

Practical Assessment, Research & Evaluation

A peer-reviewed electronic journal.

Copyright is retained by the first or sole author, who grants right of first publication to *Practical Assessment, Research & Evaluation*. Permission is granted to distribute this article for nonprofit, educational purposes if it is copied in its entirety and the journal is credited. PARE has the right to authorize third party reproduction of this article in print, electronic and database forms.

Volume 25 Number 9, December 2020

ISSN 1531-7714

Comparability of Survey Measures in Hard to Reach Populations: Methods and Recommendations

Allyson Kelley, *Allyson Kelley & Associates PLLC*
Thomas McCoy, *University of North Carolina Greensboro*
Aryn Fisher, *Allyson Kelley & Associates PLLC*
Morgan Witzel, *Rocky Mountain Tribal Leaders Council*
Bethany Fatupaito, *Allyson Kelley & Associates PLLC*
Desiree Restad, *Northern Cheyenne Tribal Board of Health*

The goal of this paper is to discuss practical recommendations for collecting data with hard to reach populations and data comparability. We also discuss the importance of piloting and community involvement in the process using an example from the Tribal Prevention Initiative (TiPI), a culturally-based substance abuse prevention program for American Indian youth from six tribes in Montana and Wyoming. We piloted a survey based on standardized Youth Risk Factor Behavioral Surveillance Survey questions (YRBSS) in tribal communities to document substance use and norms. The revised TiPI survey resulted in a 16-question survey (N=711). We compared TiPI data with the YRBSS data from 2015 to 2019. Descriptive statistics (frequency (n) and percent (%) or mean (M) and standard deviation (SD)) were used to analyze data. Results indicate that substance use is generally lower among TiPI youth than YRBS youth in reservation and urban locations with the exception of middle school marijuana use and urban binge drinking. To improve public health and document progress toward healthy future generations, communities must be aware of the unique challenges of using national surveys like the YRBSS as comparison data, and the strengths of primary data collection driven by program needs.

Introduction

Professionals often use administrative or secondary data from national and state surveys to document problems, identify needs, and develop policies. This practice is common in public health where the researchers first used administrative data in 1973 to understand patterns of hospitalization and variations in treatment based on a patient's place of residence (Wennberg & Gittelsohn, 1973). Administrative data are collected by governments and health care providers and often used to track or record information. Primary data are collected by researchers or programs often in-person or for a specific purpose while secondary data are collected by someone else and

used for another purpose (Allen, 2017). Shaghghi and colleagues (2011) define hard-to-reach populations as those that are difficult to reach/recruit into research or public health programs due to their geographic location or their social and economic situation. Marpsat and Razafindratsima (2010) further elucidate that a hard to reach population has population members that are hard to identify due to no sampling frame or that the persons do not want to disclose they are members of the population of interest. Data collection with hard to reach populations may result in poor data quality due to missing data, under-reporting, inconsistencies, data errors, invalid data, non-standardization of vocabulary, and inappropriate fields (Chen, Hailey, Wang, & Yu, 2014).

In this paper we will explore the challenges of hard-to-reach populations and how to evaluate the comparability of existing data.

Professional Standards

While the issue of data comparability using primary or administrative data sources is not new, it remains important for researchers and evaluators. Two professional standards support this line of inquiry: program evaluation standards and Standards for Education and Psychological Testing (AERA) (2014). Yarbrough and colleagues developed program evaluation standards and propriety standards that ensure data collection is responsive and inclusive to stakeholders and communities (2010). In addition, these standards ensure accuracy in data collection and evaluation. AERA's standards support the elements of standardization and instrument design.

Efforts to create uniformity in public health data have been attempted since 1969 with Johns Hopkins University, the National Center of Health Services Research and Development, and the National Center for Health Statistics, with slow progress (Trevino, 1988). Issues regarding data comparability have been noted, including variability in the wording of questions and utilizing numerous prompts to answer a question, creating difficulty in comparing data sets. Key challenges to progress include data collection methodology; what data should be collected, how it should be collected, and from whom. Further, problems related to race and ethnicity, inadequate sample sizes, lack of representation of hard-to-reach groups, and missing data are common challenges in the public health data milieu (Bilheimer & Sisk, 2008; Ericksen, 1997).

Practical Applications of Comparing Health-Related Data

Organizations like the Centers for Disease Control and Prevention (CDC) are leading nation-wide efforts to monitor health-related behaviors that contribute to death and disability in youth and adults (Virnig & McBean, 2001) using primary data collection methods. Data from CDC efforts are used to document baseline prevalence of various health measures, and then used to measure changes over time, inform research, direct policy or funding opportunities that address identified health needs. Benefits to using CDC data and standardized questions in surveys are numerous; they

have been tested for validation and consistency, they are approved by funding agencies as substitute measures, and they are relatively easy to compare across years, age groups, and geographies. Challenges with using CDC data include limited sampling from hard-to-reach populations, elective participation by schools and states in some programs, and removing questions from surveys making comparisons over time impossible. National surveys developed by the CDC such as the Youth Risk Behavior Surveillance System (YRBSS) are used to document needs, direct funding, and inform policy (Foti, Balaji, Shanklin, 2011)—yet these surveys often lack representativeness, cultural responsiveness, and have elective participation (e.g., not all schools participate, and participating schools can elect to leave out some measures) (Ericksen, 1997). Since 1991, the CDC has conducted the YRBSS every other year with a representative sample of students from national, state, and local schools to understand more about the factors that contribute to disease and disability among youth and young adults (CDC, ND).

Comparing data from existing data sets like the YRBSS with program data can be a powerful way to document disparities among populations and further understanding about differences in risk behaviors among populations. Yet at the same time, caution must be taken when comparing samples of disparate populations such as American Indian and non-American Indian populations (Pirkis, Irwin, Brindis, Patton & Sawyer, 2003). In this practical application we explore the methodological considerations for comparability of YRBSS and program specific data for substance use in American Indian youth.

Documenting rates of substance abuse among youth is an important first step in addressing factors that place them at risk and identifying behaviors that promote healthy lifestyles during adolescence (Kelley, Restad, & Killsback, 2018). However, finding reliable estimates for small and unique populations like American Indian youth, a hard-to-reach population, is often difficult (Ericksen, 1997).

The goal of this paper is to discuss methods and practical recommendations for comparing data collected from a hard-to-reach populations. We emphasize two areas related to data comparability and hard-to-reach populations: 1) practices for collecting data with small populations and the importance of piloting and community involvement in the process,

and 2) how to evaluate the comparability of existing data (e.g., from administrative or secondary data). We illustrate issues and recommendations from our experiences with American Indian youth involved in the Tribal Prevention Initiative and data from the Youth Risk Factor Behavioral Surveillance Survey (YRBSS).

Methods

Data Practices for Racial and Ethnic Minority Groups

Population based surveys such as the US Census or the National Survey on Drug Abuse are one of the most common ways to collect public health information and are useful in collecting data among racial and ethnic minority groups like American Indians and Alaska Natives. However, inadequate sample sizes, missing data, lack of standardized survey questions, and limited collection of racial and ethnic information threaten the accuracy of data collected with these surveys. A key challenge is that self-report surveys, like the US Census or National Survey on Drug Abuse list race/ethnicity choices that respondents do not identify with. In some cases, they may identify with more than one race or ethnicity category and do not accurately report that information.

Medical records and administrative data systems are common ways to collect substance abuse data on American Indian and Alaska Native groups. Utilizing these data sources have proven to be challenging because collected information is based on self-report data or observation by funeral directors or clinicians—if an individual is part of more than one race or ethnicity group, this may result in racial misclassification and biased rates among racial and ethnic groups (Reijneveld & Stronks, 1999). Further, most medical records and administrative data do not include race or ethnicity. For example, a review of results from the National Hospital Discharge Survey show that 33% of discharges did not include any information on race and 50% did not include information on ethnicity (Cherry, Woodwell & Rechtsteiner, 2007).

Methodological considerations for standardizing surveys with hard-to-reach populations are similar across disciplines. Standardizing questions and ensuring that racial and ethnic groups are adequately represented in national surveys will help improve data

comparability and representation of racial and ethnic minority groups. First, by asking the same questions across time and groups (including the same question responses), this would allow for direct comparison by year and geography. Second, educating professionals (teachers, educators, clinicians, paraprofessionals, and students) about the importance of documenting racial and ethnic classification will likely increase the quality and availability of data available. These data are critical for understanding differences in health status, educational attainment, risk and protective factors, and other contextual factors that should be considered when developing programs, policies, funding, and outreach that targets specific groups based on identified needs. Third, although the wording of standardized survey questions may be changed to address the literacy level and cultural context of a population, the response options and the frequency should not be changed. In sum, standardization of surveys is important for comparing survey responses across groups and the lack of standardization is a major challenge when comparing program-level data with state or national data.

The topic of racial and ethnic underrepresentation in the US is not new. Efforts by the US Federal Government to address the lack of public health data on racial and ethnic disparities have been ongoing. In 1997, the Office of Management and Budget (OMB) developed categories for reporting race and ethnicity—American Indian or Alaska Native, Asia, Native Hawaiian or other Pacific Islander, black, or white, with separate Hispanic ethnicity classifications. Although this marked progress, the OMB effort failed to accurately represent racial and ethnic minority groups. In a 2006 National Health Care Disparities Report of Healthy People 2010, authors wrote that statistically reliable estimates were not available for health disparities data, and more than 75% of the quality and access measures for American Indian and Alaska Native populations were unreliable (AHRQ, 2011). One of the primary issues cited was that the race/ethnic minority group was not accurately represented in the data. When hard to reach populations like American Indian and Alaska Natives are not accurately represented, this presents a major challenge with comparing and using data. In this section, we demonstrate the application of data comparability and data use. We used data from the Tribal Prevention Initiative (TPI), a 5-year substance

abuse prevention program funded by the Substance Abuse and Mental Health Services Administration (SAMHSA).

Data Sources

TiPI data were collected between January 2016 and August 2019 in six tribal locations in two states. Comparison data were extracted from the 2019 CDC YRBSS, matching data from the Montana YRBSS. One tribe was located in Wyoming but the YRBSS was not conducted in Wyoming beginning in 2016, which resulted in no data available for comparison. Because all but one tribe were located in Montana, we used only Montana YRBSS data.

The program included 711 youth between the ages of 12-20 involved in culturally based prevention activities supported by the SAMHSA prevention program. The prevention program approach and impacts have been published previously (Kelley, Restad, & KILLSBACK, 2018; Kelley, Witzel, & Fatupaito, 2019). We used 16 questions designed to measure substance use, social support, self-esteem, and family/peer norms around substance abuse. Questions came from the CDC YRBSS survey and were deemed reliable and valid for the population by SAMHSA (SAMHSA, 2018). To check this, surveys were piloted in communities with target populations before data collection. Piloting the survey resulted in limited changes. Some youth felt that the wording of the questions was too difficult, others felt it was too long, and some felt the questions were not applicable to them. One person said, “Questions would be more applicable to someone at the age of 20 years old as opposed to someone who is 12 years old.” Another person asked, “What does it mean if someone drinks 5 or more times a day?” This individual felt awkward and uncomfortable. Because these questions were required by the funding agency, we did not change them. We did, however, create a survey protocol for individuals administering the survey that reinforce the concept that it was voluntary, they could skip questions they did not want to answer, and we also defined binge drinking (5 or more drinks in a row). Descriptive statistics (frequency (*n*) and percent (%) or mean (*M*) and standard deviation (*SD*)) were used to analyze data. Weighted results are given for YRBSS data that take into account its survey design features.

Results

A total of 711 youth participated in the TiPI program. Table 1 provides characteristics of the program sample. Here, tribe size varied between 2% (Tribe 3) to 29% (Tribe 1) of the total sample, with 93% of youth on or near a Reservation and 6% urban. Participants were 14.4 years old on average (*SD*=1.7), with 56% between 11-14 years old. Slightly more than half were female (54%).

Table 1. Sample characteristics of the TiPI Program participants (*n*=711)

Characteristic	<i>n</i> (%) or <i>M</i> ± <i>SD</i>
Tribe	
Tribe 1	203 (28.6)
Tribe 2	161 (22.6)
Tribe 3	14 (2.0)
Tribe 4	44 (6.2)
Tribe 5	92 (12.9)
Tribe 6	144 (20.3)
Other	44 (6.2)
missing	9 (1.3)
Location type	
Urban	44 (6.2)
Reservation	658 (92.5)
missing	9 (1.3)
Age (years)	14.4 ± 1.7
11-14 years old	397 (55.8)
15-21 years old	300 (42.2)
missing	14 (2.0)
Gender	
Female	380 (53.5)
Male	327 (46.0)
missing	4 (0.6)

Table 2 provides prevalence of substance use and norms from the TiPI program overall and for the largest participating tribe, Tribe 1, as well as weighted prevalence from Montana high school YRBSS data. Binge drinking (7.4% vs. 16.8% YRBSS) and substance use were lower for TiPI program participants compared to YRBSS including for: marijuana (20.5% vs. 21.1% YRBSS), prescription drug misuse (3.3% vs.

12.8% YRBSS lifetime use), and inhalants/sniffing glue (3.2% vs. 8.0% YRBSS lifetime use). However, only lifetime substance use was available from YRBSS measures (relative to past 30 day within TiPI) except for binge drinking and marijuana use. Unfortunately, responses from norm questions were not available from the YRBSS relative to those asked in the TiPI program.

Table 2. Substance use and norms from TiPI program and YRBSS 2019 data

Measure	TiPI Program (n=711)	TiPI Program Tribe 1 (n=203)	YRBSS All Montana HS weighted %
<i>Substance use</i>			
Any Illegal drug use, past 30 days (p30)	104/693, 15.0%	36/197, 18.3%	n/a
Any days 5+ alcoholic drinks, p30	52/700, 7.4%	16/201, 8.0%	16.8%
Any marijuana use, p30	143/699, 20.5%	54/196, 27.6%	21.1%
Any synthetic marijuana use, p30	22/697, 3.2%	11/195, 5.6%	6.5% (life)
Any prescription drug misuse, p30	23/696, 3.3%	7/194, 3.6%	12.8% (life)
Any meth use, p30	6/696, 0.9%	2/194, 1.0%	2.4% (life)
Any other illegal drugs*	13/697, 1.9%	4/195, 2.0%	4.3% (life)
Any inhalants/sniffing use, p30	22/696, 3.2%	2/195, 1.0%	8.0% (life)
<i>Norms</i>			
Risk of harm drink 5+ drinks per week			n/a
No Risk	62 (8.7)	22 (10.8)	
Slight Risk	74 (10.4)	19 (9.4)	
Moderate Risk	199 (28.0)	39 (19.2)	
Great Risk	239 (33.6)	66 (32.5)	
Don't know	110 (15.5)	53 (26.1)	
missing	27 (3.8)	4 (2.0)	
Family feel drink 1+ drinks every day			n/a
Neither Approve Nor Disapprove	42 (5.9)	15 (7.4)	
Somewhat Disapprove	45 (6.3)	7 (3.4)	
Strongly Disapprove	512 (72.0)	141 (69.5)	
Don't know	87 (12.2)	35 (17.2)	
missing	25 (3.5)	5 (2.5)	
Close friends feel drink 1+ drinks daily			n/a
Neither Approve Nor Disapprove	110 (15.5)	39 (19.2)	
Somewhat Disapprove	129 (18.1)	34 (16.7)	
Strongly Disapprove	236 (33.2)	60 (29.6)	
Don't know	188 (26.4)	66 (32.5)	
missing	48 (6.8)	4 (2.0)	
Talked w/ family about sub. use, past yr.			n/a
Yes	422 (59.4)	133 (55.7)	
No	267 (37.6)	84 (41.4)	
missing	22 (3.1)	6 (3.0)	

Note. n for TiPI substance use data denominators < 711 because of missing data; n/a = norms item was not asked within YRBSS; * = Other illegal drugs (LSD, ecstasy)

Table 3 provides results stratified by middle-school vs. high-school aged youth along with available Montana YRBSS data for Grades 7-8. Here, middle-school binge drinking prevalence was markedly lower for TiPI program youth relative to YRBSS (4.4% vs. 7.1% YRBSS) but higher for past 30 day marijuana use

(12.3% vs. 8.5% for YRBSS). Prevalence of other substance use were lower for TiPI youth, but these were measured in the past 30 days relative to for lifetime ever use for YRBSS; underlining the important of comparable time-frame choices for data congruence in hard-to-reach populations.

Table 3. Substance use and norms from TiPI program and YRBSS 2019 data by school type

Measure	TiPI Program High School (n=300; 42.2%)	TiPI Program Middle School (n=397; 55.8%)	YRBSS Montana Grade 7-8
<i>Substance use</i>			
Any Illegal drug use, past 30 days (p30)	67/296, 22.6%	36/383, 9.4%	n/a
Any days 5+ alcoholic drinks, p30	34/296, 11.5%	17/390, 4.4%	541/7,610, 7.1%
Any marijuana use, p30	94/295, 31.9%	48/390, 12.3%	643/7,580, 8.5%
Any synthetic marijuana use, p30	12/296, 4.1%	9/387, 2.3%	297/7,546 3.9% (life)
Any prescription drug misuse, p30	16/296, 5.4%	6/386, 1.6%	771/7,600 10.1%
Any meth use, p30	4/296, 1.4%	1/386, 0.3%	110/7,602, 1.4% (life)
Any other illegal drugs*	8/296, 2.7%	4/387, 1.0%	93/7,453, 1.2% (life)
Any inhalants/sniffing use, p30	10/295, 3.4%	11/387, 2.8%	829/7,571 10.9%
<i>Norms</i>			
Risk of harm drink 5+ drinks per week			
No Risk	23 (7.7)	38 (9.6)	n/a
Slight Risk	27 (9.0)	45 (11.3)	
Moderate Risk	94 (31.3)	103 (25.9)	
Great Risk	115 (38.3)	121 (30.5)	
Don't know	36 (12.0)	70 (17.6)	
missing	5 (1.7)	20 (5.0)	
Family feel drink 1+ drinks every day			
Neither Approve Nor Disapprove	20 (6.7)	22 (5.5)	n/a
Somewhat Disapprove	23 (7.7)	22 (5.5)	
Strongly Disapprove	220 (73.3)	285 (71.8)	
Don't know	30 (10.0)	53 (13.4)	
missing	7 (2.3)	15 (3.8)	
Close friends feel drink 1+ drinks daily			
Neither Approve Nor Disapprove	56 (18.7)	53 (13.4)	n/a
Somewhat Disapprove	58 (19.3)	68 (17.1)	
Strongly Disapprove	88 (29.3)	147 (37.0)	
Don't know	89 (29.7)	92 (23.2)	
missing	9 (3.0)	37 (9.3)	
Talked w/ family about sub. use, past yr.			
Yes	182 (60.7)	235 (59.2)	n/a
No	110 (36.7)	150 (37.8)	
missing	8 (2.7)	12 (3.0)	

Note. n/a = norms item was not asked within YRBSS; * = Other illegal drugs (LSD, ecstasy)

Finally, Table 4 presents data by youth on or near a Reservation (92.5% of TiPI youth) relative to an urban setting (6.2%). Here, binge drinking (6.3% vs. 15.2% YRBSS) and past 30 day marijuana use (20.1% vs. 39.8% YRBSS) were substantially lower for TiPI

Reservation youth relative to the Montana YRBSS findings, while higher for binge drinking (23.8% vs. 17.8% YRBSS urban) and comparable for marijuana use (22.7% vs. 23.7% YRBSS urban) for youth in an urban setting.

Table 4. Substance use and norms from TiPI program and YRBSS 2019 data by Urban vs. Reservation

Measure	TiPI Program Urban (n=44; 6.2%)	TiPI Program Reservation (n=658; 92.5%)	YRBSS Montana HS Urban	YRBSS Montana HS Reservation
<i>Substance use</i>				
Any Illegal drug use, past 30 days	9/42, 21.4%	93/642, 14.5%	n/a	n/a
Any days 5+ alcoholic drinks, p30	10/42, 23.8%	41/650, 6.3%	110/618, 17.8%	150/990, 15.2%
Any marijuana use, p30	10/44, 22.7%	130/646, 20.1%	146/615, 23.7%	390/979, 39.8%
Any synthetic marijuana use, p30	3/44, 6.8%	19/644, 3.0%	58/619, 9.4% (life)	144/982, 14.7% (life)
Any prescription drug misuse, p30	2/44, 4.5%	20/643, 3.1%	96/622, 15.4% (life)	181/985, 18.4% (life)
Any meth use, p30	0/44, 0.0%	6/643, 0.9%	32/615, 5.2% (life)	40/968, 4.1% (life)
Any other illegal drugs*	1/44, 2.3%	12/644, 1.9%	38/619, 6.1% (life)	41/983, 4.2% (life; e)
Any inhalants/sniffing use, p30	3/44, 6.8%	19/643, 3.0%	89/620, 14.4% (life)	105/987, 10.6% (life)
<i>Norms</i>				
Risk of harm drink 5+ drinks per week				
No Risk	0	61 (9.3)	n/a	n/a
Slight Risk	3 (6.8)	69 (10.5)		
Moderate Risk	9 (20.5)	188 (28.6)		
Great Risk	29 (65.9)	208 (31.6)		
Don't know	2 (4.5)	106 (16.1)		
missing	1 (2.3)	26 (4.0)		
Family feel drink 1+ drinks every day				
Neither Approve Nor Disapprove	2 (4.5)	40 (6.1)	n/a	n/a
Somewhat Disapprove	4 (9.1)	40 (6.1)		
Strongly Disapprove	34 (77.3)	473 (71.9)		
Don't know	1 (2.3)	83 (12.6)		
missing	3 (6.8)	22 (3.3)		
Close friends feel drink 1+ drinks daily				
Neither Approve Nor Disapprove	2 (4.5)	106 (16.1)	n/a	n/a
Somewhat Disapprove	5 (11.4)	121 (18.4)		
Strongly Disapprove	24 (54.5)	211 (32.1)		
Don't know	10 (22.7)	175 (26.6)		
missing	3 (6.8)	45 (6.8)		
Talked w/ family about sub. use, past yr.				
Yes	33 (75.0)	383 (58.2)	n/a	n/a
No	10 (22.7)	254 (38.6)		
missing	1 (2.3)	21 (3.2)		

Note. SU=substance use; * = Other illegal drugs (LSD, ecstasy); n/a = norms item was not asked within YRBSS

Discussion

Accurate data is necessary for documenting needs, evaluating effectiveness, and understanding differences. But not all data are equally valid and reliable. Results from the TiPI example indicate that substance use is lower among TiPI youth than YRBSS youth with the exception of middle school marijuana use and urban binge drinking. These results are consistent with what we initially thought about TiPI youth involved in program activities, where involvement in culture-based prevention activities may result in lower substance use (Kelley, Restad, & Killsback, 2018). They are also similar to previous research that reported higher substance use rates in American Indian urban youth than reservation youth or non-native youth in the US (Rutman, Park, Castor, Tualii & Forquera, 2008; Lawrence, Pamepl, & Mollborn, 2015). Urban Native American youth also experience social and cultural stressors including acculturation, urbanization, marginalization, and discrimination that may explain differences observed (Hawkins, Cummins, & Marlatt, 2004).

Data tells us that substance abuse among youth is a significant public health challenge (Hawkins, Cummins, & Marlatt, 2004) with American Indian youth being placed at higher risk for substance use than others (Whitesell et al, 2014). Northern Plains reservation communities, including states of Iowa, Minnesota, Montana, North Dakota, South Dakota, Wisconsin, and Wyoming, are hard to reach areas, and substance use among American Indian youth in these locations is often higher than other reservation communities in the US (Beals et al, 2003).

From this process, we offer practical recommendations for collecting data with hard-to-reach populations. First piloting surveys and engaging the community in the process is critical for working with any hard-to-reach population. This process can help build trust and ensure that language is consistent and responsive to community needs.

Strengths

When working with hard-to-reach populations, it is important that community members are involved with survey development. Piloting surveys with community members from different backgrounds and ages ensures the survey questions are appropriate with regard to reading level and survey length. Although we

were not able to change all of the questions in the survey based on the pilot results, their involvement and engagement reflected a shift in how programs and researchers should work with racial and ethnic groups in the future. The local communities informed TiPI administration of what needed modification and what did not. Additional questions were added to provide the community with answers specifically around cultural connectedness, youths interest in cultural education, and substance use. Communities used results from these questions to document needs, develop additional programs, and apply for funding.

Limitations

There are limitations that must be considered in this example. First, the data collection setting is different for TiPI and YRBSS data collection. TiPI data were collected using a convenience sampling strategy, in a social setting during or after a program activity, where participation was voluntary. When youth were completing the survey in a large group, they would rush through the survey and then go back to a program activity. When the survey was completed in a classroom setting or in smaller groups, youth were more focused, they asked questions, and surveys were more complete. In contrast, YRBSS data is collected in a school setting and often a requirement. We also realized that social desirability, historic distrust of researchers and data collection in general (Kelley, et al., 2013) and other contextual factors could have contributed to responses that were not truthful. Second, the representativeness and generalizability of the data is limited to students that complete the surveys. With YRBSS data, they can only be applied to students that attend schools, this does not take into account youth who are not in school, those who are too young to be in school, or youth that drop out of school. With TiPI data, they only represent youth who participated in program activities that were between the ages of 12-20. Finally, the convenience sampling methods used for both TiPI and YRBSS are problematic because they do not represent the entire population, and it is nearly impossible to address hidden biases in the population sampled (Etikan, Musa, Alkassim, 2016).

Conclusions

TiPI youth report lower substance use than American Indian youth completing the YRBSS. When TiPI began, we were unsure what the comparisons

would look like. We asked ourselves, “How do the youth responses compare to the issue of substance abuse that we hear about, read about, and see first-hand in Indigenous communities?” We wanted to know if TiPI data was an accurate measure of substance use in the youth surveyed and the prevalence of substance use compared with the YRBSS data. This assumes that YRBSS data are accurate, which may not be true in general. Comparing data sets helped us understand differences in substance use rates in an American Indian population based on data attributes and measures.

In sum, educators, evaluators, program developers, policy makers, and professionals are often required to collect data that documents progress toward desired outcomes. Often this task involves developing an evaluation plan or assessment process that includes multiple data sources, methods, and indicators. Often, funding agencies like SAMHSA allow substitute data sources like the YRBSS as a proxy for primary data collection (SAMHSA, 2018). Even with permission from funding agencies, it is important consider the comparability of data, if it is valid and reliable. In the TiPI program example, we learned that communities want ownership of their data; they want to be involved in developing pertinent survey questions, administering surveys, and sharing results with their communities and tribal leaders. Although there are challenges with primary data collection in small populations that have historically distrusted Western survey models, the challenge is worth it. To improve health and wellbeing of hard to reach populations while documenting progress toward healthy future generations, we must be aware of the unique challenges of using national surveys like the YRBSS as comparison data, and the strengths of primary data collection driven by program needs.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

- Allen, M. (2017). *The sage encyclopedia of communication research methods* (Vols. 1-4). Thousand Oaks, CA: SAGE Publications, Inc doi: 10.4135/9781483381411
- Beals, J., Spicer, P., Mitchell, C. M., Novins, D. K., Manson, S. M., & AI-SUPERFPF Team. (2003). Racial disparities in alcohol use: Comparison of 2 American Indian reservation populations with national data. *American Journal of Public Health*, 93(10), 1683-1685.
- Bilheimer, L. T., & Sisk, J. E. (2008). Collecting adequate data on racial and ethnic disparities in health: The challenges continue. *Health Affairs*, 27(2), 383-391.
- Brener, N. D., Kann, L., McManus, T., Kinchen, S. A., Sundberg, E. C., & Ross, J. G. (2002). Reliability of the 1999 youth risk behavior survey questionnaire. *Journal of Adolescent Health*, 31(4), 336-342.
- Cherry, D. K., Woodwell, D. A., & Rechtsteiner, E. A. (2007). *National ambulatory medical care survey: 2005 summary*. Available from: <https://stacks.cdc.gov/view/cdc/81260>
- Chen, H., Hailey, D., Wang, N., & Yu, P. (2014). A review of data quality assessment methods for public health information systems. *International Journal of Environmental Research and Public Health*, 11(5), 5170–5207. <https://doi.org/10.3390/ijerph110505170>
- Ericksen, E. P. (1997). Problems in sampling the Native American and Alaska Native populations. *Population Research and Policy Review*, 16(1-2), 43-59.
- Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics*, 5(1), 1-4.
- Friedman, D. J., Cohen, B. B., Averbach, A. R., & Norton, J. M. (2000). Race/ethnicity and OMB Directive 15: implications for state public health practice. *American Journal of Public Health*, 90(11), 1714. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1446393/pdf/11076237.pdf>
- Friedman, D. J., Cohen, B. B., Averbach, A. R., & Norton, J. M. (2000). Race/ethnicity and OMB Directive 15: implications for state public health practice. *American Journal of Public Health*, 90(11), 1714.
- Foti, K., Balaji, A., & Shanklin, S. (2011). Uses of Youth Risk Behavior Survey and School Health

- Profiles data: applications for improving adolescent and school health. *Journal of School Health*, 81(6), 345-354.
- Hawkins, E. H., Cummins, L. H., & Marlatt, G. A. (2004). Preventing substance abuse in American Indian and Alaska native youth: promising strategies for healthier communities. *Psychological Bulletin*, 130(2), 304.
- Kelley, A., Belcourt-Dittloff, A., Belcourt, C., & Belcourt, G. (2013). Research ethics and indigenous communities. *American Journal of Public Health*, 103(12), 2146-52.
- Kelley, A., Restad, D., & Killsback, J. (2018). A Public Health Approach: Documenting the Risk and Protective Factors of Suicide Ideation in One American Indian Community. *Psychological Services*, 15, 325-331.
- Kelley, A., Witzel, M., & Fatupaito, B. (2019). Preventing Substance Use in American Indian Youth: The Case for Social Support and Community Connections. *Substance Use & Misuse*, 54, 787 - 795.
- Lawrence, E. M., Pampel, F. C., & Mollborn, S. (2014). Life course transitions and racial and ethnic differences in smoking prevalence. *Advances in Life Course Research*, 22, 27-40. doi: 10.1016/j.alcr.2014.03.002
- Marpsat, M., & Razafindratsima, N. (2010). Survey methods for hard-to-reach populations: introduction to the special issue. *Methodological Innovations Online*, 5(2), 3-16.
- Pirkis, J. E., Irwin Jr, C., Brindis, C., Patton, G. C., & Sawyer, M. G. (2003). Adolescent substance use: beware of international comparisons. *Journal of Adolescent Health*, 33(4), 279-286.
- Reijneveld, S. A., & Stronks, K. (1999). The impact of response bias on estimates of health care utilization in a metropolitan area: the use of administrative data. *International Journal of Epidemiology*, 28(6), 1134-1140.
- Rutman, S., Park, A., Castor, M., Taualii, M., & Forquera, R. (2008). Urban American Indian and Alaska native youth: Youth risk behavior survey 1997-2003. *Maternal and Child Health Journal*, 12(1), 76. doi: 10.1007/s10995-008-0351-3
- Shaghghi, A., Bhopal, R. S., & Sheikh, A. (2011). Approaches to Recruiting "Hard-To-Reach" Populations into Re-search: A Review of the Literature. *Health Promotion Perspectives*, 1(2), 86-94. <https://doi.org/10.5681/hpp.2011.009>
- Substance Abuse and Mental Health Services Administration (2018). *Center for Substance Abuse Prevention Community Outcomes Module for Partnerships for Success Grants* [Video Webinar Presentation]. Retrieved from: <https://spars.samhsa.gov/sites/default/files/PFSReferenceGuide.pdf>
- Trevino, F. M. (1988). Uniform minimum data sets: in search of demographic comparability. *American Journal of Public Health*, 78(2), 126-127.
- US Department of Health and Human Services. (2009). *Trends in Indian health*. Department of Health and Human Services, 2002-2003.
- Virnig, B. A., & McBean, M. (2001). Administrative data for public health surveillance and planning. *Annual Review of Public Health*, 22(1), 213-230.
- Wennberg, J., & Gittelsohn, A. (1973). Small area variations in health care delivery: a population-based health information system can guide planning and regulatory decision-making. *Science*, 182(4117), 1102-1108
- Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2010). *The program evaluation standards: A guide for evaluators and evaluation users*. Sage Publications.

Citation:

Kelley, A., McCoy, T., Fisher, A., Witzel, M., Fatupaito, B., & Restad, D. (2020). Comparability of Survey Measures in Hard to Reach Populations: Methods and Recommendations. *Practical Assessment, Research & Evaluation*, 25(10). Available online: <https://scholarworks.umass.edu/pare/vol25/iss1/9/>

Corresponding Author

Allyson Kelley
Allyson Kelley & Associates
69705 Lake Drive
Sisters OR 97759

email: kelleyallyson [at] gmail.com