Published in final edited form as:

Am J Eval. 2016 September ; 37(3): 326–343. doi:10.1177/1098214015603502.

Audit Culture: Unintended Consequences of Accountability Practices in Evidence-Based Programs

Jill Owczarzak¹, Michelle Broaddus², and Steven Pinkerton²

¹Department of Health, Behavior, and Society, Johns Hopkins School of Public Health, Baltimore, MD, USA

²Center for AIDS Intervention Research, Department of Psychiatry and Behavioral Medicine, Medical College of Wisconsin, Milwaukee, WI, USA

Abstract

Evaluation has become expected within the nonprofit sector, including HIV prevention service delivery through community-based organizations (CBOs). While staff and directors at CBOs may acknowledge the potential contribution of evaluation data to the improvement of agency services, the results of evaluation are often used to demonstrate fiscal prudence, efficiency, and accountability to funders and the public, rather than to produce information for the organization's benefit. We conducted 22 in-depth, semistructured interviews with service providers from four agencies implementing the same evidence-based HIV prevention intervention. We use the lens of "audit culture" to understand how the evaluation and accountability mandates of evidence-based program implementation within HIV prevention service provision affect provider–client relations, staff members' daily work, and organizational focus in natural settings, or contexts without continuous support and implementation monitoring. We conclude with recommendations for improving the use and methods of evaluation within HIV prevention service delivery.

Keywords

audit culture; evidence-based programs; HIV prevention; community-based organizations

Program evaluation has become expected within the nonprofit sector across a range of fields, including health and social service provision through community-based organizations (CBOs; Carman, 2007; Hwang & Powell, 2009). Evaluation is necessary to assess and improve the effectiveness of public health programs and demonstrate that funds are being used appropriately and effectively (Napp, Gibbs, Jolly, Westover, & Uhl, 2002). For HIV prevention, each year the Centers for Disease Control and Prevention (CDC), state and local health departments, and philanthropic organizations spend hundreds of millions of dollars to support HIV prevention programming. Therefore, program evaluation is critical for accountability, program planning, and quality assessment (Glassman, Lacson, Collins, Hill,

Corresponding Author: Jill Owczarzak, Department of Health, Behavior, and Society, Johns Hopkins School of Public Health, 624 N. Broadway, Hampton House Room 739, Baltimore, MD 21205, USA. jillowczarzak@jhu.edu.

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

& Wan, 2002). As others have argued (Chouinard, 2013a, 2013b), the standard approach to evaluation remains the collection of impartial, evidence-based, and objective information in the form of quantifiable measurements in order to satisfy accountability requirements. Collecting and reporting these data are generally understood to be critical to organizational survival and assessing the effectiveness of public health intervention (Chouinard, 2013a, 2013b; Dodd & Meezan, 2003; Harper, Contreras, Bangi, & Pedraza, 2003; Miller & Cassel, 2000; Niba & Green, 2005; Taveras et al., 2007). At the same time, numbers-based reporting and accountability practices affect the ways in which service providers interact with clients, their roles within organizations, and the work context more broadly.

The consequences of this focus on quantitative measures of numbers of clients served and the paperwork that it generates can result in what some have termed “audit culture.” Audit culture can be defined as norms and practices of assessment through which accountability and “good practices” are demonstrated (Strathern, 2000b; Vannier, 2010). In the context of audits, program success correlates with meeting quantifiable targets, documenting the attainment of these goals, and sharing these measures with management and funders. Funding recipient organizations must produce and fulfill contracts, evaluations, performance measures, and target indicators for donor agencies. For donors, audits purportedly ensure that funds are well used and policies are implemented according to donor mandates; for CBOs, compliance with audits can ensure future financial support (Vannier, 2010). Audit systems respond to concerns with quality assurance, “operational risk,” and “crises of trust” in public institutions. They are also seen as a way to provide accountability from distant bureaucracies (Shore, 2008). The audit ostensibly makes visible good practices through established auditable criteria that are examined and verified by funders, governments, and international entities (Vannier, 2010).

While accountability aims to ensure implementation with fidelity, provision of services to those in need, and opportunities for program improvement, when the techniques and values of accountancy become central organizing principles in governance and management, new kinds of relationships, habits, and practices are created (Shore, 2008, p. 279; Vannier, 2010, p. 283). The system of audits and the bureaucratic processes it entails instill new norms of conduct into the workforce. One of these consequences is the ways in which employees perceive themselves and are perceived by others:

It encourages them to measure themselves and their personal qualities against the external “benchmarks,” “performance indicators” and “ratings” used by the auditing process ... where accountability is conflated with elaborate policing mechanisms for subjecting individual performance to the gaze of external experts, and where every aspect of work must be ranked and assessed against bureaucratic benchmarks and economic targets. (Shore, 2008, p. 281)

As with the nonprofit sector in general, audit culture within HIV prevention has emerged from the introduction of systems of accountability, including between international donors and local nongovernmental organizations (Adams, 2013). Some have argued that current monitoring and evaluation approaches based primarily on measurable, well-defined indicators may unintentionally cause organizations to distance themselves from aspects of their work that are difficult or impossible to measure by the means of evaluation required by

donors (Holma & Kontinen, 2012). For example, when the appropriate questions are not asked in evaluation, activities, and programs that aim to increase empowerment, foster social change, and develop collaborative relationships may go unrecognized. Moreover, meeting compliance numbers and audit measures and creating and filing paperwork associated with audits (policy documents, audit reports, and research findings) may become the target of work themselves and distract employees from providing quality services (Hull, 2012; Oldani, 2010; Strathern, 2000a). Finally, when the stakes of an audit are high, audit procedures and quantifiable results can replace qualitative understandings of or exhaustive investigations about agent performance, participant experience, and overall quality (Kipnis, 2008; Reynolds, 2014). They can also widen the gap between administrators and direct service providers, as it becomes the primary means through which providers are evaluated, program performance is assessed, and funding and programming decisions are made (Hull, 2012).

In this article, we explore the effects of a standardized, numbers-based approach to program evaluation on the organizational context, interagency relationships, and client–staff relations of agencies that implement the same packaged, standardized evidence-based HIV prevention intervention. Our interest in the effects of this type of numbers-based evaluation efforts and reporting requirements emerged from a larger study of how four organizations in natural settings implement the HIV prevention intervention “Sisters Informing Sisters on Topics about AIDS” (“SISTA”), an evidence-based HIV prevention program for African American women. SISTA is a five-session small group intervention for young African American women (DiClemente & Wingood, 1995). In the original efficacy trial of SISTA, women between 18 and 29 years who completed the intervention demonstrated increased consistent condom use, greater sexual communication, and greater sexual assertiveness compared to control groups (DiClemente & Wingood, 1995). Therefore, the CDC designated SISTA as an intervention with sufficient evidence for dissemination to CBOs (Lyles et al., 2007) and included it in the Diffusion of Effective Behavioral Interventions (DEBI) program, a compendium of evidence-based interventions recommended for widespread dissemination (Collins, Harshbarger, Sawyer, & Hamdallah, 2006; Dworkin, Pinto, Hunter, Rapkin, & Remien, 2008). SISTA filled a significant gap in the compendium of HIV prevention programs by targeting a highly vulnerable population that had been neglected from earlier prevention efforts. We focused on SISTA because it is one of the CDC’s most popular evidence-based interventions: Through 2008, over 400 CBOs and 900 individuals had received SISTA training in the United States.

Over time, private, state, local, and private funders have required that HIV prevention service providers implement evidence-based interventions, such as those included in the DEBI program. Funders’ continued support for these programs is often contingent upon agencies documenting that funds are being used efficiently and program recruitment goals are being met. For agencies implementing DEBI interventions and funded by diverse entities, program impact and process monitoring focus on recording and reporting various numerical data, including number of clients to be reached (by race, ethnicity, gender, and age), number of clients served, and number of clients who receive each intervention session (Chen, 2001; Dodd & Meezan, 2003; Glassman et al., 2002). In this article, we use the lens

of “audit culture” to understand the consequences of numbers-based evaluation and accountability to funders on frontline service providers.

The framework of audit culture provides a useful context to organize our exploration of the ways in which evaluation activities affect intervention facilitators’ and agency directors’ experiences in implementing a standardized HIV prevention intervention. Specifically, within the lens of audit culture, we describe how systems of evaluation and accountability resulted in (1) pressures to “chase numbers” of clients to participate in the intervention rather than focus on client empowerment; (2) a focus on eligibility for services based on narrow (and often changing) demographic profiles of clients rather than a holistic, needs-based approach to service delivery to a community; and (3) a system of documentation that was seen as useful and important, yet burdensome and not fully utilized to actually improve service delivery.

Method

Data Collection

From 2011 to 2013, we conducted in-depth, semistructured interviews with 22 service providers from four agencies that implement SISTA, located in four cities in three geographically distinct regions of the United States (northeast, Midwest, and south). These four agencies were recruited to participate in a larger study evaluating the effectiveness of the SISTA program as implemented by frontline service providers. We identified agencies to participate in the study based on a list provided by the CDC of organizations that it directly funded to implement SISTA. We also contacted the state-level AIDS program coordinators in the 25 U.S. states with the greatest population of African Americans based on the 2000 census for names and contact information for organizations that they funded to implement SISTA. We contacted each agency on the lists (over 60 in total) to determine whether they were still implementing SISTA, had the funds and intention to implement SISTA over the next 5 years, conducted HIV counseling and testing, and enrolled at least 100 African American women into both the SISTA program and counseling and testing services each year. In total, 31 agencies met the minimal inclusion criteria. We directly contacted either the director or prevention director of each agency to tell them about the study and gauge potential interest in participating. In total, four agencies met inclusion criteria and agreed to participate in the study. Reasons to decline participation included plans to discontinue offering SISTA in subsequent years, anticipated staff shortages to fulfill the goals of the research project, and existing research collaborations and ongoing evaluation efforts that duplicated the study’s goals.

Participants

Of the four agencies that participated in the study, three were comprehensive AIDS service organizations that offered treatment, care, support, and prevention services to broad populations but that mainly served ethnic and racial minority communities. The fourth organization conducted HIV prevention programs, including SISTA, on a limited scale and also conducted HIV counseling and testing. Most of its work focused on capacity building both domestically and internationally. Four directors (one from each agency) and 18

frontline staff in nondirector positions who facilitated or helped with SISTA completed the interview. In general, facilitators are responsible for recruiting participants, conducting the program (preparing materials, moderating group discussions, leading activities, and conveying information), and implementing most evaluation tasks (e.g., pre and postintervention risk assessments and collection of demographic information). We interviewed all SISTA facilitators at each agency. Facilitators in the study sample were not evenly distributed across agencies due to differences in agency size and scope. Agency 1, for example, was a statewide organization that conducted SISTA in several cities and, therefore, had a larger cadre of SISTA facilitators. In contrast, Agency 2 worked in only one city and had only one full-time facilitator on staff and a contractual, as-needed relationship with a second facilitator. The distribution of interview excerpts below reflects this aspect of the study sample.

Of the 22 interview participants, 18 were female and 4 were male; 3 of the 4 agency directors were men; 17 of the 18 intervention facilitators were women; 14 identified as African American; 5 as Hispanic; and 3 as White. SISTA is typically delivered by African American women, as reflected in our study sample. Over half (13) had completed college in a range of disciplines, including psychology, business, education, and social work and related fields. The median time interviewees had worked at the agencies was 2.5 years, but the range of time varied significantly. One person in a director-level position, for example, had been with his organization for 15 years; five interviewees had been with their agencies for less than 6 months. Two participants worked at their respective agencies for several years, took breaks, and then returned to work at the agency. The 2.5-year median tenure reflects the high rate of turnover at these agencies: Over half of original interview participants in nondirector-level positions were no longer working at their agencies 2 years after the interviews were conducted.

Procedures

All interviews were conducted by the first and second authors and digitally recorded; interviews lasted between 1 and 2½ hr. Interviews were conducted in private rooms at the person's place of employment; one interview was conducted over the phone. The interviews were semistructured, based on an interview guide with a set of topics to be covered during the interview depending on the individual's position in the agency (i.e., director or facilitator) but allowed flexibility to probe participant responses and explore topics in greater depth (Bernard, 2011). Among other topics, individuals in director-level positions and SISTA facilitators were asked questions regarding the logistics of program adoption and implementation, the role of funding in determining what programs they conduct, experience with "packaged" program curricula and interventions, agency decisions about adopting SISTA, how effective they think the program is, how the program is evaluated, and processes of monitoring and documenting program outcomes.

Analysis

All interview recordings were transcribed verbatim and entered into a computer-based text file. Transcripts were then transferred to the qualitative data analysis software program MAXQDA (VERBI GmbH, 2011) to be coded and sorted. Transcripts were analyzed by the

first and second authors for emergent themes using principles of grounded theory analysis (Strauss & Corbin, 1990). That is, analysis took place deductively and inductively by exploring major domains related to the study's overall aims but remained open to unanticipated themes, patterns, and relationships. The coding tree was developed using an iterative and collaborative process to ensure reliability and consistency (Carey, Morgan, & Oxtoby, 1996; Ryan, 1999). Qualitative data analysis proceeded through a multistep process that included open coding and axial coding. In open coding, the first two authors independently read the same transcript and identified preliminary coding categories. During this open-coding process, both a priori and inductive codes were generated and applied. A priori codes included codes that reflected key topics and content area in the interview guide, such as evaluation strategies and organizational context. Inductive codes, or themes that emerged from the data but not explicitly asked in the interview, included cooperation, competition, burden, and paperwork. We then formed an initial coding tree, and the first two authors each individually coded the same interview using this initial coding scheme. After discussion, the coding tree was revised, and a different interview transcript was coded using the revised coding tree. The process was repeated until all team members were satisfied with the final coding tree. The first and second authors then independently coded the remaining transcripts with the final coding tree, periodically checking for consistent use of codes.

Transcripts were first coded by agency location, participant's gender, ethnicity, and participant's title or role in the agency (e.g., director, paid staff, and volunteer). Then, the documents were coded with text codes that reflected key analytical topics including personal history of involvement in HIV-related issues, HIV training and education experience, target populations and agency services, experiences with SISTA, funding, relationships with other organizations including state AIDS programs and other local prevention providers, organizational characteristics, and program evaluation activities. During axial coding, categories and themes were explored in relation to each other, and broader themes that combined and transcended existing codes were identified. This further analysis explored the relationship between issues related to program funding, evaluation activities, accountability and reporting requirements, participant recruitment goals and strategies, and relationships between agencies. Due to the small sample size, our analysis focused on the interviews as a complete set, although some differences did emerge between agencies and by role (director and facilitator). These differences are noted below where appropriate.

Within our data collection methods and analytic approaches, we employed several strategies to address the four dimensions of trustworthiness in qualitative research: credibility, dependability, confirmability, and transferability (Lincoln & Guba, 1985; Miles, Huberman, & Saldana, 2014; Shenton, 2004). We attempted to establish credibility, or how congruent the findings are with reality, by assuring participants that their responses would be confidential and not shared with supervisors or funder; iterative questioning, in which participants were often probed for additional detail and/or asked to rephrase previous answers; familiarity with the culture of the organizational setting, as this is the second study the first author has conducted with frontline service providers disseminating prepackaged interventions; and finally negative case analysis, with specific attention paid to examples counter to the prevailing theme. Dependability refers to whether the research process is stable over time, and we address this aspect of trustworthiness through detailed description

of the methodology and sampling strategy used. We address conformability, or a reasonable freedom from researcher biases, through the iterative, team-based process used to develop and refine the coding scheme described above as well as feedback from the third author who was not involved in the interviewing or coding process. Finally, we addressed transferability, or whether the conclusions of the study have larger import, by providing detailed background information to establish the context of the study within the larger examination of the effectiveness of SISTA and similar prepackaged interventions evaluated.

Results

Qualitative data analysis revealed three major themes related to evaluation, audit culture, and their effects on the work of agencies delivering HIV prevention interventions: chasing numbers to meet funder requirements for service, adherence to funder requirements for client eligibility, and documentation. In meeting funder requirements, participants described themselves as “chasing numbers,” and audit culture manifests itself through the need to continually respond to externally defined priorities based on epidemiological trends, shapes the relationship between agencies in a given region, and contributes to a stressful work environment for employees. In adhering to eligibility requirements of the population to be served, aspects of audit culture compel providers to view their clients in demographic rather than need-based terms in order to meet funder requirements and demonstrate accountability. Finally, documentation results in a high level of “paperwork,” and the collection of participant and program data as part of accountability systems becomes a source of anxiety, staff burden, and disruption in interactions with clients. We explore each of these themes in greater detail below.

Chasing Numbers: Meeting Funder Requirements

Numbers drive HIV prevention service delivery in two ways. First, funders and state health departments use epidemiological data to identify new priority populations and local demographic changes. In response, agencies must continually evolve their service profiles, implement new programs, and apply for competitive grants based on guidance from funders and state health departments to reflect epidemiological trends. Agencies apply for grants to target new populations and new health and social issues regardless of their previous experience or mission in order to maintain their viability. Numbers-based programming decisions create a sense that organizations must rapidly begin and end particular programs based on externally defined priorities:

In this field, it's kind of like you're always playing catch up. You're always chasing the numbers. The numbers dictate exactly what you're going to be doing at that time. When I first got here syphilis was crazy, so we put all of our efforts into the syphilis thing. And once the syphilis numbers started to drop a little bit it's like we kind of [stopped]. (Agency 3, Facilitator 1)

Second, agencies need to demonstrate through participant recruitment goals that these priority populations are being reached. The effects of the focus on achieving participant recruitment goals vary depending on the local landscape of competing or cooperating agencies and population demographics. On the one hand, the emphasis on meeting specific

targets can lead to an increased sense of competition. On the other hand, meeting targets can lead to collaboration and coordination of services. Both aspects of recruiting participants to meet goals mean that organizations must constantly monitor what other agencies in a given area are doing.

In contexts where multiple organizations receive funding from different sources to implement the SISTA program, competition for clients is intense. Different organizations contact the same community centers, public housing developments, and drug treatment centers to identify potential SISTA participants. A limited participant pool and narrow inclusion criteria create challenges to meeting recruitment goals:

The main challenge is having SISTA implemented by other people in the community ... Sometimes we will call and let people know what it is that we are offering and what we are doing, but at the same time they will say, “We have another program from your area that is coming in to do it.” So it’s a really big barrier. It’s hard to get the group when you have a lot of people doing the same program in the area. (Agency 1, Facilitator 7)

This competition can undermine efforts to partner with other agencies to recruit participants. One participant described a meeting that she had attended with another agency, in which the director expressed hesitancy to collaborate:

[The director said] ... “I understand that you need help beginning the program. But,” she said, “many of the things you do—HIV testing or IDIs [interventions delivered to individuals]—we do here, so we have always been hesitant to affiliate with you guys because you’ll be doing the services for us. And we already have it, so why would we?” (Agency 1, Facilitator 6)

Overlap in priority populations, participant recruitment pools, and services provided reduces the likelihood that agencies will work together to target particular geographic areas or populations. Numbers-based competition between agencies can have implications for both agency sustainability and individual job stability. HIV prevention funding typically is allocated through population- and program-specific budgeting, which exacerbates the sense of competition between agencies. Agencies submit funding applications with specific recruitment goals and receive funding to match this scope of program reach. This funding supports individual jobs and incentivizes agencies to continually seek new funding related to new population priorities. When population priorities change, agencies shift their focus in order to secure funding and remain viable, sometimes regardless of their previous experience with that population or the existence of other agencies in the region already working with that group:

So I think as a result of that there has been a lot of tension between many of the agencies garnering—trying to garner dollars for support for jobs. And then as a result—we see this all across the country— many agencies who weren’t working with certain populations are now trying to work with those populations to support their jobs. And that is really it: They may not have any experience at all working with African American high risk drug using women but [they say], “Now we are going after that population because we need to go after the dollars that are out for

that population.” And so I think again there’s been limited collaboration as a result of that. (Agency 2, Director)

Funding cuts at the local, state, and federal levels for prevention programming exacerbate these tensions and increase pressure on CBOs to create and meet participant recruitment goals.

Meeting recruitment goals that align with funding priorities and recognizing the limited population pool may result in collaboration between agencies if working together is seen as mutually beneficial:

We are number-driven to a very small extent but we are not number-driven like the other agency who has the PBC [Protocol-Based Counseling] and CRCS [Comprehensive Risk Counseling and Services] funding. They have an astronomical amount of numbers that they have to get ... We’re just a different agency. We collaborate a lot with each other. But they have a different population than we do. Now some of it overlaps, obviously, but there are certain communities that are comfortable with them and there are certain communities that are comfortable with us. So sometimes you know we are asked to go out [to a neighboring town]. We basically never turn [them] down. But every once in a blue moon if there is something going on here [and] every single person is doing what they have to do here, we don’t have a person just to go do testing because that’s not what we are here for. So we’ll say, “Hey, [another agency] needs to go do that. They need to do that. They need the numbers. We don’t need the numbers.” Or, sometimes what we have done when we have gone out to [this other city], we’ll help them with their paperwork because we don’t need the number RHs. They do. (Agency 3, Facilitator 2)

Collaboration is possible if the client populations of two agencies are seen as unique enough to not create overlap. In this particular case, this facilitator and agency were aware that *all agencies’* success is numbers-driven to some extent and, therefore, helping each other meet recruitment goals was one strategy to ensure continued financial support.

State-level agencies such as health departments or regional coordinating groups can help alleviate this sense of competition between agencies:

Well, because there are other agencies—because we are doing [SISTA] as a region—so as long as we meet these numbers as a region. So there are other agencies. Like, so I just went through Healthy Relationships—that is another DEBI that certain agencies are doing. So, because we are going to be doing them and broken down in that month, we will be doing them as a region. We are coming up together: this person from this agency is going to be the lead, but the person from this other agency is going to help. So that is how we meet those goals. (Agency 2, Facilitator 2)

However, such attempts to coordinate activities can actually intensify a sense of competition between agencies when the stakes for achieving numbers goals are high and agencies mistrust each other. Health departments in the state where one agency was located tried to

facilitate collaborative participant recruitment strategies for the three agencies in the region funded to do SISTA:

I: How was that meeting?

P: Well, one of the agencies sort of like wanted—I don't want to say like “steal” but it kind of sounds like that, the way that she was approaching it. Kind of just stealing our contacts and what other information we had. It was more towards them and this meeting was supposed to—for us to come over and help each other. After the meeting we decided that is not a good idea to collaborate. Plus they're offering different incentive amounts and we thought that would just collide so we decided not to [work with them]. (Agency 1, Facilitator 8)

Rather than developing collaborative systems, the meeting increased suspicion and tension between the agencies, a reflection of the high stakes tied to achieving recruitment goals and competing for limited HIV prevention funds tied to specific populations.

In addition to changing the relationship between agencies within a given region and between funders and grantee organizations, numbers-driven performance measures significantly affect the work environment for individual employees. When job performance is often based on the ability to successfully meet recruitment goals, with little consideration for the contextual factors that might affect their ability to recruit and retain participants, facilitators perceived the potential consequences for continued employment:

I: So there is a lot of pressure there for you?

P: Yeah and then the same thing is also if you don't meet your grant [numbers], then the state is not going to renew your grant. Like, why would they give the grant to the office again if it wasn't met? So there's the pressure that if after June—once the grant is completed—if I didn't get it done or if I barely get it done, or what if I did get it done but it wasn't successful. So it doesn't count. Like I got nothing done. Then would I even have a job after June? (Agency 1, Facilitator 8)

From this perspective, an employee's work only “counts” if it achieves specific recruitment goals. Moreover, producing “deliverables” competes with other work that agencies conduct but is not included in these goals:

Like you have a certain number of deliverables, which is doable with the time allotted but then you tack on tons of paperwork, and you tack on these events that we have to do, and then there is office stuff that we have to do as an agency, and there are events that we have to do as an agency. And so what was workable? If it's the only thing I have to do, I'll bang out my grants like nothing. That's not going to be a problem. But there is a lot more that's not accounted for in this—in what's required of people. And [auditors] don't want to hear excuses. They're just like, “Oh, you can do it.” ... But it's stressful because you're constantly wondering what number's looming over your head. Literally! I have my deliverables posted over my desk. (Agency 1, Prevention Director)

Numbers-oriented audits and accountability systems can divert attention away from the services and activities that agencies provide that are not easily quantifiable, including the time and effort it takes to build relationships with potential participants and collaborating organizations, addressing clients' immediate needs on an individual basis, and regular staff and organizational management. Overall, as a result of an emerging audit culture arising from the pressures to meet numbers-based accountability measures, agencies that ostensibly share common goals to improve the health and well-being of their communities by addressing local client needs must navigate a landscape of competition or negotiated cooperation and measure the value of their work against benchmarks established by funders.

Program Eligibility and Recruitment Goals

As described earlier, most packaged, evidence-based HIV prevention programs target narrowly defined populations that reflect the original research on which they are based. These well-defined populations are the basis for the “deliverables” that agencies are required to meet in order to obtain and retain funding. The emphasis on meeting “deliverables” in terms of a set number of program participants that meet specific demographic or HIV risk parameters fundamentally alters the ways in which agencies position themselves in the community and how facilitators relate to their (potential) participants.

SISTA, as originally designed and tested, is intended for young, heterosexual African American women. The intervention itself includes culturally relevant themes and uses specific imagery, concepts, and language tailored to an African American target audience—including notions of sisterhood and female leadership in the family—as the impetus for increasing women's sense of personal and community responsibility in the fight against HIV. Because the empirical study that demonstrated the effectiveness of SISTA restricted the intervention to African American women between 18 and 29 years (DiClemente & Wingood, 1995), the CDC recommends this intervention only for this particular subgroup of women.

Facilitators understood that SISTA meets the needs of African American women, a highly vulnerable population for which no comparable programs exist. Moreover, many facilitators are themselves African American or other minority women and, therefore, relate to the program's message of gender and ethnic pride on a personal level. However, in the context of audit culture, the need to meet deliverables established on SISTA's narrow inclusion criteria clashes with the reality of delivering the program in real-world settings with much more diverse clients:

Before we used to do it, [there] was not an age range. And they made it mandatory as of last year to do it from 18 to 29. And we find out that people who are a little older than that—than 29—they need it as well. They need it. And even the retention of the group, it used to be bigger before and it was not so challenging to finish the grant when it comes to the numbers. We [are] making the numbers but it's not as it used to be before. Women who are older than 29—they need it and ... I don't think it's fair not to do the SISTA project with them, just because you are 2 or 3 years older than 29. But we still do it. We still do it but it is a challenge. It is. And when you have the group and because you are 2 or 3 years older you can't participate—

how can you? You know it's hard to say "No, you can't." ... We try to address it by doing something else, but at the end, you know, they see it. (Agency 1, Facilitator 1)

This facilitator clearly understands the importance of recruitment goals and why limits are in place. However, she ultimately concludes that it is "unfair" to deny participation in the program to any women who wanted to join simply because she did not "count" toward recruitment goals.

Prohibiting certain individuals from participation in the program contrasted with facilitators' belief that no one who could potentially benefit from an HIV prevention program or who expressed interest in participating should be denied access:

P: Well, I think that any intervention that is geared towards HIV prevention or geared toward some type of empowerment: anybody that needs those types of skills or needs that type of information or education, it's needed for. But they don't adjust it to it being a blanket population. They have to pie us up in charts. And then this person gets the money because this is what's needed here. But there [are] so many people that go underserved. Now that doesn't mean that I don't believe in it, that I don't trust the people but that's just the way of the world.

I: How do you think the system would function better? What would you do to change it?

P: Make SISTA for everybody and every age ... Whoever wants it, come get it. (Agency 1, Facilitator 3)

Facilitators argued in favor of a more inclusive approach to recruitment by asserting that SISTA addresses issues universal to all women, regardless of race or age:

I feel like [SISTA] shouldn't be restricted to a certain age. Women of all ages go through all things ... And also that it's only open to Black and Hispanic. I feel like women is women. All women go through things no matter what it is. It shouldn't have no limitation on it. (Agency 1, Facilitator 7)

This view that SISTA should be open to all women contrasts sharply with the narrowly defined "deliverables" emphasis of audit culture and the underlying foundation of the SISTA intervention itself. The system of deliverables—in the words of the above facilitator—"pied up" the population and did not allow them to include people they believed could benefit from the program.

At the same time, the people for whom the program is intended may not be interested in it. They may, however, be interested in other programs the agency conducts that are not part of its "deliverables":

I think when we set a number—"You have to deliver this many unique clients"—it puts pressure on providers to find people to fit the mold ... And so it's frustrating because sometimes the things that you know you can get people to come out for is unstructured [programs]. So then what ends up happening is you try to mix and match. So you get people to come in for these unstructured things like ice cream

socials or we call them condom parties or whatever, and then you work the intervention into that. But that ends up taking more time. And so people—because time is limited—people are kind of like, “Grrrr.” And sometimes when they see you working in the intervention they are like, “I’m out of here.” (Agency 1, Prevention Director)

This provider tried to use a creative strategy—hosting an informal party—to get potential SISTA participants into the agency, with the hope that they would stay for the program that would allow them to “count” toward their deliverables. Meeting deliverables allows agencies to sustain programs and continue to serve clients. Strict deliverables can constrain providers and do not give them the flexibility they think they need to respond to shifting priorities among clients. Clients may not be interested in certain programs or a broader audience than originally outlined may want to participate. Therefore, due to the focus on (often narrowly defined) demographic profiles required by funders for eligibility for services, within an audit culture providers are restricted in their ability to respond to the needs of their community as a whole.

Documentation

As the earlier reference to “tons of paperwork” suggests, numbers-based accountability often requires data gathering and systematic documentation. These data and documentation are used in audits and reports submitted to funders to ensure that money is spent as allocated and programs operate as intended. In most cases, this oversight is accomplished by and generates a proliferation of paperwork (Hull, 2012). For the SISTA program, data collection and the resultant paperwork occurred through several mechanisms, depending on the agency. All SISTA sessions come with standardized fidelity forms that either SISTA facilitators or observers (i.e., supervisors or external auditors) could complete after each session. These fidelity forms each session’s activities with check boxes to indicate whether they were included or not, as well as space for additional notations. Some agencies require their facilitators to complete these forms, others have supervisors conduct occasional session observations and complete the forms, and others do not use the fidelity forms at all. A second form of data collection and paperwork is pre and postintervention assessment forms. These forms typically collect individual-level demographic and HIV risk behavior data from program participants immediately before and after the program. A third form of documentation includes recording and tracking the number of participants in each session and program retention rates. All this information is collected in addition to personal data, such as contact information.

In addition to shaping relationships among agencies and between funders and grantee organizations, within a context of audit culture data collection and the resultant paperwork shape the day-to-day work environment of agency employees. Generally, facilitators and directors recognized the potential value in data collection as a way of knowing whether programs work as intended, recruit the right participants, and effectively used financial resources (Owczarzak, 2012). Numerical data collection, if done properly, could be used to support arguments for continued financial support of successful programs or as a program improvement tool:

I don't think agencies have done a good job with capturing the right kind of evaluative data to show the effectiveness of these DEBIs. And even with SISTA—I think that even though the tools are there, people often—as we know—collect all of this data and the data is sitting in a box somewhere in the closet and they don't ... utilize the data to justify the need for the service. (Agency 2, Director)

Despite recognizing the potential value of these data, participants described paperwork itself as “disappearing,” or languishing on computers, in binders, or in boxes “in the closet.” When labeled as “paperwork,” the data and its purpose become vague and mysterious and a source of anxiety and burden:

I: You mentioned the SISTA fidelity forms. Who do you turn these into and who looks at them?

P: No one. They get put into a binder and then everyone gets audited and tell us that we did them wrong.

I: Audited by the state?

P: No. By multiple sources. We get audited by the state, we get audited by [large institution] and everyone tells us that we did it wrong. So, [large institution] will audit it and tell you this is what's incorrect and could be problematic in terms of your stream of funding in the future. Whereas when the state comes and audits, it's a little bit more serious because they are auditing to determine whether or not they are actually going to fund you. [Large institution] is more of like support, like “This is how you can fix things before anything is an issue,” and when the state comes they are coming to really take a close look of what you are doing. (Agency 1, Prevention Director)

This quote explicitly describes the audit culture arising out of the need for accountability measures, indicating that although audits are ostensibly used to establish that funding is being used for the general benefit of communities, “fixing” the documentation itself instead becomes the focus of time and effort of providers. Rather than serving as a useful tool to improve the program, both directors and facilitators viewed the data collected as part of an audit in negative terms, with implications for individual employees' jobs and the agency's funding stability. Importantly, these examples illustrate a lack of any kind of “ownership” of the data that are collected through evaluation forms and a lack of clarity about its intended use. On the one hand, data collection forms are viewed as superfluous because “no one” looks at them. On the other hand, they are seen as crucial due to their importance in funding decisions, and as potentially important if time and autonomy were available to change service delivery as a result of evaluation.

In addition, the requirement to collect and record information from program participants and the proliferation of “paperwork” that results potentially interferes with rather than enhances already overburdened staff. The audits and accountability systems associated with SISTA require facilitators to ask clients questions about their sexual history, sexual health, HIV-related risk behaviors, and relationships. These questions are asked in preintervention assessments and formative focus groups, before facilitators have the opportunity to establish

rapport with clients. Moreover, participants may not understand why such personal information is being collected and may therefore not complete intake forms fully and accurately. Likewise, some participants need additional personal assistance completing the forms due to low literacy skills, thus requiring more time. As a result, facilitators would have to convene their SISTA groups for an additional hour in order to allow time to complete paperwork or ask participants to come back after completing the paperwork in order to participate in the group session.

Given the unpredictable nature of participants' lives, including child care and work responsibilities or eviction from homes or the residential substance use treatment facilities through which they were recruited, participants would often drop out of the program due to this additional burden created by data collection and paperwork:

P: It's too much paperwork—like a lot of paperwork, like paperwork like oh my gosh. The pretest that they have to take. They have to do stages of change. We have to input them and then export them out to the state.

I: *You* have to input them? You have to enter them?

P: I have to enter them and then export them. We have to do all the data forms for each client so that the state can receive it. Everything that the state was doing, it was brought back to the health educators. So it's the paperwork for SISTA ... You see the mess—you see our binders. So when they tell us "paperwork," we would be like, "Oh my gosh! I hope it's not extra paperwork because these ladies are not going to stay!" Because what scares them is all of this stuff: "Why do you need all of this information? Why do you need this?" Or, "Are you going to give this information to someone else?" That's their concern. It's just not like a sign-in sheet that you can sign your first name and your last name and your date of birth. No. They are asking you questions, like "How many times has the penis gone into your vagina in this month?" (Agency 1, Facilitator 2)

This participant highlights a tension between the need to ensure fidelity and appropriate numbers of women served, and providers' actual ability to provide services to clients, as paperwork requires a significant time burden for both providers and clients. Additionally, although data about participants' risk behaviors could be used to determine whether the program was achieving its desired outcomes, it also could be seen as an intrusion on participants' privacy, a burden, and a disincentive to participation. Although the need to use documentation to demonstrate accountability was seen as important, collecting data from participants and completing the required paperwork became something "to do" rather than something of value, consistent with audit culture.

Discussion

In this article, we used the lens of audit culture to explore the impact of numbers-based evaluation and accountability strategies on the organizational context of CBOs involved in HIV prevention service delivery. Although accountability may help ensure that the target population is served, the program implemented with fidelity, and the expected outcomes achieved, these processes can lead to an audit culture. From this study, we found that audit

culture results in the perception of chasing numbers and paper pushing, competition between agencies, and doubts about collaboration with other community groups. We illustrated that this approach to evaluation encourages a prioritization of meeting recruitment goals, contributed to a sense of competition between agencies, and created a reluctance to work together to best serve a particular geographic region or target population. Likewise, a narrowly defined target population contributed to a sense among providers that meeting numbers goals took priority over providing other services to at-risk clients. It also contributed to a sense among providers that they were unable to serve people in need of HIV prevention services because they did not fit into categories that counted toward deliverables. Facilitators in our study expressed anxiety about meeting recruitment goals and the implications for failure to meet these goals for both their own job security and the viability of their agency's programs more generally. Participants described data collected as part of these evaluations as "disappearing" after collection. That is, it was often unclear to them how the data they collected would be used, if at all. Rather than incorporating feedback and training, the numbers-based monitoring and evaluation we described had few—if any—feedback mechanisms, in contrast to other feedback they received from supervisors or colleagues about how they delivered the intervention or problem solving related to interactions with clients. The SISTA providers we interviewed saw the reporting requirements and recruitment goals as punitive rather than supportive and expressed a sense of diminished autonomy due to pressure to meet numbers-based goals above all other priorities. As we and others have argued (Altman, 1994; Owczarzak, 2010), service providers often come to the field of HIV prevention due to personal connections with the disease or affected communities, a desire to help others, or a social justice mission. Audit culture may conflict with and undermine these values, such as in the case of denying program participation to certain individuals because they would not help them meet their deliverables.

While the focus on quantifiable results may be useful for determining whether programs serve intended clients, for many government-funded programs, number of clients served is an inadequate assessment of impact. Demonstrating changes in performance or behavior is also an important dimension of assessing program impact. The evaluation activities for SISTA referred to under the "Documentation" section above include assessment of knowledge and behavior change over time (i.e., HIV knowledge, condom use, condom negotiation self-efficacy, etc.), in addition to demographics of clients served, and fidelity checklists occasionally completed by supervisors attending sessions of SISTA. However, our results suggest that for these service providers, funders focused most on numbers of eligible clients served, rather than actual implementation of SISTA, and that fidelity checklists were not actually used to provide feedback to providers to improve service delivery.

Although this article focused on a single evidence-based program, the evaluation and monitoring context we described is not unique to HIV prevention efforts. For example, in a study with 178 CBOs in three fields (social services, developmental disabilities, and community development), Carman (2007) determined that over 90% of organizations collected and reported numbers-based evaluation data (e.g., number of participants, demographics, and financial expenditures) and that their evaluation activities were not focused on improving service delivery. Rather, they used these evaluation data in order to

obtain licenses, accreditation, or certification or to meet funders' requirements. The potential for lack of providers' previous experience with newly identified populations of importance, interagency competition, and burdens of documentation leading to decreased staff retention and job satisfaction are all potentially applicable in other service delivery contexts.

Considering the consequences arising from traditional assessment strategies suggested by our findings, alternative strategies to monitoring and evaluation should be considered to better incorporate accountability measures, increase staff commitment to evaluation, and promote better retention and satisfaction. Program evaluation can include a variety of strategies and information sources, including both qualitative assessment of participant experience in programs and quantitative measurement of program reach and reported benefits (Patton, 2002). More recently, in response to critiques of standard evaluation strategies, new evaluation strategies have emerged, including participatory (Smits, 2013), collaborative (O'Sullivan, 2007), and realistic approaches (Holma & Kontinen, 2011). These newer evaluation strategies take into account the political nature of social problems and value-based definitions of these problems and critically examine the premise behind public health programs designed to address them. Aarons et al. (2009) found that fidelity monitoring predicted greater staff retention over time, even while controlling for job attitudes and autonomy. Importantly, fidelity monitoring was achieved using a supportive/coaching approach, to avoid an explicit focus on monitoring. Trained professionals who were not the providers' supervisors, and picked as experts by their colleagues, were assigned as "ongoing consultants" to providers. They attended two sessions a month with each provider and then prepared feedback and additional training as needed directly to the provider (not to the provider's supervisor). These consultants were specifically trained to tailor and individualize their feedback and build rapport. The nature of the fidelity monitoring used in this study may point toward recommendations for monitoring and evaluation approaches that balance accountability and program implementation fidelity while still supporting staff and reducing excessive burden of measurement.

Second, studies have also documented staff disappointment with outcome measurement as the key evaluative technique. Service providers often assert that this approach fails to capture important ways that nonprofits address social problems, does not reflect the experimental and highly responsive nature of nonprofit work, and may not capture all the work that nonprofits do (Benjamin, 2012). Although staff and directors at CBOs may acknowledge the potential contribution of numbers-based evaluation data to the improvement of agency services, the results of evaluation are often used to demonstrate fiscal prudence, efficiency, and accountability to funders and the public, rather than to produce information for the organization's benefit (Chouinard, 2013a, 2013b; Miller & Cassel, 2000). Holma and Kontinen (2011) have proposed evaluations beyond whether a project has met its numbers-oriented objects to understand what made a project successful or not, which funders could consider as part of the evaluations of CBO activities. For example, they could incorporate narratives of participants' and providers' experiences with a particular program (Holma & Kontinen, 2011). Additionally, providers may find information regarding what participants say about the benefits of a program or their experiences with it more valuable than the number of individual program participants (Patton, 2002).

When monitoring and evaluation are seen as sources of stress and the results of evaluation and monitoring efforts are potentially punitive (i.e., failure to achieve audit measures as a basis for dismissal), program facilitators and other direct service providers may not view evaluation as central to their work and a valuable learning opportunity. Developing performance and evaluation measures that better reflect the experiences and goals of agencies and facilitators could potentially improve staff commitment to monitoring and evaluation requirements. As we and others have argued elsewhere, service providers often have different ways of assessing program effectiveness beyond the number of clients reached (Holma & Kontinen, 2012; Owczarzak & Dickson-Gomez, 2011; Owczarzak & Dickson-Gomez, 2012). Developing systems in which these experiences and observations are taken into account and supplement numerical data may increase the perceived value of monitoring and evaluation to providers and decrease the negative consequences from an emerging audit culture. However, such approaches would require a coordinated effort on the part of both funders and service providers to identify and support the most mutually beneficial evaluation strategies and the expertise needed to implement them.

Our study has several limitations. One potential limitation is the specific focus on SISTA providers and the effects of audit culture on the organizational context of service delivery from the perspective of a single public health intervention for a specific target population. Second, we did not examine the effects of evaluation and accountability systems on staff turnover specifically. Audit culture as manifest in the prioritization of numbers-based accountability systems may have implications for employee satisfaction, staff burnout and turnover, and quality of care. Therefore, more research is needed to examine how different forms of monitoring may be harnessed to improve staff perceptions and ultimately decrease staff turnover. Additionally, due to a relatively small sample size and homogeneity in participant characteristics, we did not have the ability to explore potential differences in perceptions of audit culture by participant characteristics such as length of employment, gender, or race, which should be explored in future research. Next, further investigation of the potential for agencies that implement programs to overlapping sets of clients is outside the scope of this article. However, it is an important area to investigate further in order to explore how to maximize collaboration and mitigate competition between agencies. Finally, we cannot determine whether the aspects of audit culture described above have actual detrimental effects on clients, and we have no evidence on whether numbers-based evaluation approaches help or hinder service delivery for these communities overall. For example, from funders' perspectives, struggles to find enough eligible clients may suggest that SISTA is not needed, although this is in direct contrast with frontline service providers' perspectives on their own communities; or, minimal enrollment requirements may ensure more people being served than if they did not exist.

Evaluation and monitoring activities, including numbers-based accountability measures, are necessary to assess and improve the effectiveness of public health programs, demonstrate that funds are used effectively and appropriately, and assess program quality. While numbers-based evaluation activities may provide accountability to funders, other potential benefits of more broadly defined evaluation activities are not being realized when this evaluation strategy is given highest priority. Program evaluation that includes broader techniques and approaches can be useful for program improvement, planning for future

programs, and achieving agency goals more broadly (e.g., fulfilling a social justice mission to ensure the most vulnerable populations are receiving appropriate and effective services). Audit culture curtails these other potential uses of data and evaluation, with top-down decision making from funders based primarily on achieving numbers-based goals. This approach may limit what populations to serve, which programs to implement, and what approach to service provision to take. Therefore, more creative approaches to evaluation are warranted that not only encourage accountability but also support staff and reduce evaluation burden to reduce consequences arising from an audit culture. Reimagining the purpose of evaluation from “accountability” to “learning” creates opportunities for a reconsideration of evaluation and audit methodologies (Oswald & Taylor, 2010; Torres & Preskill, 2001).

Acknowledgments

We thank this project’s coordinator, Cassandra Wright, who contributed immensely to this study’s success through her work contacting providers and other participants and maintaining the integrity of research protocols.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research was supported by R01-MH089828 and P30-MH52776, both from the National Institutes of Mental Health.

References

- Aarons GA, Sommerfeld DH, Hecht DB, Silovsky JF, Chaffin MJ. The impact of evidence-based practice implementation and fidelity monitoring on staff turnover: Evidence for a protective effect. *Journal of Consulting and Clinical Psychology*. 2009; 77:270–280. [PubMed: 19309186]
- Adams, V. Evidence-based global public health: Subjects, profits, erasures. In: Biehl, J., Petryna, A., editors. *When people come first: Critical studies in global health*. Princeton, NJ: Princeton University Press; 2013. p. 54–90.
- Altman, D. *Power and community: Organizational and cultural responses to AIDS*. London, England: Taylor and Francis; 1994.
- Benjamin L. Nonprofit organizations and outcome measurement: From tracking program activities to focusing on frontline work. *American Journal of Evaluation*. 2012; 33:431–437.
- Bernard, HR. *Research methods in anthropology: Qualitative and quantitative approaches*. 5th. Lanham, MD: AltaMira; 2011.
- Carey JW, Morgan M, Oxtoby MJ. Intercode agreement in analysis of responses to open-ended interview questions: Examples from tuberculosis research. *Cultural Anthropology Methods*. 1996; 8:1–5.
- Carman J. Evaluation practice among community-based organizations: Research into the reality. *American Journal of Evaluation*. 2007; 28:60–75.
- Chen H. Development of a national evaluation system to evaluate CDC-funded health department HIV prevention programs. *American Journal of Evaluation*. 2001; 22:55–70.
- Chouinard J. The case for participatory evaluation in an era of accountability. *American Journal of Evaluation*. 2013a; 34:237–253.
- Chouinard J. The practice of evaluation in public sector contexts: A response. *American Journal of Evaluation*. 2013b; 34:266–269.
- Collins C, Harshbarger C, Sawyer R, Hamdallah M. The diffusion of effective behavioral interventions project: Development, implementation, and lessons learned. *AIDS Education & Prevention*. 2006; 18:5–20. [PubMed: 16987085]
- DiClemente RJ, Wingood GM. A randomized controlled social skills trial: An HIV sexual risk-reduction intervention among young adult African-American women. *Journal of the American Medical Association*. 1995; 274:1271–1276. [PubMed: 7563531]

- Dodd S, Meezan W. Matching AIDS service organizations' philosophy of service provision with a compatible style of program evaluation. *Journal of Gay & Lesbian Social Services*. 2003; 15:163–180. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=9729009&site=ehost-live&scope=site>.
- Dworkin SL, Pinto RM, Hunter J, Rapkin B, Remien RH. Keeping the spirit of community partnerships alive in the scale up of HIV/AIDS prevention: Critical reflections on the roll out of DEBI (Diffusion of Effective Behavioral Interventions). *American Journal of Community Psychology*. 2008; 42:51–59. [PubMed: 18612809]
- Glassman M, Lacson R, Collins B, Hill C, Wan C. Lessons learned from the first year of implementation of the Centers for Disease Control and Prevention's standardized system for HIV prevention programs. *AIDS Education and Prevention*. 2002; 14:49–58. [PubMed: 12092936]
- Harper GW, Contreras R, Bangi A, Pedraza A. Collaborative process evaluation: Enhancing community relevance and cultural appropriateness in HIV prevention. *Journal of Prevention & Intervention in the Community*. 2003; 26:53–69. Retrieved July 22, 2014, from http://www.tandfonline.com/doi/abs/10.1300/J005v26n02_05.
- Holma K, Kontinen T. Realistic evaluation as an avenue to learning for development NGOs. *Evaluation*. 2011; 17:181–192.
- Holma K, Kontinen T. Democratic knowledge production as a contribution to objectivity in the evaluation of development NGOs. *Forum for Development Studies*. 2012; 39:83–103.
- Hull E. Paperwork and the contradictions of accountability in a South African hospital. *Journal of the Royal Anthropological Institute*. 2012; 18:613–632.
- Hwang H, Powell W. The rationalization of charity: The influences of professionalism in the non-profit sector. *Administrative Science Quarterly*. 2009; 54:268–298.
- Kipnis A. Audit cultures: Neoliberal governmentality, socialist legacy, or technologies of governing? *American Ethnologist*. 2008; 35:275–289.
- Lincoln, YS., Guba, EG. *Naturalistic inquiry*. Beverly Hills, CA: Sage; 1985.
- Lyles CM, Kay LS, Crepaz N, Herbst JH, Passin WF, Kim AS, DeLuca JB. Best-evidence interventions: Findings from a systematic review of HIV behavioral interventions for US populations at high risk, 2000–2004. *American Journal of Public Health*. 2007; 97:133–143. [PubMed: 17138920]
- Miles, MB., Huberman, AM., Saldana, J. *Qualitative data analysis: A methods sourcebook*. 3rd. Thousand Oaks, CA: Sage; 2014.
- Miller R, Cassel JB. Ongoing evaluation in AIDS-service organizations: Building meaningful evaluation activities. *Journal of Prevention and Intervention in the Community*. 2000; 19:23–39.
- Napp D, Gibbs D, Jolly D, Westover B, Uhl G. Evaluation barriers and facilitators among community-based HIV prevention programs. *AIDS Education & Prevention*. 2002; 14:38–48. [PubMed: 12092935]
- Niba MB, Green JM. Major factors influencing HIV/AIDS project evaluation. *Evaluation Review*. 2005; 29:313–330. Retrieved July 22, 2014, from <http://www.ncbi.nlm.nih.gov/pubmed/15985522>. [PubMed: 15985522]
- O'Sullivan R. Collaborative evaluations: A step-by-step model for the evaluator. *American Journal of Evaluation*. 2007; 28:381–382.
- Oldani M. Assessing the 'relative value' of diabetic patients treated through an incentivized, corporate compliance model. *Anthropology and Medicine*. 2010; 17:215–228. [PubMed: 20721758]
- Oswald K, Taylor P. A learning approach to monitoring and evaluation. *IDS Bulletin*. 2010; 41:114–120.
- Owczarzak J. Activism, NGOs, and HIV prevention in postsocialist Poland: The role of 'anti-politics'. *Human Organization*. 2010; 69:200–211. [PubMed: 25308987]
- Owczarzak J. Evidence-based HIV prevention in community settings: Provider perspectives on evidence and effectiveness. *Critical Public Health*. 2012; 22:73–84.
- Owczarzak J, Dickson-Gomez J. Providers' perceptions of and receptivity toward evidence-based HIV prevention interventions. *AIDS Education and Prevention*. 2011; 23:105–117. [PubMed: 21517660]
- Patton, MQ. *Qualitative research and evaluation methods*. Thousand Oaks, CA: Sage; 2002.

- Reynolds LJ. 'Low-hanging fruit': Counting and accounting for children in PEPFAR-funded HIV/AIDS programmes in South Africa. *Global Public Health*. 2014; 9:124–143. Retrieved February 3, 2015, from <http://www.ncbi.nlm.nih.gov/pubmed/24498970>. [PubMed: 24498970]
- Ryan GW. Measuring the typicality of text: Using multiple coders for more than just reliability and validity checks. *Human Organization*. 1999; 58:313–322.
- Shenton A. Strategies for ensuring trustworthiness in qualitative research projects. *Education for Information*. 2004; 22:63–75.
- Shore C. Audit culture and illiberal governance: Universities and the politics of accountability. *Anthropological Theory*. 2008; 8:278–297.
- Smits P. An assessment of the theoretical underpinnings of practical participatory evaluation. *American Journal of Evaluation*. 2013; 29:427–442.
- Strathern, M. Abstraction and decontextualisation: An anthropological comment or: E for ethnography. *Virtual society? Get real!*. 2000a. Retrieved from <http://virtualsociety.sbs.ox.ac.uk/GRpapers/strathern.htm>
- Strathern, M., editor. *Audit cultures: Anthropological studies in accountability, ethics, and the academy*. London, England: Routledge; 2000b.
- Strauss, A., Corbin, J. *Basics of qualitative research: Grounded theory and procedures and techniques*. Newbury Park, CA: Sage; 1990.
- Taveras S, Duncan T, Gentry D, Gilliam A, Kimbrough I, Minaya J. The evolution of the CDC HIV prevention capacity-building assistance initiative. *Journal of Public Health Management & Practice*. 2007; 13:S8–S15.
- Torres R, Preskill H. Evaluation and organizational learning: Past, present, future. *American Journal of Evaluation*. 2001; 22:387–395.
- Vannier CN. Audit culture and grassroots participation in rural Haitian development. *Political and Legal Anthropology Review*. 2010; 33:282–305.
- VERBI GmbH. MAXQDA: The art of data analysis. 2011. Retrieved from www.maxqda.com