Document Title: Investigating the Scope of Measurement Error in Calls-for-Service as a Measure of Crime

Author(s): David A. Klinger

Document Number: 171946

Date Received: 1997

Award Number: 95-IJ-CX-0023

171946

171946

INVESTIGATING THE SCOPE OF MEASUREMENT ERROR IN CALLS-FOR-SERVICE AS A MEASURE OF CRIME

DAVID A. KLINGER
ASSISTANT PROFESSOR OF SOCIOLOGY
UNIVERSITY OF HOUSTON

This research project was funded by the National Institute of Justice, Office of Justice Programs, U.S Department of Justice, Grant Number 95-IJ-CX-0023. Points of view or opinions expressed in this report are those of the author and do not necessarily represent the official position or policies of the U.S. Department of Justice.

Among the most pressing problems facing criminal justice officials and researchers who study crime and justice issues is a lack of knowledge about how much crime occurs in the United States. This problem exists because both of the major measures of crime -- The Uniform Crime Reports (UCR) and the National Crime Survey (NCS) -- contain substantial distorting biases. At least two factors contribute to the measurement bias in UCR data: 1) citizens' decisions about whether to notify the police about criminal activity and 2) police decisions about whether to take reports when citizens do inform them that crimes have occurred (e.g., Black, 1970; Warner and Pierce, 1993). Sources of bias in the NCS include, but are not limited to, citizens' failure to report crimes to interviewers, and problems that are common to survey research; for example, errors associated with interview effects and other response biases (e.g., Bailey, Moore, and Bailer, 1978; Biderman and Lynch, 1991).

In an attempt to overcome the limitations inherent in NCS and UCR data, some criminologists have recently proposed a "new" indicator of crime; using records of citizens' telephone calls-for-service to police dispatch centers to measure crime at the address (Sherman, Gartin, and Buerger, 1989), neighborhood (Bursik, Grasmick, and Chamlin, 1989; Warner and Pierce,1993), and city (Bursik and Grasmick, 1993) levels. Proponents of using calls-for-service (CFS) to measure crime assert that such data is superior to NCS data because they are not subject to the response and other problems inherent in surveys. CFS proponents argue further that the measure is superior to UCR data because it reflects callers' descriptions of criminal events *before* officers arrive on-scene, thus eliminating bias introduced in UCR data by police behavior.

Indeed, two of the CFS measures' strongest advocates maintain that as long as researchers are careful to weed-out multiple calls regarding the same incident, CFS data are "biased only by citizens' willingness to report crimes" (Warner and Pierce, 1993).

Consideration of the matter, however, indicates that there may be other sources of distorting biases in CFS data besides that introduced by citizens' willingness to call the police. One such problem is that many crimes come to the attention of the police via means besides phone calls to dispatch centers. For example, citizens sometimes report criminal activity directly to officers on patrol (and at precinct houses) and officers often observe criminal activity while patrolling their beats (e.g., Reiss, 1971). Because CFS records, by definition, reflect only those events that citizens' phone-in to police dispatch centers, they do not include any crimes that come to police attention via other means. The undercount of criminal activity that this produces in CFS data is particularly problematic vis-a-vis UCR crime counts; officers may take reports of such crimes, which means that there is a universe of crimes that cannot appear in CFS data, but which can appear in UCR data.

A second obvious problem with using CFS to measure crime is that callers' descriptions to police operators may not accurately reflect the legal nature of the events they are reporting -- as Reiss (1971:11) puts it, callers can provide "misleading" information -- which may also bias CFS data. One reason citizens may provide inaccurate information to police telephone operators is that they may not accurately apprehend the legal character of the events they phone-in, for as Reiss (1971:77) notes, "many citizens have only a vague understanding of the difference between civil, private, and criminal matters." Even if citizens do have a sound understanding of the legal nature of the problems they report, errors can still enter CFS records. First, citizens are sometimes (due to low linguistic skills or high levels of excitement; Reiss, 1971:11) unable to clearly articulate the character of the problems they report. Moreover, citizens may purposely mislead police operators, describing, for example, events as more legally serious than they know them to be in order to prompt a faster police response (e.g., Whitaker, 1979). Finally, the legal character of events may change between the time callers report a problem to police operators and the time patrol officers arrive on-scene (see below).

Citizen inaccuracy can lead to three types of error in CFS crime counts. First, call-in data may fail to record crimes that have occurred, when calls classified as non-criminal events actually involve some type of criminal activity (**false negatives**). This can occur when callers misunderstand the nature of the events they are reporting (e.g., the "suspicious person" who is burglarizing a neighbors house) and when non-criminal incidents at the time the call is placed escalate into crimes by the time officers arrive on-scene (e.g., "domestic disputes" that escalate to assaults). A second type of error in CFS data is that callers can classify non-criminal behavior as criminal (**false positives**). In this instance, citizens report criminal activity when none, in fact, occurred (e.g., a person reports a "burglary" in progress when he mistakes a neighbor who has locked himself out of his house for a burglar trying to pry open a rear window). Third, call-in data can misclassify the nature of criminal activities that do occur, when callers describe incidents that involve one type of crime (e.g., theft) as involving another type (e.g., robbery) (**crime misclassification**). Callers may *overstate or inflate* the seriousness of crimes (e.g., reports robbery instead of theft) or they may *understate* their seriousness (e.g., reports theft instead of robbery). In sum, in addition to the problem of crimes that come to the attention of the police via means besides citizens' calls to 911 operators, calls-for-service records may include three other types of errors. They may omit crimes in incidents that citizens call-in to the police, they may include crimes that never actually occurred, and they may misclassify the nature of crimes that actually do occur.

Two prominent CFS advocates, Bursik and Grasmick (1993; see also Bursik, Grasmick, and Chamlin, 1990) recognize that citizens will sometimes make inaccurate reports to police telephone operators. They consider misclassifications to be the sole source of measurement error in CFS data, however. Bursik and Grasmick never mention the possibility that CFS records can miss some crimes due to false negative errors. Where false positives are concerned, they maintain that over-reporting due to such errors can be eliminated from CFS data by having police officers immediately report back to headquarters whether the crime calls to which they are dispatched involve criminal activity. When officers report that there is no violation of law, dispatchers can "correct" their logs, thereby ridding CFS records of false positives. It would appear that this process may actually exacerbate measurement error in CFS data, however. Introducing into CFS data patrol officers' reports about whether telephone complaints of crime are "founded" or "unfounded" would appear to build into

CFS data precisely the type of error CFS advocates wish to avoid -- errors resulting from discretionary decisions by police that distort estimates of crime and criminal behavior.

This problem has been incorporated into the only extant study that explicitly seeks to assess the validity of CFS as a measure of crime; Bursik and Grasmick's (1993) study of the correspondence between UCR and CFS crime counts in Oklahoma City, where patrol officers "correct" dispatch records when they respond to "unfounded" criminal radio calls. In addition to using a suspect means of addressing false positive errors, seeking to validate CFS records by comparing them to UCR data does nothing to address the problems of false negative and misclassification errors. The current study seeks to take a more in-depth look at the validity of CFS data as an indicator of crime by examining how all types of citizen inaccuracy, as well as the fact that many crimes come to police attention via means besides calls to dispatch centers, may bias crime counts based on citizens' calls for police service.

DATA FROM AN OBSERVATIONAL STUDY OF THE POLICE

In 1977 researchers from the Workshop in Political Theory and Policy Analysis at Indiana University and the Center for Urban and Regional Studies at the University of North Carolina at Chapel Hill conducted the Police Services Study (PSS). The PSS examined numerous facets of policing in 60 neighborhoods served by 24 police departments. During the study, researchers interviewed approximately 200 randomly-selected residents of each study neighborhood, collecting information about numerous aspects of the respondents' experiences, opinions, and socio-demographic characteristics. The study also included systematic observations of police officers at work wherein trained observers accompanied officers on 900 patrol shifts (15 per study neighborhood), recording select information on 5,688 police-citizen interactions that occurred during these shifts in observation schedules.

Observers used what the researchers called "problem codes" to record information about the nature of each of the 5,000+ encounters in the PSS at three points in time: 1) prior to actual police-citizen contact (based on, e.g., the initial broadcast for encounters that came from radio calls and the first information about non-dispatched encounters that came to the involved officers' attention, such as the apparent traffic law violation for vehicle stops), 2) at the point of initial police contact with citizens, and 3) at the close of encounters. Observers could use up to three of the 200+ codes to describe each encounter at each of the three points in time. Observers could thus use the problem codes to, for example, describe a radio call as initially dispatched as involving "screams" (113) at a "family trouble" (029) call. With the problem codes then, PSS observers recorded the nature of the problems officers dealt with 1) as they appeared when first presented to police officers, 2) as they appeared when officers first arrived on-scene, and 3) in terms of what they actually entailed as determined during encounters.

As detailed below, these problem codes afford the opportunity to examine the nature and structure of errors in CFS data due to both citizen inaccuracy and crimes that come to police attention via means besides calls for service. Because the problem codes included multiple codes pertaining to single types of crimes (e.g., robberies of private citizens, financial institutions, and other commercial establishments all are represented by different codes), the first step taken to ready the data for this task was to group together the problem codes that pertained to each of the following crimes: 1) sexual assault, 2) all other assault, 3) robbery, 4) burglary, 5) vehicle theft, 6) all other theft, 7) vandalism, and 8) trespass. The sample of encounters were then divided into two groups; 1) those that were initiated by a radio call (N=2,853) and 2) those that were not (N=2,835). The final problem code data (i.e., that pertaining to the actual problem the police encountered) was then used to classify each of the 5,688 encounters as either belonging to one of the eight crime categories, or as some other type of matter. Encounters were classified as criminal if the final problem code information indicated that the encounter actually involved any of the eight aforementioned types of crime. Following the same procedures, the initial problem codes pertaining to the 2,853 encounters that were initiated by radio calls were then used to classify the initial dispatch description of these encounters as one of the eight crime categories, or some other type of matter.

With the data grouped by problem types (i.e., as one of the eight types of crimes or some other kind of matter) and sub-divided into dispatched and non-dispatched encounters, obtaining counts of non-reported crimes and of each of the three types of errors due to citizen inaccuracy are rather straight-forward. First, because the final problem codes represent the actual problems that officers encountered, a count of the non-dispatched encounters whose final coding corresponds to one of the eight crime types under study will yield a count of the crimes that patrol officers encounter via means besides calls to police dispatch centers (for simplicities sake, such crimes will henceforth be called "non-reported" crimes). Second, because the initial problem codes for the encounters that were initiated by radio calls represent the description of the problem that would be recorded in police dispatch logs, comparisons of the initial and final problem code classifications for the dispatched encounters will yield counts of errors due to citizen inaccuracy. Counts of false positive and misclassification errors can be obtained by comparing the initial and final codes for those dispatched encounters that were initially classified as involving a crime. A count of false positives can be obtained by conducting a similar comparison for those encounters initiated by radio calls that actually involved some crime.

DESCRIPTIONS OF ERRORS

The first analytical step taken was to develop information about the number of non-reported crimes (i.e., those that came to police attention via means other than radio calls). As displayed in Table 1, an examination of the final problem codes for the 2,835 encounters that were not initiated by radio dispatches disclosed that 288 of them involved one of the eight types of crime under study. A closer look at these 288 cases disclosed that they included 11 sexual assaults, 84 non-sexual assaults, seven (7) robberies, 54 burglaries, 19 vehicle thefts, 53 other thefts, 34 trespasses, and 26 acts of vandalism. In sum, a look at the data discloses that PSS officers encountered a crime that was *not* initiated by a citizens' call for service in almost one of every three patrol shifts they worked (288 criminal encounters/ 900 shifts).

Table 1: Non-Dispatched Crimes	
Assault	84
Sex Assault	11
Robbery	7
Theft	53
Auto Theft	19
Burglary	54
Vandalism	26
Trespass	34
All Crimes	**288**

Where errors due to citizen inaccuracy (i.e., false positive, false negative, and crime misclassification) are concerned, the first order of business was to determine how many of the 2,853 encounters that were initiated by radio call were initially classified as involving one of the eight crime categories and how many of them were classified as actually involving some type of crime. The data disclose 905 calls that were initially classified as crimes, and 883 calls that actually involved crimes.

As shown in Table 2, the comparison of initial and final problem classifications for the 905 cases that were dispatched as crimes disclosed 185 encounters that involved no actual crime (i.e., 185 false positives) and another 65 cases that were dispatched as one type of crime, but turned out to involve another sort of crime (i.e., 65 misclassification errors). Thus, more than one in four of the calls initially classified as involving a particular crime actually involved either another sort of crime (seven percent) or no crime at all (20%). A closer look at the data in Table 2 discloses that the patterns of error varied somewhat across crime types. Where overall accuracy is concerned, vandalism calls were most accurate (81% correct) and trespass calls least accurate (55% correct). Where the two types of errors in calls when citizens report some criminal activity are concerned, vandalism calls were most likely to be false (37%; with assaults a close second at 30%), and robbery calls most likely to involve crime misclassification (16%). In sum, the data show that the dispatch records of the agencies in the PSS contain a goodly number of false positive and misclassification errors.

Table 2: False Positive and Misclassification Errors by Crime Type (Based on 905 Radio Calls Initially Dispatched as Crimes)

	False Positive	Misclassification	Correct	Total Calls
Assault	49(30%)	4(3%)	110(68%)	163
Sex Assault	3(16%)	2(11%)	14(74%)	19
Robbery	3(12%)	4(16%)	18(72%)	25
Theft	36(16%)	16(7%)	181(78%)	233
Auto Theft	7(21%)	3(9%)	24(71%)	34
Burglary	62(22%)	23(8%)	198(70%)	283
Vandalism	11(10%)	10(9%)	89(81%)	110
Trespass	14(37%)	3(8%)	21(55%)	38
All Crimes	**185(20%)**	**65(7%)**	**655(72%)**	**905**

As shown in Table 3, the comparison of initial and final classifications for the 883 encounters initiated by radio calls that did involve a crime discloses 165 cases that were dispatched as non-criminal events (i.e.,165 false negative errors). As was the case with false positive and misclassification errors, the pattern of false negative errors varied across crime types. As indicated in Table 3, PSS officers responded to no calls that involved vehicle thefts that were initially classified as non-criminal matters, but a large portion of the assault (40%) and trespass (39%) cases they encountered at radio calls were broadcast as non-criminal events.

Table 3: False Negative Errors by Crime Type (Based on 883 Dispatched Calls that Actually Involved Crimes)

	False Negative	Correct	Total Calls
Assault	79(40%)	119(60%)	198
Sex Assault	4(20%)	16(80%)	20
Robbery	2(8%)	23(92%)	25
Theft	23(11%)	195(89%)	218
Auto Theft	0	27(100%)	27
Burglary	14(6%)	206(94%)	220
Vandalism	26(20%)	105(80%)	131
Trespass	17(39%)	27(61%)	44
All Crimes	**165(19%)**	**718(81%)**	**883**

A key question about the non-reported crimes and errors due to citizen inaccuracy is the extent to which they actually bias CFS crime counts. If the errors are distributed randomly or evenly such that false positives and misclassification errors that overstate the seriousness of crime are comparable in frequency and magnitude with non-reported crimes, false positives, and misclassifications that understate the seriousness of crimes, they will cancel each other out in aggregated CFS data. In other words, it is *theoretically* possible that inaccuracies in dispatch data might not actually bias CFS crime counts.

Additional analysis was conducted to assess how the various inaccuracies observed in the current data affect the dispatch-based crime counts. First, for each of the eight crime types, the encounters that were initiated by radio calls that actually involved criminal activity were compared with the encounters initiated by radio calls that were initially broadcast as crimes. This procedure yielded error counts among the encounters that were initiated by citizens' calls-for-service wherein false positive and over count misclassification errors cancel out false negative and undercount misclassification errors. Second, a similar procedure was conducted that compared *all* encounters that actually involved some criminal activity (i.e., both dispatched and non-reported crimes) with those encounters initiated by radio calls that were initially classified as involving some crime. This procedure yielded error counts for the entire 5,688 encounters wherein false positive and over count misclassification errors cancel out false negative errors, undercount misclassification errors, and non-reported crimes.

Table 4: Error Count in Dispatch Data by Crime Type (Based on Encounters Initiated by Citizens Calls for Service)			
	Encounters Dispatched as Crimes	Dispatched Criminal Encounters	Mis-count in Dispatch Data
Assault	163	198	-35 (-21%)
Sex Assault	19	20	-1 (-5%)
Robbery	25	25	0 (0%)
Theft	233	218	+15 (+6%)
Vehicle Theft	34	27	+7 (+21%)
Burglary	283	220	+63 (+22%)
Vandalism	110	131	-21 (-19%)
Trespass	38	44	-6 (-16%)
All Crimes	**905**	**883**	**+22 (+2%)**

Table 4 above summarizes the first comparison. Here, the calls that were dispatched as criminal matters are presented in the first column of numbers, dispatched calls that actually involved some criminal activity are presented in the second column, and a count of the difference between encounters dispatched as crimes and dispatched encounters that actually involved crimes is presented in the final column. The first point of interest in this table is that the overall difference between dispatched and actual crime calls for service is rather modest; the count of total calls broadcast is just 2% higher than the count of dispatched calls that involved some crime (22/905). The second point of interest is that this difference does not hold across the eight crimes types. While the dispatch and actual robbery counts are a perfect match, there is substantial discord between dispatch and actual counts of the other types of crimes.

The discrepancies are greatest for assaults and burglaries: The number of calls broadcast as assaults *undercounts* the number of calls for service that actually involved assaults by 21%, while dispatch numbers *over count* burglaries by 22%. In sum, the current data indicate that *among encounters that were initiated by citizens calls for service*, dispatch records do not provide accurate counts of many types of crimes. The data further show that both the magnitude of the inaccuracies and the direction they take (i.e., over vs. undercount) vary substantially across types of crime.

Table 5: Overall Error Count in Dispatch Data by Crime Type (Based on all Encounters)

	Encounters Dispatched as Crimes	Encounters Involving Crimes	Mis-count in Dispatch Data
Assault	163	282	-119 (-60%)
Sex Assault	19	31	-12 (-63%)
Robbery	25	32	-7 (-28%)
Theft	233	272	-39 (-17%)
Vehicle Theft	34	46	-12 (-35%)
Burglary	283	273	+10 (+ 4%)
Vandalism	110	157	-47 (-43%)
Trespass	38	78	-40 (-105%)
All Crimes	**905**	**1171**	**-266 (-30%)**

As shown in Table 5 above -- which displays the comparison of dispatched and actual crime numbers for *all encounters, regardless of how they were initiated* -- the pattern of bias is even more pronounced when one accounts for the fact that many crimes come to the attention of the police via means besides citizens' calls to police dispatch centers. In table 5, the first column of numbers includes the calls that were dispatched as criminal matters, the second column contains the sum of dispatched calls that actually involved some criminal activity *and* non-reported crimes, while the final column presents a count of the difference between encounters dispatched as crimes and encounters that actually involved crimes.

The first point of interest in Table 5 is that the total number of calls that were dispatched as crimes is 30% lower than the actual number of total crimes that officers encountered. In other words, the calls for service records of the agencies that participated in the PSS undercount by nearly one-third the total number of the eight types of crimes under study that PSS officers encountered on patrol. The second point of interest is that the undercount held for all crimes except burglaries, which had a 4% over count. More striking still where the undercount pattern is concerned is that for several types of crimes the scope of the undercount is substantially greater than the overall rate of 30%; officers encountered 35% more vehicle thefts, 43% more vandalisms, 60% more assaults, 63% more sexual assaults, and 105% more trespass cases than were broadcast. In sum, except for burglaries, the PSS data indicates that dispatch records provide counts of crime that are substantially biased by a systematic undercounting of the number of crimes that officers encounter on patrol and that the magnitude of the undercount varies substantially across types of crimes.

ERRORS IN NEIGHBORHOOD CONTEXT

The primary empirical end to which CFS advocates have used dispatch data is the study of neighborhood-level correlates of crime (e.g., Bursik, Grasmick, and Chamlin, 1990; Warner and Pierce, 1993). That the PSS data show substantial discrepancies between crime counts based on radio calls and criminal activity that officers encounter in the field, however, raises concerns about using dispatch data for this purpose. If errors in neighborhood-level CFS crime counts are associated with other neighborhood-level factors, then research using dispatch data to examine the correlates of neighborhood crime rates may produce misleading results. Attention now turns to assessing how the errors observed in the PSS data are associated with other features of the five-dozen neighborhoods where the study was conducted.

The first step in this aspect of the investigation was to calculate separate net error counts for each of the 60 study neighborhoods. Many of the encounters in the PSS occurred outside the bounds of these neighborhoods, however, so the number of cases available for calculating neighborhood-based net error counts is substantially lower than the number of cases that have been used thus far in the analysis. For example, only 526 of the 905 calls initially classified as crimes occurred in study neighborhoods. The small number of cases of cases per neighborhood (X=8.8), coupled with the small N for many of the eight types of crimes, precludes the development of meaningful neighborhood-specific net error counts for each crime type.

Consequently, the neighborhood-specific error counts developed were based on the eight types of crimes pooled together.

Two such counts were produced for each neighborhood; one that pertained only to those encounters that were initiated by radio calls, and one that pertained to all encounters, regardless of how they were initiated. The first was obtained by subtracting the number of false positive errors in each neighborhood from the number of false negative errors in each neighborhood among those encounters that were initiated by radio dispatches. The second neighborhood-specific error count was obtained by subtracting the number of false positives from the sum of the number of false negatives and non-reported crimes among all encounters.

Because PSS observers accompanied officers patrolling each neighborhood for 15 shifts, the numbers produced from the above-described procedures represent the net number of errors in each neighborhood *per 15 officer-shifts*. Because the number of false positive errors can be larger than the number of false negative errors (and non-reported crimes) for a single neighborhood, these numbers can take both negative and positive values. For the count pertaining to dispatched encounters, positive values indicate more false positive errors as compared to false negative errors, negative values indicate the converse, while zeros indicate that the errors cancel each other out. For the count pertaining to all encounters, positive values indicate that the number of false positive errors is greater than the sum of false negative errors and non-reported crime, negative values indicate the converse, while zeros indicate that the two sets of numbers cancel each other out.

Table 6: Frequency Distribution of Net Errors by Neighborhood (Among Encounters Initiated by Radio Calls)		
Value	Frequency	Percent
4	2	3
3	5	8
2	4	7
1	16	27
0	14	23
-1	13	22
-2	3	5
-3	0	0
-4	2	3
-5	0	0
-6	1	2
Total	60	100

As shown in Table 6 above, the neighborhood-specific net error count among radio-dispatched encounters ranges from 4 to -6. In other words, among those encounters initiated by radio calls, the largest net over count of crime in dispatch records for any single neighborhood during 15 officer-shifts was four, while the largest net undercount was six. Similarly, the range of 2 to -9 in Table 7 below indicates that among all encounters, regardless of how they were initiated, the largest net over count was two, and the largest net undercount was nine. Where the shapes of the distributions of the two measures are concerned, they are likewise notably different: The false positive and false negative errors in the dispatch-only count either canceled each other out or came within a single case of doing so in a substantial majority (43) of the cases, indicating that most of the neighborhoods had little of no net error among the encounters that were initiated by radio calls during the 15 officer-shifts studied. Conversely, where the measure that included all encounters is concerned, fewer than half (26) of the neighborhoods had either no net errors or a single one. Thus, the majority had a net positive error score of two or higher, indicating that dispatch data would provide notable undercounts of the total number of crimes that occurred during the 15 officer-shifts examined in the PSS research.

Table 7: Frequency Distribution of Net Errors by Neighborhood (Based on all Criminal Encounters)

Value	Frequency	Percent
2	2	3
1	6	10
0	6	10
-1	13	22
-2	11	18
-3	4	7
-4	3	5
-5	3	5
-6	7	12
-7	3	5
-8	1	2
-9	1	2
Total	60	101

With the two neighborhood-specific error counts in hand, attention turns to examining how they are associated with several features of the study neighborhoods: their socio-demographic composition, residents experiences with and attitudes about crime, and residents experiences with and attitudes about calling the police. To facilitate this, the following neighborhood-level measures were obtained by aggregating the responses of the 200-odd residents in each study neighborhood who participated in the aforementioned (on page six) survey that was conducted as part of the PSS:

1) *Age*. The mean age of the respondents.

2) *Sex*. The proportion of female respondents.

3) *Race*. The proportion of white respondents.

4) *Education*. The mean number of years respondents attended school.

5) *Income*. The mean of seven response categories that grouped annual family income into $5,000 dollar blocks, with below $5,000 as the lowest response category, and above $30,000 the highest.

6) *Housing stock*. The proportion of domiciles that are single family residences.

7) *Home ownership*. The proportion of respondent families who own, or are buying, their domicile.

8) *Residential stability*. The mean number of years that respondents have lived in the study neighborhood.

9) *Victimization rate*. The mean number of criminal victimizations that respondents reported experiencing in the study neighborhood during the 15 months preceding the survey.

10) *Fear of crime*. An index crafted from respondents' estimates of the likelihood that they will be robbed, that their residence will be burglarized, and that their residence will be vandalized.

11) *Likelihood of calling the police.* The mean of respondents' answers to a question about the likelihood that residents of their neighborhood will call the police when they observe suspicious activity.

12) *Calls to the Police.* The mean number of times respondents reported calling the police during the preceding year.

13) *Speed of police response.* The mean of respondents' answers to a question about how rapidly the police in their neighborhood typically respond to calls for service.

The zero-order correlations between each of these measures and the two error counts are displayed in Table 8 below. Before turning to these associations, however, some comments about the relationship between the two error counts are in order. The strength of the association between these two measures ($r = .63$; $p<.05$) indicates that while there is a strong relationship between the two error counts, they share only about 40% of their variance in common. This illuminates an important point that was implicit in the above discussion of the data presented in Tables 6 an 7: The picture of neighborhood-specific errors in dispatch data differs according to whether the indicator of errors one uses accounts for all crimes that come to the attention of the police, or whether it accounts only for those crimes that are initiated by radio calls.

As one would expect given the relationship between the two error counts, the pictures they provide of how errors are associated with other neighborhood characteristics converge at some points and differ at others. The dispatch-only count is significantly (alpha=.05) associated with age, both housing measures, victimization level, and fear of crime, while the overall error count is a significant correlate of the victimization level, fear of crime, and the speed of police response. The significant

Table 8: Correlation Matrix of Neighborhood Characteristics

	1)	2)	3)	4)	5)	6)	7)	8)	9)	10)	11)	12)	13)	14)	15)
Disp. Errors	--														
All Errors	.63*	--													
Age	.31*	.11	--												
Sex	.07	.03	.25*	--											
Race	-.01	.11	.08	-.20	--										
Education	-.00	.20	-.28*	-.33*	.67*	--									
Income	-.00	.12	-.29*	-.33*	.49*	.75*	--								
Housing Stock	.26*	.21	.11	.00	.41*	.45*	.47*	--							
Home Ownership	.30*	.22	.05	.04	.30*	.38*	.41*	.95*	--						
Pop. Stability	.10	-.20	.51*	.14	-.14	-.51*	-.14	-.07	-.07	--					
Victimization	-.34*	-.25*	.39*	-.13	-.23	-.23	-.11	-.29*	-.34*	-.17	--				
Fear of Crime	-.32*	-.34*	-.15	-.01	-.59*	-.59*	-.48*	.55*	-.57*	-.04	.75*	--			
Call Odds	.07	.14	.01	-.31*	.63*	.69*	.68*	.59*	.55*	-.09	-.40*	.64*	--		
Call Police	-.02	.03	-.35*	-.58*	.28*	.38*	.25*	.05	.12	-.22	-.05	.28*	.33*	--	
Response Time	-.17	-.28*	-.03	.35*	-.55*	-.66*	-.48*	-.38*	-.42*	.11	.52*	-.71*	-.70*	-.52*	--

* p < .05

zero-order correlations indicate that dispatch records are less likely to undercount and more likely to over count the number of crimes *that officers encounter at radio calls* in neighborhoods with older residents, more single-family dwellings, and higher home ownership rates, while they are more likely to undercount and less likely to over count crimes *at encounters that were initiated by radio calls* in neighborhoods with higher victimization levels and where residents are more fearful of crime. Where the overall error count is concerned, the zero-order correlations indicate that dispatch records are more likely to undercount and less likely to over count *the total number of crimes that come to the attention of the police* in neighborhoods with higher victimization levels, where residents are more fearful of crime, and where residents believe that the police respond more slowly to their calls for service.

DISCUSSION AND CONCLUSION

The re-analysis of data from the PSS undertaken here raises several questions about the validity citizens' calls to police dispatch centers as an indicator of crime. First, because dispatch records, by definition, count only those crimes that citizens report to the police by telephone, they do not include crimes that come to the attention of officers by other means. In the current data, study officers encountered substantially more -- 30% -- crimes than were called-in to police dispatch centers. Moreover, the data indicate that there is a remarkable degree of variability in the extent of the discrepancies between dispatch crime counts and what officers encounter across types of crimes. For example, while the dispatch figures undercount most types of crime -- by

105% for trespass, 60% for assaults, and 17% for thefts -- they actually over count by four percent the number of burglaries that officers encountered. This variability in the direction and magnitude of error across crime types is also evident when the crimes under study are limited to those that come to patrol officers attention via calls for service. For example, the number of assaults that officers encountered at radio calls is 21% greater than the number of radio calls broadcast as assaults, while the number of radio calls broadcast as burglaries overstates the number of burglaries that officers encountered at radio calls by 22%.

The variability in errors across types of crimes has important implications for understanding neighborhood-level error counts and how they are associated with other neighborhood characteristics. Recall that the neighborhood-level error counts indicate the number of net errors across all eight types of crimes. Because the error rates for specific crimes in the entire sample were often markedly different from the rate for all crimes pooled together, it is possible that the associations between neighborhood-level error counts for specific crimes and other aspects of neighborhoods are markedly different from the correlations reported above. For example, it may be that the observed tendency for dispatch data to undercount crimes more substantially in neighborhoods where residents are more fearful of crime holds only for some crimes, or that it is stronger for some crimes than for others. Conversely, it may be that some type(s) of crimes are significantly associated with some neighborhood characteristic(s) that were not associated with the error counts based on all crimes pooled together.

The fundamental reason why the current study was unable to examine how error counts for specific crimes are correlated with other neighborhood characteristics is the limited number of shifts that PSS observers spent with officers patrolling each study neighborhood; 15. This translates into five 24-hour officer-days, or 1/73 of a full years-worth of coverage by a single patrol unit for each study neighborhood. This has important implications for the current findings regarding the neighborhood-level correlates of the overall error counts reported herein. It is impossible to know the degree to which the counts developed from the 15 shifts accurately represent the nature of net overall errors based on all crimes in the study neighborhoods during the mid-1970s. It is almost certain that the raw neighborhood-specific net error numbers that would be observed over a markedly longer period of time would be substantially different from the numbers observed in the current study. On the other hand, the general shape of the distribution of overall errors across neighborhoods may be quite similar to what was observed in the current data. If so, the correlations reported in Table 8 would accurately reflect how dispatch errors (for all crimes pooled) in the jurisdictions where the PSS was conducted are associated with other features of the study neighborhoods. Conversely, the error counts may misrepresent the true nature of how overall errors are distributed across the study neighborhoods, which means that the observed associations do not provide a sound picture of the relationship between errors and other features of the study neighborhoods. In sum, due to the limited number of shifts per-neighborhood, the current findings about how errors in pooled dispatch data are correlated across physical space must be viewed as tentative.

While limitations in the PSS data render the current findings tentative, the fact that the study disclosed gross discrepancies between crime counts based on citizens calls to dispatch centers and what officers actually encountered in the field, and that these discrepancies are correlated with several features of neighborhoods raises serious questions about the validity of dispatch records as a measure of crime. Thus this re-analysis of data from the PSS points to the need for additional research on the nature and consequences of errors in CFS data. Such research should pursue at least two different avenues. *First,* researchers must develop more complete knowledge of the scope and extent of measurement errors in CFS crime counts. The present study indicates that CFS data typically underestimates the number of crimes that come to the attention of the police, but that the degree of undercount varies substantially across crime types (from 17% to 105%) and that CFS data actually *over counts* some types of crime (i.e., burglary). Each year police departments across the U.S. receive tens of millions of calls for service. The current study is based on less than 3,000 calls. Additional research (hopefully based on larger samples) can provide important information about the degree to which the current findings accurately represent how errors in CFS crime counts vary across types of crime.

Second, research must more thoroughly examine the relationship between CFS errors and the characteristics of spatial aggregates. As alluded to above, we need data based on more than the five officer-days per neighborhood than was available here in order to thoroughly asses how errors in dispatch data -- both overall and crime-specific -- are correlated with other features of neighborhoods. Moreover, we need to explore the problem of correlated error in other geographic aggregates. As noted at the outset of the this report, researchers have advocated the use of CFS data to measure crime in spatial units besides neighborhoods (i.e., individual addresses [Sherman, Gartin, and Buerger, 1989] and cities [Bursik and Grasmick, 1993]) . Bailey's (1985) research has shown, however, that the correlations between crime rates and other social phenomena are often different at different levels of spatial aggregation (i.e., states, SMSA's, and cities). It may be that the correlations between errors in CFS crime counts and other variables likewise vary across levels of spatial aggregation. Consequently, future research must examine this possibility.

To conclude, the current study raises serious questions about the validity of calls for service data as an indicator of crime, which only future research can resolve. Thus while its advocates assert that CFS data provides a useful alternative to the standard measures of crime, this report raises a caution flag and indicates that a good deal of additional research is needed before criminal justice practitioners and researchers wholeheartedly embrace calls for service as an empirically valid measure of crime.

BIBLIOGRAPHY

Bailey, Leroy, Thomas F. Moore. And Barbara A. Bailer.
1978 An interviewer's variance study of the national crime study city sample. Journal of the American Statistical Association 73: 16-23.

Bailey, William C.
1985 Aggregation and disaggregation in cross-sectional analyses of crime rates: The case of states SMSA's and cities. Paper presented at the Annual Meetings of the American Society of Criminology, San Diego, CA.

Biderman, Albert D. and James P. Lynch.
1991 Understanding Crime Statistics: Why the UCR Diverges From the NCS. New York: Springer-Verlang.

Black, Donald J.
1970 Production of crime rates. American Sociological Review 35:733-748.

Bursik, Robert J., Harold G. Grasmick, and Mitchell B. Chamlin.
1990 The effect of longitudinal arrest patterns on the development of robbery trends at the neighborhood level. Criminology 28:431-450.

Bursik, Robert J. And Harold G. Grasmick.
1993 The use of multiple indicators to estimate crime trends in American cities. Journal of Criminal Justice 29:509-516.

Reiss, Albert J.
1971 The Police and the Public. New Haven: Yale University Press.

Sherman, Lawrence W., Patrick R. Gartin, and Michael W. Burger.
1989 Hot spots of predatory crime: Routine activities and the criminology of place. Criminology 27:27-55.

Warner, Barbara D. And Glenn L. Pierce.
1993 Reexamining social disorganization theory using calls for service as a measure of crime. Criminology 31: 493-518.

Whitaker, Gordon P.
1979 Methods Report MR-20. Bloomington, Ind., Workshop in Political Theory and Policy Analysis.

UNIVERSITY *of* HOUSTON

Department of Sociology Houston, TX 77204-3474

June 3rd, 1996

Jeff Ross, Ph.D.
United States Department of Justice
Office of Justice Programs
633 Indiana Avenue
Washington, DC 20531

Dear Dr. Ross:

Enclosed please find two copies of the final report for my grant *Investigating the Scope of Measurement error in Calls-For-Service as a Measure of Crime* (Grant# 95-IJ-CX-0023). Thank you for your assistance during the past year. I enjoyed working with you.

Sincerely,

D. Klinger

David Klinger
Assistant Professor