Professions, 2*, 477-486.
- Benoit, C., Jansson, M., Millar, A., & Phillips, R. (2005). Community-academic research on hard-to-find populations: Benefits and challenges. *Qualitative Health Research, 15*, 263-282.
- Birman, D. (2005). Ethical issues in research with immigrants and refugees. In J. E. Trimble & C. B. Fisher (Eds.), *Handbook of ethical research with ethnocultural populations and communities* (pp. 155-177). Thousand Oaks, CA: Sage.
- Boykins, A., & Mynatt, S. (2007). Assault history and follow-up contact of women survivors of recent sexual assault. *Issues in Mental Health Nursing, 28*(8), 867-881.
- Boykins, A., & Mynatt, S. (2007). Assault history and follow-up contact of women survivors of recent sexual assault. *Issues in Mental Health Nursing, 28*(8), 867-881.
- Campbell, R., Davidson, W. S., Dorey, H., Bennett, K., Hagstrom, J., & Bonter, F. (2000). *Resources for evaluating sexual assault service delivery programs*. Okemos: Michigan Public Health Institute.
- Campbell, R., Patterson, D., Adams, A. E., Diegel, R., & Coats, S. (2008). A participatory evaluation project to measure SANE nursing practice and adult sexual assault patients' psychological well-being. *Journal of Forensic Nursing, 4*, 19-28.
- Campbell, R., Patterson, D., Bybee, D., & Diegel, R. (2007, October). *Prosecution rates for adult sexual assault cases: A ten year analysis before and after the implementation of a SANE program*. Paper presented at the International Forensic Nursing Scientific Assembly, Salt Lake City, UT.
- 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: A Review Journal, 6*, 313-329.
- Campbell, R., Sefl, T., Barnes, H. E., Ahrens, C. E., Wasco, S. M., & Zaragoza-Diesfeld, Y. (1999). Community services for rape survivors: Enhancing psychological well-being or increasing trauma? *Journal of Consulting and Clinical Psychology, 67*, 847-858.
- Campbell, R., Sefl, T., Wasco, S. M., & Ahrens, C. E. (2004). Doing community research without a community: Creating safe space for rape survivors. *American Journal of Community Psychology, 33*, 253-261.
- Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*, 397-418.
- Cousins, J. B., & Earl, L. M. (Eds.). (1995). *Participatory evaluation in education: Studies in evaluation use and organizational learning*. London: Falmer.
- Cousins, J. B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation, 80*, 5-23.
- Derhammer, F., Lucente, V., Reed, J., & Young, M. (2000). Using a SANE interdisciplinary approach to care of sexual assault victims. *Journal on Quality Improvement, 26*, 488-495.
- Elman, R. A. (2005). *Confronting the sexual abuse of women with disabilities*. VAWNet Document, National Online Resource Center on Violence Against Women. Minneapolis: University of Minnesota. Retrieved from http://new.vawnet.org/Assoc_Files_VAWnet/AR_SVDDisability.pdf
- Ericksen, J., Dudley, C., McIntosh, G., Ritch, L., Shumay, S., & Simpson, M. (2002). Clients' experiences with a specialized sexual assault service. *Journal of Emergency Nursing, 28*, 86-90.
- Fleishman, J. A., Mor, V., Cwi, J. S., & Piette, J. D. (1992). Sampling and accessing people with AIDS: Implications for program evaluation. *Evaluation & the Health Professions, 15*, 385-404.
- Garland, A. F., Saltzman, M. D., & Aarons, G. A. (2000). Adolescent satisfaction with mental health services: Development of a multidimensional scale. *Evaluation and Program Planning, 23*, 165-175.
- Gondolf, E. W., & Deemer, C. (2004). Phoning logistics in longitudinal follow-up of batterers and their partners. *Journal of Interpersonal Violence, 19*, 747-765.
- Heckathorn, D. D. (1997). Respondent-driven sampling: A new approach to the study of hidden populations. *Social Problems, 44*, 174-199.
- Heckathorn, D. D. (2002). Respondent-driven sampling II: Deriving valid population estimates from chain-referral samples of hidden populations. *Social Problems, 49*, 11-34.
- Heckathorn, D. D., Broadhead, R. S., Anthony, D. L., & Weakliem, D. L. (1999). AIDS and social networks: Prevention through network mobilization. *Sociological Focus, 32*, 159-179.
- Herman, J. L. (1992). *Trauma and recovery*. New York: Basic Books.
- Holmes, M. M., Resnick, H. S., & Frampton, D. (1998). Follow-up of sexual assault victims. *American Journal of Obstetrics and Gynecology, 179*, 336-342.

- International Association of Forensic Nurses. (2008). *Database of the International Association of Forensic Nurses*. Available from <http://www.forensicnurse.org>
- Lebow, J. (1982). Consumer satisfaction with mental health treatment. *Psychological Bulletin*, *91*, 244-259.
- Lebow, J. (1983). Client satisfaction with mental health treatment: Methodological considerations in assessment. *Evaluation Review*, *7*, 729-752.
- Liamputtong, P. (2007). *Researching the vulnerable: A guide to sensitive research methods*. London: Sage.
- Littel, K. (2001). Sexual assault nurse examiner programs: Improving the community response to sexual assault victims. *Office for Victims of Crime Bulletin*, *4*, 1-19.
- Michael, R. P., & Zumpe, D. (1983). Sexual violence in the United States and the role of season. *American Journal of Psychiatry*, *140*, 883-886.
- Miller, R. L., & Campbell, R. (2006). Taking stock of empowerment evaluation: An empirical review. *American Journal of Evaluation*, *27*, 296-319.
- Miller, R. L., & Campbell, R. (2008). *Methodological development for conducting research with disenfranchised, hidden, and hard-to-reach populations*. Manuscript submitted for publication.
- Miller, R. L., Forte, D., Wilson, B. D. M., & Greene, G. J. (2006). Protecting sexual minority youth from research risks: Conflicting perspectives. *American Journal of Community Psychology*, *37*, 341-348.
- Molnar, J., & Stup, B. (1994). Using clients to monitor performance. *Evaluation Practice*, *15*, 29-35.
- Muhib, F. B., Lin, L. S., Stueve, A., Miller, R. L., Ford, W. L., Johnson, W. D., et al. (2001). A venue-based method for sampling hard-to-reach populations. *Public Health Reports*, *116*, S216-S222.
- Neigher, W. D., & Schulberg, H. C. (1982). Evaluating the outcomes of human service programs: A reassessment. *Evaluation Review*, *6*, 731-752.
- O'Connell, A. A. (2000). Sampling for evaluation: Issues and strategies for community-based HIV prevention programs. *Evaluation & the Health Professions*, *23*, 212-234.
- Patterson, D., Campbell, R., & Cabral, G. (2007, October). *The humanizing effect: How SANE programs are helpful to victims throughout the stages of the legal system*. Paper presented at the International Forensic Nursing Scientific Assembly, Salt Lake City, UT.
- Pett, M. A. (1997). *Nonparametric statistics for health care research*. Thousand Oaks, CA: Sage.
- Riger, S., Bennett, L., Wasco, S. M., Schewe, P. A., Frohmann, L., Camacho, J. M., et al. (2002). *Evaluating services for survivors of domestic violence and sexual assault*. Thousand Oaks, CA: Sage.
- Rosenbaum, A., & Langhinrichsen-Rohling, J. (2006). Meta-research on violence and victims: The impact of data collection methods on findings and participants. *Violence and Victims*, *21*, 404-409.
- Sanders, L. M., Trinh, C., Sherman, B. R., & Banks, S. M. (1998). Assessment of client satisfaction in a peer counseling substance abuse treatment program for pregnant and postpartum women. *Evaluation and Program Planning*, *21*, 287-296.
- Stake, R. E. (2004). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage.
- Stake, R. E., & Abma, T. A. (2005). Responsive evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 376-379). Thousand Oaks, CA: Sage.
- Sullivan, C., & Alexy, C. (2001). *Evaluation outcomes of domestic violence service programs: Some practical considerations and strategies*. Harrisburg, PA: VAWNet, a project of the National Resource Center on Domestic Violence and the Pennsylvania Coalition Against Domestic Violence. Available from <http://www.vawnet.org>
- Sullivan, C. M., & Cain, D. (2004). Ethical and safety considerations when obtaining information from or about battered women for research purposes. *Journal of Interpersonal Violence*, *19*, 603-618.
- Townsend, S. M., & Campbell, R. (in press). Organizational correlates of secondary traumatic stress and burnout among sexual assault nurse examiners. *Journal of Forensic Nursing*.
- Ware, J. E., & Davies, A. R. (1983). Behavioral consequences of consumer dissatisfaction with medical care. *Evaluation and Program Planning*, *6*, 291-297.
- Warner, J. L., Berman, J. J., Weyant, J. M., & Ciarlo, J. A. (1983). Assessing mental health program effectiveness: A comparison of three client follow-up methods. *Evaluation Review*, *7*, 635-658.