Evaluation of Classification Error in a Survey on Sexual Assault Among College Students

Marcus E. Berzofsky[1], Chris Krebs[1], Christine Lindquist[1]
Michael Planty[1], Lynn Langton[2]
[1]RTI International, 3040 Cornwallis Road, Research Triangle Park, NC 27709
[2]Bureau of Justice Statistics, 810 Seventh Street NW, Washington, DC 20531. The views expressed in this paper are those of the author and do not reflect the views or position of Bureau of Justice Statistics or the U.S. Department of Justice.

Abstract
This paper evaluates the level of classification error that may exist when surveying college students regarding sexual assault. How sexual violence is defined within a survey can greatly impact how a student responds to the victimization prevalence items. If a student misunderstands the meaning of the definition or does not feel the definition fits their exact circumstances, the result may be classification error – a student misclassifying their true sexual assault status. To address this concern, the Campus Climate Survey and Validation Study (CCSVS), conducted in the spring of 2015 at nine post-secondary institutions, incorporated multiple indicators of sexual assault. Using these multiple indicators, this paper shows the results of a latent class analysis (LCA). LCA allows the measurement component of the parameters to be isolated thereby estimating the false positive – an indication a person was victim when their true status is not a victim – and a false negative rate – an indication a person is not a victim when their true status is a victim. Our findings show that a false negative classification error is larger problem than a false positive. Overall, the reported victimization rates in the CCSVS are most likely slight underestimations of the true victimization prevalence at each school.

Key Words: Latent class analysis, sexual assault, college students, classification error, measurement error

1. Introduction

1.1 Background

The validity of individual reports of sexual assault among college students is a current subject of great debate (Krebs and Lindquist 2014; Yoffe 2016). Adding to the difficulty for schools is the need to understand the magnitude and nature of the problem across the population at large. Schools recognize that official reports through the Clery act do not fully enumerate the amount of sexual assault occurring on their campus (Krebs et al., in press). Therefore, schools have turned to climate surveys as a way to provide more valid and reliable information about the problem of student sexual violence.

One criticism of these climate surveys is that respondents will not provide accurate responses – or that there will be *measurement error*. Measurement error is the difference between the true value of a measurement and the value obtained during the measurement process (Lessler & Kalsbeek 1992). Measurement errors can arise because of poorly designed questions or questionnaires, interviewer and respondent characteristics and behaviors, data collection modes (in-person, telephone, web, etc.), the survey setting (e.g., private or non-private) and many other factors (Biemer and Lyberg 2003).

Further, accurate measurement is complicated by the sensitive and personal nature of sexual violence. Sensitive topics have the potential to induce measurement error in one of two ways: (1) a false positive response or (2) a false negative response (Berzofsky, Biemer, & Kalsbeek 2014). A false positive response occurs when a non-victim responds in the affirmative to being a victim. This may occur if the respondent reports an incident from outside the reference period (i.e., telescopes the incident), the respondent misunderstands the definition of sexual assault to include actions the survey designers did not intend, or because the respondent wants to inflate the estimate because of the topics importance. A false negative response occurs when a victim responds in the negative to experiencing a sexual assault. Possible reasons for a false negative response include when a respondent is afraid to report an incident, misunderstands the definition, or has forgotten or blocked out an incident.

Thus, climate surveys should incorporate best practices to minimize the potential for measurement error. These practices include using a web-based self-administered mode which allows respondents to take the survey on their own time, in private and without an interviewer present. Even with such accommodations, climate surveys are still subject to several potential sources of measurement error. With the greatest source coming from the survey instrument.

1.2 Validating Climate Surveys

Validating a survey measure -- the extent that the survey measures what it is intended to measure -- can either be done through *external* or *internal* methods. External validation consists of using the same survey instrument across multiple populations to determine if the question wording provides expected (although not necessarily similar) estimates. Internal validation consists of embedding multiple items within the survey instrument to verify that a respondent answers the items in a consistent manner.

External validation is often a preferred method because it does not require additional survey items to be added to a questionnaire. However, schools have independently developed their climate survey instruments resulting in inconsistent definitions of sexual assault. For example, the University of Michigan (2015) used the very broad definition of "non-consensual (also known as unwanted) kissing or touching, or vaginal or anal penetration; and sexual harassment." The general inclusion of sexual harassment in the definition makes it broader than that used by Rutgers (2014), "Sexual violence refers to a range of behaviors that are unwanted by the recipient and include remarks about physical appearance; persistent sexual advances that are undesired by the recipient; unwanted touching; and unwanted oral, anal, or vaginal penetration or attempted penetration," which could result in respondent thinking only about the specific examples listed in the definition. MIT's (2014) definition of "attempted or completed non-consensual touching or kissing, oral sex or sexual penetration by force, threat, or incapacitation," was even more restrictive in terms of scope. However, the instrument did not explicitly define the terms, leaving the potential for misinterpretation. Subsequently, this variability across studies limits our ability for external validation.

Internal validation, on the other hand, includes methods ranging from consistency checks across related survey items (e.g., reported height and weight have a reasonable relationship) to more sophisticated modeling techniques such as latent class analysis (LCA; Biemer 2011). LCA uses embedded replication – the inclusion of multiple survey questions – to estimate the consistency of responses for a latent construct. When the latent construct is dichotomous, LCA can be used to estimate the classification error rates (i.e., the false positive and false negative rates), as well as an adjusted estimate that accounts for potential measurement error.

LCA is useful for assessing measurement error in surveys on sensitive topics, because it does not require a knowledge of the true response (Biemer 2011). (?)Additionally, because of the use of embedded replication, LCA can assess the quality of different question wording (Biemer & Wiesen 2002; Krueter, Yan, and Tourangeau 2008; Biemer & Berzofsky 2011). Examples of the prior use of LCA include, assessing sexual assault among an inmate population (Berzofsky, Biemer, & Kalsbeek 2014), crime victimization among the general population (Edwards, Berzofsky, & Biemer 2017), drug use (Biemer & Wiesen 2002), and employment status (Biemer, 2004). However, this technique has not been used on a college population to assess sexual assault.

1.3 Research Questions

This paper presents the results of using LCA to asses the validity of a survey of campus sexual assault. The following research questions guided the LCA:

1. Can the validity of questions on sexual assault be ascertained?
2. What is the level of measurement error in survey items on sexual assault?
3. Does measurement error vary across schools or subgroups of students?
4. Are there specific student characteristics that have higher rates of measurement error?

2. Methods

2.1 Campus Climate Survey and Validation Study

The analysis was conducted using data from the Campus Climate Survey and Validation Study (CCSVS; Krebs et al. 2016), which was conducted at 9 post-secondary institutions during the 2014 – 15 academic year. Conducted by RTI International and sponsored by the U.S. Bureau of Justice Statistics (BJS) and the Office of Violence Against Women, the CCSVS collected data from approximately 23,000 male and female students, using a self-administered web survey. Participating schools provided student rosters with the students' name, school ID and sex, as well as other administrative information on the students that could be used for nonresponse bias analysis. Within each school, a stratified simple random sample of students was drawn, stratifying on the sex of the student.

The CCSVS contained questions related to the prevalence, incidence and nature of sexual assault during the current academic year, as well as questions related to the climate on campus generally and pertaining specifically to sexual assault. For female students, the CCSVS was statistically powered to produce precise estimates of sexual assault. For male students, the CCSVS was powered to climate measures. As such, the LCA to assess measurement error was restricted to the approximately 15,000 female respondents.

2.2 Latent Class Analysis

LCA was developed by Lazarsfeld & Henry (1968) to classify and estimate latent constructs that cannot be directly measured in a survey. More recently, for cross-sectional surveys, survey methodologists have used LCA to measure classification error (measurement error for categorical variables) when the number of latent classes is fixed

(Biemer 2011). To conduct LCA, the data need to contain the following minimum requirements:

- Categorical independent and dependent variables,
- At least 3 indicators of the latent construct,
- Indicators asked of all respondents – not just victims, and
- Locally independent indicators (Berzofsky, Biemer, & Kalsbeek 2014).

In LCA, the latent construct is denoted by *X*, the indicators from each embedded replication are denoted by *A*, *B*, *C*, *D*, etc., and grouping variables – subdomain variables within which classification error is homogeneous – are denoted by G_i for subdomain *i*. For a dichotomous latent construct and indicator, the classification probability is written as $P(A=a|X=x)$ where $P(A=1|X=2)$ represents the false positive rate and $P(A=2|X=1)$ represents the false negative rate.

Indicators, developed from each embedded replication, need to represent the latent construct, but cannot be asked in identical ways. Examples of how to alter question wording include: (1) asking a series of behaviorally specific acts or actions related to the latent construct, (2) asking a single yes/no question about whether the latent construct occurred, (3), asking when the last time the latent construct occurred with an option for never, and (4) asking how many times the latent construct occurred with zero as an option.

A latent class model (LCM) consists of two components: *structural* and *measurement* (Bassi, Hagenaars, Croon, & Vermunt, 2001; Hagenaars, 1998). The structural component estimates the latent construct overall and by each grouping variable. This estimate is considered the unbiased "true" estimate (Biemer 2011). The measurement component estimates the classification error for each indicator overall and by each grouping variable.

For a single grouping variable, G, an LCM with four indicators can be written as:

$$P(GABCD) = P(X|G)P(A|GX)P(B|GX)P(C|GX)P(D|GX)$$

where $P(X|G)$ is the structural component and $P(A|GX)P(B|GX)P(C|GX)P(D|GX)$ is the measurement component.

2.3 LCA in the CCSVS

The CCSVS instrument was developed to contain four indicators of sexual assault. The latent construct was defined as the student's sexual assault status during the current academic year. Prior to the asking each of the indicators, each respondent was provided with, and had to acknowledge understanding, the behaviorally specific definition of sexual assault used for the survey.

Figure 1 presents the question wording for the first two indicators and **Figure 2** presents the question wording for the third and fourth indicator. Survey Item P1(*Indicator A*) is a dichotomous question asking about unwanted sexual contact during the current academic year. Survey Item P2 (*Indicator B*) asks the number of times the respondent experienced unwanted sexual contact during the current academic year. Survey Item LCA2 (*Indicator C*) asks a series of behavioral specific questions. The respondent needed to answer yes or no to each question. A response of "yes" to any of the questions classified the respondent as a victim of sexual assault. Survey Item LCA3 (*Indicator D*) asks about the last time during the respondent's life that they experienced unwanted sexual contact. A respondent could answer “never” if they never experienced unwanted sexual contact. If the respondent indicated a date which fell within the current academic year, then the respondent was classified as a victim of sexual assault.

Survey Item P2 was used to measure the published rate of sexual assault. This item was selected as the main indicator of sexual assault prior to data collection because it was the only indicator which provided the number of experiences of unwanted sexual contacts. Because this number was necessary for the follow-up incident reports, Survey Item P2 was the only measure that would allow consistency between the overall estimate and the follow-up items (see Krebs et al. (2016) for details on the measurement process).

P1. **Since the beginning of the current academic year in** [FILL: **August/September**], **2014**, has anyone had unwanted sexual contact with you?

- Yes
- No

P2. How many separate incidents of unwanted sexual contact have you experienced **since the beginning of the current academic year in** [FILL: **August/September**], **2014**?

- 0 incidents [IF P2 = 0 IINCIDENTS, SKIP TO LCA2]
- 1 incident
- 2 incidents
- 3 incidents
- 4 incidents
- 5 or more incidents

Figure 1: Indicator A and Indicator B used in the CCSVS. Source: Campus Climate Survey Validation Study (CCSVS), 2015

LCA2. Just to confirm, **since the beginning of the current academic year in** [FILL: **August/September**], **2014**, has anyone had any of the following types of unwanted sexual contact with you (i.e., sexual contact without your consent and that you did not want to happen?

	Yes	No
a. **Forced touching of a sexual nature** (forced kissing, touching of private parts, grabbing, fondling, rubbing up against you in a sexual way, even if it is over your clothes)	○	○
b. **Oral sex** (someone's mouth or tongue making contact with your genitals or your mouth or tongue making contact with someone else's genitals)	○	○
c. **Anal sex** (someone putting their penis in your anus)	○	○
d. [RESPONSE WILL NOT DISPLAY IF D3=MALE] **Sexual intercourse** (someone putting their penis in your vagina)	○	○
e. **Sexual penetration with a finger or object** (someone putting their finger or an object like a bottle or a candle in your [IF D3= FEMALE OR TRANSGENDER ORSOMETHING ELSE OR MISSING, FILL: "vagina or"] anus	○	○

Previous Next

LCA3. Thinking about your **whole life**, when was the last time you experienced unwanted sexual contact?

Never	Month			Year		
☐	Select an answer...	˅	[DROP DOWN LIST JAN-DEC]	Select an answer...	˅	[DROP DOWN LIST 2015-2005 OR EARLIER

Figure 2: Indicator C and Indicator D used in the CCSVS. Source: Campus Climate Survey Validation Study (CCSVS), 2015

The CCSVS included set of demographic variables that were considered for grouping variables. These variables include age, year of study, race/ethnicity, sexual orientation, and gender identity. Using the model fitting approach proposed by Berzofsky, Biemer and Kalsbeek (2014) to reduce the number of grouping variables and identify the most significant, sexual orientation (heterosexual or lesbian, gay, bisexual or other), year of study (first, second, third, or fourth), and school were identified as the best grouping variables.

The analysis was conducted using LatentGold 5.0 (Vermunt & Magidson 2005). While the CCSVS had a very low rate of item nonresponse, the indicators and grouping variables still had some missing data. To ensure that all cases could be incorporated in the analysis, the full information maximum likelihood (FIML) approach was used, incorporating the approach detailed in Edwards, Berzofsky, and Biemer (2017) to determine the missing data mechanism (missing at random or missing not at random) and ensure proper FIML application.

3. Results

Figure 3 presents the estimated false positive (over-reporting sexual violence) and false negative (under-reporting sexual violence) rates by indicator. The false positive rates were all less than 1.5% with Indicator A and Indicator B having the smallest rate (0.6%). The false negative rate ranged between 7.3% (Indicator C) and 21.2% (Indicator D). Indicator B had a false negative of 9.9% while Indicator A had a rate of 17.9%.

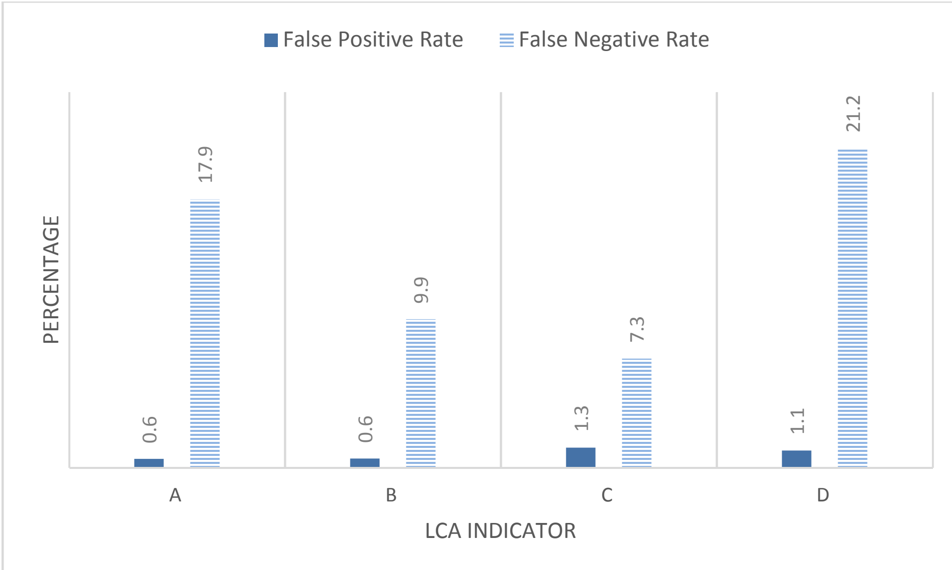

Figure 3: False positive and false negative rates based on LCA among undergraduate females, by indicator, 2014 – 2015 academic year. Source: Campus Climate Survey Validation Study (CCSVS), 2015

Figure 4 presents the estimated unbiased estimate based on the LCA and the reported estimate of sexual assault among undergraduate females by school. Overall, the unbiased LCA estimate was larger than the reported estimate (10.7% vs. 10.2%). Except for school 3, unbiased estimates were higher than the reported estimates for all schools.

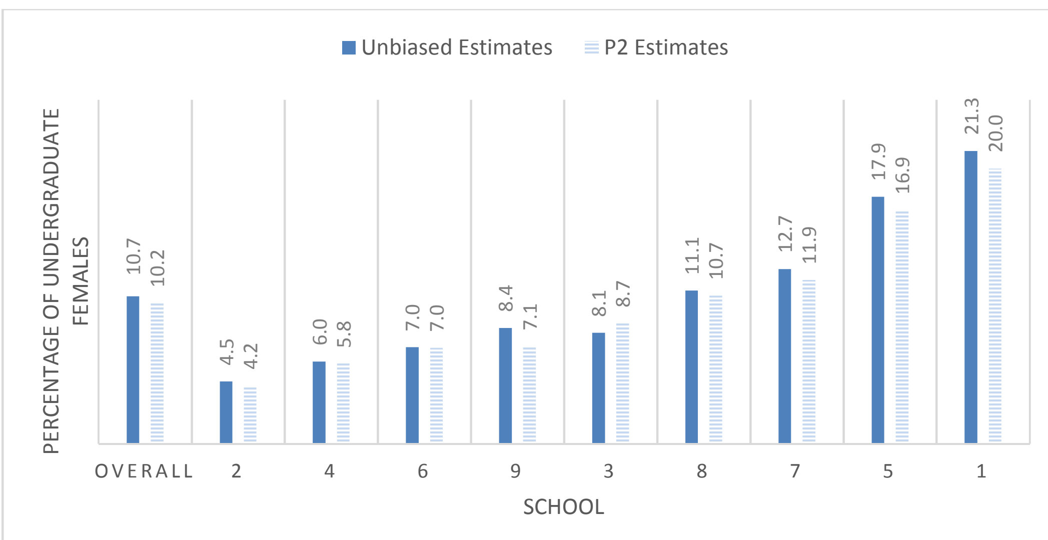

Figure 4: Unbiased LCA and primary estimates of sexual assault for undergraduate females, 2014 – 2015 academic year, by school. Source: Campus Climate Survey Validation Study (CCSVS), 2015

Figure 5 presents the unbiased LCA estimate and reported estimate by student year of study. While the unbiased LCA estimate was larger for each year of study, the magnitude of the difference was greatest for first year (12.8% vs. 12.1%) and second year (11.6% vs.

10.6%) female students. The differences for third year (10.1% vs. 9.9%) and fourth year (8.9% vs. 8.8%) students were minimal.

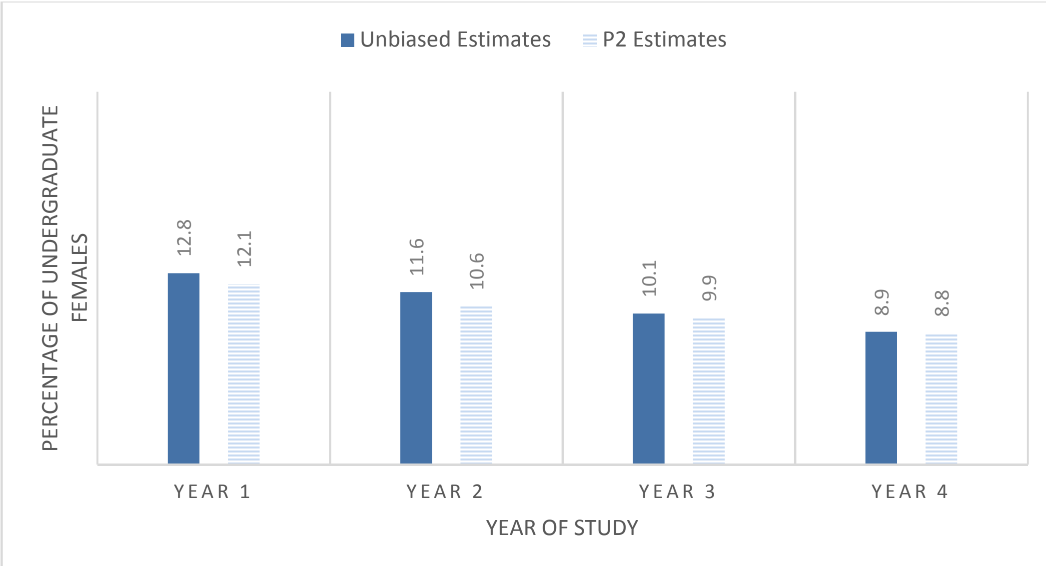

Figure 5: Unbiased LCA and primary estimates of sexual assault for undergraduate females, 2014 – 2015 academic year, by year of study. Source: Campus Climate Survey Validation Study (CCSVS), 2015

Figure 6 presents the unbiased LCA estimate and reported estimate by sexual orientation. The unbiased LCA estimates were greater for both sexual orientation types. However, the magnitude of the difference was greater for lesbian, gay, bisexual and other orientation respondents (19.1% vs. 16.6%) than heterosexual respondents (9.8% vs. 9.3%).

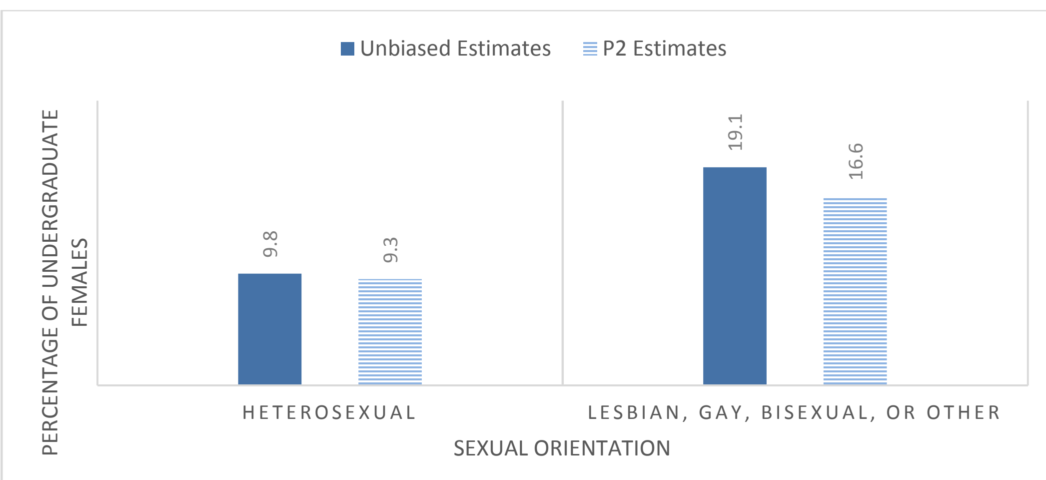

Figure 6: Unbiased LCA and primary estimates of sexual assault for undergraduate females, 2014 – 2015 academic year, by sexual orientation. Source: Campus Climate Survey Validation Study (CCSVS), 2015

5. Discussion

Findings on the classification error rates estimated for the CCSVS – a relatively large false negative rate compared to a near zero false positive rate – follow a similar pattern to other LCAs of sensitive topics, such as sexual assault among inmates (Berzofsky, Biemer & Kalsbeek 2014). This finding indicates respondents are more likely to omit reporting a sexual assault rather than report an incident that did not occur. However, even though the magnitude of the false negative rates was higher than the false positive rates the unbiased estimate is only 0.5% larger (10.7% versus 10.2%, figure 3). This is because, when estimating the unbiased estimate, the false negative rate is applied to the set of respondents who indicated a sexual assault and the false positive rate is applied to the set of respondents who did not indicate a sexual assault. Since the number of respondents who indicated a sexual assault was much smaller than those who did not, the influence of the false negative rate is not as great as its difference with the false positive rate.

When comparing the classification error rates for each of the indicators, Indicator B (based on Survey Item P2) had the smallest false positive rate and the second smallest false negative rate. This suggests that asking a respondent the number of times they experienced unwanted sexual contact, after clearly defining what is meant by unwanted sexual contact, during the academic year induces the smallest amount of measurement error. This provides internal validation for the use of Survey Item P2 to estimate the prevalence of sexual assault.

Indicator C also performed well with the lowest false negative rate (7.3% compared to 99% for Indicator B). This indicator, based on Survey Item LCA2, was derived from a set of behaviorally specific items. This finding is in line with other studies showing that questions about specific acts or actions the respondent may have experienced do a better job of eliciting affirmative responses than items that ask more broadly about the latent construct (Berzofsky, Biemer, & Kalsbeek 2014).

Indicator D performed the worst in terms of classification error rates. Indicator D, based on Survey Item LCA3, required respondents to enter a specific month and year for when the unwanted contact occurred. The level of specificity may have resulted in the higher level of classification error, which should be considered prior to requesting respondents to provide specific dates for a sensitive event like sexual assault.

In terms of respondent characteristics, there appear to be differences in the amount of measurement error depending on a student's year of study and sexual orientation. First and second year students were more likely to report inconsistent responses, which yielded larger unbiased estimates, compared to third and fourth year students. While the exact reason for this difference cannot be determined from this analysis, this finding does suggest that care needs to be taken to ensure underclassmen fully understand the questions and feel comfortable providing an accurate response. Furthermore, non-heterosexuals have a higher rate of classification error than heterosexuals. This indicates that non-heterosexual students are less willing to provide an accurate response or have more difficulty interpreting the questions than heterosexual respondents. Moreover, for both year of study and sexual orientation, the domain levels with higher rates of sexual assault had larger classification error rates.

5. Conclusions

Campus climate surveys at colleges and universities have been used to allow institutions to understand the magnitude and nature of sexual assault on their campus. The validity of these studies is paramount to ensure that the administration, student body, and public can trust accuracy of the study results. This paper presents the first internal validation of a campus climate study. LCA was applied to the CCSVS to measure both the classification (measurement) error rates and the unbiased prevalence rates after accounting for measurement error.

Findings were aligned with other studies on classification error for sensitive topics, which found that the false negative rate outweighed the influence of the false positive rate on the unbiased estimates. This resulted in unbiased estimates which were larger than the reported estimates. However, overall, the classification error had a minimal impact on reported CCSVS findings. In addition, the indicator used for reporting sexual assault had the lowest levels of classification error compared to the other three indicators included in the analysis.

Finally, the study showed that classification error rates are not uniform across subdomain levels. Specifically, first and second year students and non-heterosexual students had higher classification error rates than other students. They also had higher rates of sexual assault. While it's not possible with this analysis to determine the exact reason for this difference, survey designers should consider methods to reduce the level of classification error among students of these characteristics.

Acknowledgements

The authors would like to thank BJS for sponsoring this research. However, we would like to note that the views expressed in this paper are those of the authors only and do not reflect the views or position of BJS or the Department of Justice.

References

Bassi, F., Hagenaars, J. A., Croon, M. A., & Vermunt, J. K. (2001). Estimating true changes when categorical panel data are by uncorrelated and correlated classification errors: An application to unemployment data. *Sociological Methods and Research, 29*, 230-268.

Berzofsky, M. E., Biemer, P. P., & Kalsbeek, W. D. (2014). Local dependence in latent class analysis of rare and sensitive events. *Sociological Methods and Research, 43*(1), 137–170. doi:10.1177/0049124113506407

Biemer, P. (2004). An analysis of classification error for the revised Current Population Survey employment questions. *Survey Methodology*, *30*(2), 127–140.

Biemer, P.P. (2011). *Latent Class Analysis of Survey Error*. Hoboken, NJ: John Wiley & Sons.

Biemer, P.P, & Berzofsky, M. (2011). Some issues in the application of latent class models for questionnaire design. In J. Madans, K. Miller, G. Willis, & A. Maitland (Eds.). *Questionnaire evaluation methods*, Hoboken, NJ: John Wiley & Sons.

Biemer, P. P. and Lyberg, L. 2003. *Introduction to Survey Quality*, John Wiley & Sons, Hoboken, N.J.

Biemer, P. P., & Wiesen, C. (2002). Latent class analysis of embedded repeated measurements: An application to the National Household Survey on Drug Abuse. *Journal of the Royal Statistical Society: Series A, 165*(1), 97-119.

Edwards, S., Berzofsky, M. E., & Biemer, P. P. (2017) Effect of missing data on classification error in panel surveys. *Journal of Official Statistics, 33,* 551–570. https://doi.org/10.1515/jos-2017-0026

Hagenaars, J. A. (1998). Categorical causal modeling: Latent class analysis and directed log-linear models with latent variables. *Sociological Methods and Research, 26*, 436-487.

Krebs, C. & Lindquist, C. (2014, December 15). Setting the Record Straight on '1 in 5.' *Time Magazine*. http://time.com/3633903/campus-rape-1-in-5-sexual-assault-setting-record-straight/

Krebs, C., Lindquist, C., Langton, L., Berzofsky, M. E., Planty, M., Asefnia, N., et al. (in press). The value and validity of self-reported survey data on the rape experiences of college students. *Violence Against Women*.

Krebs, C. P., Lindquist, C. H., Berzofsky, M. E., Shook-Sa, B. E., Peterson, K. C., Planty, M., et al. (2016). Campus Climate Validation Survey: Final technical report. *Bureau of Justice Statistics Research and Development Series*, *NCJ 249545*. Washington, DC: Bureau of Justice Statistics, Department of Justice. http://www.bjs.gov/content/pub/pdf/ccsvsftr.pdf

Kreuter, F., Yan, T., & Tourangeau, R. (2008). Good item or bad--Can latent class analysis tell? The utility of latent class analysis for the evaluation of survey questions. *Journal of the Royal Statistical Society: Series A, 171*(3), 723-738.

Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.

Lessler, J. T., & Kalsbeek, W. (1992). *Nonsampling error in surveys*. New York: Wiley.

Massachusetts Institute of Technology (2014). *Survey Results: 2014 Community Attitudes on Sexual Assault*. Retrieved from Web: http://web.mit.edu/surveys/health/MIT-CASA-Survey-Summary.pdf

Rutgers University (2015). Rutgers University—New Brunswick Campus Climate Assessment. file:///C:/Users/krebs/Downloads/rutgers_campus_climate_assessment_process_-final_report.pdf

University of Michigan (2015). University of Michigan Campus Climate Survey on Sexual Misconduct. http://publicaffairs.vpcomm.umich.edu/wp-content/uploads/sites/19/2015/04/Complete-survey-results.pdf

Vermunt, J. K., & Magidson, J. (2005). *Technical guide to Latent Gold 4.0: Basic and advanced*. Belmont, MA: Statistical Innovations.

Yoffe, E. (2015, September 24). The Problem with Campus Sexual Assault Surveys. *Slate*. http://www.slate.com/articles/double_x/doublex/2015/09/aau_campus_sexual_assault_survey_why_such_surveys_don_t_paint_an_accurate.html