Do Peep Shows "Cause" Crime? A Response to Linz, Paul, and Yao

Richard McCleary and James W. Meeker
University of California, Irvine

Government regulation of adult entertainment businesses, including peep shows, must be aimed at mitigating adverse secondary effects such as crime. To determine whether San Diego's regulations meet this Constitutional threshold, Linz, Paul, and Yao (2006) compared police calls-for-service (CFSs) in peep show and control areas. Finding no significant difference, they concluded that San Diego has no legitimate rationale for regulating any aspect of peep shows. We disagree not only with the Linz et al. finding, but also with the logical adequacy of their conclusion. Their finding is a methodological artifact, in our opinion, and their conclusion is a fallacy. Before explaining our opinion, however, we disclose two facts.

First, although Linz et al. acknowledged that their article was based on an earlier paper, they did not acknowledge a still earlier report (Linz & Paul, 2002) commissioned by the plaintiffs in a lawsuit (Mercury Books *v.* City of San Diego, U.S. District Court, Southern District of California, 00-CV2461). This omission does not necessarily invalidate the Linz et al. finding, but the article's ancestry may be a material fact in judging its "suitability, credibility, and validity for publication" in a peer-reviewed journal (Horton, 2004, p. 821).

Second, we were retained by the defendant in that lawsuit, the City of San Diego, to write a rebuttal report (McCleary & Meeker, 2003). We have no connection to the Community Defense Counsel of Scottsdale, Arizona, however, or to any "politically conservative religious-based organization devoted to the strict regulation or elimination of sex businesses" (Linz et al., 2006).

With these facts disclosed, the Linz et al. finding is a methodological artifact of their novel design. This empirical error is compounded by a common hypothesis testing fallacy. Although Cook and Campbell (1979, p. 30) discussed both errors as threats to statistical conclusion validity, neither is well understood, at least in reference to criminological research.

The Null Hypothesis and Type II Errors

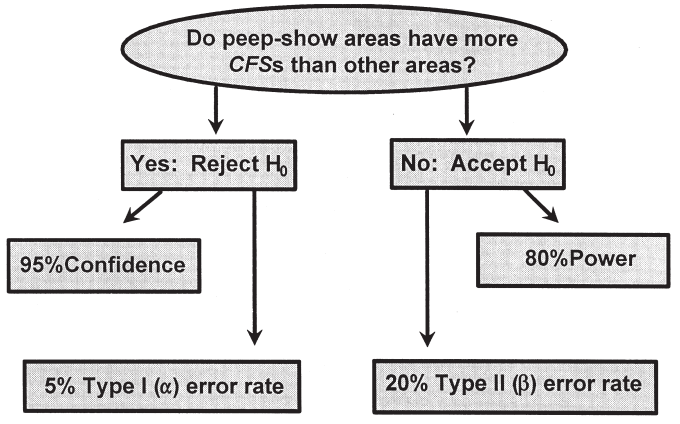
Linz et al. (2006) found that peep show areas had 210.4 (or 15.7%) more CFSs than other areas. Whereas any urban police department would judge a 15.7% difference in CFSs to be *substantively* significant, Linz et al. argued that the difference was not *statistically* significant and, thus, that there was no "reliable evidence of differences in crime levels between the control and test areas" (Linz et al.). This argument reflects a misunderstanding of the logic of hypothesis testing.

Figure 1 diagrams a test of the null hypothesis, H_0. A Yes answer to "Do peep show areas have more CFSs?" rejects H_0 but runs the risk of a Type I (or α-type) error. By a convention dating to Fisher (1925), H_0 is rejected only if the probability of a Type I error is $\alpha < 0.05$. Since $\alpha = 0.533$ in this instance, Linz et al. decided not to reject H_0. Had Linz et al. stopped here, their conclusion might be defensible. But Linz et al. argued further that not rejecting H_0 implies that H_0 is true – that "crime levels" around peep shows are no higher than the levels in other areas.

The risk of a Type II (or β-type) error invalidates this argument. By a convention dating to Neyman and Pearson (1928), H_0 is accepted only if the probability of a Type II error is $\beta < 0.2$. Figure 2 plots the Type II error function for the parameters reported (*i.e.*, $n_1 = n_2 = 19$ areas, $\alpha = .05$, and s = 304.5 CFSs. The function was calculated with *PASS 6.0* [Hintze, 2001].). Since the probability of a Type II error for the 15.7% difference is $\beta = .508$, there is no empirical basis for believing that H_0 is true. In this instance, H_0 is neither rejectable (because $\alpha \geq 0.05$) nor acceptable (because $\beta \geq .2$); the test of H_0 is inconclusive.

Although Linz et al. (2006) acknowledged that their Type II error rate exceeded the conventional critical level,

Figure 1. Logic of the null hypothesis test.

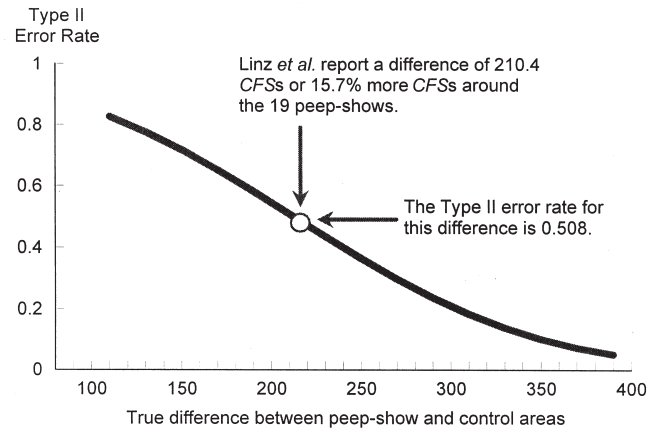


Note. By convention, H_0 is rejected if $\alpha < 0.05$. If $\alpha > 0.05$, H_0 can be accepted if $\beta < 0.2$; but otherwise, the test of H0 is inconclusive.

Note. Although much of the research described here was conducted while we were retained by the City of San Diego, the opinions expressed are our own. Christopher Burton provided research assistance.

Address correspondence to Richard McCleary, School of Social Ecology, University of California, Irvine, Irvine, CA, 92697; e-mail: mccleary@uci.edu.

Figure 2. Type II error function.



Note. The 15.7% difference reported by Linz et al. (2006) has a Type II error rate of $\beta = .508$, vs. the conventional value of $\beta < .2$ required to accept H_0.

they argued that the convention is unrealistic. Citing the authority of Erdfelder, Faul and Buchner (1996), they proposed to test H_0 with a critical value of $\beta < 0.44$ for $\alpha < 0.44$ (vs. $\beta < 0.2$ for $\alpha < 0.05$).

Since $\beta = .508$ (see Figure 2), even if one agreed with these arbitrary critical values, the test of H_0 would be inconclusive. Linz et al. missed this point. But we do *not* agree with the arbitrary critical values of β and α proposed by Linz et al. Other than the limited circumstances described by Erdfelder, Faul and Buchner (1996), varying the critical values of α and β from test to test invites anarchy, rendering the idea of statistical-conclusion validity meaningless. Indeed, the general authorities on Type II errors cited by Linz et al. (Cohen [1988], pp. 3-4; Lipsey [1990], pp. 38-40) endorse the conventional critical levels of $\beta < 0.2$ for $\alpha < 0.05$.

Police CFSs as a Measure of Crime Risk

A corollary threat accrues from the use of CFSs to measure crime. For each 911 call, a police dispatcher records the time, location, etc. and forwards these data to a responding patrol unit. The resulting CFS record gives a crude picture of police service demand and, indirectly, of crime risk. The correlation between crime risk and CFSs is weak at best, however.

Table 1 demonstrates one source of error in the crime-CFSs correlation. Of the 607,903 CFSs analyzed by Linz et al., 17% were unfounded, duplicated, or cancelled. More than half (55%) were disposed of without report; patrol units could find no victim, witness, or evidence of a crime. Only 20% resulted in an arrest or report. Furthermore, many crimes are not reported through 911 calls but are discovered through routine patrolling or special unit activity. These crimes are not counted as CFSs and constitute another source of error.

Table 1. CFSs by Final Disposition

Disposition	Frequency	%
Arrest or Report (Incident)	119,250	19.6
Cancelled, Duplicated, or Unfounded (No Incident)	104,443	17.2
No Report Filed	332,014	54.8
Other or Unknown	52,196	8.3
Total	607,903	100

Whatever the source, measurement error in CFSs has dire consequences for significance tests. Linz et al. tested H_0 with a *t*-statistic. Using their numbers, we write this as

$$t = \frac{1552.6 - 1342.2}{334.5} = .629\,\rho \text{ (where } \rho = \sigma_{CFS} / \sigma_{CRIME})$$

Linz et al. assumed that $\rho = 1$, or that CFSs are a perfectly reliable measure of crime. Under that assumption, the value of $t = .629$ is too small to reject H_0. CFSs are *not* a perfect measure of crime, however. McCleary and Meeker (2003) estimated that ρ lies in the interval $.25 < \rho < .30$. Consequently, t lies in the interval $2.10 < t < 2.52$. These t-values reject H_0. Had Linz et al. used a more reliable measure of crime, they would have concluded that peep shows pose a significant public safety hazard.

Conclusion

Both threats to statistical-conclusion validity stem from the use an unorthodox measure of crime risk. Linz et al. (2006) justified this use of CFSs with this claim:

> Many criminologists have employed ... [CFSs] to police dispatch centers to measure crime at the address (Sherman et al., 1989), neighborhood (Bursik et al., 1990; Warner & Pierce, 1993) and city (Bursik & Grasmick, 1993) levels. According to its proponents, the CFSs measure offers a more valid description of aggregate levels of crime than either police records collated in the FBI's Uniform Crime Reports (UCR) or victimization data collected in the National Crime Survey (Linz et al., 2006).

When computerized 911 systems were introduced, criminologists experimented with calls-for-service. The articles cited by Linz et al. are from this historical era. Experimentation soon revealed the errors in CFSs (Klinger & Bridges, 1998).

Modern criminologists do *not* use CFSs to measure crime or crime risk. In 2000-2004, the official journals of the two national criminology professional associations, *Criminology* and *Justice Quarterly*, published 245 articles. Of the 100 that analyzed a crime-related statistic, 98 analyzed UCRs and/or surveys; two analyzed CFSs, but even in these two cases, CFSs were not used to measure crime or crime risk.

Since Linz et al. are not criminologists, they were unaware of this convention and its rationale. Criminologists do not use CFSs to measure crime or crime risk because relative to the conventional measures, CFSs are unreliable. Tests of H_0 are biased consequently in a way that supports the plaintiffs' argument.

These threats to statistical-conclusion validity, in our opinion, are poorly understood outside criminology. We do not endorse other aspects of the Linz et al. quasi-experimen-

tal design, however. Furthermore, although we used only one of several statistical results to illustrate the consequences of uncontrolled threats to statistical-conclusion validity, our point applies to all statistical results reported by Linz et al.

REFERENCES

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences, 2nd Ed.* Englewood Cliffs, NJ: L. E. Erlebaum Associates.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston, MA: Houghton-Mifflin.

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

Fisher, R. A. (1925). *Statistical methods for research workers*. Edinborough: Oliver & Boyd.

Hintze, H. (2000). *Power analysis and sample size, Version 6* (PASS). Kayesville, UT: NCSS, Inc.

Horton, R. (2004). A statement by the editors of *The Lancet*. *The Lancet, 363*, 999.

Klinger, D., & Bridges, G. S. (1997). Measurement errors in calls-for-service as an indicator or crime. *Criminology, 35*, 529–541.

Linz, D., & Paul, B. (2002). *A secondary effects study relating to hours of operation of peep show establishments in San Diego, California*. Report to the San Diego City Attorney's Office.

Linz, D., Paul, B., & Yao, M. Z. (2006). Peep show establishments, police activity, public place, and time: A study of secondary effects in San Diego, California. *The Journal of Sex Research, 43(2)*, 182–193.

Lipsey, M. (1990). *Design sensitivity: Statistical power for experimental research.* Thousand Oaks, CA: Sage.

McCleary, R., & Meeker, J. W. (2003, March 12). *A methodical critique of the Linz-Paul report: A report to the San Diego City Attorney's Office*.

Neyman, J., & Pearson, E. S. (1928). On the use and interpretation of certain test criteria for purposes of statistical inference. *Biometrika, 20A*, 175–240.