Identifying "mischievous" responders through Latent Class Analysis

Hanno Petras[1] Michele Ybarra[2]
[1]American Institutes for Research (AIR)
[2]Center for Innovative Public Health Research (CiPHR)

24th Annual Meeting
Using Prevention Science to Promote Health Equity and Promote Well-Being
May 31st –June 2nd , 2016
San Francisco, CA

Abstract

Sensitive questions about behaviors and attitudes are common in population surveys. Self-report bias may affect rates. This could be a result of *over-reporting* of desirable behaviors or *under-reporting* of undesirable behaviors. Similar to outliers, misreporting has the potential to significantly bias statistical results and, unlike random measurement error, increases in sample size will not reduce the problem. Misreporting is more likely to occur if questions:

- Are perceived as intrusive
- Raise fears about the potential repercussions of disclosing the information
- Trigger desirability concerns

In this presentation, we will focus on individual reports of having engaged in sexual violence perpetration, including sexual harassment (in-person and through internet), rape, attempted rape, coercive sex, and sexual assault. Although response bias is controllable to some extent through research and questionnaire design (e.g., conducting data in a private space, obtaining a Certificate of Confidentiality), it cannot be eliminated entirely. Consequently, a common strategy to identify and control for misreporting is to use dedicated scales which add non-substantive questions to a survey. This assumes however, that these scales accurately identify mis-reporting. In this presentation, we will present results from an ongoing study using person-centered latent variable models to identify suspected misreporting among groups of participants.

Objective

This study builds upon the results presented by Ybarra & Petras in the previous poster. Taking the 3-class solution for both age groups as a point of departure, this poster will focus on the following aims: 1) Inspect evidence for under- and over reporting in the sample; 2) Assess whether these findings differ for 16-18 year versus 19+ year olds; 3) Evaluate whether these findings differ by gender.

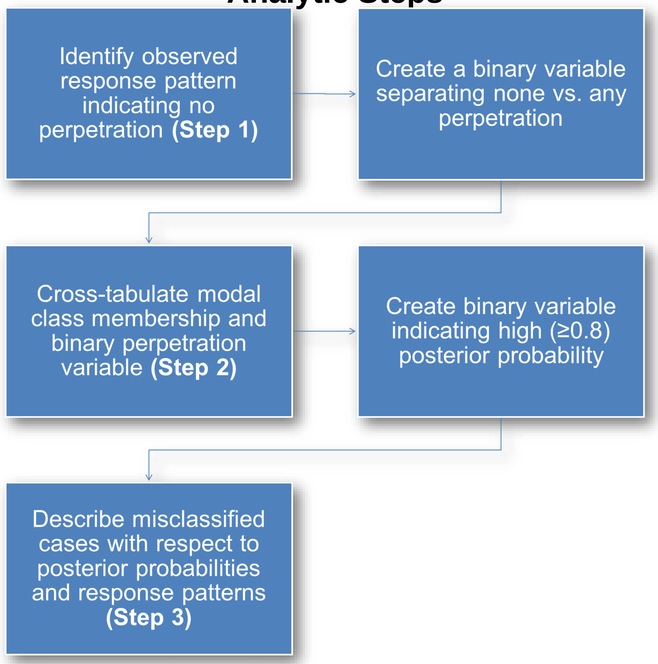
Methods

Latent Class Analysis is a well utilized latent variable method to describe covariation in observed categorical variables through categorical latent variables, i.e., latent classes.

Individuals are assigned to latent classes in a probabilistic fashion given their observed response patterns utilizing Bayes` Theorem:

$$\Pr(C = c) = \frac{\Pr(C = c) x Pr(Y_i | C = c)}{\sum_{c=1}^{C} \Pr(C = c) x Pr(Y_i | C = c)}$$

Analytic Steps



* Thank you for your interest in this presentation. Please note that analyses included herein are preliminary. More recent, finalized analyses may be available by contacting CiPHR.

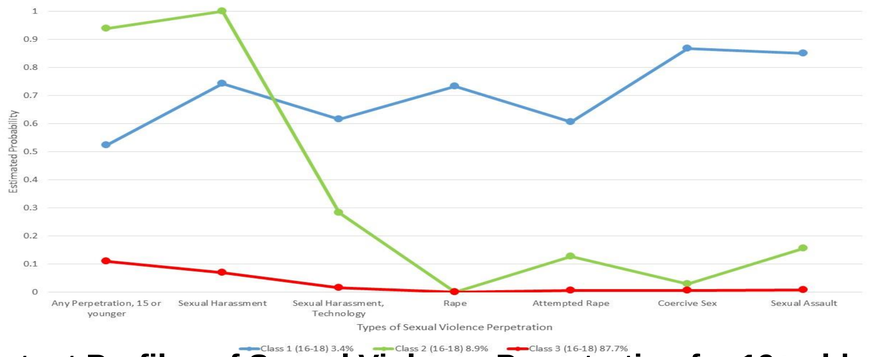
Identifying "mischievous" responders through Latent Class Analysis
Hanno Petras[1] Michele Ybarra[2]
[1]American Institutes for Research (AIR)
[2]Center for Innovative Public Health Research (CiPHR)
24th Annual Meeting
Using Prevention Science to Promote Health Equity and Promote Well-Being
May 31st –June 2nd , 2016
San Francisco, CA

Results

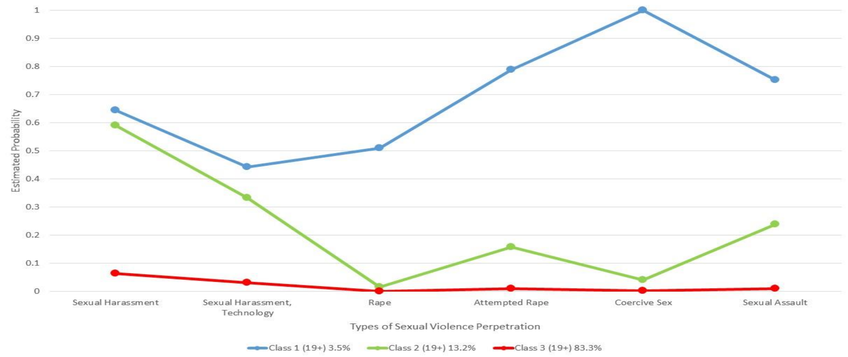
STEP 0

Below Figures show the three latent profiles of sexual violence perpetration (high, sexual harassment, low-no perpetration) for the two age groups. Cases are classified with high levels of accuracy (16-18 years old: 0.912; 19+ year olds: 0.895).

Latent Profiles of Sexual Violence Perpetration for 16-18 olds



Latent Profiles of Sexual Violence Perpetration for 19+ olds



STEP 1

Among individuals 16-18 years old, four response patterns (RP) were observed which did not indicate any engagement in sexual violence perpetration. These four patterns accounted for 76.4% of youth in this age group. Among individuals 19 years or older, two RPs were observed that did not indicate any engagement in sexual violence perpetration.

Observed Response Patterns without report of any perpetration

	Perp	SH	SHT	RP	ARP	COS	SA	Freq
16-18	0	0	0	0	0	0	0	207
	9	0	0	0	0	0	0	411
	0	9	9	9	9	9	9	38
	9	0	9	0	0	0	9	1
19+	NA	0	0	0	0	0	0	415
	NA	9	0	0	0	0	0	2

STEP 2

Below, the overlap between observed (O) perpetration patterns (yes vs. no) and class assignment (A) is shown. Variable "A" was constructed by summing across the two latent perpetration profiles. While specificity is high in both groups, sensitivity is low: 0.33 for 16-18 year olds and 0.6 for 19+ year olds.

16-18 Year Olds

O=Observed, A=Assigned	A=Yes	A=No
O=Yes	66	137
O=No	0	657

19+ Year Olds

O=Observed, A=Assigned	A=Yes	A=No
O=Yes	49	81
O=No	0	417

Identifying "mischievous" responders through Latent Class Analysis

Hanno Petras[1] Michele Ybarra[2]
[1]American Institutes for Research (AIR)
[2]Center for Innovative Public Health Research (CiPHR)

24th Annual Meeting
Using Prevention Science to Promote Health Equity and Promote Well-Being
May 31st –June 2nd , 2016
San Francisco, CA

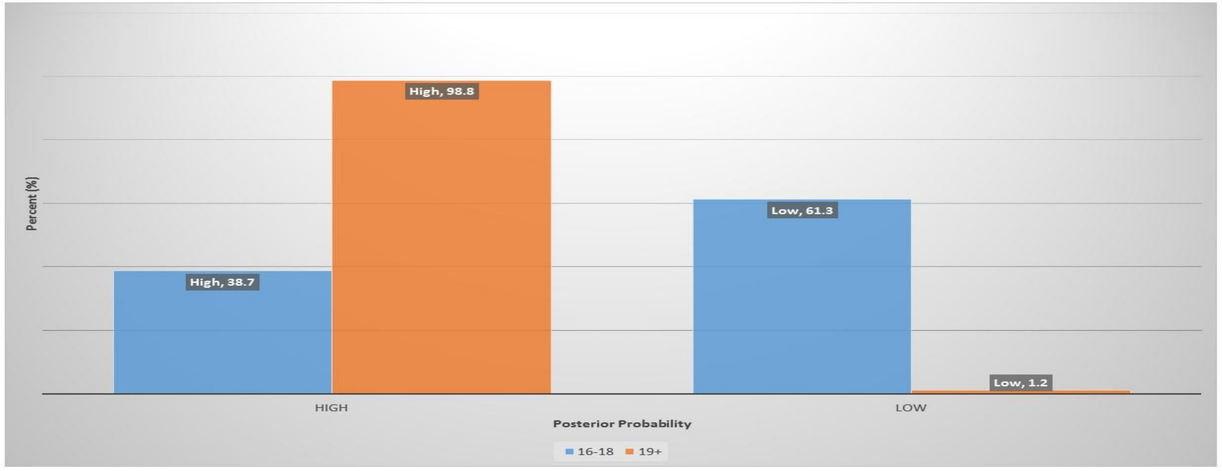
Results (cont.)

Step 3

In Step 2, we identified that 137 (15.9%) of 16-18 year olds and 81 (14.8%) of 19+ year olds were assigned to the low perpetration group despite the fact that their observed response patterns indicated perpetration.

In Step 3, we will now inspect whether they were assigned with high or low probability. A high probability **may be** interpreted as indication of over-reporting, while a low probability may indicate partial under-reporting.

Among 16-18 year olds, only 53 cases (38.7%) were assigned to the low perpetration profile with high posterior probabilities. Among youth with low posterior probabilities, the most frequent observed response pattern (76.2%) indicated that they reported engagement in sexual harassment through the internet only. In the older group, only one person was assigned with low posterior probabilities and this person reported engagement in sexual harassment (in person and through internet).



Conclusions

It is possible to classify individuals with respect to their reported perpetration behavior, but not without classification error. Importantly, while specificity for both age groups is high, identifying perpetrator based on their reported behavior is more difficult (i.e., lower sensitivity was noted).

For 16-18 year olds, support for both under- and over reporting was found, while for the older individuals, over-reporting appeared to be the larger problem. This finding is not unlike what has been found for self reported delinquency.

The next iterations of analysis will involve the following steps:

1) Identify significant predictors of biased reporting ➔Bias Index
2) Compare results to models controlling for measurement error ➔ Factor Mixture model
3) Incorporate information from Desirability measure available in the next wave of data collection.

References

Flaherty, B. P. (2002). Assessing reliability of categorical substance use measures with latent class analysis. *Drug and Alcohol Dependence, 68*, S7-20.

Yan, T., Kreuter, F., & Tourangeau, R. (2012). Latent class analysis of response inconsistencies across modes of data collection. *Social Science Research, 41*, 1017-1027.

Costa, P. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 6*, 882-888.

The project described was supported by Award Number R01HD083072 from the National Institute of Child Health and Human Development (NICHD). The content is the sole responsibility of the authors and does not necessarily represent the official views of NICHD.