IDENTIFYING AND UNDERSTANDING INDIGENOUS WAYS OF EVALUATING PHYSICAL ACTIVITY PROGRAMS

Erica Blue Roberts, PhD, MHS, James Butler III, DrPH, MEd, and
Kerry M. Green, PhD

Abstract: Indigenous evaluation frameworks have not been investigated in the context of American Indian and Alaska Native (AI/AN) physical activity programs, an important area given the relationship between effective physical activity programs and quality of life among these populations. To address this gap, staff members of AI/AN physical activity programs were interviewed to explore their understanding of and experiences with evaluation. Findings suggest that Indigenous evaluation is perceived as narrative and holistic, Indigenous knowledge is used in program decision making, though it is not always acknowledged as evaluation, and there is not a universally desired way to evaluate AI/AN physical activity programs.

INTRODUCTION

Indigenous and tribal evaluation models have been created based on traditional ways in which Indigenous peoples assign merit or worth to activities or programs of importance to them (LaFrance & Nichols, 2009). These models have been designed to build the capacity of American Indian and Alaska Native (AI/AN) communities and external evaluators to implement culturally responsive evaluation methods that recognize and value Indigenous knowledge, respect tribal sovereignty, and recognize the historical misuse and imposition of research and evaluation conducted *on* and not *with* AI/AN communities (Kawakami, Aton, Cram, Lai, & Porima, 2007; LaFrance, 2004; Tribal Evaluation Workgroup, 2013). However, Indigenous evaluation approaches have rarely been assessed within the context of physical activity programs, and the types and sources of capacity necessary to implement these methods have not been identified.

Physical Activity among AI/ANs

An analysis of the 2000-2010 Behavioral Risk Factor Surveillance System data found that, when compared to their White counterparts, AI/AN males reported higher rates of obesity (33.9% vs. 23.3%, respectively) and diabetes (15.1% vs. 7.3%, respectively); AI/AN females also reported higher rates of obesity (35.5% vs. 21%, respectively) and diabetes (14.3% vs. 5.8%, respectively; Cobb, Espey, & King, 2014). Physical activity has been found to protect against type 2 diabetes, reduce the prevalence of obesity and its associated indicators (e.g., body mass index [BMI], percentage of body fat) and complications, and improve perceived quality of life among these populations (Coble & Rhodes, 2006; Foulds, Warburton, & Bredin, 2013; Kriska et al., 2003; Poltavaski, Holm, Vogeltanz-Holm, & McDonald, 2010). In addition, AI/ANs have reported a greater number of mentally unhealthy days when compared to all other racial and ethnical groups, and a 2010 study found that regular physical activity was associated with fewer days of poor mental health among this population (Poltavaski et al., 2010).

Recently, there has been an increase in funding to support the design and implementation of interventions directed at promoting an active lifestyle and addressing the rates of chronic disease among AI/ANs (Teufel-Shone, Fitzgerald, Teufel-Shone, & Gamber, 2009). Federal agencies such as the Indian Health Service (IHS), the Centers for Disease Control and Prevention, and the Executive Office of the President have launched physical activity-promoting initiatives for AI/ANs, including the Special Diabetes Program for Indians, Healthy Weight for Life, and Let's Move in Indian Country. Additionally, nongovernment and philanthropic organizations such as the Notah Begay III Foundation and the Nike N7 Fund have supported similar programs. However, meaningful program evaluation is lacking at the local level to inform program improvement and sustainability, measure effectiveness, and communicate impact to the communities being served (Fleischhacker, Roberts, Camplain, Evenson, & Gittelsohn, 2015; Foulds et al., 2013; Teufel-Shone et al., 2009).

Teufel-Shone and colleagues' (2009) systematic review of peer-reviewed articles and the grey literature identified 64 physical activity programs being implemented across Indian Country, and found that only 42% measured impact. Another recent systematic review of the peer-reviewed literature indicated that only 8 of 19 interventions focusing on physical activity among AI/AN youth described the use of formative assessment, and only 8 included process or outcome evaluation (Fleischhacker et al., 2015).

This lack of evaluation is significant, given that evaluation not only can lead to program improvement and better health outcomes, but also is linked to program sustainability, as many grant mechanisms require evaluation and reporting during the funding period (i.e., tribal organizations must demonstrate success to ensure continued funding), particularly where funding mechanisms are limited (Teufel-Shone et al., 2009). In addition, funding entities often require standardized evaluation methods that provide comparable data across grantees, but do not attend to the unique, local differences among AI/AN communities, and they assume that there can be a universal way to evaluate programming. However, it also is important to identify and evaluate factors valued by local AI/AN communities (Grover, 2010). Therefore, evaluation that is localized and based on the values of AI/AN communities as well as funding agencies and other scientific entities is paramount to ensuring true success and effectiveness of programs for improving health and well-being.

Indigenous Knowledge

Indigenous knowledge (or epistemology) is, in its most basic form, the ways in which Indigenous peoples (e.g., AI/ANs; First Nations and Aboriginal peoples) come to know the world. This knowledge is created uniquely within each Indigenous community; however, the following three processes often transcend communities: empirical knowledge, traditional knowledge, and revealed knowledge (Barnhardt & Kawagley, 2005; Lavallee, 2009). Empirical knowledge is representative of observations taken from different vantage points over time in real-life settings (LaFrance & Nichols, 2009; Lavallee, 2009). Traditional knowledge is based on the history and experiences of the community and is passed down through generations (LaFrance & Nichols, 2009; Lavallee, 2009). Revealed knowledge is discovered through dreams, visions, and spiritual practices (LaFrance & Nichols, 2009; Lavallee, 2009). In addition, scholars have identified other characteristics of Indigenous ways of knowing that are common across communities. For example, Indigenous knowledge often is identified as favoring holistic thinking, subjectivity, and social betterment, in contrast to Western science-based practice that values linear and hierarchical thinking, and privileges objectivity (a distance between the evaluator/researcher and the participant) and improving the individual participant (Cavino, 2013; LaFrance, 2004; LaFrance, Kirkhart, & Nichols, 2015; LaFrance & Nichols, 2010; Tribal Evaluation Workgroup, 2013). Further, Western science often utilizes standardized quantitative

instruments that emphasize validity and reliability, as compared to qualitative methods, which are generally more accepted and more meaningful to AI/ANs, given their oral traditions (Chouinard & Cousins, 2007; LaFrance, 2004). In the context of program evaluation, Indigenous knowledge can be used when examining the merit or worth of a program (LaFrance & Nichols, 2009).

Existing Indigenous Evaluation Models

To increase and promote the inclusion of Indigenous knowledge in evaluation, as well as to address the history of exploitative, intrusive, and invasive evaluation imposed upon Native populations, Indigenous evaluation frameworks and models have been created (Cavino, 2013; Hodge, 2012; LaFrance & Nichols, 2010; Tribal Evaluation Workgroup, 2013). In 2012, the U.S. Department of Health and Human Services, Administration for Children and Families, Children's Bureau convened a Tribal Evaluation Workgroup, composed of staff members of tribal child welfare programs, university researchers that have worked with AI/AN communities, technical assistance providers, and others with expertise in evaluation, to create a tool that improves AI/AN evaluation capacity through the incorporation of Indigenous knowledge, culture, and tradition: the *Roadmap for Collaborative and Effective Evaluation in Tribal Communities* (Tribal Evaluation Workgroup, 2013). Within the context of tribal child welfare programs, the purpose of the roadmap was to guide the development and implementation of evaluation with external partners, serve as a reference in grant applications, assist in training future researchers and evaluators at Tribal Colleges and Universities (TCUs), inform evaluation requirements established by federal and state funders, and improve Tribal Institutional Research Review Boards' guidelines (Tribal Evaluation Workgroup, 2013). The workgroup agreed that the roadmap should serve to establish a "new narrative" for evaluation that empowered AI/ANs to "move beyond fear and distrust of evaluation" and build relationships with various stakeholders that support meaningful and culturally inclusive knowledge exchange (Tribal Evaluation Workgroup, 2013, p. 6).

The circular shape of the roadmap represents the cyclical process of improving programming through evaluation (Figure 1; Tribal Evaluation Workgroup, 2013, p. 7). The overarching concepts of 'relationship building' and 'knowledge and skill building' overlap to reflect their interdependence and equal importance, with the goal of 'building a new narrative'

symbolically in the center as the central focus of the roadmap. These concepts are encompassed by 'values' of Indigenous communities, and further underpinned by 'historical context', which shapes existing and future practices with this population. Embedded within the roadmap are the names of various types of stakeholders that play significant roles in the process of evaluation with AI/ANs (Tribal Evaluation Workgroup, 2013).

Figure 1
Roadmap for Co-creating Collaborative and Effective Evaluation to Improve
Tribal Child Welfare Programs
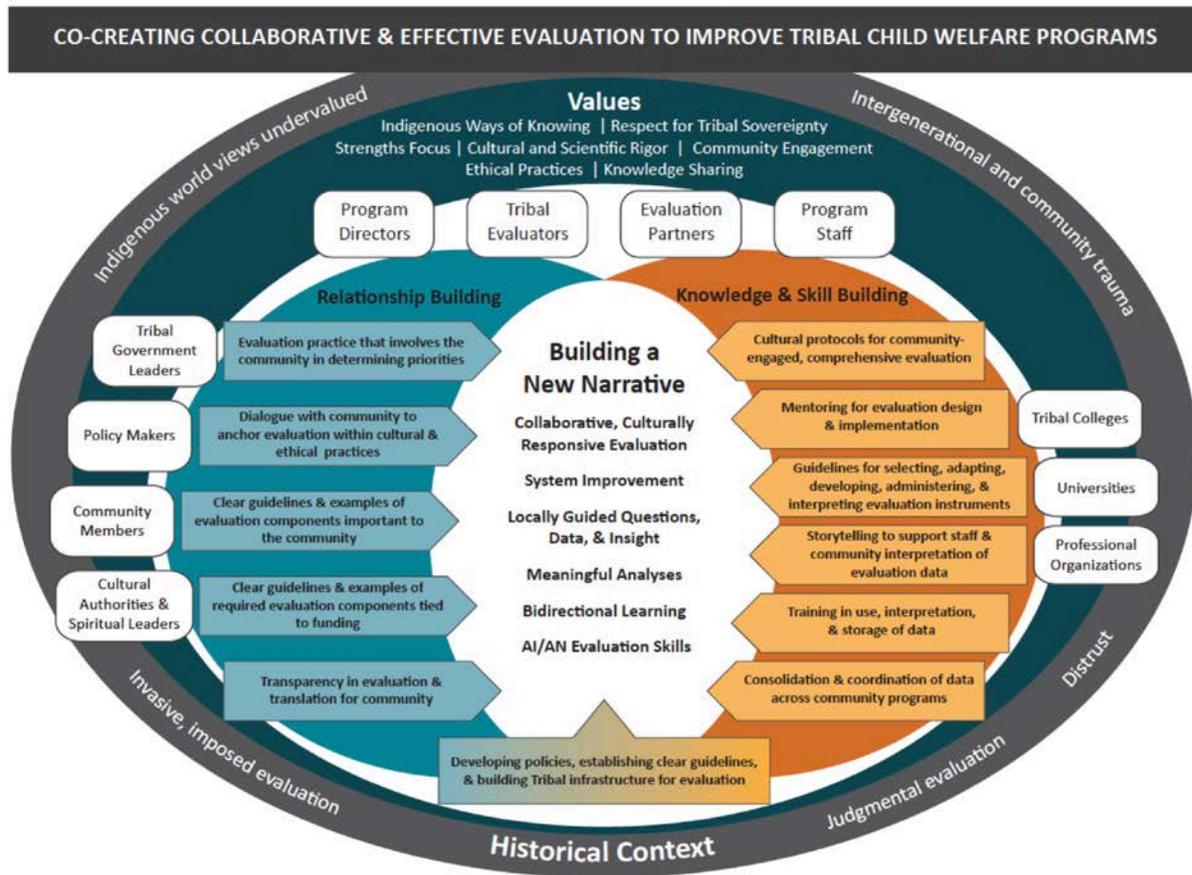
Another approach created to support and promote sound evaluation in AI/AN communities is the Indigenous Evaluation Framework (IEF; LaFrance et al., 2015). This framework was developed by the American Indian Higher Education Consortium (AIHEC) and was informed by the knowledge and wisdom of over 100 AI/AN Elders, cultural experts, scientists, and educators (LaFrance & Nichols, 2009). The AIHEC convened these individuals over the course of four 1-day focus group sessions during which they discussed ideas and goals

for evaluation within the context of Indigenous ways of knowing (LaFrance & Nichols, 2009). The result was the IEF, which identifies four core values integral to implementing evaluation with AI/ANs: being a people of place, recognizing gifts, honoring family and community, and respecting sovereignty (LaFrance & Nichols, 2010). Based on these values, the framework suggests that evaluation with AI/AN programs should not focus on generalizability; should utilize a strength-based approach to assess achievement; should emphasize the role of the community throughout the entire process; and should include and respect traditional, empirical, and revealed knowledge (LaFrance & Nichols, 2010). The ideas underpinning the framework were pilot tested at seven TCUs and one elementary school, as the framework was focused on educational evaluation (LaFrance & Nichols, 2009).

The aforementioned frameworks attend to the call from scholars for a decolonization of evaluation. Decolonization is an approach that serves to counter colonization and the privileging of Western knowledge over Indigenous knowledge (Smith, 2012). It can privilege Indigenous knowledge, voices, and experiences, and "re-write and re-right [the Indigenous] position" in society (Denzin & Lincoln, 2008; Lavallee, 2009; Smith, 2012, p. 29). In the current social context, the continued view of Western knowledge as the only means of generating logical ideas perpetuates imperialism, colonialism, and power differentials between Western and Indigenous peoples (Smith, 2012). In the domain of health, scholars suggest that colonization "occurs when majority society…assumptions about health and disease are applied to develop programs to alleviate poor health of Indigenous peoples caused by colonization" (Bartlett, Iwasaki, Gottlieb, Hall, & Mannell, 2007, p. 5). In relation to research and evaluation, Western science is identified and regarded as superior and is widely taught and practiced, and, while it may seek to include Indigenous perspectives in the data, findings are often used to change the lives of Indigenous peoples based on Western assumptions and standards of what is progress or what is good/bad and right/wrong (Bartlett et al., 2007; Smith, 2012).

While these frameworks provide guidance for conducting culturally sensitive, culturally driven, and decolonized evaluation with AI/ANs, they are not specific to the evaluation of any certain type of program or intervention (e.g., nutrition, physical activity). Further, scholars have not explored how AI/AN organizations conducting evaluations understand the components of these frameworks. Moreover, to date, these frameworks have not been investigated in the context of physical activity. To explore how Indigenous evaluation models and knowledge can be

applied to AI/AN physical activity programs and identify a best practice for evaluating AI/AN physical activity programs, the first author, a doctoral candidate, conducted an exploratory qualitative study to answer the following research questions: (1) How do AI/AN organizations define and describe Indigenous knowledge-based evaluation in the context of physical activity programs? (2) How interested are AI/AN organizations in using Indigenous knowledge-based evaluation to evaluate their AI/AN physical activity programs? (3) What organizational capacity is necessary for conducting evaluation grounded in Indigenous knowledge? and (4) What barriers do AI/AN organizations face in implementing Indigenous knowledge-based evaluation to evaluate their physical activity programs? This paper presents the findings and discusses what program evaluation looks like for AI/AN organizations implementing grant-funded physical activity programs.

METHODS

The first author, who self-identifies as a White and AI female and was a doctoral candidate at the time this work was done, conducted this study for her dissertation research, which was based on her experiences as an evaluator for AI/AN public health programs. In her role as an evaluator, she noticed the challenges faced in implementing meaningful and useful evaluation for tribal communities and wanted to contribute to the scholarship and practice of culturally responsive and Indigenous evaluation with and for AI/ANs. Aligned with the practice of reflexivity (Krefting, 1999), the first author identified her assumptions going into the study: first, that individuals working on the evaluation of grant-funded AI/AN physical activity programs desired to use Indigenous evaluation, could recognize Indigenous evaluation, and would chose Indigenous evaluation over Western science-based evaluation; and second, that the evaluations required by the programs' funding agencies might influence how individuals perceived or understood Indigenous and Western science-based evaluation.

Study Design

To explore Indigenous evaluation in the context of AI/AN physical activity programs, the research team conducted an exploratory qualitative study grounded in a decolonizing framework (i.e., an approach that emphasizes the use of culturally based, iterative, and process-oriented

methods whereby knowledge is gathered through respectful listening, the study participants are involved in the review of data and interpretations, and bidirectional learning and empowerment occur; Bartlett et al., 2007). The research team included the first author, a graduate research assistant, and the first author's academic advisors. The study was approved by the University of Maryland's Institutional Review Board, and informed consent was obtained from all participants.

Sample

The research team conducted in-depth interviews with individuals working at AI/AN organizations implementing externally funded physical activity programs ($n = 17$). When conducting decolonizing research, it is recommended that a diverse sample of participants be recruited to ensure the viewpoints of many within the community are heard (Bartlett et al., 2007). Therefore, we purposefully sampled AI/AN organizations conducting physical activity programs based on their funding source (e.g., federal grant, state grant), geographic location (e.g., IHS area), and type of community (e.g., urban, rural). From these organizations, we targeted individuals working most closely on the evaluation of the physical activity program as prospective participants.

The research team chose to include non-AI/AN participants in the study, given that no literature has suggested that only AI/ANs can practice Indigenous evaluation and because Indigenous evaluation has been described as an approach that is "practiced by evaluators who value building strong relationships with those involved in the evaluation" (LaFrance, 2004, p.43). This definition does not require that Indigenous evaluators be Indigenous; rather, they should have an understanding of and appreciation for the tribal context and base their evaluative decisions on the relationships they have built with those involved.

Participants were staff members at 17 different AI/AN organizations. Their physical activity programs were located in the following 9 IHS areas: Alaska, Albuquerque, Bemidji, California, Great Plains, Nashville, Navajo, Oklahoma City, and Phoenix. Most programs were implemented in reservation communities ($n = 11$), with the remaining in rural nonreservation areas ($n = 4$), urban areas ($n = 1$), and both reservation and urban communities ($n = 1$). The majority reported the source of their physical activity grants as federal agencies ($n = 14$); other grants came from nonprofit organizations ($n = 6$), and/or the state ($n = 3$). Some programs ($n = 6$) were funded by multiple sources. Participants identified a wide range of physical activity

interventions implemented through their programs, reaching both AI/AN youth and adults and spanning from environmental or policy change to community-wide health promotion to individual behavior change. These interventions included community events (e.g., hosting 5k runs/walks, health fairs; $n = 12$); partnership with a fitness center or gym ($n = 9$), youth after-school programs, in-school physical education, summer camps, and/or sports leagues ($n = 7$); group exercise classes ($n = 5$); and personal training or individual case management ($n = 5$; see Table 1).

Table 1
Program Characteristics

	Number (17)	Percentage
Program Locale		
AI/AN Reservation	11	64.7%
Rural nonreservation	4	23.5%
Urban	1	5.9%
Both urban and reservation	1	5.9%
Grant Mechanism[a]		
Federal agency	14	82.4%
Nonprofit organization	6	35.3%
State agency	3	17.6%
Intervention Type[b]		
Hosts community events	12	70.6%
Provides fitness/gym space	9	53%
Coordinates youth physical activity programs	7	41.2%
Provides group exercise classes	5	29.4%
Provides personal training/case management	5	29.4%

[a] Six participants identified more than one funding source; therefore, the percentages do not add to 100%. [b] Participants noted that their interventions have many of the identified components; therefore, the percentages do not add to 100%.

The 17 interviews included 18 individuals (2 individuals participated in an interview together). Of those 18 participants, 17 provided additional demographic information (see Table 2). These participants held many (often multiple) roles and/or job titles within their organizations, including Program/Project Director ($n = 9$), Program Coordinator ($n = 6$), Health Educator ($n = 3$), and Personal Trainer ($n = 2$). Approximately half of the participants had worked at their organizations for less than 5 years, and most ($n = 15$) reported at least some training in evaluation. Nearly half of the participants ($n = 8$) self-identified as AI, AN, or Native Hawaiian. In addition, participants came from a variety of backgrounds, including development, exercise science, nursing, fitness training, accounting, and physical therapy.

Table 2
Participant Characteristics

	Number (17)[a]	Percentage
Role/Job Title[b]		
Program/Project Director	9	52.9%
Program Coordinator	6	35.3%
Health Educator	3	17.7%
Personal Trainer	2	11.8%
Evaluation Training		
Yes	15	88.2%
No	2	11.8%
Years at Organization		
1-5 years	8	47.1%
6-10 years	4	23.5%
11+ years	5	29.4%
Self-identified Race		
American Indian/Alaska Native/Native Hawaiian	8	47.1%
Non-Native[c]	9	52.9%

[a] Of the 18 individuals who participated in the 17 interviews, 17 provided demographic information. [b] 20 job title are identified for 17 participants because one participant held 2 job titles, and 1 participant held 3 job titles. [c] Non-Natives self-identified as White (n = 7), Spanish and Asian (n =1), and unspecified (n = 1).

Data Collection

The research team developed a semi-structured interview guide based on the research questions above that included 13 questions under the following four themes: program information, identifying success and Indigenous evaluation methods, evaluation experiences, and capacity and barriers. In line with decolonizing research methods, the initial questions in the in-depth interview guide did not mention the Western terms *evaluation*, *reporting*, or *data*, with the express notion that the use of these terms may elicit responses aligned with Western culture rather than based on Indigenous knowledge (Bartlett et al., 2007). The guide was pilot tested with three AI/AN public health practitioners, during which they were asked to provide feedback on the appropriateness, intrusiveness, completeness, understandability, and respondent burden of the questions. The guide was revised based on the feedback. The telephone interviews were conducted by the first author and were digitally recorded.

As a standard for identifying a sufficient qualitative sample size, saturation is defined as "the point at which all questions have been thoroughly explored in detail and no new concepts or themes emerge in subsequent interviews" (Trotter, 2012, p. 399). In this study, saturation was reached at 17 participants; after that point, no additional interviews were conducted (Trotter,

2012). Researchers have identified a sample size of 15 participants as appropriate and sufficient for qualitative research in the field of public health using expert sampling designs (Trotter, 2012), and qualitative decolonizing research with Indigenous peoples has been conducted with 16 participants (Lavallee, 2009); thus, we decided this sample size was sufficient. Participants were given a $20 Amazon gift card in appreciation for their time and insight.

Data Analysis

Thematic analysis was used to analyze the qualitative data. The study included member checking and ongoing reflexivity (described below) throughout the research process to further the bidirectional learning emphasized in decolonizing research (Bartlett et al., 2007). A graduate research assistant transcribed all interviews verbatim, and the first author reviewed the transcripts for quality assurance by comparing them to the recordings for accuracy and revising as needed. Identifiable (e.g., personal and tribal) information was removed from all transcripts and, as a form of member checking, transcripts were sent to the associated interview participants for their voluntary review. Member checking is a method used in decolonizing research that serves to improve the credibility of qualitative research; it involves matching the researcher's data and analysis with the study participant's interpretations (Bartlett et al., 2007; Krefting, 1999). For the 17 interviews, nine participants confirmed that they reviewed their transcripts. Of those, seven approved the transcript as it was and two provided edits. One participant's edits were minor and involved changing a word that was misunderstood on the recording. A second participant perceived the member check as an opportunity to revisit the responses and make changes and requested more involved edits, which the first author felt altered the authenticity of the initial response. After discussions about the purpose of the member check, the first author and participant agreed on a version of the transcript that was mostly true to the original version, but included some edits where the participant completed a thought or comment that was not completed during the interview, and where the participant's first response did not make sense. Transcripts were imported into NVivo qualitative software to facilitate organization and coding.

As a first step of the thematic analysis, the first author created a codebook based on the research questions and themes that emerged during the interviews (Braun & Clarke, 2006; DeCuir-Gunby, Marshall, & McCulloch, 2010). To assess the reliability of the codebook through a group-consensus approach (without quantification), the graduate research assistant and first author:

- coded 2 transcripts (selected based on their medium length)
- met with a faculty researcher experienced in qualitative research to review and discuss the codebook and its application,
- reviewed and discussed any inconsistencies in coding, and
- decided together on appropriate revisions (DeCuir-Gunby et al., 2010).

The codebook was revised, and the first author and graduate research assistant coded two new transcripts, met to discuss discrepancies, and collaboratively finalized the codebook. From there, the first author coded all 17 transcripts, and the graduate research assistant coded a random sample of 30% of the transcripts ($n = 5$) to assess the validity of the coding. Instances of Indigenous knowledge collection were identified through the coding process. To capture participants' use of traditional knowledge to help evaluate their physical activity programs, the first author coded instances when participants identified gathering information through relationships, experiences, interpretations, Native teachings, and community learning. To identify when participants used empirical knowledge to help evaluate their programs, the first author coded instances when participants identified collecting information and/or obtaining knowledge through their observations of the program participants that occurred over time. All coding was agreed upon, and the codebook was used to guide a deductive approach to data analysis while allowing for inductive analysis to take place as codes emerged (Braun & Clarke, 2006).

Through an examination of the coded text, concepts were generated to represent the various dimensions of each theme, recognizing the continuity and variability of each theme within and across the interviews (Daly, 2007; Ryan & Bernard, 2003). Based on these concepts, a narrative was created, representative of the preliminary findings and written as a journal entry from an individual evaluating an AI/AN physical activity program. This narrative was sent to all

participants as the second member check, and 7 confirmed that they read the narrative, approved its content, and reported that it represented their interviews well. The remaining 10 participants did not confirm receipt or provide feedback on the narrative.

Preliminary findings also were summarized for dissemination to an expert panel, as a form of peer examination (Krefting, 1999). Individuals were asked to serve as members of the expert panel based on their professional expertise in AI/AN public health research. Two expert panelists and one academic mentor experienced in qualitative research with minority populations reviewed the preliminary findings and provided feedback, identifying areas for further deconstruction.

RESULTS

Although the participants were recruited based on their employment at AI/AN organizations, they were not speaking on behalf of their organizations; thus, the results reflect their individual perceptions and not necessarily the views of their organizations. Three themes arose: 1) Indigenous approaches to evaluating AI/AN physical activity programs are perceived as narrative and holistic, 2) Indigenous knowledge is used in AI/AN physical activity program decision making, but sometimes is not acknowledged as evaluation, and 3) there is not a universally desired way to evaluate AI/AN physical activity programs.

Theme 1: Indigenous Approaches to Evaluating AI/AN Physical Activity Programs are Perceived as Narrative and Holistic

When participants described how they broadly understood "Indigenous evaluation" or "Indigenous approaches to evaluation," the most common perception was that these approaches are holistic and narrative in nature. Four described Indigenous evaluation as holistic (i.e., focused on the impact of the program on the person or the community as a whole). Language such as "the whole gestalt of the experience" and "how it made them feel" was used to convey their understanding of this approach to evaluation. Participants supported this understanding by saying: "[Indigenous knowledge-based evaluation includes] more qualitative measures like spiritual health, or...patient happiness, satisfaction..." and

[Indigenous knowledge-based evaluation is] being able to tell their story. So it's not a hard physical, "Oh we saw somebody's nutrition knowledge improve." So much as, "We hear there's someone [that told] a story about something that they did or something that improved in their life."

Six individuals identified Indigenous knowledge as being narrative or oral in nature, which, to them, generally referred to storytelling. These participants conceptualized Indigenous evaluation using language like "their voices," "their stories about that work," "subjective," and "telling you how they felt during the time they were there" to explain how they understood the process and components of Indigenous evaluation.

According to participants, approaches to evaluation that included narrative and oral methods were perceived as being culturally appropriate and well received by AI/ANs ($n = 10$), and they included these methods when describing their understanding of "Indigenous evaluation." For example, one participant explained perceptions of Indigenous evaluation and how narrative methodology fit the AI/AN context by saying:

I feel like the diversity in Indian Country is so much that you can't really...like things are so different from community to community that you need that sort of human voice, and those sort of unique stories to kind of get a grasp of the character and the difficulties associated with like a particular community in Indian Country. And that storytelling sort of helps you capture that uniqueness or not capture it but at least get a feel for it. And, in a lot of ways inspire you to work with it rather than to sort of just implement a curriculum or a program that you feel like you know works and for some reason it may not work with that particular community. So storytelling allows you to sort of discern what those differences may be and take them into account when you actually implement a plan.

Additionally, when explaining why narrative and oral methods may be well received, a participant said:

Well, I think that us as Native people, you know, have been so used to filling out forms and applying for this, and applying for that, and that check boxes somewhat feel like that and to them I think it's a huge turn off. And I even think that maybe, you know, not asking people to write, but to document what they say is gonna be a better approach.

To capture how participants experienced Indigenous evaluation in the context of their physical activity programs, the first author coded instances when they identified using a method of evaluation that they felt was Indigenous and/or shared the qualities of Indigenous evaluation. Such approaches included focus groups ($n = 3$), talking circles ($n = 3$), interviews ($n = 3$), storytelling ($n = 1$), digital storytelling ($n = 1$), and a pictorial survey ($n = 1$). Two additional participants noted the use of talking circles in their organizations, but to evaluate behavioral health initiatives, not physical activity programs. Participants described the use of these methods for all levels of evaluation (i.e., formative, process, outcome), identifying how they had used focus groups, interviews, and talking circles to capture information about community needs, how existing programs could be improved, and how programs have impacted people's lives. Digital storytelling was described as a way to disseminate information about programmatic success to community members. One participant described hosting talking circles during group bike rides, saying "Usually on our bike rides we'll talk. When I get a big group, a big group will talk about you know a lot of things. 'Okay, what is it you wanna do?'" Another participant described using storytelling:

Yes we have [used storytelling to look at how well our program is doing]. I got a nice letter from like a lady who lost 150 pounds over, and you know, really praised the trainers and the facility in making her feel comfortable. Actually I've gotten a couple of those real success stories, and I've asked if we can share them, and I've done that.

Theme 2: Indigenous Knowledge is Used in AI/AN Physical Activity Program Decision Making but Sometimes is Not Acknowledged as Evaluation

In line with the IEF, the majority of participants ($n = 14$) discussed the inclusion of Indigenous knowledge (e.g., traditional, empirical) in their evaluation. Participants ($n = 12$) identified instances where knowledge collection occurred naturally (i.e., through the traditional

or common ways of knowing for the community) due to the dynamic of the culture (e.g., close-knit, oral), during which program attendees shared success stories, positive program outcomes, or program feedback. Many participants (n = 10) also noted that, in small, close-knit communities, the knowledge held by program staff members allowed them to evaluate their programs through observation. These naturally occurring approaches, grounded in Indigenous ways of knowing, were used for formative, process, and outcome evaluation. One participant described capturing personal stories, saying:

> Our community is you know, we pretty much know (laughs) one another. So it's a little easier for us to know who's been in our programs and who hasn't. And you know we're so tight together that we know, I shouldn't say 'know', but we see kinda what the behaviors are. So, in that way we're able to kind of give ideas of who we feel would, you know, we could possibly interview. And then just community members that are just, you know, they're always sharing their stories already with us.

However, when asked how they evaluated their program, 5 of the 14 participants who referenced traditional and empirical knowledge collection did not explicitly identify it as a method of evaluation. Rather, it arose informally during the interview as a way they knew how to improve their program. For example, one participant's organization had not done a lot of evaluation for its exercise classes, but the participant noted a way in which staff members collect feedback about the classes—suggesting that this knowledge collection was not vieed as evaluation:

> We haven't really done a whole lot of evaluation for our exercise classes. I know it tends to be like a lot of people that are attending - they like to tell us you know what times might work for them, or this class might be good for them... Um, that's just mainly word of mouth...more just a conversation piece...

Another participant described using traditional and empirical knowledge about the community to make a modification to the program, but did not identify this process as evaluation:

> We've opened [our program] up to have a buddy system, because for example, maybe one of our Indian Health Service beneficiaries lives alone and doesn't have transportation, but her non-Native neighbor is her support person. Well, we've invited those buddies to come in and participate so that we can continue to

support the health choices of our beneficiary participants. [Interviewer: How did you come up with that?] It was just knowing who our people were that we were serving, and trying to reduce the barriers to their access.

Theme 3: There is Not a Universally Desired Way to Evaluate AI/AN Physical Activity Programs

Participants provided a variety of responses when describing their ideal or desired way to evaluate their physical activity programs. They reported interest in collecting success stories ($n = 3$), using apps or other systems to track physical activity levels over time ($n = 2$), collecting baseline data on measures of health status ($n = 2$), conducting audits of electronic health record data ($n = 2$), collecting digital stories ($n = 1$), measuring health indicators ($n = 1$), using a tool like the Fitness Gram to assess physical fitness at one discrete timepoint ($n = 1$), using social media ($n = 1$), conducting a self-assessment survey or a survey to track physical activity over time ($n = 2$), using a " robust check-in process" (to collect data beyond a sign-in sheet) ($n = 1$), and using evidence-based practices that are culturally relevant ($n = 1$).   For example, one participant explained an ideal method of evaluation:

I think we would, if money and resources weren't an issue, we would have a very well-established and engineered infrastructure for basically creating profiles for each one of our participants, that sort of tracked their physical activity throughout the year.

Participants also identified a wide range of outcomes that they would like to measure, including program satisfaction ($n = 3$), program participants' physical activity levels ($n = 3$), BMI ($n = 3$), hemoglobin A1C ($n = 2$), blood pressure ($n = 2$), cholesterol ($n = 2$), program attendance ($n = 2$), feedback on the program in general ($n = 2$), program reach ($n = 1$), program retention ($n = 1$), blood sugar ($n = 1$), connectivity with other AI/ANs ($n = 1$), participants' perceived impact of the program ($n = 1$), intermediate steps of behavior change related to physical activity and health outcomes ($n = 1$), engagement in traditional AI/AN physical activity ($n = 1$), exercise capacity ($n = 1$), physical fitness levels ($n = 1$), and wellness ($n = 1$). These responses include both Western science-based outcomes (e.g., BMI, exercise capacity) and Indigenous knowledge-based outcomes (e.g., connectivity with AI/ANs, wellness). One participant reported wishing to collect the following information for program evaluation:

When we established the leagues themselves, it was more to provide two different things—one, physical activity for the youth, that was number one. And number two, provide that connectivity with other Native youth in the area. What we found is that the Native youth especially like in some of our more remote reservations— the only thing that they know is, you know, experience and exposure on those reservations. So when we actually do games and activities, those kids have to come to other reservations and get exposure to what reservation life is like for the ones [in] that the location they're going to. So that's kind of what we would hope to do. Now, have we done any type of an assessment and know whether or not we're actually, you know, connecting with those two points—the answer unfortunately is no at this point.

DISCUSSION

This study's findings contribute to the recent focus on enhancing Indigenous and culturally responsive evaluation (Bledsoe & Donaldson, 2015; LaFrance & Nichols, 2010; Tribal Evaluation Workgroup, 2013) but investigate evaluation in the context of physical activity programming, which had not been explored previously. Findings highlight how evaluation interests include Indigenous methods, but specifics vary across AI/AN organizations implementing physical activity programs. Results also stress the importance of attending to these unique local interests when establishing evaluation plans and requirements.

This study found that AI/AN physical activity program staff members felt that narrative and oral evaluation methods were culturally appropriate for evaluating physical activity programs, given the ways in which Indigenous peoples know and understand the world, and considering the negative history of research and evaluation among AI/ANs. These findings support the Roadmap for Collaborative and Effective Evaluation in Tribal Communities, which stresses the importance of valuing oral tradition and respecting historical context when evaluating programs in Indian Country (Tribal Evaluation Workgroup, 2013). These findings also align with the IEF (LaFrance & Nichols, 2010), which acknowledges that the traditional AI/AN core value of community may lead program staff members to look beyond individual achievement to the more holistic outcomes of community health and well-being. As the

Roadmap was created in the context of tribal child welfare programs, and the IEF was developed with a focus on educational evaluation (LaFrance & Nichols, 2010; Tribal Evaluation Workgroup, 2013), this study extends to AI/AN evaluation of physical activity programs.

In congruence with the cultural traditions of AI/ANs, participants used a variety of narrative and oral methods to evaluate their physical activity programs, including talking circles, interviews, and digital storytelling. While other studies have noted the lack of scientifically rigorous evaluation for AI/AN physical activity programs (Teufel-Shone et al., 2009), this study is the first to identify the use of culturally appropriate and valuable evaluation methods to capture the impact of physical activity programs.

Findings from this study suggest that evaluation training provided to staff members at AI/AN activity programs could be improved by including Indigenous ways of knowing as valuable methods of data collection. Although knowledge gathered through Indigenous ways of knowing informs decision making and the perceptions of effectiveness of AI/AN physical activity programs, participants shared that these ways of knowing often are not recognized as evaluation. This finding may be due to the participants' training in evaluation, which likely was grounded in Western science and did not consider these methods to be scientifically rigorous. While it was not surprising that narrative and oral ways of knowing were being used, this study highlights the need to make evaluators and funding agencies aware of their use and importance in evaluation and reporting.

Because valuable program data can be collected through narrative and oral methods, it is important that these ways of knowing are understood and valued by funding agencies, as they may reveal important information that is not detectable through Western science-based approaches. For example, Cochran et al. (2008) describe an instance when Inuit whalers identified the presence of whales by "listening for the sound of their breathing," which was distinctly different from the counting method used by the International Whaling Commission (p. 24). The Inuits' method was criticized as inaccurate because their estimates did not match the Commission's; however, they were later "verified by successive aerial surveys" (Cochran et al., 2008, p. 24). In addition, this study's identification of traditional and empirical data gathering, which may or may not be systematic, highlights the need for defining when evaluation occurs. The Roadmap identifies the *evaluative culture* of AI/ANs, noting that these populations have always used traditional ways of knowing to determine what is and is not working, and share

these decisions using oral traditions (Tribal Evaluation Workgroup, 2013). The Roadmap recommends that evaluators acknowledge these ways of knowing and include them in rigorous evaluation designs, suggesting that they would then be considered evaluation (Tribal Evaluation Workgroup, 2013). Therefore, we suggest that it would be beneficial for funding agencies to increase their understanding of Indigenous ways of knowing and to acknowledge, include, encourage, and accept the systematic application of these processes in evaluation designs and reporting for physical activity programs.

Finally, given that funding agencies often require standardized measurement and evaluation across grantees, we sought to identify a best practice for evaluating AI/AN physical activity programs that was culturally appropriate and also could be useful for funding agencies. However, the findings indicate that there is no universally desired approach. Participants identified more than 10 different methods for collecting evaluation data (e.g., electronic tracking, surveys, digital storytelling), and nearly 20 different physical activity-related outcomes. These findings speak to the diversity of AI/AN communities, and consequently their programs, programmatic goals, and ways of knowing. The IEF supports this finding, recognizing that AI/ANs are "people of a place," and, as such, what is appropriate in one community may not be easily translated to another (LaFrance & Nichols, 2010). It is recommended that evaluators and funding agencies seek to understand the cultural differences among AI/AN communities, recognize the contextual differences of their physical activity programs and evaluation capacity, and attend to these differences in evaluation designs and requirements (Chouinard & Cousins, 2007; LaFrance & Nichols, 2010).

To further improve the practice of AI/AN physical activity program evaluation and build upon the findings presented, future research is needed to examine use in Indian Country of the evaluation methods and physical activity-related outcomes of interest to study participants, and to validate these measures and methods with AI/AN populations when appropriate. In addition, future research should explore whether systematically collecting traditional and empirical knowledge to evaluate physical activity programs would weaken the cultural appropriateness of the evaluation, through the perspectives of the AI/AN community members and evaluators.

Limitations

This study is unique in its exploration of culturally responsive evaluation of AI/AN physical activity programs, and contributes to the growing dialogue about the inclusion of cultural context in evaluation practice (Bledsoe & Donaldson, 2015). However, it is not without limitations. The in-depth interviews were conducted by telephone, which may have impacted the participants' feelings of comfort and willingness to share. To address this limitation, the first author recruited participants through her professional network with the hope that, if the contact "approved" of her and the study, the prospective participant would feel more comfortable participating. Face-to-face interviews in future research could address this limitation.

CONCLUSION

Both building from and contributing to the research on program evaluation in AI/AN communities, this study identifies perceptions of Indigenous evaluation, use of Indigenous evaluation for physical activity programs, and the approaches to evaluating physical activity programs desired by program staff members. Aligned with the recommendations of the Roadmap for Collaborative and Effective Evaluation in Tribal Communities, and based on the findings from this study, we recommend bidirectional learning related to physical activity program evaluation, whereby AI/AN program staff members receive training on systematic approaches to evaluation that are culturally appropriate and validated among AI/ANs, and outside entities (e.g., universities, funding agencies) learn about Indigenous ways of knowing specific to individual communities (Tribal Evaluation Workgroup, 2013). The findings from this study suggest that universal or standardized evaluation across many unique AI/AN physical activity programs may not be appropriate, as desired ways of capturing programmatic success varied among participants. Future research should seek to expand these findings across the broader population of AI/AN communities conducting physical activity programs, and should continue to examine culturally responsive methods for evaluating AI/AN physical activity programs that are rigorous and systematic, and incorporate Indigenous ways of knowing.

REFERENCES

Bartlett, J., Iwasaki, Y., Gottlieb, B., Hall, D., & Mannell, R. (2007). Framework for Aboriginal-guided decolonizing research involving Metis and First Nations persons with diabetes. *Social Science & Medicine, 65*, 2371-2382. http://dx.doi.org/10.1016/j.socscimed.2007.06.011

Barnhardt, R., & Kawagley, A.O. (2005). Indigenous knowledge systems and Alaska Native ways of knowing. *Anthropology & Education Quarterly, 36*(1), 8-23. http://dx.doi.org/10.1525/aeq.2005.36.1.008

Bledsoe, K. L., & Donaldson, S. I. (2015). Culturally responsive theory-driven evaluation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 3-27). Charlotte, NC: Information Age Publishing, Inc.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*, 77-101. http://dx.doi.org/10.1191/1478088706qp063oa

Cavino, H. M. (2013). Across the colonial divide: Conversations about evaluation in Indigenous contexts. *American Journal of Evaluation, 34*(3), 339-355. http://dx.doi.org/10.1177/1098214013489338

Chouinard, J., & Cousins, J. (2007). Culturally competent evaluation for Aboriginal communities: A review of the empirical literature. *Journal of MultiDisciplinary Evaluation, 4*(8), 40-57. Retrieved from http://www.uaf.edu/ces/info/reporting/programevals/aboriginal-eval.pdf

Cobb, N., Espey, D., & King, J. (2014). Health behaviors and risk factors among American Indians and Alaska Natives, 2000-2010. *American Journal of Public Health, 104*(Suppl 3), S481-S489. http://dx.doi.org/10.2105/AJPH.2014.301879

Coble, J. D., & Rhodes, R. E. (2006). Physical activity and Native Americans: A review. *American Journal of Preventive Medicine, 31*(1), 36-46. http://dx.doi.org/10.1016/j.amepre.2006.03.004

Cochran, P., Marshall, C., Garcia-Downing, C., Kendall, E., Cook, D., McCubbin, L., & Grover, R. (2008). Indigenous ways of knowing: Implications for participatory research and community. *American Journal of Public Health, 98*(1), 22-27. http://dx.doi.org/10.2105/AJPH.2006.093641

Daly, K. (2007). *Qualitative methods for family studies & human development*. Thousand Oaks, CA: Sage Publications, Inc.

DeCuir-Gunby, J. T., Marshall, P. L., & McCulloch, A. W. (2010). Developing and using a codebook for the analysis of interview data: An example from a professional development research project. *Field Methods, 23*(2), 136-155. http://dx.doi.org/10.1177/1525822X 10388468

Denzin, N. K., & Lincoln, Y. S. (2008). Introduction: Critical methodologies and Indigenous inquiry. In N. K. Denzin, Y. S. Lincoln, & L. T. Smith (Eds.), *Handbook of Critical and Indigenous Methodologies* (p. 1-20). Thousand Oaks, CA: Sage Publications, Inc.

Fleischhacker, S., Roberts, E., Camplain, R., Evenson, K., & Gittelsohn, J. (2015). Promoting physical activity among Native American youth: A systematic review of the methodology and current evidence of physical activity interventions and community-wide initiatives. *Journal of Racial and Ethnic Health Disparities.* Advance online publication. http://dx.doi.org/10.1007/s40615-015-0180-1

Foulds, H. J., Warburton, D. E., & Bredin, S. S. (2013). A systematic review of physical activity levels in Native American populations in Canada and the United States in the last 50 years. *Obesity Reviews, 14*(7), 593-603. http://dx.doi.org/10.1111/obr.12032

Grover, J. G. (2010). Challenges in applying Indigenous evaluation practices in mainstream grant programs to Indigenous communities. *The Canadian Journal of Program Evaluation, 23*(2), 33-50. Retrieved from http://www.cjpe.ca/secure/23-2-033.pdf

Hodge, F. S. (2012). No meaningful apology for American Indian unethical research abuses. *Ethics & Behavior, 22*(6), 431-444. http://dx.doi.org/10.1080/10508422.2012.730788

Kawakami, A., Aton, K., Cram, F., Lai, M., & Porima, L. (2007). Improving the practice of evaluation through Indigenous values and methods: Decolonizing evaluation practice-returning the gaze from Hawai'i and Aotearoa. *Halili: Multidisciplinary Research on Hawaiian Well-Being, 4*(1), 319-348. Retrieved from http://www.ksbe.edu/spi/hulili/

Krefting, L. (1999). Rigor in qualitative research: The assessment of trustworthiness. In A. Miliniki (Ed.), *Cases in qualitative research: Research reports for discussion and evaluation* (pp. 173-181). Los Angeles: Puscale Publications.

Kriska, A., Saremi, A., Hanson, R., Bennett, P., Kobes, S., Williams, D., & Knowler, W. (2003). Physical activity, obesity, and the incidence the type 2 diabetes in a high-risk population. *American Journal of Epidemiology, 158*(7), 669-675. http://dx.doi.org/10.1093/aje/kwg191

LaFrance, J. (2004). Culturally competent evaluation in Indian Country. *New Directions for Evaluation, 102*, 39-50. http://dx.doi.org/10.1002/ev.114

LaFrance, J., Kirkhart, K. E., & Nichols, R. (2015). Cultural views of validity: A conversation. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 49-72). Charlotte, NC: Information Age Publishing, Inc.

LaFrance, J., & Nichols, R. (2009). *Indigenous Evaluation Framework: Telling our story in our place and time*. Alexandria, VA: American Indian Higher Education Consortium. Retrieved from http://indigeval.aihec.org/Pages/Documents.aspx

LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *The Canadian Journal of Program Evaluation, 23*(2), 13-31. Retrieved from http://www.evaluationcanada.ca/secure/23-2-013.pdf

Lavallee, L. (2009). Practical application of an Indigenous research framework and two qualitative Indigenous research methods: Sharing circles and Anishnaabe symbol-based reflection. *International Journal of Qualitative Methods, 8*(1), 21-40. Retrieved from https://ejournals.library.ualberta.ca/index.php/IJQM/

Poltavski, D., Holm, J., Vogeltanz-Holm, N., & McDonald, L. (2010). Assessing health-related quality of life in Northern Plains American Indians: Prominence of physical activity as a health behavior. *American Indian and Alaska Native Mental Health Research, 17*(1), 25-48. http://dx.doi.org/10.5820/aian.1701.2010.25

Ryan, G., & Bernard, H. (2003). Techniques to identify themes. *Field Methods, 15*, 85-109. http://dx.doi.org/10.1177/1525822X02239569

Smith, L. T. (2012). *Decolonizing methodologies: Research and Indigenous peoples* (2nd ed.). New York: Zed Books Ltd.

Teufel-Shone, N. I., Fitzgerald, C., Teufel-Shone, L., & Gamber, M. (2009). Systematic review of physical activity interventions implemented with American Indian and Alaska Native populations in the United States and Canada. *American Journal of Health Promotion, 23*(6), S8-S32. http://dx.doi.org/10.4278/ajhp.07053151

Tribal Evaluation Workgroup. (2013). *A roadmap for collaborative and effective evaluation in tribal communities*. Washington, DC: Children's Bureau, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from https://www.acf.hhs.gov/sites/default/files/cb/tribal_roadmap.pdf

Trotter, R. T., 2nd. (2012). Qualitative research sample design and sample size: Resolving and unresolved issues and inferential imperatives. *Preventive Medicine, 55*(5), 398-400. http://dx.doi.org/10.1016/j.ypmed.2012.07.003

AUTHOR INFORMATION

Dr. Roberts is a Research Associate with James Bell Associates, Inc. She is the corresponding author and can be reached at 3033 Wilson Blvd, Suite 650, Arlington, VA, 22201, roberts@jbassoc.com, or (410) 733-6381.

Dr. Butler is an Assistant Professor in the Department of Behavioral and Community Health at the University of Maryland School of Public Health.

Dr. Green is an Associate Professor in the Department of Behavioral and Community Health at the University of Maryland School of Public Health.