

**A METHODOLOGICAL CRITIQUE OF THE LINZ-PAUL REPORT:
A REPORT TO THE SAN DIEGO CITY ATTORNEY'S OFFICE**

Richard McCleary, Ph.D.
James W. Meeker, J.D., Ph.D.

March 12, 2003

Antonio Ohe, B.A., Christopher Hernandez, and James A. Chiampi provided research assistance for this report, including library searches and data management.

Table of Contents

	Executive Summary	
I.	Introduction	1
	A. Overview of the Methodological Flaws in the Linz-Paul Report	1
	B. Outline of this Report	5
II.	Measurement Problems in the Linz-Paul Report	8
	A. The Fatal Flaw: <i>CFSs Do Not Measure Crime</i>	9
	B. <i>CFS Addresses Are Not “Crime Event” Addresses</i>	13
	C. “Raw” <i>CFSs Ignore Crucial Differences in Public Safety Risk</i>	15
	D. Burglary <i>CFSs Illustrate the Threats to Validity</i>	17
	E. The 2-6 AM Period	
III.	The Linz-Paul Statistical Analyses	25
	A. Statistical Power	25
	B. How to Raise the Statistical Power of an Analysis	30
IV.	Design Issues	32
	A. How Linz and Paul Chose Peep-Show and Control Areas	34
	B. How Peep-Show and Control Areas Should Have Been Selected	36
	C. “Fishing” in the Linz-Paul Report	38
V.	Conclusions	40
	A. The <i>Daubert</i> Standards	
	Appendix: Technical Details of the Linz-Paul Dataset	45

Tables and Figures

Table IIA	Regression of <i>CFSs</i> on Population and Crime
Table IIB	<i>CFSs</i> by Final Disposition
Table IIC	<i>CFSs</i> by Incident-Type
Table IID	Burglary <i>CFSs</i> by Source and Disposition
Figure II	Crime- <i>CFS</i> -Response Timeline
Figure IIIA	False-positive and False-negative Rates
Figure IIIB	False-negative Rates for the Linz-Paul Report
Figure IVA	The Linz-Paul Experimental and Control Areas
Figure IVB	Distribution of <i>CFSs</i> Across the Areas
Table IVC	Designs of Three Secondary Effect Studies
Figure A1	Locations of the 19 Peep-Shows
Figure A2	“The Crypt,” 4094 30 th Street
Figure A3	Overlapping Peep-Show Sites (Central)
Figure A4	Overlapping Peep-Show Sites (Balboa Ave)

I. Introduction

The expert witness report by Linz and Paul¹ analyzes a small subset of calls-for-service (CFSs) made to the San Diego Police Department (SDPD) during 1997-2001.

Based on analyses of these data, Linz and Paul find:

- (1) That “crime levels” at and around 19 peep-shows are no higher than “crime levels” in comparable areas of San Diego; and
- (2) That the “amounts of crime” occurring between 2 AM and 6 AM in San Diego are not higher than the amounts occurring in other 4-hour periods.

Based on these findings, Linz and Paul conclude that the City of San Diego has *no* valid public safety rationale for regulating any aspect of peep-shows, much less for regulating the hours in which peep-shows can operate.

To assess the validity of the Report’s findings and conclusions, we have read the Linz-Paul Report and other relevant literature. We have reanalyzed the original data and analyzed auxiliary data. Based on our readings, on our analyses, and on our experience in this field, it is our opinion that the Linz-Paul Report’s methodology fails to meet the standards of scientific rigor mandated by *Daubert*.² Accordingly, the Report’s findings and conclusions are invalid and should be *inadmissible* under Rule 104(a).

I.A. Overview of the Methodological Flaws in the Linz-Paul Report

The shortcomings of the Linz-Paul Report span all three elements of scientific

¹Daniel Linz, Ph.D. and Bryant Paul, M.A. *A Secondary Effects Study Relating to Hours of Operation of Peep Show Establishments in San Diego, California*. September 1st, 2002. In the text, we will refer to this report as “the Linz-Paul Report,” as “Linz and Paul,” or as “the San Diego study.”

² *Daubert v Merrell Dow Pharmaceuticals* 509 US 579 (1993). Hereafter we refer to this U.S. Supreme Court decision as “*Daubert*.”

method (or “methodology”), including (1) the measures of public safety collected for the study; (2) the statistical models used to analyze the public safety measures; and (3) the quasi-experimental design used to interpret the analytic results. This ordering reflects the relative seriousness of the methodological flaws. Some flaws are fatal in that they invalidate *all* of the Report’s findings. While not necessarily fatal, other flaws invalidate specific findings.

(1) Measurement problems. The most serious flaws by far involve the use of *CFSs* to measure public safety risk. There is virtually no precedent in the criminology literature generally or in the secondary effects literature specifically for using *CFSs* to measure crime. A review of national criminology journals over the last three years, for example, finds not even one published article where *CFSs* are used to measure crime. A review of more than a dozen unpublished secondary effects studies, including those by the authors of the San Diego study, leads to the same conclusion.³ Every study that we have reviewed, whether published or unpublished, has used Uniform Crime Reports (*UCRs*) to measure public safety risk. Like all California police departments, moreover, the SDPD collects and disseminates *UCRs*.

Since *CFSs* do *not* measure crime, and since all of Linz-Paul Report’s findings and conclusions are couched in terms of crime, this flaw invalidates *all* of the Report’s findings and conclusions. Although one need go no further, corollary methodological

³ This includes prior reports by Linz and Paul. *E.g.*, in a 2001 report on adult businesses in Ft. Wayne, IN Linz and Paul use *arrests* to measure crime. Their rationale for using arrests: “Only those incidents for which arrests were made and not based on unfounded charges were included in the study.” (*Measurement of Negative Secondary Effects Surrounding Exotic Dance Nightclubs in Fort Wayne, Indiana*, p.14).

flaws, involving mistaken assumptions about *CFSs*, invalidate specific conclusions.

These include:

- (1) Assuming that *CFS* addresses are the locations of crimes;
- (2) Assuming that the precipitating incident and *CFS* occur simultaneously;
- (3) Assuming that *CFSs* pose identical public safety risks;

But in fact, each of these assumptions is mistaken. The *CFS* data obtained by Linz and Paul from the SDPD are insufficient to specify the locations, times, or natures of the precipitating incidents (“crimes”). In detail,

- (1) The address recorded on a *CFS* is the caller’s address. If X calls the SDPD to complain about a disturbance at Y’s house, for example, the address on the *CFS* belongs to X.
- (2) The time recorded on a *CFS* is the time that the SDPD received the call. Nearly 40 percent of the *CFSs* in the Linz-Paul dataset were precipitated by property crimes, primarily burglary and auto theft. These crimes are often discovered and reported hours (or even days) after they occur.
- (3) Nearly half of the *CFSs* in the Linz-Paul dataset posed no public safety threat at all (because they were unfounded, cancelled, duplicated, false alarms, *etc.*). Of the 20 percent that were precipitated by “crimes,” public safety risk ranged from low (*e.g.*, vehicle recovery) to high (*e.g.*, crime-in-progress). The raw, total *CFS* index used by Linz and Paul treated all of these *CFSs* equally.

These incorrect assumptions invalidate (1) the Linz-Paul Report’s “hot-spot” findings;

(2) the findings about “crime” in the 2-6 AM period; and indeed, (3) *all* other findings.

(2) Statistical problems. If one could ignore the fatal methodological flaws posed by the use of *CFSs* to measure “crime,” the manner in which these data were analyzed poses another fatal methodological flaw. Although a superficial reading of the Linz-Paul Report leaves the impression that the peep-show and control areas generated equal numbers of *CFSs*, a more careful reading reveals that the peep-show areas generated 15.7 percent more *CFSs* than the control areas. Noting that this difference is “statistically nonsignificant,” Linz and Paul conclude that the peep-shows generate no more *CFSs* than other businesses. The Report fails to disclose that the probability associated with this conclusion is only 50 percent. This means that the Linz-Paul statistical analyses have roughly the same validity as a coin-flip.

(3) Design problems. The problems of statistical analysis reflect fundamental shortcomings of the Linz-Paul Report’s “quasi-experimental” design. Compared to other secondary effects studies, the quasi-experimental design of the San Diego study is idiosyncratic in at least two respects:

(1) Whereas virtually all prior studies look for secondary effects in circular areas around a peep-show, the San Diego study confines the search to a strip on either side of a peep-show but on the same street.

(2) Whereas virtually all prior studies use comparable businesses or areas for controls, the San Diego study uses contiguous strips of addresses.

Per se, these design idiosyncracies pose serious threats to internal validity the Linz-Paul Report’s conclusions. But worse, the joint debut of so many design innovations

raises the specter of “fishing.”

In the jargon of scientific research, “fishing” refers to the practice of replicating a study several times. With just a few variations in measurements, statistical models, and quasi-experimental designs, a cynical researcher can capitalize on chance to produce any desired result. “Fishing” need not imply dishonesty or cynicism. On the contrary, scientific method recognizes that “fishing” can occur without the researcher’s intent or awareness. In experimental research, science controls “fishing” by means of explicit design structures, including placebos, blinding, *etc.* In quasi-experimental research, where these structures cannot be used, “fishing” is controlled by means of rigidly enforced design conventions. Departures from convention must be explained and justified. If they are not explained, the critical scientific reader must assume that findings and conclusions are an artifact of “fishing.”

I.B. Outline of Our Report

The goal of this report is to explicate the methods used in the Linz-Paul Report, including measurement, data analysis, and quasi-experimental design. Our explication will demonstrate that the Linz-Paul Report’s methods do not meet current standards of validity as the standards apply to social science generally and criminology specifically. By implication, the Linz-Paul Report’s methods also fail the standards mandated by *Daubert*. With respect to *Daubert*, shortcomings in any of the three methodological areas is sufficient to invalidate the Linz-Paul Report’s findings and conclusions.

The salient methodological flaw in the Linz-Paul Report is the use of total, raw *CFSs* to measure “crime.” The correlation between *CFSs* and conventional measures

of crime is exceptionally weak. In Section II, we demonstrate this fact by estimating the correlation between San Diego's 2001 Uniform Crime Reports (UCRs) and *CFSs* for the SDPD's 103 neighborhood beats. The small estimated correlation suggests that 90 percent of the variance in *CFSs* is due to factors other than crime. This is only one reason why criminologists do not use *CFSs* to measure crime. What do criminologists use to measure crime? Crime itself.

After demonstrating the weak *CFS*-crime correlation, we detail methodological problems posed by the Linz-Paul Report's misuse of *CFSs*. Because the addresses assigned to *CFSs* record the location of complainants, for example, *CFSs* cannot be used to analyze "hot spots." The Report's conclusion that the number of *CFSs* to 19 peep-show addresses is lower than the number of *CFSs* other addresses, thus, says nothing about the public safety risks of these businesses.

Likewise, because *CFSs* and their precipitating incidents ("crimes") are not simultaneous events, the temporal distribution of *CFSs* says little about the temporal distribution of public safety risk (or even, the demand for police service) in San Diego. An incident that occurs at noon, for example, may result in a *CFS* at 4 AM. The Linz-Paul Report's conclusion that crime is lower during the 2-6 AM period is invalidated by the weak temporal relationship between precipitating incidents and *CFSs*.

Finally, all of the Linz-Paul Report's conclusions are based on analyses of raw, total *CFSs*. Raw *CFSs* ignore germane differences in the precipitating incidents and outcomes. Only 20 percent of the *CFSs* analyzed by Linz and Paul resulted in a crime report or an arrest, for example, while another 20 percent were cancelled, duplicated, or

unfounded. This methodological problem is aggravated when raw *CFSs* are counted for a summary index. A *CFS* precipitated by an armed-robbery-in-progress and a *CFS* precipitated by a loud party are very different things. Counting them as two equivalent *CFSs*, as the Linz-Paul Report does, invalidates any conclusions about public safety secondary effects.

In Section III, we address the idiosyncratic statistical models used in the Linz-Paul Report to analyze *CFSs*. The most salient statistical problem arises when the Report concludes that areas around the 19 peep-shows and comparable control areas have the same number of *CFSs*. Now in fact, Linz and Paul find that peep-show areas generate 15.7 percent more *CFSs* than control areas; but they claim that the difference is “statistically nonsignificant.” Though quoting *Daubert* on the importance of an “error rate,” Linz and Paul fail to report an error rate for the statistical test underlying their conclusion. In fact, the error rate for their test is approximately 50 percent. Having the same validity as a coin-flip, the Linz-Paul Report’s conclusions are “much ado about nothing.”

The statistical problem described in Section III is a function of the study’s quasi-experimental design which, in many important respects, is unique. It is unprecedented in both the published criminology literature or in the unpublished secondary effects literature. In Section IV, we describe the relevant design idiosyncracies and discuss the problem of “fishing.” When a design can be picked from even a modest menu of options, the “significance” or “nonsignificance” of a finding is meaningless. The sheer number of design idiosyncracies in the Linz-Paul Report are sufficient to invalidate the

Report's conclusions.

In the concluding Section V, we summarize and explain how the methodological criticisms relate to the *Daubert* standards. Problems of measurement (Section II), data analysis (Section III), and design (Section IV) are sufficient to invalidate any and all of the Report's findings and conclusions.

II. Measurement Problems in the Linz-Paul Report

Measurement is the *sine qua non* of science. If public safety risk (or crime) cannot be measured adequately, it cannot be studied scientifically. The adequacy of a measurement is summed up in the properties of *reliability* and *validity*.⁴ To illustrate reliability, Linz and Paul counted 2,551 *CFSs* to University Avenue addresses within 2000 feet of Midnite Books Hillcrest.⁵ A recount would probably not yield the same number because even simple counts vary randomly.⁶ If the count-recount difference is reasonably small and random, however, the measurement is reliable and adequate for scientific research. When measurement errors are *nonrandom*, on the other hand, they cannot be "averaged out." They accumulate into biases.

The property of validity is associated with nonrandom measurement errors. In

⁴ For definitions, see H.M. Blalock's *Measurement and Conceptualization in the Social Sciences* (Sage, 1982). See also *Quasi-Experimentation: Design and Analysis Issues for Field Settings* by T.D. Cook and D.T. Campbell (Houghton-Mifflin, 1979). Cook and Campbell discuss reliability and validity of measurement under the concept of "construct validity."

⁵ Linz and Paul, Table 1, p. 20.

⁶ In his classic *On the accuracy of economic observations, 2nd Edition* (Princeton: Princeton University Press, 1965), Nobel laureate O. Morgenstern expressed this idea as "*Incipit numerare, incipit errare!*." Begin to count, begin to make mistakes!

general, the reliability of a measurement can be increased by design or by statistical adjustment. The validity of a measurement is fixed and cannot be increased. The validity of crime measurements depends generally on three factors:

- (1) The particular crime incidents that are counted (or not counted);
- (2) The method used to count crime incidents; and
- (3) The manner in which crime incidents are aggregated into a composite index

Linz and Paul use a subset of raw police *CFSs* to measure crime. This idiosyncratic choice of measures has no precedent and *per se* invalidates their conclusions.

II.A. The Fatal Flaw: *CFSs* Do Not Measure Crime

Throughout their Report, Linz and Paul speak of “*CFSs*” and “crimes” as if these two terms were synonymous. For example, in describing their methodology:

The calls for service were then plotted using a computerized mapping program. All calls were plotted based on the longitude and latitude co-ordinates provided by the city’s crime analyst. Comparisons of the number of crime incidents were then made for the inner and outer areas. Comparisons were also made for the number of crime incidents occurring between the hours of 2 a.m. and 6 a.m. and those occurring throughout the entire 24 hours of the day.⁷

Again, we had requested the city to produce data for all service calls for alleged crimes occurring within the specified areas. The City informed us that its CAD database was the only place it kept track of service calls for crimes by category, and it was giving us all of the data for all the crime service-calls kept in its CAD database.⁸

In reporting their results, Linz and Paul continue to use “*CFSs*” and “crimes”

⁷ Linz and Paul, p. 10, lines 3-8.

⁸ Linz and Paul, p. 11, lines 18-22.

interchangeably. For example:

... the amount of crime within the inner and outer areas is nearly identical. For 10 of the peep show locations, crime incidents are *higher* in the inner 1000-foot areas than in the outer areas. For nine of the locations, crime is *lower* in the inner areas compared to the outer areas.⁹

... criminal activity is *not* disproportionately *greater* in the areas surrounding the peep show establishments during these hours. In fact, the study data shows that criminal activity at peep show establishments in San Diego is proportionately *less* during the 2 a.m. to 6 a.m. time period.¹⁰

The salient fatal flaw in the Linz-Paul Report is that “CFSs” and “crimes” (or crime-like incidents) are *not* strongly correlated. One reason for the weak CFS-crime correlation is that CFSs exclude crimes that are discovered through internal channels. Excluded crimes include:

- ◆ Crimes discovered through routine patrolling;
- ◆ Crimes discovered through directed (or proactive) patrolling;
- ◆ Crimes discovered by specialized unit activity.

Another reason for the weak CFS-crime correlation is that CFSs include calls that are not really crime incidents. These include:

- ◆ Duplicate and cancelled CFSs;
- ◆ CFSs precipitated by false alarms;¹¹

⁹ Linz and Paul, p. 14, lines 10-12.

¹⁰ Linz and Paul, p. 16, lines 20-22.

¹¹ Alarm-initiated CFSs in the Linz-Paul dataset have “FINAL_TYPE” values ending in “A.” Most of the alarm-initiated CFSs in the data set are precipitated by robberies (211A) or burglaries (459A). Assuming that valid alarm-initiated CFSs result in an arrest, a report, or a recovery, the false alarm rates for burglary and robbery were 99.1 and 98.6 percent respectively.

◆ “All Units” CFSs.

In fact, as we demonstrate below, most of the 607,903 CFSs analyzed for the Linz-Paul Report fall into this category; less than 20 percent of the CFSs were precipitated by crime incidents. It is no surprise then, that in San Diego, as in most large urban areas, CFSs and crimes are only weakly correlated.

Given its nominal purpose— to determine whether peep-shows have criminogenic properties – the Linz-Paul Report should have analyzed crimes, not raw CFSs. The vast criminology literature has not even one precedent for using raw CFSs to measure crime. Criminologists invariably measure crime with Uniform Crime Reports (UCRs) or sample surveys of victims.¹² The smaller, unpublished secondary effects literature has also typically used UCRs or analogous crime statistics.¹³ This is not to say that CFSs are not a useful statistic. On the contrary, all urban police departments, including the

We return to this issue in Section IID below.

¹² See, e.g., *Measuring Crime* (D.L. MacKenzie, P.J. Baunach, and R.R. Roberg, State University of New York Press, 1990). The criminological literature is consistent on this point. A search of four national criminology journals (*Justice Quarterly*, *Criminology*, *Criminal Law and Criminology*, and *Journal of Quantitative Criminology*) for the last three years found not one study that used CFSs to measure crime.

¹³ This includes studies conducted by Linz and Paul (e.g., *Measurement of Negative Secondary Effects Surrounding Exotic Dance Nightclubs in Fort Wayne, Indiana*, February 13, 2001) or others at their commission (e.g., *Are Adult Dance Clubs Associated with Increases in Crime in Surrounding Areas? A Secondary Crime Effects Study in Charlotte, North Carolina* by K.C. Land, J.R. Williams, and M.E. Ezell). The Fort Wayne study uses arrests; the Charlotte study uses crimes. An apparent exception are “reports” that list CFSs to liquor license addresses (e.g., *A Study of CFSs to Adult Entertainment Establishments which Serve Alcoholic Beverages* by Capt. Ron Fuller and Lt. Sue Miller, Fulton County, GA Police Dept., June 13th, 1997). CFSs are traditionally used in liquor license reviews. The “report” is nothing more than a computer print-out, however.

SDPD, collect these data for use in budgeting.¹⁴ But no police department uses *CFSS* to measure crime or public safety. Criminologists and police departments alike use *crime* to measure *crime*.

TABLE IIA - REGRESSION OF <i>CFSS</i> ON POPULATION AND CRIME				
Variable	Weight	t	F_{22,99}	Correlation
Constant	281.55	1.15	16.03	
Population	26.16	3.27		
Part I UCR Violent Crimes	7.68	4.38		r = .249
Constant	926.47	1.09	8.19	
Population	64.95	2.11		
Part I UCR Property Crimes	7.34	6.98		r = .294

San Diego's 2001 UCRs demonstrate the inadequacy of *CFSS*s as a measure of crime. The SDPD publishes Part I UCRs by 103 beats (or neighborhoods). We downloaded these data from the SDPD web site and used them as independent (or explanatory) variables in two multiple regression models. The regressions, reported in Table IIA, were then used to estimate the correlation between *CFSS*s and crimes. In the first model, the *CFSS*s corresponding to Part I UCR "violent" crimes (homicide, robbery, rape, and assault) were regressed on the actual numbers of UCR crimes recorded for

¹⁴ These valid uses of *CFSS*s are discussed in undergraduate policing texts. See, e.g., *Police Administration* by O.W. Wilson and R. McLaren (McGraw-Hill, 1978); *Police and Society* by R.R. Roberg, J. Crank and J. Kuykendall, (Wadsworth, 1999) or *Police Administration* by C. Swanson, L. Territo, and R. Taylor (Macmillan, 1993). All of these texts make the same points that we make about *CFSS*s.

the beats in 2001, the beats' 2001 population, and dichotomous indicator variables for each of the 22 SDPD districts. The second model was identical except that Part I UCR "property" crimes (auto theft, larceny, and burglary) were used.

The parameter estimates in Table IIA show that, for both "violent" and "property" crimes, *CFSs* and Part I UCRs are indeed correlated. This is not surprising. Virtually all weak crime indicators are correlated. Readers who are not familiar with crime statistics may be surprised (or even shocked) at how relatively weak these correlations are, however. Specifically

- ◆ For "violent" crimes, the *CFS*-UCR correlation (.249) means that only 6.2 percent of the variance in *CFSs* among beats is explained by crime; *93.8 percent is due to factors unrelated to crime.*
- ◆ For "property" crimes, the *CFS*-UCR correlation (.294) means that only 8.6 percent of the variance in *CFSs* among beats is explained by crime; *91.4 percent is due to factors unrelated to crime.*

This is only one reason why criminologists do not use *CFSs* to measure crime.¹⁵ Linz and Paul ignore this *dictum* and, as a result, their conclusions are invalid.

II.B. *CFS* Addresses Are Not "Crime Event" Addresses

Using *CFSs* to measure crime is *per se* a fatal flaw. But suppose that Linz and Paul had somehow been able to select out the minority of the 607,903 *CFSs* in their database that were precipitated by a crime; and that they had also been able to include

¹⁵ For another reason, see our "Uniform Crime Reports as organizational outcomes." (*Social Problems*, 1982, 29:361-372.). In this article, we show that a simple personnel change in an urban police department resulted in a thirty percent reduction in *CFSs*.

crimes that were discovered without *CFSs* . Even then, many of their specific findings would be invalidated by assumptions that they make about *CFSs*. The most egregious example concerns their analysis of “hotspots.

... we conducted an additional “hotspot” analysis. This analysis would allow us to pinpoint the exact source of the crimes. The analyses The analyses commenced by looking at the ten “inner” areas where there was a greater amount of crime than in the control “outer” areas. Within each of these ten inner areas, we then identified the 15 street addresses that had the greatest individual number of calls for service. We reasoned that if the primary source of crime events is the peep show establishment, the street address for that business should appear at the top of this list as the highest-ranking “hotspot.”¹⁶

The fallacy in this reasoning is that the address recorded on a *CFS* is not necessarily the location of the precipitating incident.¹⁷ The *CFS* address tells the patrol unit where

¹⁶ Linz and Paul, p. 15, lines 6-12.

to find the caller. If X calls the SDPD to complain about a disturbance at Y's house, in a majority of cases, the *CFS* goes to X's address. By the Linz-Paul logic, however, the "crime event" occurred at X's address.

If the proprietor of an business is familiar with this geo-coding convention, *CFS*s can be manipulated to make the business look more or less in need of police service or regulation. To build a case for more police services, the proprietor can complain to the police about problems that might otherwise be handled informally. Or to hide a public safety hazard, on the other hand, the proprietor can handle many problems informally, thereby recording fewer *CFS*s and making the business seem safer than it actually is. This is why criminologists do not use *CFS*s for "hotspot" analyses.¹⁸

II.C. "Raw" *CFS*s Ignore Crucial Differences in Public Safety Risk

Assuming Linz and Paul were somehow able to select out those *CFS*s that were precipitated by a crime or crime-like incident *and* that had valid addresses, specific findings would be invalidated by another false assumption about *CFS*s. Linz and Paul analyze raw total *CFS* and this assumes that all *CFS*s are equal or interchangeable. In fact, however, *CFS*s differ in their underlying public safety risks. Two variables in the Linz-Paul dataset are relevant to this concern.

¹⁷ The dataset described in Linz and Paul, pp. 10-12, consists of fourteen variables for 607,903 cases. Eight of the fourteen variable (STNO, STDIR, STNAME, STTYPE, XSTDIR, XSTNAME, XCOORD, and YCOORD) refer to the location of the *CFS*. See our Appendix for details of these variables.

¹⁸ For another reason, see our "Uniform Crime Reports as organizational outcomes." (*Social Problems*, 1982, 29:361-372.). In this article, we show that a simple personnel change in an urban police department resulted in a thirty percent reduction in *CFS*s.

Table IIB reports the final “dispositions” (or outcomes) of the 607,903 *CFSs* analyzed by Linz and Paul.¹⁹ As shown, only one in five was cleared by arrest or report, indicating that the *CFS* was precipitated by a crime or crime-like incident. An equal proportion turned out to be unfounded, duplicated, or cancelled, indicating that the *CFS* was *not* precipitated by a crime. Finally, three in five *CFSs* – the majority – were disposed of with no report. Lacking a report, we can only speculate on the nature of these *CFSs*. But it is highly unlikely that these *CFSs* were precipitated by crimes or crime-like incidents.²⁰

	Frequency	Percent
Cleared by Report	88,215	14.6
Cleared by Arrest	31,035	5.1
Cancelled or Duplicated	71,686	11.8
Unfounded	32,757	5.4
No Report Filed	332,014	54.8
Other or Unknown	52,196	8.3

Table IIC reports the type (or precipitating incident) of the 607,903 *CFSs* analyzed in the Linz-Paul Report.²¹ *CFS*-types are listed by relative frequency. The

¹⁹ This variable is called “DISPO_R” in the Linz-Paul dataset.

²⁰ When a *CFS* ends without a report, the responding patrol unit could not locate a victim, a complainant, or an informant. With obvious exceptions, one assumes that these *CFSs* were not precipitated by unique crimes or crime-like incidents. An “All Units” *CFS* illustrates this situation. These *CFSs* instruct patrol units to watch for a suspect or vehicle. Strictly speaking, Linz and Paul should have analyzed only those *CFSs* that ended in an arrest or report.

²¹ This variable is called “FINAL_TYPE” in the Linz-Paul dataset.

most common precipitating incident is disturbing the peace, which accounts for 43.2 percent of all CFSs, followed closely by Part I UCR “property” crime, which accounts for 37.7 percent. The least common crimes are prostitution and child molestation, which account for only 0.5 percent of the total. Although judgements of seriousness involve subjectivity, the most serious CFSs involve Part I UCR “violent” crime (8.4 percent of the total), shots fired at a house or car (1.3 percent), and DWIs (1.2 percent).²²

	Frequency	Proportion
Disturbing the Peace	262,365	43.2 %
UCR Part I Property Crimes	229,149	37.7 %
UCR Part I Violent Crimes	50,838	8.4 %
Vandalism	19,974	3.3 %
Narcotics Activity	11,489	1.9 %
Public Drunkenness	10,368	1.7 %
Shots Fired at a Home or Auto	7,962	1.3 %
Driving While Intoxicated	7,023	1.2 %
Ambulance Call (Overdose)	2946	0.5 %
Indecent Exposure	2636	0.4 %
Prostitution	1662	0.3 %
Child Molestation	1491	0.2 %

Aggregating CFSs without regard to disposition or type is akin to adding “apples and oranges.” To be sure, an “unfounded” CFS and an “arrest” CFS sum to two CFSs. Likewise, a “robbery” CFS and a “public drunkenness” CFS sum to two CFSs. The totals are uninterpretable, however. Thus, when Linz and Paul conclude that:

²² Although child molestation is a serious crime, we assume that these CFSs involve police officers accompanying Child Protective Services caseworkers on home visits. In California, molestation incidents are reported direct to Child Protective Services, so the CFS does not ordinarily indicated a response to a molestation incident *per se*.

... there is not only no indication of a disproportionately high number of crimes during that period, but substantially fewer crimes than would be expected.²³

We cannot interpret the conclusion. Depending on the mix of dispositions (unfounded, arrest, *etc.*) and types (violent crime, DWI, *etc.*) a *rise* or *fall* in total CFSs may be good or bad. No one knows.

TABLE IID - BURGLARY CFSs BY SOURCE AND DISPOSITION

	Frequency	% of Total
UCR Crime?	279,987	46.1 %
Burglary?	147,127	24.2%
Alarm Call?	110,111	18.1%
False Alarm?	109,135	18.1%

II.D. Burglary CFSs Illustrate the Threats to Validity

UCR burglaries illustrate the most common threats to validity posed by equating CFSs and crime. As Table IIA demonstrates, burglary CFSs and UCR burglaries have only a weak correlation. The weak correlation is due in part to the fact that only a small of burglaries are reported to the police.²⁴ More important, however, *most* of the UCR

²³ Linz and Paul, p. 16, lines 16-17.

²⁴ Biderman and Reiss (On exploring the 'dark figure' of crime. *The Annals of the American Society of Political and Social Science*, 1967, 374:1-15) suggest that fewer than half of all U.S. burglaries are reported to the police. The exact figure for San Diego is unknown, of course.

burglaries reported to the SDPD are initiated by security alarms, virtually all of which turn out to be false alarms. The numbers and proportions reported in Table IID describe this phenomenon. As reported in Table IID, 46.1 percent of the 607,903 *CFSs* in the Linz-Paul dataset were precipitated by incidents that were initially classified as UCR crimes; of these, 52.5 percent were initially classified as UCR burglaries; 74.8 percent of these burglary *CFSs* were initiated by alarms; and 99.1 percent of these alarm-initiated burglary *CFSs* proved to be false alarms. Overall, 18.1 percent of the *CFSs* counted as “crimes” in the Linz-Paul Report were, in fact, burglary false alarms.²⁵ The implications for the validity of their conclusions should be obvious.

II.E “Crime” in the 2-6 AM Period

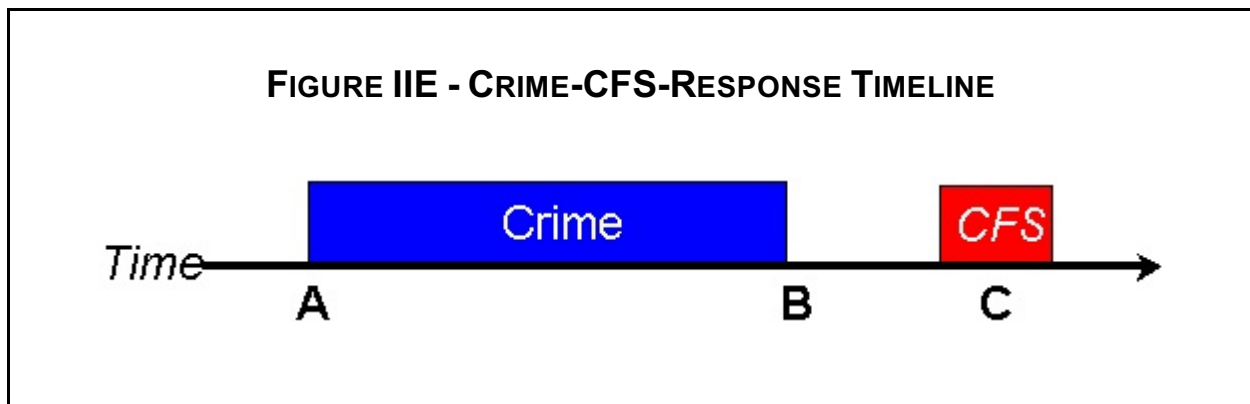
The methodological flaws described in the preceding sections culminate in the case of the Linz-Paul analyses of “crime” during the 2-6 AM period. Comparing raw total *CFSs* during the 2-6 AM period with those made at other times, Linz and Paul conclude that:

On average, Table 3 shows only 11 percent of crime events are occurring in the “inner” areas during the 2 a.m. to 6 a.m. period contrary to the expectation of 16.6 percent. This indicates that criminal activity is not disproportionately greater in the areas surrounding the peep show establishments during these hours. In fact, the study data shows that criminal activity at peep show establishments in San Diego is

²⁵ Burglary *CFSs* not initiated by alarms were presumably initiated by calls from victims or witnesses. Non-alarm burglary *CFSs*, which accounted for 25.2 percent of all burglary *CFSs*, were slightly better. Only 64.5 percent were false. When *CFSs* initiated by UCR robberies are considered, the proportion of false-alarms in the Linz-Paul dataset rises to 21.9 percent of the total.

proportionately less during the 2 a.m. to 6 a.m. period.²⁶

Since *CFSs* are a wholly invalid measure of “crime events” or “criminal activity,” this conclusion is a *non sequitur*. Ignoring the question of what *CFSs* measure, however, this conclusion assumes that the *CFS* and the precipitating incident (or crime) occur simultaneously. If this assumption is incorrect, the analyses of *CFS*-times reported by Linz and Paul are logically insufficient for their conclusion.



The hypothetical timeline in Figure IIE illustrates the logical flaw. In this figure, a crime begins at time-A and ends at time-B; the SDPD receives a *CFS* at time-C. Since the only times in the Linz-Paul dataset are time-Cs, inferences about “crime” during any period assumes that A, B, and C occur within a few minutes of each other. Although this assumption may be warranted for *some* crimes, for *most* crimes, the assumption is unwarranted and incorrect.

UCR property crimes, for example, which make up 37.7 percent of the *CFSs* in the Linz-Paul dataset (see Table IIC), are reported *not* when they occur, but when they

²⁶ Linz and Paul, p. 16, lines 17-22.

are discovered. Most residential burglaries occur during the day, for example, when household residents are at work or school; they are reported when household residents return home. This is often hours later. The Linz-Paul Report's conclusion that there are "substantially fewer crimes" during the 2-6 AM period assumes otherwise. Their conclusion is valid only if crimes and CFSs occur simultaneously.

Even if crime were uniformly distributed across the 24-day, of course, there could still be a compelling pragmatic rationale for closing peep shows during the 2-6 AM period. First, policing is more difficult during night-time hours. Second, among all night-time hours, the hours immediately after bar-closings pose the greatest difficulty and danger for any police department.²⁷ At the risk of over-simplification, as a practical matter, policing during the 2-6 AM period consists largely of shepherding bar patrons to their homes. Given the heightened traffic and crime mortality risks during these hours, this is an extremely difficult time for urban police departments generally and the SDPD in particular.

Although this substantive issue is troublesome, the logical fallacy underlying the Linz-Paul conclusion is even more difficult to understand. In effect, Linz and Paul examine "crime" when peep shows are closed, and conclude that, because crime is not higher when peep shows are closed, there is no rationale for closing peep shows. Of course, the only empirically valid way to arrive at this conclusion would be to examine

²⁷ See "Additional evidence that taverns enhance nearby crime" by D.W. Roncek and M.A. Pravatiner (*Social Science Research*, 1989, 73:185-188). Analyzing crime – not CFSs – in San Diego, the authors demonstrate that the risk of violent crime victimization is more than four times higher in blocks with liquor licenses. The hours of greatest risk occur immediately before and shortly after closing-hour.

crime before and after the implementation of closing-time regulation.

III. The Linz-Paul Statistical Analyses

To assess the criminogenic properties of the 19 peep-shows, Linz and Paul tabulate the number of total, raw *CFSs* at addresses within 1000 feet on either side of the peep-shows and the number in a 2000-foot “control” segment. Ignoring questions about the way that Linz and Paul define “control” segments, Linz and Paul assess the statistical significance of the difference with Mann-Whitney tests. This use of Mann-Whitney tests is unprecedented not only in the criminological literature generally but in the secondary effects literature specifically. There are two reasons why Mann-Whitney tests are not used in these literatures. First, the test requires unrealistic assumptions about the data.²⁸ Second, the test has very low statistical power.

III.A. Statistical Power

A superficial reading of the Linz-Paul Report suggests that areas around the 19 peep-shows had the same number of *CFSs* as the control areas. But a more thorough reading reveals that the peep-show areas generated 15.7 percent *more CFSs* than the control areas. Whereas any urban police department would consider a 15.7 percent difference in *CFSs* *substantively* significant, Linz and Paul claim that the difference is not *statistically* significant. They then argue that a

... statistically nonsignificant result and must be interpreted, as meaning *that* there is no significant difference between these two

²⁸ *E.g.*, see pp. 259-264 of H.M. Blalock’s *Social Statistics, 2nd Revised Ed.* (McGraw-Hill, 1979) or any other statistics text. The Mann-Whitney test is a non-parametric – and hence, low power – test that assumes *independence*. The *CFS* data analyzed by Linz and Paul are *not* independent either geographically or temporally.

averages – an indication that the level of criminal activity for the inner area is equal to the level of criminal activity for the outer area.²⁹

This conclusion reflects a basic misunderstanding of error rates in statistical tests. The Linz-Paul Report cites as a prerequisite for scientific validity the calculation of an “error rate.” Specifically,

The error rate is the degree of chance a scientist will allow. In the social sciences, it is conventional to set the error rate at five percent or less (*i.e.*, 95 time out of 100 the results could not be obtained by chance).³⁰

But there are *two* types of error rates in any scientific study. This passage of the Linz-Paul Report refers to the “false-positive” error rate. The Report fails to mention the

²⁹Linz and Paul, p. 15, lines 1-3.

³⁰Linz and Paul, p. 6, lines 10-12.

complementary “false-negative” error rate and this is an unfortunate oversight.³¹

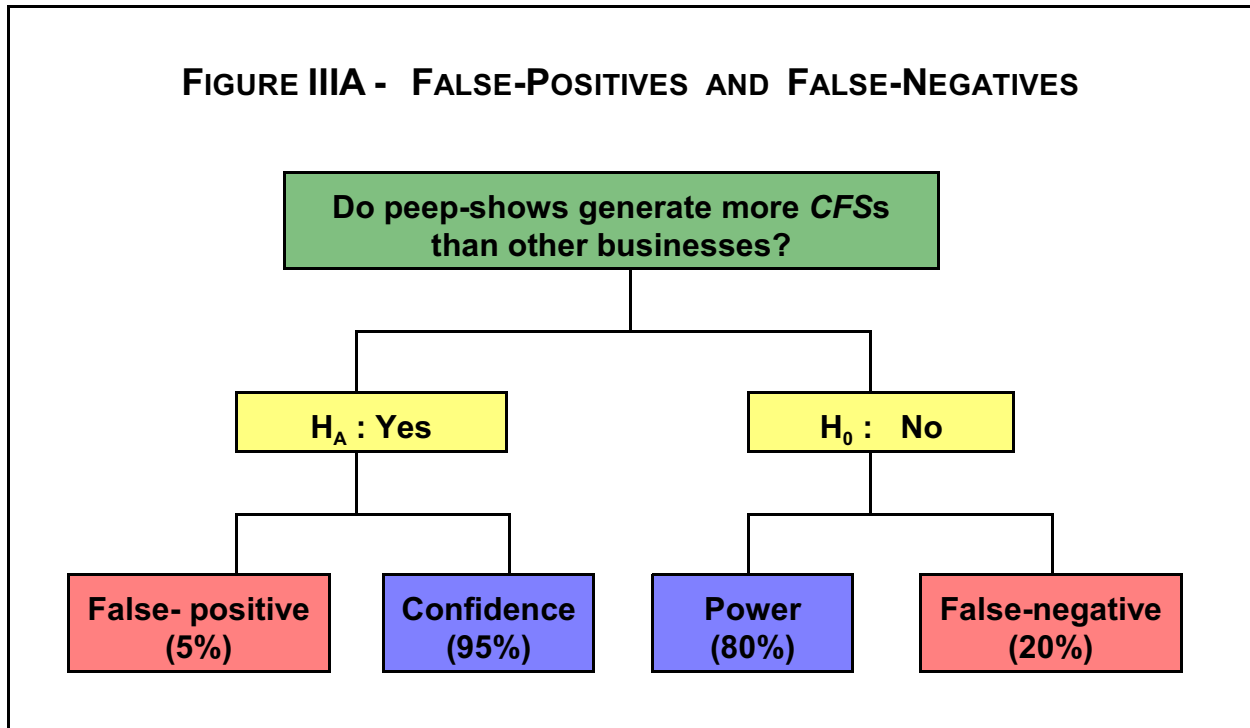


Figure IIIA illustrates the relationship between false-positives and false-negatives. The research question “Do peep-shows generate more *CFSs* than other businesses?” has two possible answers, “Yes” or “No.” The data will support one of the two answers.³² But since data vary from sample to sample, the data can point to the wrong answer. Incorrect decisions (or errors) are painted red in Figure IIIA; correct

³¹ In statistical hypothesis testing, a false positive is called a “Type I” or “alpha-type” error. A false negative is called a “Type II” or “beta-type” error. The terms “false positive” and “false negative,” which come from the field of public health screening, are widely used in popular discourse. We use the terms “false positive” and “false negative” for descriptive simplicity.

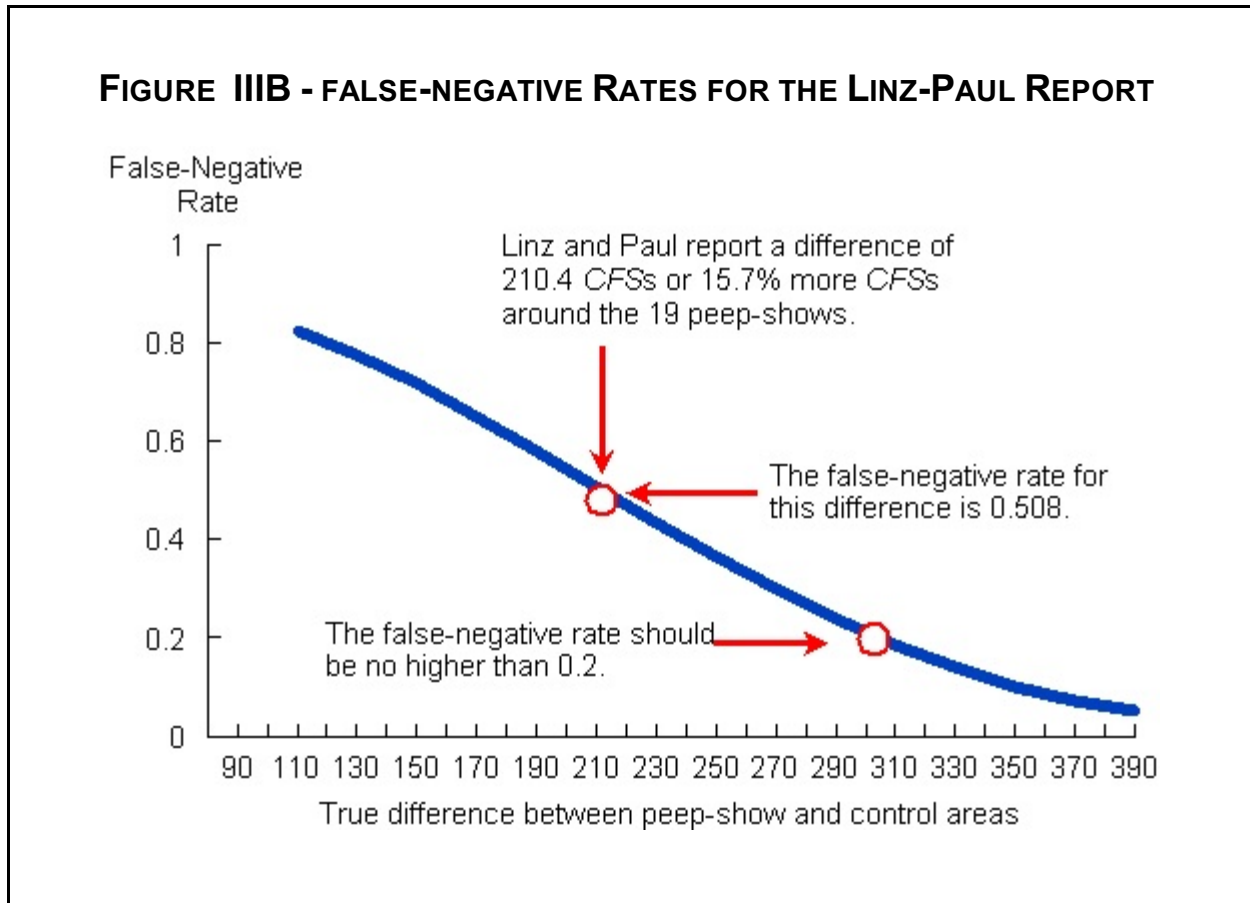
³² *I.e.*, by accepting or rejecting the null hypothesis. The null hypothesis (“No, peep-shows do not generate more *CFSs*.”) and the alternative hypothesis (“Yes, peep-shows do generate more *CFSs*.”) are labeled H_0 and H_A in Figure IIIA.

decisions are painted blue. In practice, of course, the scientist never knows for certain whether his or her answer is correct. But for didactic purposes, assume that the correct answer is known. If “Yes” turns out to be wrong, the answer is a false-positive. If “No” is wrong, the answer is a false-negative. The conventional false-positive and false-negative rates are 5 and 20 percent.³³ Complements of the false-positive and false-negative rates, “confidence” and “power,” are 95 and 80 percent respectively. These conventional levels imply that “Yes” decisions are correct 95 percent of the time, “No” decisions are correct 80 percent of the time respectively.

Since Linz and Paul answered “No” to the research question, the false-positive rate is wholly irrelevant. Despite its irrelevance, they report a false-positive rate of five percent. They do not report a false-negative rate, however, even though that error rate is highly relevant. Fortunately, Linz and Paul do report that the 210.4 CFSs difference has a t-statistic of 0.629. We can use this number to estimate the false-negative rates

³³ The most comprehensive authority on this issue is Chapter 22 of *The Advanced Theory of Statistics, Vol. 2, 4th Ed.* by M. Kendall and A. Stuart (Charles Griffin, 1979). This authority requires a strong background in mathematics, however. J. Cohen’s *Statistical Power Analysis for the Behavioral Sciences, 2nd Ed.* (L.E. Erlbaum Associates, 1988) and M. Lipsey’s *Design Sensitivity: Statistical Power for Experimental Research.* (Sage Publications, 1990). Both Cohen (pp. 3-4) and Lipsey (pp. 38-40) set the conventional false-positive and false-negative rates at .05 and .2, respectively. These rates can be set lower, of course. The convention also sets the ratio of false-positives to false-negatives at 4:1, implying that false-positives are “four times worse than” false-negatives. The 4:1 convention dates back at least to 1928 (J. Neyman and E. Pearson, “On the use and interpretation of certain test criteria for purposes of statistical inference.” *Biometrika*, 1928, 20A:175-240). It reflects a view that science should be conservative. In this instance, for example, the 4:1 convention works in favor of the peep-shows. When actual decision error costs are known, the actual ratio is used.

for a range of effects.³⁴



Plotted in Figure IIIB, these estimates show that the 210.4 CFSs (or 15.7 percent) difference reported by Linz and Paul has a false-negative rate of .508. Thus, if the true difference were 15.7 percent, the Linz-Paul analyses would miss the effect nearly 51 percent of the time! If the true difference were somewhat smaller – say, ten

³⁴ These rates were calculated by a software package called *Power Analysis and Sample Size, Version 6* (PASS); PASS is distributed by the NCSS Corporation. Our calculations assume sample sizes of 19 peep-show and 19 control areas; a false-positive rate of 0.05; and a standard deviation 304.5 CFSs. These are the values that were reported explicitly or implicitly by Linz and Paul.

percent – the Linz-Paul analyses would miss the difference nearly 75 percent of the time. Since the conventional false-negative rate in the social sciences is .2, the reader might wonder, How large would the true difference have to be before the Linz-Paul analyses could detect it with 80 percent statistical power? The true difference would have to be at least 304.5 *CFSs* (22.7 percent) before the Linz-Paul analyses could detect the difference with conventional power.

III.B. Why Is the Power so Low in the Linz-Paul Report?

The mathematics of statistical power is so obtuse that few scientists – and even fewer non-scientists – understand the concept or its importance in statistical research.³⁵ Statistical power is the “dirty little secret” of social science research in the sense that anyone with a modest research background can design a study so as to guarantee “a statistically nonsignificant result.”³⁶ Science guards against such abuses by requiring that researchers publish false-negative rates; or alternatively, as in this case, data sufficient for skeptics to calculate the false-negative rate.

Although Linz and Paul published the data needed to calculate a false-negative rate, their failure to report the unacceptably low power level is puzzling.³⁷ Nonetheless,

³⁵ *E.g.*, “I attributed this disregard of power to the inaccessibility of a meager and mathematically difficult literature...” (p. 155, “A power primer.” J. Cohen, *Psychological Bulletin*, 1992, 112:155-159).

³⁶ In purely logical terms, *not* finding something does *not* prove its nonexistence. The second most widely cited sentence in Isaac Newton’s *Principia Mathematica* acknowledges this point: *Negativa non Probanda*.

³⁷ But not surprising. Other reports by Linz and Paul exhibit the same cavalier attitude toward statistical power. See, *e.g.*, Table 3 on p. 22 of *Measurement of Negative Secondary Effects Surrounding Exotic Dance Nightclubs in Fort Wayne, Indiana*. None of the false-

having demonstrated the weak statistical power of their analyses, readers will have two questions. First, why is the statistical power of the Linz-Paul quasi-experimental design so low? Second, is there is any practical way to raise the design's statistical power? Answers to both questions involve *sample size*.

In this instance, because there are only 19 peep-shows, the sample size seems fixed. There is no statistical rationale for limiting the number of control units to 19, however.³⁸ So raising the number of control units would be an economical way to raise statistical power. How many controls? Assuming the same parameter values used to generate Figure IIIB, including two controls for each peep-show would guarantee the nominal 80 percent statistical power level. The idiosyncratic way that the Linz-Paul Report defines control areas limits this option, of course. We revisit this problem in Section IV.

On its face again, the variability of the true difference between peep-shows and controls would also seem set by circumstance. But in fact, extraneous background noise can be reduced by statistical adjustment with multiple regression. Our Table IIA above is an example. Regression adjustment would have eliminated the need for "matching," for example, would have allowing for a larger number of controls, and would have reduced background variability so as to raise statistical power.

negative rates for the t-statistics in this Table are smaller than the convention .2 level.

³⁸ If there were only 38 total units, 19 peep-shows and 19 (perfectly matched) controls will optimize power (See, e.g., "Neyman allocation" in W.G. Cochran's *Sampling Techniques, 3rd Edition*. Wiley, 1977. In this present case, of course, the controls are *not* perfectly matched and there are literally thousands of potential control areas in San Diego.

IV. Design Issues

“Design” refers generally to the set of methods, or methodology, used to collect, analyze, and interpret data. We have already criticized two aspects of the Linz-Paul Report’s design. In Section II, we noted that, because *CFSs* do not measure crime, any inferences about crime are invalid. In Section III, we noted that, ignoring the inherent problems with *CFSs*, the Report’s statistical analyses had unacceptably high false-negative rates. In this Section IV, we criticize aspects of the Report’s design that are intended to facilitate interpretation of the data.

The Linz-Paul Report is based on a “quasi-experimental” design. Using the conventional notation of Campbell and Stanley, a general quasi-experimental design can be diagrammed as³⁹

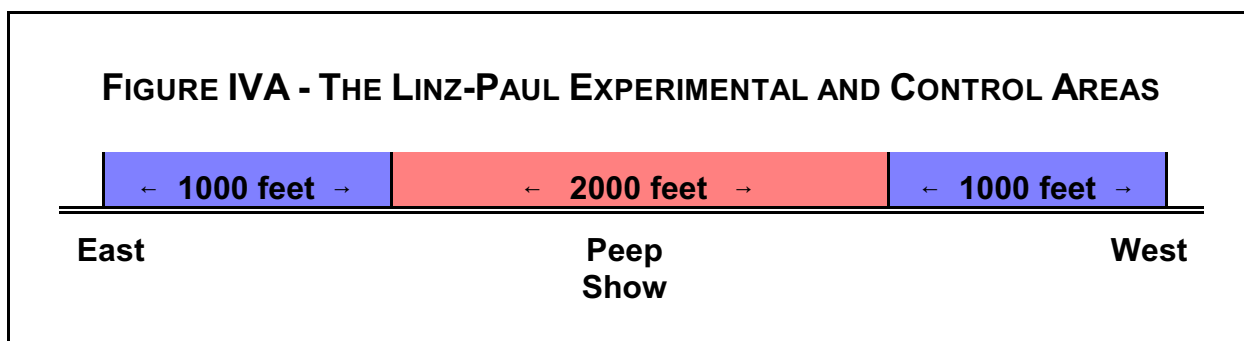
Experimental Units	(O)	X	O
Control Units	(O)	.	O

where “O” denotes an observation, or measurement, and “X” denotes a variable that distinguishes the experimental and control units. In this case, “X” represents the presence of a peep-show establishment in an area; and the “O”s are the total number of *CFSs* recorded in an area during the 1997-2001 period. The first wave of “O”s are

³⁹ The design authority cited by Linz and Paul is *Experimental and Quasi-Experimental Designs for Research* by D.T. Campbell and J.C. Stanley (Skokie, IL: Rand-McNally, 1966). A more recent authority by the same authors is *Quasi-experimentation: Design and Analysis Issues for Field Settings* by T.D. Cook and D.T. Campbell (Chicago: Rand-McNally, 1979).

set off in parentheses to indicate that this design feature is optional.⁴⁰ Had Linz and Paul included observations both *before* and *after* the peep-shows opened, the design would have been stronger.

The internal validity of a quasi-experimental design depends on the similarity of experimental and control units. Secondary effects studies have optimized the similarity of peep-show and control units by “statistical adjustment” and by “matching.” Under ideal circumstances, the two methods yield identical results, minimizing the biasing effects of differences between peep-show and control areas. The Linz-Paul Report used neither method but, rather, assumed that there were no important differences between peep-show and control areas.

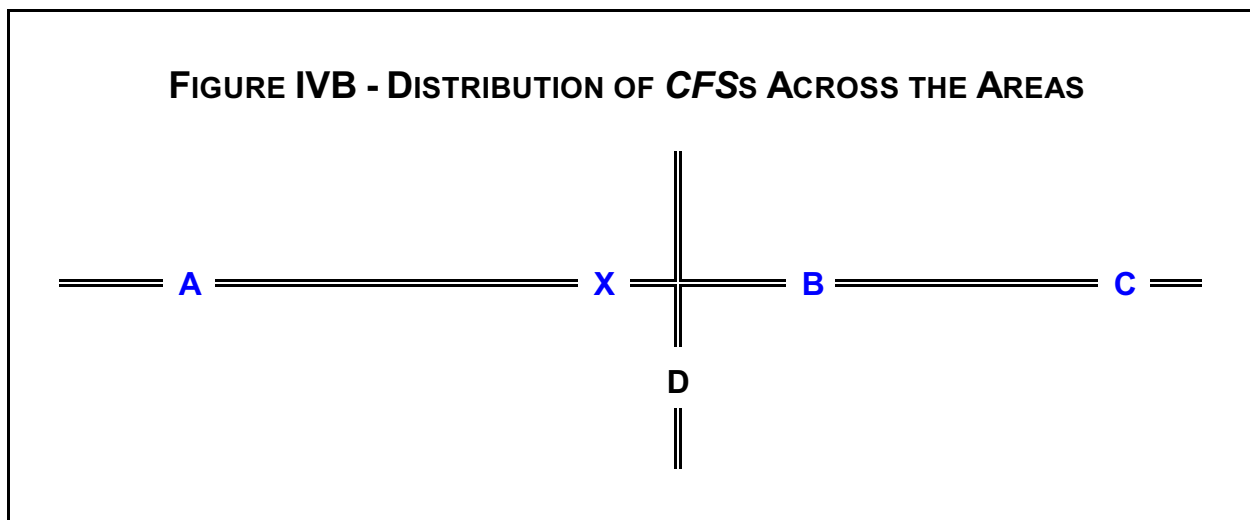


IV.A How Linz and Paul Chose Peep-Show and Control Areas

The most troubling aspect of the Linz-Paul quasi-experimental design is the way that experimental and control areas are defined. The definitions are illustrated by the hypothetical map in Figure IVA. The double line is a 4000 foot segment of an east-west

⁴⁰ Linz and Paul could have strengthened their design by including observations *before* and *after* the peep-shows opened. This design option was apparently not possible. In a before-after quasi-experimental design, each peep-show area would serve as its own control. We used this design element in our 1991 Garden Grove study.

street. A peep-show lies at the center of this segment and addresses within 1000 feet in either direction constitute an experimental (or “inner”) area; it is painted red to denote this status. The control (or “outer”) area, which is painted blue, lies 1000 feet on either side of the experimental (or “inner”) area. CFSs to addresses in the red-painted experimental area are attributed to the peep-show; CFSs to addresses in the blue-painted area are attributed to the control areas. If the null hypothesis is correct – *i.e.*, if peep-show and control areas generate the same number of CFSs – the “inner” and “outer” areas should have similar numbers of CFSs.



But this raises a dilemma. Figure IVB diagrams a hypothetical distribution of CFSs around a peep-show.⁴¹ The double line represents intersecting streets. On the east-west street, X marks the location a peep show; A, B, and C mark the locations of CFSs made from addresses on that street; D marks the location of a CFS to an address around the corner from the peep-show, on the intersecting north-south street. As the

⁴¹ For a real illustration, see Figure A2 in the Appendix.

crow flies, distances from the peep-show to CFSs A, B, C, and D are 1000, 500, 1250, and 350 feet, respectively.

The Linz-Paul design attributes A and B to the peep-show. They are painted red in Figure IVB to denote this fact. C is painted blue to denote the fact that the Linz-Paul design attributes it to the control area. D is attributed to neither and this is a problem. Because D is closer to the peep-show than either A or B, it should be *more* relevant to the research question than either A or B. Because A and C lie roughly equal distances from the peep-show, moreover, they should be treated as roughly equivalent pieces of evidence. Instead, they are treated as diametrically opposing pieces of evidence while D is treated as wholly irrelevant.

The problem posed by the exclusion of D is the more obvious problem of the two. It is easily solved by defining peep-show and control areas as two-dimensional *circles*. In fact, this is a design standard for secondary effect studies. The Linz-Paul Report appears to be the only secondary effects study that defined experimental and control areas as one-dimensional street segments (*i.e.*, as in Figure IVB). We have no idea *why* this weaker, non-standard design was used in this San Diego study. Perhaps this design flaw was dictated by limited time and/or resources. Whatever the reason, however, the use of one-dimensional strips violates both the general principles of design and the standards of secondary effect studies.

IV.B How Peep-Show and Control Areas Should Have Been Selected

Referring to Figure IVB again, although A and C are roughly equal distances from the peep show, the fixed 1000-foot zones put A in the peep show area and C in

the control area. This unrealistic result can be solved either by using explicit distances as a dependent variable or by defining concentric zones around each peep show and control. In fact, this too is the standard in secondary effects studies.⁴²

The use of *areas* as control units is itself a problematic departure from the general principles of quasi-experimental design, however, and from the standards of the secondary effects literature. The design standard in the secondary effects literature compares the crime risks for two *businesses* at different addresses⁴³ or compares the crime risks for the same address *before and after* an adult business opens.⁴⁴ The San Diego study could have used either of these two standard designs. In the first instance, crime risks for the 19 peep-shows could have been compared to crime risks for the 20 adult bookstores without peep-shows.⁴⁵ In the second instance, crime risks for the 19 peep-shows could have been compared before and after peep-shows opened at the addresses.

⁴² For example, the 2001 study by Land *et al.* (*Are Adult Dance Clubs Associated with Increases in Crime in Surrounding Areas? A Secondary Crime Effects Study in Charlotte, North Carolina.*) used both 500-foot and 1000-foot circles around the adult business. Our 1991 Garden Grove, CA study (*Final Report to the City of Garden Grove: The Relationship between Crime and Adult Business Operations on Garden Grove Boulevard*) used circles with radii of 250, 500, and 1000 feet.

⁴³ *E.g.*, the 2001 Charlotte, NC study.

⁴⁴ *E.g.*, the 1991 Garden Grove, CA study.

⁴⁵ Linz and Paul, p.10, lines 13-19: “We then requested a record of all *CFSs* for each beat that included within it one or more of the city’s 39 various adult entertainment businesses ... Although the city has 39 adult entertainment businesses, only 19 of those businesses are peep show establishments. Accordingly, for purposes of this study, we only analyzed the data tied to these 19 peep show establishment locations.”

Ideally, of course, the San Diego study could have used both design features. The general principles of quasi-experimental design require that businesses be compared to businesses, however, or addresses to addresses. Since this principle is also a standard design convention in the secondary effects literature, the decision to employ a weak, non-standard quasi-experimental design raises a related issue.

IV.C “Fishing” in the Linz-Paul Report

Preceding sections have described discrepancies between the design used in the Linz-Paul Report and designs used in the criminology literature, in the secondary effects literature, and even in prior research reports by Linz and Paul. If the relevance of these discrepancies is not apparent, design conventions serve two crucial functions. First, by controlling methodological variance, design conventions enhance the comparability of research findings. Second, however, more important in this instance, *strictly enforced design conventions eliminate “fishing.”*⁴⁶

⁴⁶ See pp. 42-3 in *Quasi-experimentation: Design and Analysis Issues for Field Settings* by T.D. Cook and D.T. Campbell (Chicago: Rand-McNally, 1979) for a discussion of “Fishing and the error rate problem.” Note further that *Daubert* addresses this issue implicitly in its discussion of “the known or potential rate of error.”

TABLE IVC - THE DESIGNS OF THREE SECONDARY EFFECT STUDIES

	San Diego, CA	Fort Wayne, IN	Charlotte, NC
Crime Measure	CFs	UCR Arrests	UCR Crimes
Affected area	1000-foot strip on either side of a peep-show	1000-foot radius around an adult business	500- and 1000-foot radii around an adult business
Control area	Contiguous 1000-foot strips on both sides of the affected area	1000-foot circle in a non-contiguous "matched" area	500- and 1000-foot radii around other businesses
Publication Date	September, 2002	February, 2001	July, 2001
Authors	Linz and Paul	Linz and Paul	Land <i>et al.</i>

The potential for “fishing” arises when there are several possible ways to design a research project. To illustrate, Table IVC lists summarizes the design differences for secondary effects studies in San Diego, Fort Wayne, and Charlotte.⁴⁷ Although all three of these studies were conducted during the same period by the same investigators, the design differences are striking. For example,

- ◆ Three different crime measures (CFs, arrests, and UCR crimes);
- ◆ Three different definitions of the affected areas (1000-foot strips, 1000-foot radii, and 500-foot radii); and

⁴⁷ The Fort Wayne and Charlotte studies, cited above, are *Measurement of Negative Secondary Effects Surrounding Exotic Dance Nightclubs in Fort Wayne, Indiana* and *Are Adult Dance Clubs Associated with Increases in Crime in Surrounding Areas? A Secondary Crime Effects Study in Charlotte, North Carolina*. The San Diego study is the Linz-Paul Report.

- ◆ Three different types of control areas (contiguous strips, non-contiguous “matched” circles, and other businesses).

Considering only these three elements, there are at least $(3 \times 3 \times 3 =)$ 27 different ways to design a secondary effects study. With this many design variations, the researcher can “fish” for a result that supports his or her position. Strictly enforced design conventions minimize the potential for “fishing.”

Although “fishing” artifacts are not easily calculated,⁴⁸ it should be intuitively clear that one could find design variations that “prove” either side of the question, “Do peep-shows cause crime?” There is no evidence to suggest that the findings and conclusions of the Linz-Paul Report are the product of a “fishing” expedition. On the other hand, the controversial nature of the Report’s findings and conclusions, in conjunction with the design discrepancies listed in Table IVC, invite healthy skepticism.

V. Summary and Conclusions

Based on a thorough reading of the Linz-Paul Report, on a reading of other relevant literature, on reanalyses of the original data, on analyses of auxiliary data, and on our experience in this field, it is our opinion that the San Diego study’s methodology falls short of the current standards in the social sciences, particularly criminology. The methodological shortcomings of the Linz-Paul Report span all three basic areas: (1) Measurement; (2) statistical analyses; and (3) design.

The most serious methodological flaws follow directly (or indirectly) from the use

⁴⁸ “Fishing” biases the research by inflating the false-positive and false-negative rates. Because the possible design variations are not independent, the degree of bias is difficult to calculate.

of *CFSs* to measure crime. First and foremost,

- ◆ The use of *CFSs* to measure crime is unprecedented in either the criminological or the secondary effects literatures.
- ◆ In San Diego specifically, *CFSs* explain only ten percent of the variance in crime (see Table IIA).

The absence of precedent and explanatory power would prevent the San Diego study from being published in top-ranked criminology journals. This fatal methodological flaw in the Linz-Paul Report is aggravated by erroneous assumptions about SDPD *CFSs*.

Specifically,

- ◆ *CFS* addresses do not necessarily tell us *where* precipitating incidents occurred.
- ◆ *CFS* times do not necessarily tell us *when* precipitating incidents occurred (see Figure IIE).

These anomalies of the SDPD *CFSs* invalidate the Linz-Paul Report's analyses of "hotspots" and crime during the 2-6 AM period, of course. But a more serious problem arises when Linz and Paul create an aggregate index from raw, total *CFSs*.

Specifically,

- ◆ Only 20 percent of the *CFSs* in the index resulted in a crime report or arrest; an equal number were wholly spurious (see Table IIB).
- ◆ The *CFS* index counts incidents ranging from disorderly conduct to armed robbery as equivalent (see Table IIC).
- ◆ The vast majority of *CFSs* precipitated by incidents that appear to

be UCR crimes turned out to be false alarms (see Table IID).

Any one of these demonstrably incorrect assumptions about SDPD *CFSs* is sufficient to invalidate any of the Linz-Paul Report's conclusions.

Ignoring the validity problems related to use of *CFSs* to measure crime, the manner in which the *CFSs* are analyzed pose another serious threat to validity. When Linz and Paul find that a 15.7 percent difference in *CFSs* between peep-show and control areas is "statistically nonsignificant," for example, leading to the conclusion that peep-shows are no different than other businesses, they do not consider the possibility that the "statistically nonsignificant" finding may be a false-negative. In fact,

- ◆ The probability of a false-negative finding is .508, approximately the same as a fair coin-flip (see Figure IIIB).
- ◆ A false-positive probability of .2, which is the conventional level in the social sciences, would require that peep-show areas generate 22.7 percent more *CFSs* than control areas (see Figure IIIB).

Since a 15.7 percent increase (or decrease) in *CFSs* would be *substantively* significant to police, courts, and taxpayers, the failure to find *statistical* significance is a fatal threat to validity. Put simply the Linz-Paul Report was designed so as to make it practically impossible to find a statistically significant difference between peep-show and control areas.

The problem of inadequate statistical power cannot be separated from the quasi-experimental design of the San Diego study and this raises a more fundamental issue. Compared to other secondary effects studies, including those conducted by Linz and

Paul, the San Diego design is *idiosyncratic* (see Table IVC). Design innovations are to avoided because, if for no other reason, each innovation raises potential threats to internal validity. The *simultaneous* debut of *many* design innovations raises the more serious question of whether the study was a “fishing” expedition. Although Linz and Paul may have seen the San Diego study as an opportunity to break new ground in the study of secondary effects, a critical reader should suspect that the specific new methods were selected in part because they supported a specific finding or conclusion. To allay suspicions, researchers who used new methods are required to explain the departure from convention.

V.A. The *Daubert* Standards

If the *Daubert* decision requires that scientific evidence be *both* relevant *and* valid, the Linz-Paul Report meets neither criterion. With respect to relevance, because the Report’s conclusions are couched in terms of “crime” or its synonyms, and because CFSs are not an acceptable measure of crime, the Report’s conclusions are *irrelevant* to the secondary effects debate. With respect to validity, the Report’s conclusions also fall below the minimum threshold. In their authoritative work on validity in social science research,⁴⁹ Thomas D. Cook and Donald T. Campbell distinguish four types of validity: construct, statistical conclusion, internal, and external validity. Uncontrolled threats to construct, statistical conclusion, and internal validity prove fatal to the Linz-Paul Report’s conclusions. Accordingly, as social scientists and statisticians, we find that the

⁴⁹ *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand-McNally, 1979.

Linz-Paul Report does not meet the threshold of scientific rigor in these fields.

Purely legal issues make it more difficult to arrive at the same categorical level of certainty with respect to the *Daubert* standard. Nevertheless, in the introduction to their Report, Linz and Paul give a succinct interpretation of *Daubert*:

Offering “some general observations” as to how this connection can be made, the Court provided a list of factors that federal judges could consider in ruling on a proffer of expert scientific testimony: (1.) The “key question” is whether the theory or technique under scrutiny is testable, borrowing Karl Popper’s notion of falsifiability; (2.) Although publication was not an absolute essential, the Court noted that peer review and publication increased “the likelihood that substantive flaws in methodology will be detected;” (3.) An error rate or estimate of the probability that empirical relationships are due to chance should be calculated; (4.) adherence to professional standards in using the technique in question; (5.) Finally, though not the sole or even the primary test, general acceptance could “have a bearing on the inquiry.”⁵⁰

The Linz-Paul Report is deficient by all five of criteria. In principle, for example, the Linz-Paul Report’s core null hypothesis – “Peep-shows generate no more *CF*’s than other businesses” – is falsifiable or testable, as required by criterion (1.). In practice, however, given a 50 percent false-negative rate and unlimited “fishing,” the core null hypothesis cannot be rejected.

The 50 percent false-negative rate fails criterion (3.) *per se* and, by implication, (4.) as well. The design idiosyncracies, described in Section IV, and the use of *CF*’s to measure crime, described in Section II, fail criteria (4.) and (5.). Publication in peer

⁵⁰Linz and Paul, p. 4, lines 1-9. The Popper work cited in this excerpt, *Conjectures and Refutations* (Basic Books, 1962), argues that scientific and pseudo-scientific theories differ in terms of falsifiability (or testability). Pseudo-scientific theories – Popper’s examples include Freudian psychoanalysis and Marxian history – cannot be tested because any outcome confirms the theory.

reviewed journals, criterion (2.), is related to criteria (4.) and (5.) in the sense that peer review enforces adherence to the standards of acceptance by the scientific community. Publication is not a norm in the secondary effects literature, however, and peer review plays a somewhat different in the social sciences than in the natural sciences. Given its many serious methodological flaws, it is unlikely that the Linz-Paul Report would be published in a top tier social science journal.

In summary, applying the five criteria cited in Linz-Paul Report, it is our opinion that the Report's findings and conclusions do not meet the minimum threshold mandated by *Daubert*.

Appendix: Technical Details of the Linz-Paul Dataset

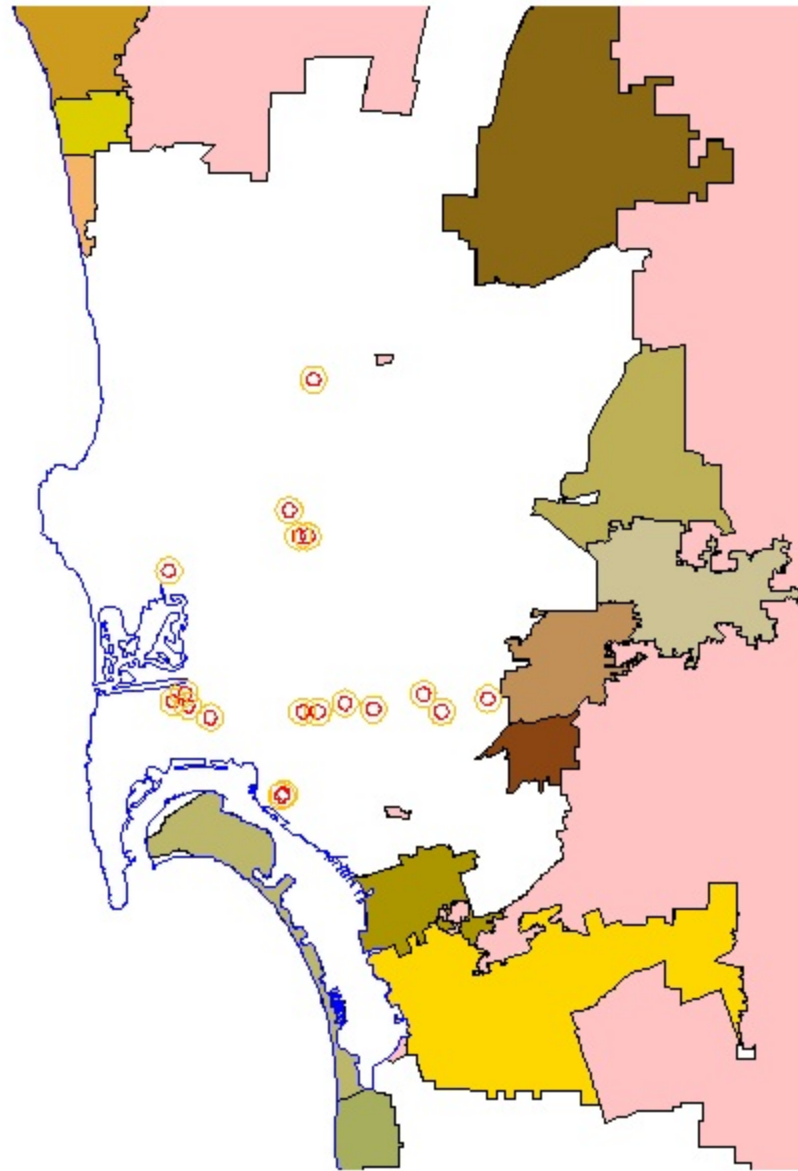
This appendix provides details of the Linz-Paul dataset which were too technical or tangential to be included in the text. A copy of the dataset described in pp. 10-12 of the Linz-Paul Report was supplied by the SDPD Information Services Division. The dataset is a MicroSoft Access file with 607,903 cases, each representing a single CFS, with the following variables:

DATE	Date of the CFS
TIME	Time of the CFS
INC_NO	A unique identifier
ENTRY_DT	Date the CFS was entered into the database
FINAL_TYPE	Type of CFS (burglary, DWI, etc.)
DISPO	Disposition of the CFS (duplicate, arrest, etc.)
STNO	Street number (address number)
STDIR	Direction of the street (N, SE, etc.)
STNAME	Name of the street
STTYPE	Type of street (avenue, street, etc.)
XSTDIR	Cross-street direction (N, SE, etc.)
XSTNAME	Cross-street name
XCOORD	East-West co-ordinate
YCOORD	North-South co-ordinate
BEAT	SDPD beat number

Approximately 5,000 CFSs had missing co-ordinates⁵¹ or addresses in conflict with the co-ordinates. These CFSs, amounting to less than one percent of the total, were excluded from all analyses. Other CFSs were excluded from our analyses (though not necessarily from the Linz-Paul analyses) if CFS-types and/or dispositions were missing.

⁵¹ These are “State Plane Co-ordinates for California Zone 6.” See Table 1, p. 14, of *The GPS Observer*, July, 1999. The unit is feet.

FIGURE A1 - LOCATIONS OF THE 19 PEEP-SHOWS

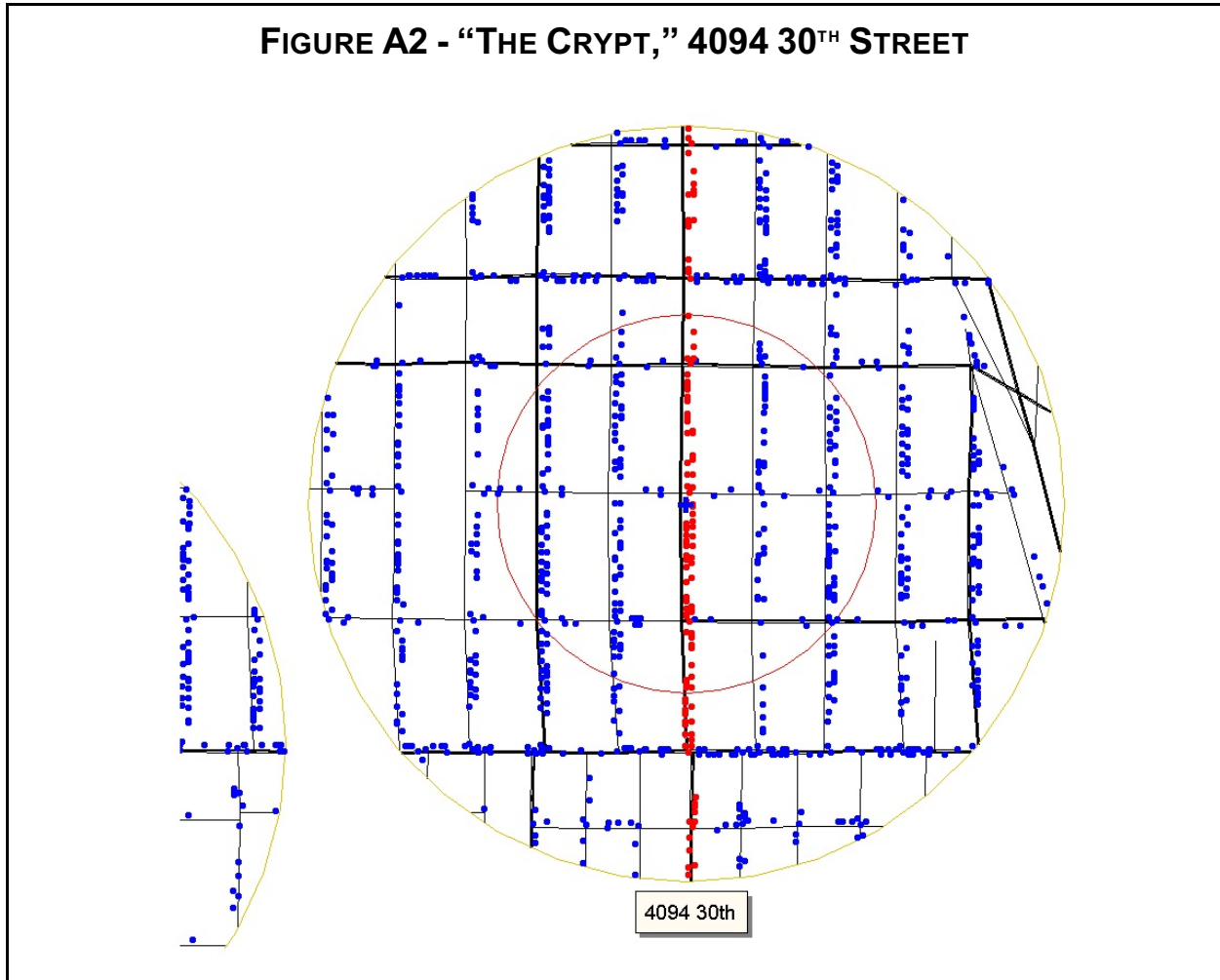


The data were analyzed with MicroSoft Access, ARCVIEW, and SPSS. On our part, the choice of software was a matter of convenience. The fact that Linz and Paul used different software cannot explain differences in findings and conclusions. ARCVIEW was generally used for mapping and geo-statistical analyses. For example, we used ARCVIEW to construct Figure A1. In that figure, the 19 peep-shows are plotted as 2000-foot diameter circles on a San Diego map. The geographical distribution of peep-shows is clearly *not* random. Rather, the peep-shows are clustered together in a few districts, often with overlapping, non-distinct boundaries. This non-random distribution raises a formidable challenge to the Linz-Paul design which we will discuss shortly.

Figure A1 is deceiving in at least one respect: Linz and Paul did not use circular areas, such as those shown in Figure A1 but, rather, used 2000-foot strips on either side of the peep-show address. This general design scheme is depicted in Figure VA. Because the design considered only *CFSs* on the same street as a peep-show, only a small subset of San Diego's 103 police beats⁵² were included in the Linz-Paul dataset. Figure A2 below illustrates one consequence of this design. In this figure, concentric circles with diameters of 1000 and 2000 feet are drawn around "The Crypt," a peep-show at 4094 30th Street. *CFSs* to 30th Street addresses are plotted in red. If a red *CFS* lies within 1000 feet of 4094 30th Street, it is attributed to the affected area; if a red

⁵² The 35 beats in the dataset includes several double counts due to beat renumbering changes during the five-year period, 1997-2001. Beat #531 became #529; #532 became #531; #533 became #526; #534 became #527; #535 became #517; #536 became #518; and #537 became #528. The SDPD system uses both the old and new numbers.

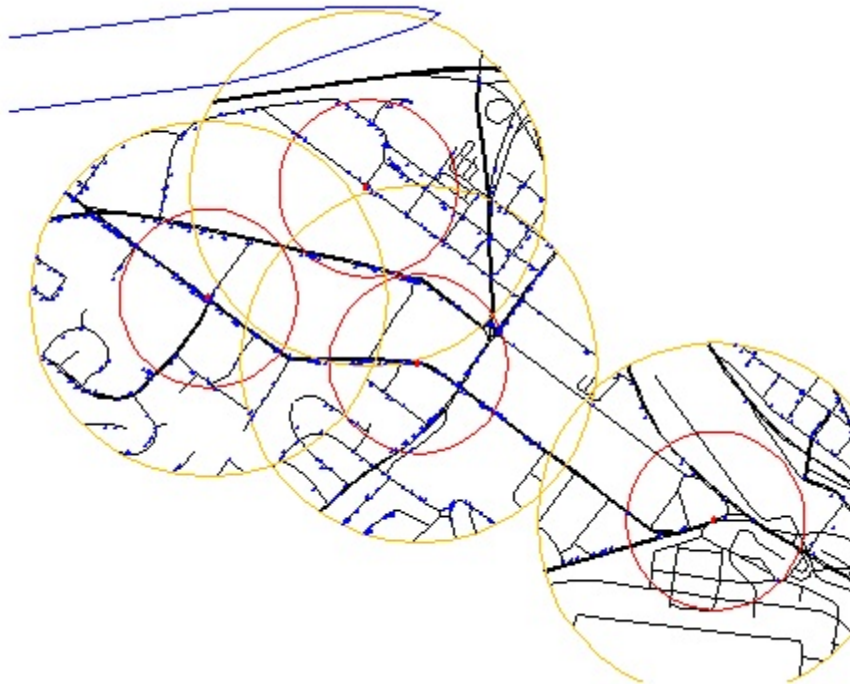
FIGURE A2 - "THE CRYPT," 4094 30TH STREET



CFS lies more than 1000 feet away but less than 2000 feet away, on the other hand, it is attributed to the control area. If the number of CFSs in the affected area exceeds the number in the control area, Linz and Paul conclude that "The Crypt" has criminogenic properties.

This ignores CFSs that may be near "The Crypt" but are not on 30th Street *per se*, of course. These CFSs are painted blue in Figure A2. Even though some of these CFSs lie within a few yards of "The Crypt," they are excluded from the Linz-Paul design and analyses. Figure A2 also demonstrates the validity consequences of an arbitrary

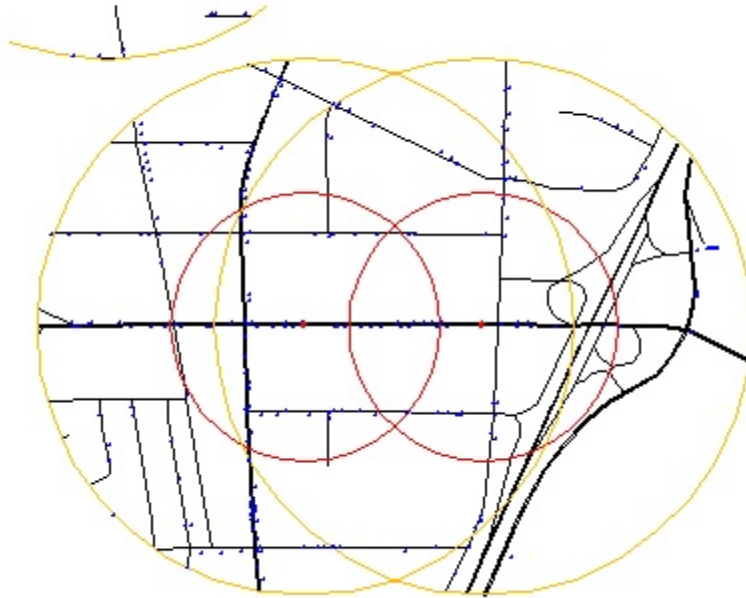
FIGURE A3 - OVERLAPPING PEEP-SHOW SITES (CENTRAL)



1000-foot zone. On the south end of the 30th Street zone, a number of red-painted CFSs are located within a few feet of the affected or control areas.

Figure A3 shows another aspect of the arbitrary 1000-foot zones used by Linz and Paul. The circles in this map are centered on “Adult Depot” at 3489 Kurtz, “Adult Superstore” at 3610 Barnett, “F Street Sports Arena” at 3112 Midway, and “Midnite Books Midway” at 3606 Midway. Overlap in the circular control zones and proximity of the affected areas create obvious methodological problems. This assumes that the affected and control areas are defined as circles, of course. If the affected and control areas are defined as one-dimensional strips (see Figure VA, *e.g.*), the methodological

FIGURE A4 - OVERLAPPING PEEP-SHOW SITES (BALBOA AVE)



is mitigated somewhat. The problem does not disappear, however. Figure A4 shows the affected and control areas around “Mercury Adult Books” at 8081 Balboa and “F Street Kearny Mesa” at 7865 Balboa. Here the affected areas overlap even when defined as one-dimensional 1000-foot strips.