Methodological Critique: Executive Summary

On the face of it, the July 13, 2015 *JAMA Pediatrics* article entitled, "Trajectory Analysis of the Campus Serial Rapist Assumption," authored by Swartout, Koss, White, Thompson, Abbey, and Bellis appears to be scientifically sound. The paper reports a study conducted with sizeable samples from both a derivation and validation dataset to fit latent trajectories of the probability of raping across the college years, using perpetrators' own reports. The article reads well and the naïve reader would likely be impressed and ready to accept the authors' conclusions – that most campus rapists are *not* predatory serial rapists but rather opportunistic, time-limited rapists. However, a closer look reveals a number of problems that call the science and conclusions of the paper into question.

Problems with the Analyses

The dataset and the statistical programming code used to conduct the analyses provided by Dr. Swartout revealed a number of irregularities. The model for the derivation dataset yielded a warning automatically generated by the Mplus software that the model was under-identified, so that key growth parameters were fixed rather than estimated. In essence, the model was unable to define or differentiate trajectory groups.

Despite reporting that the fifth timepoint was omitted from the derivation analysis sample due to low response rates, Dr. Swartout's programming code included the fifth timepoint. The consequence of using the "low response" fifth timepoint is that the analysis relied heavily on how the missing data were handled in the analysis. The authors claim (E3, 1st paragraph) that "…missing data were not related to reports of sexual violence across the study," citing the Pearson chi-square test. When the final timepoint was *excluded* from the analysis model (and the proportion of missing data was reduced), this test indicated a significant departure from this assumption. In other words, the missing data handling technique used in the *JAMA* article may have resulted in biased estimates of the probability of rape.

Another thing that changed when 4 rather than 5 timepoints were used was the shape of the trajectory curves (Figure 1). Patterns that yield 100% and/or 0% likelihoods, as in Class 2 (panel B of Figure 1), often indicate that there are too few members in a given class, causing the parameter values for that class to be "overfitted" to these specific individuals. In both the 4 and 5 timepoint models, less than 2% of the sample was assigned to each of the two "rapist" classes (sample sizes ranged from 6 to 12). The statistical power for detecting such rare latent classes is low (70%). When sample sizes are this small, the predicted probabilities tend to be unstable, likely to change with small changes to the model. This is evident when examining the classification quality, where individuals in the two smallest classes had a significant likelihood of being misclassified (4 timepoint model: 32% & 35%; 5 timepoint model: 49% & 39%).

The authors do not demonstrate that the two populations are identical and thus their approach to replication is questionable. For example, from the information given, it is apparent that the two samples are markedly different with respect to race and Hispanic ethnicity (Derivation Sample: White=68.5%;

Black=25.7%; Other=5.8%; Validation Sample: White=89.4%; Black=7.3%; Other=3.3%). Differences in data collection technique and/or the willingness of the validation sample to participate in the study over time also call into question the validity of the replication approach.

The results of the model using the validation dataset didn't match the numbers reported in the article – except the sample sizes in the trajectory classes. All the growth parameters were estimated, not fixed, and the missing data assumption held in this dataset. However, as with the derivation model, the sample sizes for the rapist trajectory classes were very small (n=19 and 39), constituting 2.3% and 4.9% of the sample. As noted above, this means that the predicted probabilities tend to be unstable and are likely to change with small changes to the model. Indeed, the likelihood of being misclassified was 16% for the derivation model and 20% for the validation model. A Monte Carlo simulation study showed that the statistical power associated with the trajectory parameters were all less than 60% and the power to distinguish between the Increasing and Decreasing trajectory classes was 44%, substantially lower than the recommended 80%.

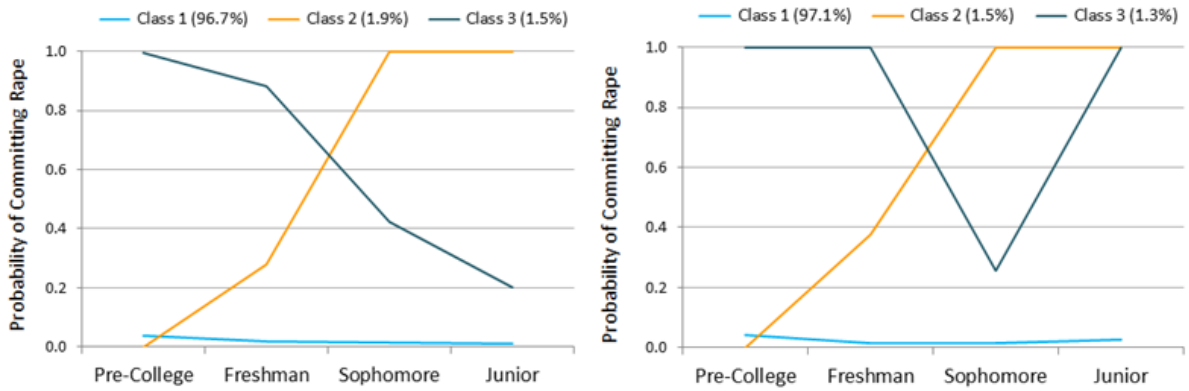

Figure 1. Different trajectory classes when 5 timepoints (left panel, to parallel the figure in the article the fifth point is not shown) versus 4 timepoints are used to model the derivation dataset. The trajectories on the left match those for the derivation dataset in the figure in Swartout et al.'s paper; those trajectories were produced only when senior year was included, but the authors explicitly said that senior year was excluded from their analyses. The trajectories on the right are what their model produces when senior year data actually are excluded; note the absence of a "decreasing" trajectory.

Swartout and Colleagues' Own Data Tell a Very Different Story

It is not unusual that a very large subpopulation in a latent trajectory analysis heavily influences the solution. To focus the analysis on college-aged rapists, the very large subgroup of non- rapists (i.e., no rape reported at any timepoint) was omitted, and the derivation and validation datasets were combined, and every timepoint was included – all of which maximized the number of rapists on which to base the trajectory analysis. Under these conditions, the same syntax provided by Dr. Swartout yielded a very different solution. In this model, the classes were more balanced (61%, 22%, 17%) and therefore more likely to describe prevalent subpopulations. In this model, only the smallest class (17%) had a pattern that described time-limited rape (Figure 2, Class 3, green line).

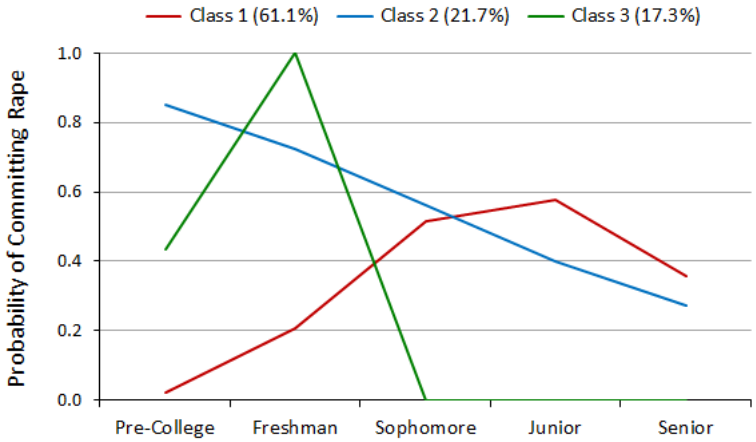

Figure 2. Trajectory classes among rapists (only) in the combined derivation-plus-validation dataset, using Swartout and colleagues' model.

By using trajectory analysis, Dr. Swartout and his colleagues imposed a restriction on the data, specifically that the probability of rape can best be described by a smooth line over time, which is a severe constraint that mismodels men who rape at non-consecutive timepoints. Further, using the pre-college timepoint, for which all participants have data, heavily influenced the growth trajectory – particularly the intercept term – on which the trajectory classes were based. Indeed, the "increasing" class, which is shown graphically in the *JAMA* article's figure, had a 0% probability of rape at pre-college, and the "decreasing" class had a 100% probability of rape at pre-college. Such heavy reliance on the pre-college timepoint, paired with the (incorrectly) assumed independence of pre-college and college rapes (E6, 1st paragraph), does not directly address rapes committed *during the college years*.

When Swartout and colleagues' untenable trajectory constraint was released, and the probabilities of rape were instead estimated at each timepoint, this resulted in a 5-class model in which 65% of the rapist sample had *at least* 40% probability of rape at two or more of the four timepoints (Figure 3).

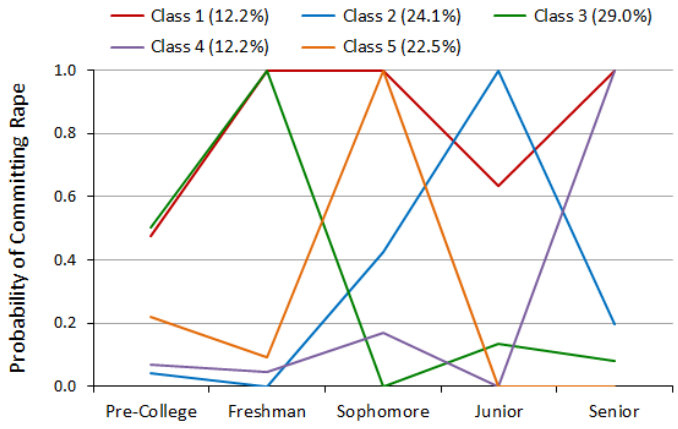

Figure 3. Five trajectory classes in the combined subsample of rapists (derivation and validation datasets), after release of the untenable smooth trajectory constraint of Swartout et al.'s model.

Importantly, these latent class models, including the *JAMA* model and the model lacking the unwarranted smooth trajectory constraint of the *JAMA* model, had estimation problems and most were under-identified.

An alternative model, logistic autoregression, tests the serial rapist assumption in a way that avoids the computationally intensive and heavily assumption-dependent latent trajectory modeling. This approach uses rape at one timepoint to predict rape at the next timepoint. *This* model converged on a solution without difficulty; it had ample statistical power, since all participants' data were used (rapist and non-rapists); and it minimized the impact of missing data, since estimates depended only on consecutive timepoints rather than the full study period. The fit of this model was excellent and each regression parameter was significant to the $p<.001$ level. Based on this model, the probability of a man raping was strongly predicted by whether he raped the previous year. In the derivation dataset, the odds of a man committing rape during Freshman year were more than 5 times higher if he had raped pre-college; the probability of raping during sophomore year was almost twice as high if a man had raped during freshman year; and the probability of raping during junior year was more than 2 times higher if a man had raped during sophomore year.

In the validation dataset, this autoregression model found that pre-college rapists are 6.7 times more likely than pre-college non-rapists to rape in the freshman year. Freshman year rapists were twice as likely as freshman non-rapists to rape in the sophomore year. Sophomore rapists were 2.6 times more likely to rape in their junior year, and junior rapists were 2.1 times more likely to rape in their senior year. In addition, these effects accumulate over time, further increasing the likelihood that man raping pre-college or as an underclassman will rape as an upperclassman (Figure 4).

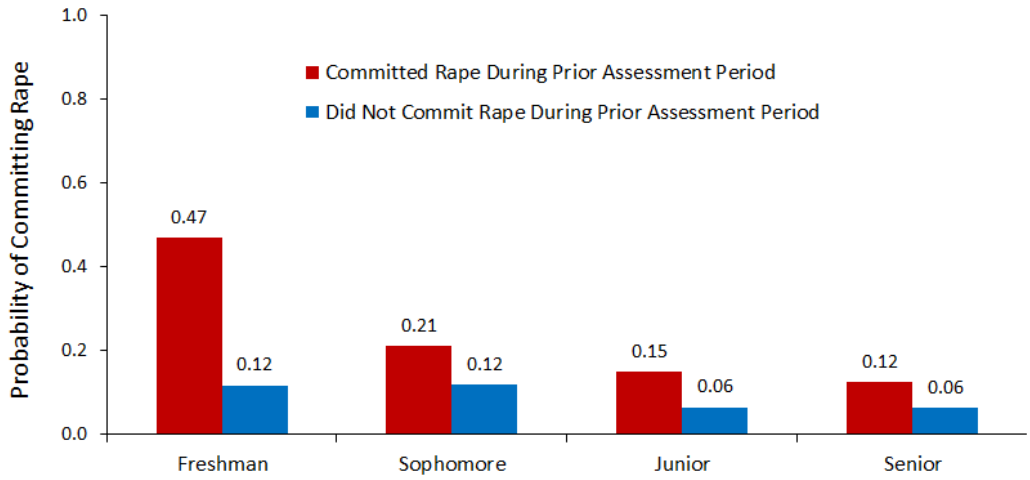

Figure 4. For the validation dataset, predicted probabilities of committing rape during one assessment period as a function of having committed rape during the prior assessment period.

Returning to the Primary Dataset: Problems with Data Integrity and Validity

Additional irregularities in the models reported in the *JAMA* article were found when analyzing the original public-use dataset from which Dr. Swartout and his colleagues created their derivation dataset; the *Longitudinal Study of Violence Against Women: Victimization and Perpetration Among College Students in a State-Supported University in the United States, 1990-1995,* ICPSR 3212). Unfortunately, the original case ID numbers were absent in the Mplus analysis dataset used for the article and, despite multiple requests, no file was provided to match the *JAMA* dataset to the public-use dataset.

Descriptively, the raw frequencies for each of three sexual assault items used by the JAMA study to operationalize rape clearly showed that the majority of college rapists rape multiple times – not only within timepoints but across them as well. (See Figure 5 for the most conservative possible estimation of the percentage of rapists who were repeat rapists, and additional pie charts that include perpetrators of attempted rapes in Dr. Hopper's presentation of simple frequency analyses.)

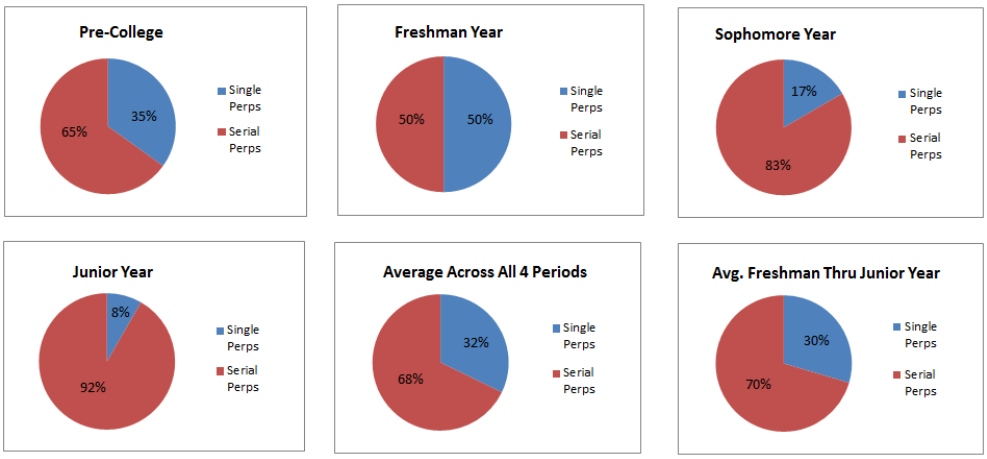

Figure 5. Percentages of rapists committing completed rapes who perpetrated more than one rape during that assessment period, assuming complete overlap (i.e., assuming that rape reported on any of the three different SES items measuring completed rape referred to the same rape(s) reported on the other two of those three SES items; see Dr. Hopper's presentation for findings when no overlap is assumed and when attempted rapes are included).

Using the data management programming code, exactly as it was provided by Dr. Swartout, the dichotomous rape variables were reconstructed (i.e., variables indicating yes vs. no for rape at particular timepoints). But the analysis using these variables did not replicate the results presented in the article, and the frequencies of the rape variable did not match the frequencies of the variables in the *JAMA* analysis dataset. Swartout's code did not draw directly from the raw frequency data that the respondents provided, but instead drew from recoded dichotomies of each of three sexual assault items operationalizing rape. A cross tabulation of the dichotomized sexual assault indicators with the raw data showed that in some cases missing data were assigned as "never" raped, which has the effect of underestimating the rapes committed.

There was a great deal of missing sexual assault data in the original dataset and the proportion of missing data in the original dataset did not match the missing data in the *JAMA* analysis dataset (*JAMA*: 25%, 47%, & 65%; Original: 25%, 60%, & 73% at Times 2, 3, & 4, respectively). In addition to study attrition, there was a substantial amount of unexplained missingness, where men participating in the study (completing at least some of the survey) did not provide data for any of the sexual assault indicators (i.e., percentages of participants within each time period with any missing data: 18%, 13%, & 29% at times 2, 3, & 4, respectively). The extent of missing data, the unexplained missing data patterns, and the missing data miscoded as never raped are serious issues that, even if the latent class analyses *were* valid, would undermine the validity of the conclusions reported in the JAMA article.

Choices that Underestimate Serial Rape and Serial Rapists

In addition to the miscoded missing values, several other decisions and actions by the authors resulted in underestimation of the numbers of serial rapes and serial rapists. For example, by creating a single dichotomous indicator of rape for each timepoint, the authors ignored multiple rapes within timepoint, thereby underestimating and "defining away" serial rapists (e.g., men who reported committing "more than 5" rapes freshman year, but then dropped out of the study, were not defined as serial rapists). By assigning a single rape indicator to multiple reports of rape across sexual assault items, the authors assumed that all responses to rape items refer to the same rape incident(s), that is, assumed total overlap across items, again underestimating serial rapists. Finally, the authors underestimated serial rapists by excluding attempted rapes.

Conclusion

The scientific integrity of the study described in the JAMA article is highly suspect at best. As it stands, the article is based on a study that uses erroneously coded data and underestimates both the prevalence of serial rapists in the data and the percentage of rapes those serial rapists report committing. Further, the models used in this study are based on untenable assumptions and ill-considered constraints. No reasonable and scientifically grounded debate over the "serial (campus) rapist assumption" can depend on this study.

Note: A full technical report is available here.

Appendix

Review of Dr. Tracy's Methodological Critique and Technical Report

I have reviewed and provided feedback on Dr. Tracy's document, "Methodological Critique: Executive Summary," and her lengthy technical report. I am impressed with Dr. Tracy's work and I believe that she has well demonstrated that the study by Swartout and colleagues, "Trajectory Analysis of the Campus Serial Rapist Assumption," has numerous problems, including data integrity, data coding, model estimation, model presentation, and comparability of the two data sets.

Hanno Petras, Ph.D.
Principal Researcher
Health & Social Development Program
1000 Thomas Jefferson Street, NW
Washington, DC 20007-3835
www.air.org/topic/social-development