Peep Show Establishments, Police Activity, Public Place, and Time: A Response to McCleary and Meeker

Daniel Linz
University of California, Santa Barbara

Bryant Paul
Indiana University

Mike Z. Yao
University of California, Santa Barbara

First, let us dispel one unseemly insinuation made by McCleary and Meeker (2006) in their response to our article—the notion that because this research was conducted as part of a lawsuit, it is not to be trusted and perhaps not even published, due presumably to some form of conflict of interest. McCleary and Meeker, previously employed by the government of San Diego, are concerned that our research has been funded by plaintiffs bringing a lawsuit against the city. In this suit, the plaintiffs alleged that their right of freedom of speech regarding sexual matters had been abridged in violation of the First Amendment. The research reported in our article was funded by plaintiffs as part of this lawsuit and this is acknowledged.

It is agreed among ethicists involved in research of all types that alleged conflicts of interest must be assessed within a specific factual context and not merely presumed to exist in the absence of facts supporting such a conclusion (National Institutes of Health, 2002). The specific context here—research conducted as part of a legal proceeding—renders the work more, not less, credible. Rather than being suspect, as McCleary and Meeker imply, the research conducted in this legal context should inspire greater confidence in the veracity of the study findings than most other situations.

As McCleary and Meeker know, we have been required by law to share all data and all written materials associated with our study with city officials. The calls-for-service (CFSs) data used in this study were generated by the city and made available to McCleary and Meeker. Had we been deposed about the study or had the case gone to court (the case was settled) and the findings presented there, we would have been required to present sworn testimony under oath. These features of research within a legal context are arguably more demanding than those associated with most peer-reviewed journals.

This said, we do not believe that our findings are methodological artifacts as McCleary and Meeker (2006) claimed. We believe there is ample precedent in the literature for the quasi-experimental design we have employed. Our study is one in a series of studies undertaken by local governments and academics that fail to find adverse secondary effects in the community for commercial enterprises involving sexual speech.

McCleary and Meeker's (2006) reply to our article is an attempt to recast our research in light of questionable statistical procedures and misleading analyses. Specifically, we will address the following problems with their reply: (a) their focus on an unreliable mean difference in CFSs between peep show and control areas illustrates a fundamental error in reasoning with statistics; (b) their application of retrospective power analysis is misleading; (c) they have inappropriately set up a system, wittingly or unwittingly, whereby the government can never lose when attempting to legislate against free speech; (d) their position is logically inconsistent with regard to the use of CFSs as a measure of crime and adverse secondary effects, and they mischaracterize the state of the field regarding the use of CFSs as a valid measure of secondary effects; and (e) their assertion that CFSs have dire consequences for significance tests is based on a set of unjustifiable assumptions.

The Null Hypothesis and Type II Errors

We showed that the mean difference between CFSs immediately surrounding peep shows and adjacent areas is not statistically significant. Thus, we concluded there is no "reliable evidence of differences in crime levels between the control and test areas" (Linz, Paul, & Yao, 2006). McCleary and Meeker asserted that this argument reflects a misunderstanding of significance tests. We could hardly disagree more. A "significant" difference is one for which the chance explanation has been rejected. Using conventional, agreed-upon statistical techniques, we were unable to reject the chance explanation.

McCleary and Meeker (2006) interpreted our failure to reject the null hypothesis as an implication of accepting H_0 to be true. This interpretation is unfounded. The goal of any null hypothesis significance test is to establish a conservative and scientific method to prevent researchers from accepting a difference found by chance (Cohen, 1994). We did not attempt to prove that crime-related activities are not higher in the immediate areas around peep show establishments than in other areas in San Diego. The goal of our

Address correspondence to Daniel Linz, Department of Communication and Law and Society Program, University of California, Santa Barbara, Santa Barbara, CA, 93106; linz@comm.ucsb.edu.

study was to test empirically whether or not the hypothesized secondary effects can be reliably found.

So eager were McCleary and Meeker to take this nonsignificant difference and call it a substantive finding that they ignored the basics of statistics. Statistical reasoning in our field is generally based on two parameters, measures of central tendency and variability. McCleary and Meeker, in a fundamentally misleading way, considered the mean but not the variance of the units being observed in our study. As Table 1 in our article clearly indicated (Linz et al., 2006), the difference in the mean numbers of CFSs is primarily driven by one outlier peepshow area. The number of CFSs (5,328) in the inner 1,000-foot areas of peepshow establishment I, which included a large shopping center, was three standard deviations higher than the group mean ($SD_{\text{peep show}}$ = 1,250.70) and almost four standard deviations higher than the grand mean (SD_{grand} = 1,022.89). When this outlier was eliminated from the analysis, there was no difference, statistically and literally, in the number of CFSs between peep show areas (M = 1,342.89, SD = 878.21) and control areas (M = 1,342.26, SD = 750.42), $t(35)$ = .002, p = .998. Further, even though this peep show was in an "outlier" area in terms of crime, there was no evidence that the adult business itself was a hotspot of criminal activity (see Linz et al., Table 2). Finally, we found that the 95% confidence interval for the mean difference between peep show areas and control areas contained zero (-468.26 < mean difference < 889), even when the outlier was included in the analysis. Upon completing these analyses, we are more committed than ever to our interpretation that there is no evidence of differences in crime levels between the control and test areas.

Retrospective Power Analyses

The entirety of McCleary and Meeker's (2006) discussion of power and the null hypothesis in their reply is moot once the outlier and confidence interval issues are addressed in the analyses reported above. However, there are other points we feel compelled to make regarding the post-hoc or retrospective power analyses they undertook.

We disagree with McCleary and Meeker and agree with those statisticians who do not believe the calculation of power retrospectively is necessary or even desirable in assessing the findings of a study (Goodman & Berlin, 1994; Lenth, 2001; Levine & Ensom, 2001; Thomas & Krebs, 1997). Further, conducting retrospective power analysis presents special problems and is misleading for the interpretation of findings in this particular area of research.

Despite its popularity, there are many statisticians who maintain that computing a power analysis after a study is completed only confuses the issue; it does not provide any additional insight beyond the results of the original statistical test (Hoenig & Heisey, 2001). The main technical point is that it can be shown that observed power is a decreasing function of the p value of the test; since we already know how to interpret p values, we don't need observed power. This pessimistic view of the value of post-hoc power analysis has been advanced by many researchers (Goodman & Berlin, 1994; Lenth, 2001; Levine & Ensom, 2001; Thomas & Krebs, 1997).

This one-to-one correspondence between the p value and power can be further illustrated by McCleary and Meeker's Figure 2 in their reply (2006). As seen in Figure 2, as the true difference between peep show and control areas decreases, the Type II or beta error rate increases. Their figure shows that as the true difference between peep show areas and control areas approaches zero (i.e., H_0), the probability of committing a Type II error approaches 1.

The fact that when the test is significant, the power is bound to be high, and when the test is non-significant, the power is bound to be low, has important implications for the debate about adult businesses in the community and adverse secondary effects. Null effects in this line of research will always be "underpowered."

Retrospective calculation of power statistics is based on three inputs: effect sizes (e.g., the size of the difference between control and test areas), n, and variances. For example, if we wanted to calculate, retrospectively, the power to detect the difference between the level of crime in an adult location versus a control (non-adult) area at the alpha .05 level, we need to consider the difference between the means of the adult and the control areas, the standard deviations for the two groups of observations and the number of observations, or the n (Lenth, 2000).

If the true difference between two means is zero, no matter how many observations and no matter what the variances for the two samples, power is .05; or, couched in terms of beta errors, if the difference is zero, this produces a false-negative rate of .95. Such a false-negative rate is too large to be ignored, according to McCleary and Meeker.

The implication for secondary effects crime analysis is that false-negative rates will always be unacceptably high as we approach a true null finding or zero difference between test (adult) and control (non-adult) means on crime effects; that is just the nature of a power calculation. McCleary and Meeker's (2006) application of power analysis to our data and in this area of research always stacks the deck against the plaintiff, who can never demonstrate that adult businesses are not associated with adverse effects. A finding of no difference between test and control areas, and thus no evidence of secondary effects for adult businesses, will, by virtue of the mathematics of power analysis, always be underpowered and thus statistically unacceptable to McCleary and Meeker. This is a convenient statistical stance to take when you are a consultant for a local government wishing to regulate adult businesses. But it is both meaningless and distracting to undertake the type of power analyses suggested by McCleary and Meeker if one's genuine interest is in determining if adult businesses are a problem for the community.

Making Alpha and Beta Errors Equal

We maintain that if one is asked to conduct a retrospective power analysis, as we were by reviewers of our manuscript, we should not be arbitrary in the undertaking and we should try to be fair about it.

Traditionally, researchers in most social science fields have accepted the notion that alpha should be set at .05 and power at 80% (corresponding to a beta of .20). This ratio of alpha to beta is based strictly on tradition. Researchers have asserted that a Type I error is four times as harmful as a Type II error (the ratio of alpha to beta is .05 to .20). This ratio is a matter of convention and nothing more. We are not convinced that this ratio makes any sense in the area of secondary effects research.

In a compromise power analysis like the analyses we undertook, the ratio q = beta/alpha specifies the seriousness of both types of errors. For instance, if alpha errors appear twice as serious as beta errors, then one can risk a beta error which is twice as large as alpha, thus q = beta/alpha = 2/1 = 2. This value is what you would then insert as the "beta/alpha ratio" in a compromise power analysis.

These choices depend on the value you associate with either outcome of the test. In our research, we considered the alpha error and the beta error to be equally serious problems. It is a serious societal problem to conclude there is a difference in crime between test and control areas when there is not (an alpha error), because making this mistake wrongfully deprives speakers of their First Amendment rights regarding sexual speech. It is equally serious to fail to detect a difference when, in fact, there is one (a beta error). To make this mistake would mean we would fail to err on the side of protecting the public from crime associated with the presence of the adult business. The value choices for this study are no different than so many debates in our society in which public safety is pitted against individual rights. McCleary and Meeker appear to want to err on the side of public safety. We believe that there is no way to distinguish between these two philosophical values; both are equally important and the errors associated with each should be set as equal.

Compromise power analyses are also recommended when the sample size is too small to satisfy conventional levels of alpha level and power, given the effect size, for reasons that are beyond a researcher's control (e.g., working with clinical populations; Erdfelder, Paul, & Buchner, 1996). This is the situation with which we were confronted in San Diego, where we were constrained by the fixed number of peep show establishments that exist in the city. In this situation, a compromise power analysis is the agreed-upon strategy.

It strikes us as histrionic to say that the practice of varying the critical values of alpha and beta from test to test "invites anarchy" and that it renders the idea of statistical conclusion validity meaningless. Indeed, the general authorities on Type II errors we cite advocated consideration of the relative seriousness of both errors (Cohen, 1988). The broader question is, therefore, why compromise analyses are missing in most research. It is only that non-standard results may occur; that is, results that are inconsistent with established conventions of statistical inference. It is hardly an invitation to anarchy to specify the relative seriousness of both types of errors. In fact, we would assert it is radically more anarchistic to ignore or recklessly manipulate the alpha levels in statistical analysis as McCleary and Meeker have done with our research.

Using Calls-for-Service as a Measure of Adverse Secondary Effects

It is necessary to draw attention to a fundamental flaw in McCleary and Meeker's (2006) logic concerning adverse secondary effects and calls-for service (CFSs) to the police. They note that we found that peep show areas, on average, had somewhat more CFSs than control areas—a finding we demonstrated to be unreliable by well-understood and agreed-upon methods. McCleary and Meeker maintained that any urban police department would nonetheless judge this difference in CFSs to be *substantively* significant. However, they also appear to take the position that by using CFSs, we render our study invalid, because this measure of crime is too flawed to be of use. How can Mccleary and Meeker square these two ideas? By their own assertion, the police should be concerned with a "substantively" significant finding in what is apparently an important measure of public safety. Yet we are to be faulted for using this same measure as an indicator of the impact of adult businesses in the community; their use alone is sufficient for invalidating our study. They cannot logically have it both ways. The implication of their position is that findings for CFSs are only worth noting when they go in the direction the government and McCleary and Meeker prefer.

McCleary and Meeker failed to acknowledge that calls-for-service are routinely used by municipalities and police in the secondary effects debate and in lawmaking concerning adult businesses. Several of the most frequently-cited secondary effects studies undertaken by municipalities and other communities as evidence of secondary effects used calls-for-service as an index of adverse secondary effects (e.g., Phoenix, Arizona, 1994; The Malin Group Study, Dallas, 1997; Austin, Texas, 1986; Fulton County, Georgia, 1997; Newport News, Virginia, 1996). These studies are relied upon by cities such as San Diego to justify ordinances limiting adult businesses. McCleary and Meeker have testified on behalf of the government concerning many city ordinances across the country. If calls-for-service are the basis of a fatal methodological error, then surely any city that relies on them would be advised to discount studies using them as justification for their ordinances. In our opinion, one reason city governments and their proxies fight so hard against the use of calls-for-service may be that their use in the 1997 Fulton County, GA, study was vetted by the U.S. 11^{th} Circuit Court of Appeals in an opinion which stands for the proposition that adult businesses are no more problematic for the community than other similarly-situated businesses that do not feature sex communication.

Contrary to McCleary and Meeker's (2006) assertions, there is also a substantial amount of published research using CFSs as an index of criminal activity and community disturbance. We have located several recent criminological publications not in the secondary effects area using CFSs data after a casual search of the criminological literature (e.g., Bursik, Grasmick, & Chamlin, 1990; Carter et. al., 2003; Cohn et. al., 2003; Rotton et. al., 2000; Verma, 1998; Warner & Pierce, 1988). The utility of calls-for-service as an indicator of crime in criminology research may be best summarized by criminologist Arvind Verma of the Department of Criminal Justice at Indiana University, Bloomington, who opined that the CAD (Computer Aided Dispatch) system—despite (some) limitations—is well-designed to ensure minimal human intervention and discretion, thereby creating a record of crime data with greater reliability. According to Verma, calls to police provide the most faithful and extensive account of what the public tells the police about crime or order maintenance.

CFSs are also a more comprehensive index of both police and crime activity, and thus a more comprehensive indicator of secondary effects, than are other indicators such as Uniform Crime Reports (UCRs). Neighborhood problems may occur that may not rise to the level of a UCR, but crime may trouble residents nonetheless. For example, it is impossible to measure so-called "blight," a highly critical aspect of the secondary effects debate with UCRs. These blighting incidents include police responses to public disorder and disturbance, drunken subjects, noise disturbances, and loud, disturbing parties. Other minor offenses, such as gambling law violations, harassment, subject pursuits, suspicious activity, suspicious vehicle threats, and trespassing, are measured by CFSs and are ignored by the UCR-based system.

In summary, the use of CFSs provides a more comprehensive and complete picture of the possible adverse secondary effects of adult businesses in a community. Finally, when considering the hours of operation questions at issue in San Diego, expenditure of police resources during late hours was an important factor in decision making about the regulation of adult businesses. It is universally agreed that police expenditure of resources is measured quite accurately by calls-for-service, making this measure especially useful in our study.

CFSs Have Dire Consequences for Significance Tests

McCleary and Meeker (2006) undertook a re-analysis of our *t* tests under the assumption that calls-for-service data are unreliable. They maintained that had we used a different measure of crime, we would have concluded that peep shows pose a significant public safety hazard. Specifically, McCleary and Meeker argued that since CFSs are not a perfect measure of crime, there should be an adjustment made to take into account that by their calculations, only 25-30% of CFSs are "actual crimes." Based on this reasoning, they applied a factor of $.25 < p < .30$ to the standard error term (the denominator of the *t*-test function). McCleary and Meeker argued that since only one fourth of the CFSs are "actual crimes" in their opinion, they must adjust our *t* statistics by a factor of 4. This analysis procedure is unconventional and statistically invalid. The essence of the problem with McCleary and Meeker's reanalysis of our *t* tests is that while they wished to adjust for measurement error, they instead adjusted for sampling error when they applied a new weight to the denominator of the test. We elaborate below.

First, let's assume that McCleary and Meeker's concern about the unreliable nature of CFSs is valid, and that their measure of "actual crime" is in itself free of measurement error. The conventional solution would be to adjust the means of the control group and test group, not to re-weight the standard error. In other words, if the error is a measurement error, then an adjustment needs to be made to the mean scores in the numerator of the *t*-test equation, not the standard error or denominator of the equation.

The implication of their formula is that the standard deviation of true crime is somehow smaller than the standard deviation of CFSs; that is, that the distribution of true crime and the distribution of CFSs are somehow different from each other. This is not defensible. McCleary and Meeker offered no justification of why that would be the case, and we cannot think of one.

If their argument is that CFSs are not an accurate measure of crime, then this measurement error should be embedded and randomly distributed in each raw score in our sample. The implication is that the same error would exist in both control and test groups. The effect of what McCleary and Meeker did in their formula was to shrink the standard error while allowing the mean differences between the control and test areas to remain unadjusted. This adjustment only to the denominator of the *t* statistic calculation may arguably be a solution for fixing sampling error, but not a sensible solution for measurement error, as was their intention.

In summary, in their re-weighting scheme, McCleary and Meeker (2006) confused two different types of error, making the scheme invalid as applied. The first type of error is measurement error, which is a methodological issue. The key question here is, "are CFSs an accurate measure of crime?" This source of error is distributed evenly within each unit in our sample and cannot be adjusted by altering the denominator in the final *t* statistic. The second type of error is sampling error, which is a statistical rather than measurement issue. It deals with the question, "is our sample of CFSs a good estimate for the CFSs within the population?" By shrinking the denominator of the *t*-test function, McCleary and Meeker implied that our sample of CFSs should be corrected as an estimate for CFSs in the population, but we don't think this is the point they are trying to make.

Linz et al. Quasi-Experimental Design

Finally, McCleary and Meeker (2006) implied that our quasi-experimental design was unique and therefore invalid. There is actually little difference in the design we employed and the quasi-experimetal designs employed by others who have conducted research in this area. The exception may be that our design was somewhat stronger than most because it employed a control area immediately adjacent to the test area, and therefore tended to equalize the areas in terms of demographic characteristics. This has the effect of increasing the sensitivity and accuracy of the comparisons between the control and peep show areas.

Conclusion

In our article, we concluded that there was no "reliable evidence of differences in crime levels between the control and test areas" (Linz et al., 2006). This is still our position. On the other hand, the tenuous assumptions and dubious statistical procedures suggested by McCleary and Meeker (2006) greatly accentuate the possibility of committing the troubling error of believing there is an effect when one is not present. The result of this recklessness may be to deprive citizens of their First Amendment right to freedom of speech guaranteed in the U.S. Constitution while not appreciably addressing the problem of crime in our communities.

References

Bursik, R. J. Jr., Grasmick, H. G., & Chamlin, M. B. (1990). The effect of longitudinal arrest patterns on the development of robbery trends at the neighborhood level. *Criminology,* 28(3), 431–50.

Carter, S. P., et. al. (2003, September). Zoning out crime and improving community health in Sarasota, Florida: "Crime prevention through environmental design." *American Journal of Public Health*, *93*(9), 1,442–1,445.

Cohn, E. G., et al. (2003, July/August). Even criminals take a holiday: Instrumental and expressive crimes on major and minor holidays. *Journal of Criminal Justice*, 31(4), 351–360.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1,304–1,312.

Cohen, J. (1994). The earth is round (p < .05). *American Psychologist*, *49*, 997–1,003.

Erdfelder, E., Paul, F., & Buchner, A. (1996). G*POWER: A general power analysis program. Behaviour. *Research Methods, Instruments, and Computers*, *28*, 1–11.

Goodman, S. N., & Berlin, J. A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of Internal Medicine, 121*(3), 200–206.

Hoenig, J. M., & Heisey, D. M. (2001). The abuse of power: The pervasive fallacy of power calculation for data analysis. *The American Statistician*, *55*, 19–24.

Lenth, R. V. (2000). *Java applets for power and sample size.* [Online]. Available http://www.stat.uiowa.edu/rlenth/Power/.

Levine, M., & Ensom, M. H. (2001). Post-hoc power analysis: An idea whose time has passed? *Pharmacotherapy, 21*(4), 405–409.

Linz, D., Paul, B., & Yao, M. Z. (2006).A secondary effects study of peep show establishments in San Diego. *The Journal of Sex Research*, *43(2)*, 182–193.

McCleary, R., & Meeker, J.W. (2006). Do peep shows "cause" crime? A response to Linz, Paul, and Yao. *The Journal of Sex Research*, *43*(2), 194–196.

National Institutes of Health. (2002, September 30). *The conflict of interest workshop summary.* Lister Hill Auditorium. Bethesda, MD: Author.

Rotton, J. et. al. (2000). Violence is a curvilinear function of temperature in Dallas: A replication. *Journal of Personality and Social Psychology*, *78*(6), 1,074–1,081.

Thomas, L., & Krebs, C. (1997). A review of power analysis software. *Bulletin of the Ecological Society of America, 78*(2), 128–139.

Verma, A. (1998). The fractal dimension of policing. *Journal of Criminal Justice*, *26*(5), 425–435.

Warner, B. D., & Pierce, G. L. (1988). Reexamining social disorganization theory using calls to the police as a measure of crime. *Criminology,* 31(4), 493–518.

Yuan, K. H., & Maxwell, S. (2005). On the post-hoc power in testing mean differences. *Journal of Educational and Behavioral Statistics*, *30*, 141–167.