In Mike Hough and Mike Maxfield (eds.)
Surveying Crime in the 21st Century
Crime Prevention Studies, vol. 22
Monsey, NY: Criminal Justice Press, 2007.

Survey Assessments of Police Performance

by

Wesley G. Skogan
Institute for Policy Research
Northwestern University

Abstract: *This article considers the British Crime Survey (BCS) as a vehicle for monitoring police performance. The BCS has two complementary foci on the police: monitoring reports of* general confidence *in the police and* tracking encounters *between police and the public. These raise substantive and methodological issues, and have implications for survey design. Among these are response validity, or the issue of whether "confidence" questions actually reflect the quality of policing on the ground. We also need to ensure that the measurement process measures what it does with maximal accuracy. This paper reviews validity and reliability issues in the context of assessing general confidence and tracking public encounters with the police. It calls for a program of methodological research to document the error structure of the data and guide improvements and decisions about key features of the survey.*

"BUT THEY ARE JUST PERCEPTIONS"

One objection to the use of surveys to monitor public confidence in police performance in any official way is that "they are just perceptions." The obvious rejoinders that perceptions are important and that they are real

Crime Prevention Studies, volume 22 (2007), pp. 165–181.

in their consequences are certainly true, but those are different matters, and the complaint raises serious issues. At root, it is a response validity question, and the query that needs answering is, "Do survey assessments of the quality of policing measure what they claim to measure?" There are reasons to believe that a significant component of the general confidence in police that is measured by surveys is only tangentially related to the empirical reality of the moment, so the question cannot be put off. If policy makers are seeking survey data to answer the question "How are we doing?" it is not clear that they will always be getting a meaningful answer. After reviewing some of this evidence, I propose that a performance assessment survey is better off focusing on change measures rather than on level measures, a decision with survey design and cost implications.

Validity of Confidence Measures

The research literature provides ample challenges to assumptions about the validity of measures of the general confidence people have in the police. One challenge is issued by research on public encounters with the police (which constitutes an important component of what the police do on the street), and the impact of those experiences on general confidence measures (which are to provide an answer to the "How are we doing?" question). Another challenge comes from research comparing public opinion with independent measures of the quality of policing. Both point to shortfalls in the interpretability of general confidence measures.

Causal Order

An important question is whether experiences that people have recently had with police affect their confidence in the police strongly enough that general confidence measures can be used as an indicator of on-the-ground performance, and especially recent improvements in performance. Research suggests that this is not necessarily the direction in which the causality of attitudes toward the police runs. A study addressing the issue of causal direction found that the effect of recent experiences with the police (as recalled in a survey) on their general confidence was not very large. A counter effect – that of people's general views of the police on how they interpreted their recent experiences – was stronger (Brandl, Frank, et al., 1994). In the authors' view, people stereotype the police and selectively perceive even their own experiences. Although Brandl and

colleagues do not make the link, since their 1994 article social psychologists have begun to stress a parallel hypothesis: that general, preexisting attitudes strongly affect how people interpret their experiences. They find that individuals read their experiences in the light of their prior expectations, perhaps more than specific recent experiences affect their expectations. In any slice of time, general attitudes affect people's evaluations of their experiences, rather than the reverse. If further research on the direction-of-causality issue (which is greatly needed) indicates that this is the case, it challenges the assumption that confidence measures are useful for answering the question "How are we doing?"

Asymmetry of Impact

There is also evidence that any impact of the public's direct experiences on their confidence in the police is asymmetrical. That is, police get little or no credit for delivering professional service, while bad experiences deeply influence views of their performance and even legitimacy. This brings into question whether positive experiences that people have recently had with police affect their confidence to such an extent that confidence measures can be used as an indicator of performance quality.

I examined the asymmetry issue in a recent article on public contacts with police (Skogan, 2006a). It is based on self-report survey data from Chicago assessing police-initiated and citizen-initiated contacts with police. Like the BCS, the survey screened for recent encounters and asked six follow-up questions about the nature of any contacts. Overall, about 20% of adults recalled being stopped by the Chicago police in a year, and half reported contacting them about some matter. A majority of these respondents recalled a favorable experience, even when they were stopped. The study addressed the relationship between the positive or negative character of those experiences and a six-question index of confidence in police performance and their effectiveness in addressing community problems. Statistically, the impact of having a bad experience was 4 to 14 times as great as that of having a positive experience. Worse, the coefficients associated with having a good experience – including being treated fairly and politely, and receiving service that was prompt and helpful – were not statistically different from zero. Bad police-initiated and bad citizen-initiated encounters both had large negative consequences, with the latter actually having a stronger effect. This does not mean that bad police work, as assessed from the public's side of encounters, was common; on the

contrary, most respondents gave the police high marks even in police-initiated traffic and foot stops. However, all of this good police work counted for little, when it came to the public's expressions of confidence in the police.

So strong was the asymmetry in the Chicago data that in another section of the 2006 paper I replicated this finding using surveys of residents of seven other urban areas located in three different countries, including a BCS urban subsample for England and Wales. The pattern was everywhere the same, in places ranging from St. Petersburg in Florida to St. Petersburg in the Russian Federation. It is consistent with findings reported for the British Crime Survey, in which the most favorable attitudes are reported by those who have had *no* recent contact with police; those who had contacts of any kind are generally unfavorable (Allen, Edmonds, Patterson, & Smith, 2006). If good practice is not reflected straightforwardly in general confidence measures, this also challenges the assumption that confidence measures answer the question "How are we doing?"

Confounding Variables

Another reason for skepticism is my experience in rating the quality of community policing programs in Chicago's neighborhoods. There, confounding variables apparently overwhelmed any evidence of direct program effects on public opinion. This study concluded that public opinion did not reflect in any straightforward way the quality of policing that we had directly observed on the ground. This also questions the assumption that confidence measures answer the question "How are we doing?"

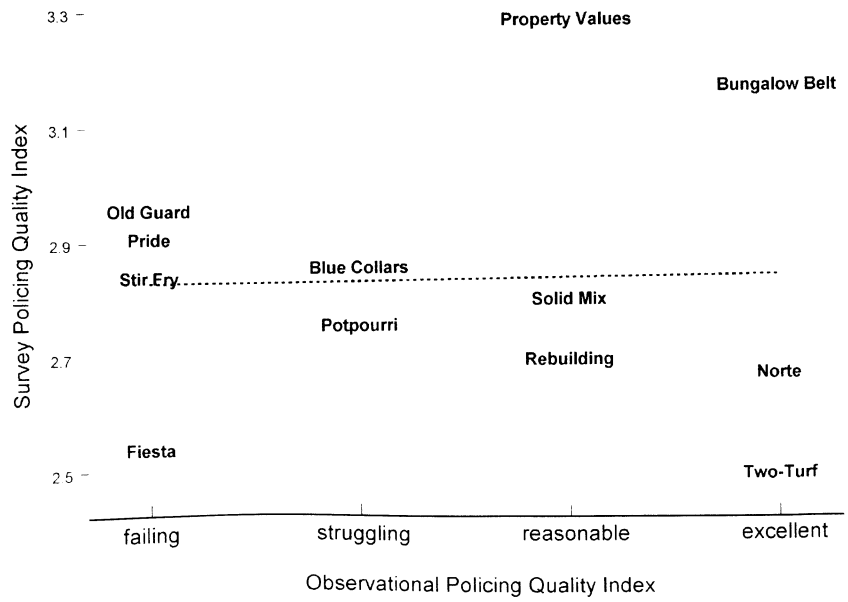
In the study, we conducted fieldwork in 12 police beats. The areas were chosen to reflect a broad range of race and class configurations. We interviewed police officers, attended beat meetings and surveyed those in attendance, rode along with officers on patrol, reviewed their formal plans identifying priority issues, and discussed the local situation with community informants. Our community policing quality ratings were developed along four major dimensions: the commitment and support of district management, beat team leadership, beat team activities, and the efforts of the district's specialized neighborhood relations unit to mobilize residents in support of community policing projects. Ratings were developed on 4 to 10 subdimensions for each of these, and a cluster analysis based on all of the ratings was used to construct a summary, four-category ranking of each beat. We labeled the twelve programs either "excellent," "reasonable,"

"struggling," or "failing." All of the details regarding the study are presented in Skogan, Hartnett et al., 1999.[1]

In parallel, we conducted a telephone survey of the study areas. A total of 1,290 households were surveyed, including at least 100 respondents in each of the 12 beats. Respondents were selected from telephone listings, and the response rate for the survey was 78%. Interviews were conducted in English and Spanish, a must in Chicago. Among the questions, respondents were asked to rate police working in their neighborhood in terms of their "responsiveness to community concerns," "dealing with the problems that really concern people in your neighborhood," and "working together with residents in your neighborhood to solve local problems." These questions targeted key goals of the city's community policing program. In each case, police were rated as "very good," "good," "fair," or "poor." Responses to these three questions were highly interrelated, and were combined together to produce an overall opinion index.

The survey results were unrelated to our laboriously arrived-at on-site observations and rankings, which are illustrated in Figure 1. On the

Figure 1: Observational and survey measures of policing quality.



horizontal axis, it arrays police units from "failing" to "excellent," based on the fieldwork. On the vertical axis, it arrays beats by the average rating score awarded by residents in our beat surveys. The names we gave the areas (all described in detail in Skogan et al. [1999]) are presented to mark where each beat fell. Figure 1 also presents a regression line, which indicates that there was no relationship between the two measures.

It is apparent in Figure 1 that opinion was not clearly linked to ratings of police on the ground. The best policing was being delivered in Two-Turf, a name we bestowed on this beat because half its residents were Mexican and half Puerto Rican (but all were poor). Police were also doing an excellent job in Bungalow Belt (a White, blue-collar area) and Norte (a stable, largely Puerto Rican neighborhood). Resident ratings of those three areas were all over the map, however, ranging from the lowest to the second-highest in the study. On the other hand, policing in Stir Fry stunk; it was our lowest-rated beat. The population there was evenly divided among Whites, African Americans, Asians, and Latinos. Policing was not much better in Pride, one of the most upscale Black communities in Chicago, and in Old Guard, which was also a middle-class African American enclave. Pride and Old Guard were the third- and fourth-rated policing operations in the eyes of the public, however. Fiesta, where police could care less about the community, was completely Mexican-American in character. Yet except for Fiesta, public ratings of the police working in these areas were relatively positive.

As this sociological tour suggests, what was really affecting opinion rankings were race and class. Property Values and Bungalow Belt were our only predominately White areas, and Old Guard and Pride were certainly the equal of Bungalow Belt when it came to their middle-class status. Their's were the four most highly rated policing units, based on public opinion. The lowest opinion ratings were from residents of three almost uniformly Latino (and mostly low-income) beats, and from poor African Americans living in Rebuilding.

Although a modest study, it may be the only one that holds popular views of the police up against an independent measure of the quality of service being delivered, in order to document the relationship between them. If our field ratings are taken as the gold standard against which to assess the convergent validity of the survey measure of policing quality, it was zero. Public opinion was measuring something, but it was not policing on the ground.

In a report published by the Home Office, I noted in the introduction that "[T]his report focuses on people who have had contact with the police because those people's experiences and opinions should give a better basis for assessing the police's efforts" (Skogan, 1994, pp. 1-2). However, direction-of-causality issues, strong asymmetry in the impact of the good and bad service delivered by police, the zero credit that the public gives to police actions even when they themselves describe them in positive fashion, and – on the evidence of the observational study – the lack of fit between independent measures of policing quality and public opinion say otherwise. They provide fuel to the view that survey-based confidence measures do not, without considerable massaging, adequately represent the reality of policing on the ground at the time. These findings are also bad news for police administrators intent on solidifying their support among voters, taxpayers, and the consumers of police services. The message of the asymmetry finding is, unfortunately, "You can't win, you can just cut your losses." No matter what you do, it only counts when it goes against you (Skogan, 2006a, p. 119).

None of these validity issues is unique to the police. There has been a tremendous amount of research on the relative merits of objective versus subjective measures of service quality by students of "urban service delivery," a subcategory in the field of public administration. They still have not figured it out. In 2003, Kelly issued a discouraging summary of the state of the art: "[D]ecades of research on citizen satisfaction have not yielded a decisive answer to the validity question – whether citizens' evaluations of service quality reflect objective changes in service quality" (Kelly, 2003, p. 857). Research cited in my 2006 (Skogan, 2006a) article indicates that asymmetry of effects is the rule rather than the exception when it comes to subjective measures of the quality of service delivered by government agencies. In the case of the police, we are interested in utilizing subjective measures of performance quality because survey indicators seemingly go directly to the heart of our concerns about policing: that it is fair and effective, and reinforces or rebuilds their legitimacy. Most of the quality measures that fall out of police information systems, on the other hand, concern inputs (money, people), activities (how fast they drive to crime scenes, in order to fill out reports), and intermediate outputs (cases solved), but they do not tell us much about important outcomes that we seek to achieve. This does not mean, however, that subjective measures can be accepted at face value.

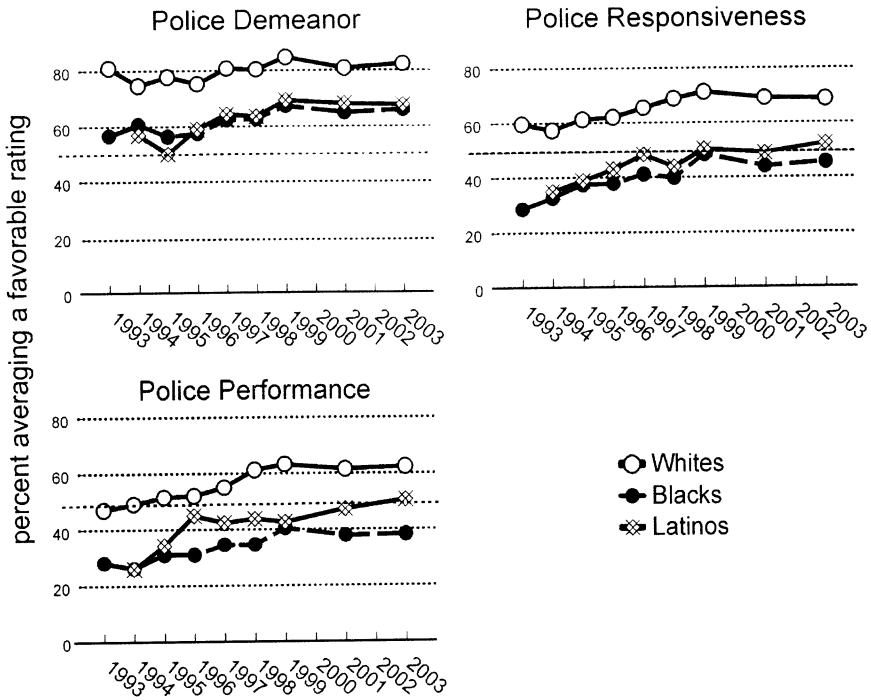
Assessing Changes in Confidence in Police

One way out of the "they are just perceptions" box may be to shift focus from *levels* of confidence to *changes* in confidence over time. With an appropriate survey design, changes over time can be benchmarked against the level of early measures in the series. Earlier measures should incorporate (for example) stable race and class differences in area predispositions toward the police. This would highlight later changes in assessments of the nature and quality of policing. Focusing on changes in confidence reflects the logic of evaluations of innovations in policing, which typically gather "before" measures that provide benchmarks for assessing shifts in "after" measures, because the first-wave measures incorporate many potentially confounding factors. Note that this recommendation is not low cost. It raises the stakes considerably in terms of the frequency with which areas should be revisited. It also has implications for local-area sample sizes, because the design would need to have the capacity of reliably identifying change over time. The BCS is already producing some area estimates of confidence in police, including for 10 Government Office Regions (Allen et al., 2006) and for 42 subregions within them (Nicholas, Povey, Walker, & Kershaw, 2005). It might be possible to push further down the geographical ladder for confidence measures, but monitoring encounters with the police at the local level – an issue to be considered in the next section – will be more demanding.

A Chicago Example

The BCS is not conducted in the context of field experiments, however, and does not require a research design that aides in inferring causation. Instead, it could become an exercise in repeated measures at the area level. An example of how opinions can change in a context in which police are innovating rapidly can be found in the results of surveys monitoring public attitudes in Chicago. An evaluation of community policing monitored public opinion in a series of surveys conducted between 1993 and 2003 (Skogan, 2006b). Relying on one or two very general satisfaction questions would reveal little about the shortfalls and accomplishments of the police; rather, I worked out a set of questions that tapped the major goals of the department for this and related projects. The resulting 10 questions balanced coverage of the substantive domains of interest and scarce questionnaire space. Because single-item measures of attitudes are mostly dominated by error variance, I strove to have at least three questions addressing

Figure 2: Trends in confidence in Chicago police.



each domain. The resulting subscales were quite reliable in light of that still limited number.

As illustrated in Figure 2, during the evaluation period there were observable changes for the better in perceptions of several aspects of police service. Opinion improved steadily between 1993 and 1999, before leveling off at a new high in the after 2000. In addition, on every measure, positive changes in opinion were apparent among Whites, African Americans, and Latinos alike. At the same time, substantial gaps between the races were apparent in the first survey, and those gaps remained about the same despite the positive trends. The opinion gaps are equivalent to the race (and class) predispositions illustrated in Figure 1, except that they can be accounted for using repeated measures.

Chicagoans were asked to rate police using 10 questions representing three opinion dimensions. The first was *demeanor*, which was measured by responses to questions asking about the politeness, concern, helpfulness,

and fairness of police in their area. Police *responsiveness* was measured by responses to questions about their responsiveness to community concerns, "dealing with problems that really concern people in your neighborhood," and "working together with residents in your neighborhood to solve local problems." Police *performance* of their traditional duties was tracked by questions about how good a job they were doing preventing crime, helping victims, and "keeping order on the streets and sidewalks." Figure 2 charts the percentage of respondents averaging a positive rating (the two best of four rating categories) on each index. Separate trend lines are presented for Whites, African Americans, and Latinos. The 1993 survey was conducted only in English, so the data point for Latinos is omitted in that year.

Before community policing began, almost two thirds of the respondents already averaged a positive score on the police demeanor index. As in our neighborhood studies, there were notable differences between the races in how they initially rated the police. In general, Whites perceived that police treated people well even in the early 1990s, and there was not much room for improvement. However, positive perceptions of police demeanor rose by about 10 percentage points among both Latino and African American respondents and ended on a high note. Perceptions of police responsiveness to community concerns improved steadily until 1999; overall, the responsiveness index rose by nearly 20 percentage points during the 1993-1999 period. Perceptions of responsiveness went up most between African Americans and Latinos, rising by almost 20 percentage points between 1993 and 1999. The views of Whites, which were more positive even before the program began, improved by about 10 percentage points. Finally, at the outset Chicagoans were mostly negative in their views of how well police performed their traditional tasks. Yet over time, the index measuring this aspect of police service improved significantly, rising from a low of 36% in 1994 to a high of 51% in 2003.

Note again that, for all of these changes, the gaps between the races closed not at all. The earliest surveys provided a baseline documenting interfacial differences in opinions of the police that persisted until the very end of the series. Yet having taken that into account, positive trends in the data were also apparent.

Sample Size and Statistical Power

Shortchanging on area-level sample sizes would be a mistake. Other things being equal, the smaller the sample size, the larger any changes in reports

of policing quality would need to be in order to be statistically reliable. Small samples could raise the bar impossibly high for local police.[2] How big is big enough? A sample size sufficient for detecting a 10% shift in general confidence for an analytic area would be about 160 respondents in each wave.[3] A 10% shift in confidence over a multiyear time frame is a plausible one. For police responsiveness (see Figure 2), the average year-to-year shift in opinion in Chicago during the 1990s was about 5%. The 2003-2004 BCS sample targeted completing 600-700 sample interviews in each police force area (and 37,000 overall), which would have allowed making small-area estimates of general confidence for about four subareas within each police force area, as they were defined at that time. A subject for discussion is the administrative level of the police at which the findings of surveys would be of interest. The priorities from that discussion could be matched to these figures, to determine if the current survey could provide the desired coverage.

Of course, if analysts are interested in the views and experiences of subgroups, the sample requirements would be the same at that level. In Chicago the most important were Latinos who could not speak English and had to be interviewed in Spanish, and they constituted about 16% of the overall sample. Based on the same assumptions, to track a year-to-year change reliably in a group this size required overall samples of about 1,000 respondents in each wave. In any survey, subgroups that are targeted to be of analytic interest will have to be chosen judiciously. Two groups of great interest are people who have contacted the police or have been stopped by them; they are considered next.

"YOU CAN'T BE 'THE FRIEND OF THE PEOPLE' AND DO YOUR JOB"

What I hope my informant, the Chicago police officer who is quoted above, meant was that cool professionalism and an aloof manner, rather than the "customer-friendly" policing that was being discussed in the front of the room, was the appropriate demeanor for "real" police. His statement is a reminder that many officers do not aspire to be friendly and accessible, believing it puts them at risk on the street. However, the quality of service rendered when police and the public come in contact is one of the things that administrators can actually hope to control. Through their training and supervision practices, departments have some capacity to shape the relationship between residents and officers working the street. Whether

police are polite or abrasive, concerned or aloof, and helpful or unresponsive to the obvious needs of the people they encounter depends importantly on actions taken by department leaders. This led a National Research Council review panel to recommend more attention to what was dubbed "process-oriented policing" (Skogan & Frydl, 2004), in addition to community- and problem-oriented policing. The trick is to identify the dimensions of behavior that are important, and where police should fall on them.

Screening and Recall Issues

Currently, the BCS hands respondents a card displaying 17 common reasons that respondents could have contacted the police in the past 12 months, to help them recall incidents. On the police-initiated side, the survey asks about vehicle and pedestrian stops in the same recall period. Follow-up questions are then asked about the most recent of each kind of contact that respondents recall having had. This procedure produces a reasonable random sample of all experiences, and presumably the most recent are most likely to be fresh in respondents' minds. However, as I note further on, the domain of police-initiated encounters is a larger one. There exists a broad range of police-initiated interactions that probably fall below the threshold of being a stop, and whether or not they are included makes a very large difference in the yield of the screener.

A key issue here is that we have no idea what the optimal length of the recall period is for accurately remembering encounters with the police. The current period is the 12 months preceding the interview, a date which is read to each respondent as part of the question. The length of the recall period is one of the most important design decisions for the victimization component of surveys, and a large number of field tests were conducted to determine how far back in the past crime victims could be relied upon to remember events. It turns out that the "time-dependent forgetting curve" is very steep, with victimizations occurring further than 3 months in the past going unrecalled with alarming frequency (for a review of these studies see Skogan, 1981). There is no reason to think that the various ways and reasons for which people encounter the police are any different from victimization with regard to time-dependent forgetting. The recall length decision is fraught with cost implications. A tighter recall window probably will yield a smaller number of more accurately recalled encounters. To ask the same questions as in the past (did they think they were treated fairly? were they satisfied with the experience?) would almost surely

require larger samples due to the decreased yield (albeit greater accuracy) of each survey interview.

There is also no reason to think that the task of retrieving reports of encounters with police is immune from many of the other shortcomings of victimization measures. We know that victims underrecall events involving private or embarrassing circumstances, for example, when they were themselves culpable to some degree, when there was a relationship between the parties, when there was no insurance claim to be made, when alcohol was involved, and when the incident was not very serious. A long string of hypotheses about differential recall rates for different kinds of police-public encounters comes easily to mind. Research in North Carolina indicates that recall biases vary by race, with African Americans known to have been stopped reporting a smaller fraction of the criterion events, thus possibly underrepresenting the extent of racial disparity in police-initiated stops (Tomaskovic-Devey, Wright, Czaja, & Miller, 2006). Based on the "asymmetry" research described earlier, I fear that negatively rated encounters are much more readily recalled than positive rated ones, overrepresenting the extent to which they are not "the friend of the people."

The legitimacy of any official use of survey-based evaluative measures of police-public encounters could depend upon a program of research that identifies the error structure of the data, and ensures that the questionnaire design is optimal for maximizing the accuracy of the data. The most important tool for doing so would be reverse record checks. They would involve selecting samples of persons known to have encountered the police and then testing questionnaire drafts to determine the rate at which the criterion encounters are recalled in interviews. The encounter samples would be selected to represent dimensions along which we expect difficulties with recall – including the length of time in the past in which they occurred – across indicators of event seriousness, when respondents were themselves culpable, and other factors. Respondents would also be chosen to represent population groups presenting a range of challenges to response validity, for example, by immigration status and age.

Sample Size

Of course, a big impediment to producing reliable estimates of encounter characteristics for smaller geographical areas is sample size. As with victimization, features of encounters and those involved in them can only be ascertained among those recalling experiences at the screener stage. Special

challenges are posed by important but less frequent forms of encounters, including pedestrian stops and those involving relatively small population groups. In the United States, nontraffic police-initiated encounters very heavily target young minority males, a group that usually proves particularly elusive in surveys. Another key factor would be what questions about encounters the survey intends to answer; just knowing the percentages involved in them is not very revealing. An example considered further on is the rate at which individuals who are stopped under various circumstances are searched.

In any event, incorporating encounters with police into the area-level agenda of the BCS would doubtless call for scrapping of the current supplement-based policing module in favor of questionnaire items that would be administered to all main and booster sample respondents. In parallel with local-area victimization estimates, only the more frequent kinds of encounters between the police and the public could be examined.

When it comes to monitoring encounters, all sample-size calculations will begin with estimates of the proportion of respondents who will have experiences with the police to report upon. Currently the BCS finds that about 20% of respondents nationwide recall initiating an encounter with the police, and about 22% recall being stopped. This will vary by area – a large victimization survey estimated that 38% of Londoners sought contact with the police in a year's time (during 1999-2000), and 12% were stopped (Fitzgerald, Hough, Joseph, & Qureshi, 2002). Some of this difference is doubtless due to methodology. When the London study added being "approached" by the police to the "stopped" count, the police-initiated contact rate rose from 12% to 24%, and I think the latter definition more closely approximates that of the BCS.

The next most important subset of respondents is those who recall being searched during a police-initiated encounter. Searches raise the ante in police-citizen encounters. In studies of satisfaction with police, being stopped has a much smaller effect than the increment of dissatisfaction added by a search. Compared to stops, searches are much more tightly defined and monitored by the Police and Criminal Evidence Act. The London survey put the search rate at 37% of pedestrian stops and 22% of vehicle stops, and notes that these rates seemed much higher than elsewhere in Britain. Any special focus on persons who are stopped in any BCS police monitor would impose an even greater burden on the sample.

Because of the importance of certain low-frequency public experiences that are heavily concentrated among harder-to-reach populations, it might

also be worthwhile considering conducting supplemental interviews with samples of persons known to have encountered the police. This would involve stepping outside of the BCS sampling frame, into the world of jurisdiction-level "customer satisfaction" surveys. The long list of problems in conducting these surveys is outside of the scope of this chapter, but they too are legion.

Evaluative Content

Miller, Bland, and Quinton (2001) review British research on factors that influence how people evaluate encounters with police, and recommend a set of evaluative dimensions. Thirty years of research on the views of crime victims and others who have called the police, and on the subjects of police investigations, have documented the importance of satisfaction with police fairness, courtesy, understanding, and capability. Satisfaction is higher when officers take adequate time to inform members of the public of how they would handle a complaint and what could be expected to come of their case. Victims who later receive a follow-up contact from police are more favorably inclined as a result, regardless of the news they receive. Highly rated officers are those who were thought to have made a thorough examination of the scene, informed victims about their situation, offered advice, listened to the parties involved, and showed concern for their plight. Satisfaction is very consistently linked to perceived response time as well. Police-initiated stops are also better received when subjects are given a good reason for the stop, and when it is conducted in a manner that does not belittle subjects in front of bystanders. Satisfaction plummets when officers make unproductive and apparently uncalled for searches, a rationale for economy in police aggressiveness. The more of these details that can be included in monitoring surveys, the more closely police managers can make use of the data to identify areas of practice that may be engendering dissatisfaction with the quality of police service.

CONCLUSION

This chapter considered two objections that doubtless will arise if serious managerial and evaluative uses are made of opinion data regarding the police. The objection that "it's just attitudes" implies that survey-based ratings are so abstracted from policing on the ground that they provide an uncertain guide to the service that is actually being delivered. There

is considerable truth to this charge, but measures of change over time are more defensible than one-time measures of levels of confidence in police. The chapter also stressed the importance of ensuring that survey methodologies are employed that produce optimally accurate recall of encounters with police, and that research be conducted to illuminate problems and improve upon our ability to accurately assess encounters that respondents may be inclined to underreport. Underlying all of these specific concerns is a larger one: that the legitimacy of including public opinion on the list of official statistical performance indicators will be enhanced by directly addressing them proactively, rather than waiting for their inevitable appearance in the political realm.

Address correspondence to: Wesley G. Skogan, Northwestern University, Institute for Policy Research, Evanston, IL 60208, U.S.A.; e-mail: skogan@northwestern.edu

NOTES

1. The book is unfortunately out of print, but an exact digital image is available at www.skogan.org
2. The "statistical conclusion validity" question in every study is whether the data are robust enough to detect an effect of reasonable size. Robustness involves both sample size and the reliability of the measures. Small samples and bad measures can lead to false conclusions.
3. As illustration. I used sample means and standard deviations for my Chicago police performance scale, forecast an expected change in the mean of 10%, set the significance level at .05 and the power of the test at .9. For this hypothetical change to be statistically reliable, samples of 158 respondents would be needed in each wave.

REFERENCES

Allen, J., Edmonds, S., Patterson, A., & Smith, D. (2006). *Policing and the criminal justice system – Public confidence and perceptions: Findings from the 2004/2005 British Crime Survey*. London: Home Office.

Brandl, S. G., Frank, J., Worden, R. W., & Bynum, T. S. (1994). Global and specific attitudes toward the police: Disentangling the relationship. *Justice Quarterly, 11*, 119-134.

Fitzgerald, M., Hough, M., Joseph, I., & Qureshi, T. (2002). *Policing for London*. Devon, England: Willan.

Kelly, J. M. (2003). Citizen satisfaction and administrative performance measures: Is there really a link? *Urban Affairs Review, 38*, 855-866.

Miller, J., Bland, N., & Quinton, P. (2001). A challenge for police-community relations: Rethinking stop and search in England and Wales. *European Journal on Criminal Policy and Research, 9*, 71-93.

Nicholas, S., Povey, D., Walker, A., & Kershaw, C. (2005, July). *Crime in England and Wales 2004/2005*. London: Home Office Statistical Bulletin 11/05.

Skogan, W. G. (1981). *Issues in the measurement of victimization*. Washington, DC: U.S. Government Printing Office.

Skogan, W. G. (1994). *Contacts between police and the public: A British Crime Survey report*. Home Office Research Series. London: Her Majesty's Stationery Office.

Skogan, W. G. (2006a). Asymmetry in the impact of encounters with police. *Police & Society, 6*, 99-126.

Skogan, W. G. (2006b). *Police and community in Chicago: A tale of three cities*. New York: Oxford University Press.

Skogan, W. G., & Frydl, K. (2004). *Fairness and effectiveness in policing: The evidence*. Washington, DC: National Academies Press.

Skogan, W. G., Hartnett, S. M., DuBois, J., Comey, J. T., Kaiser, M., & Lovig, J. H. (1999). *On the beat: Police and community problem solving*. Boulder, CO: Westview.

Tomaskovic-Devey, D., Wright, C. P., Czaja, R., & Miller, K. (2006). Self-reports of police speeding stops by race: Results from the North Carolina reverse record check survey. *Journal of Quantitative Criminology, 22*, 279-298.

SURVEYING CRIME IN THE 21ST CENTURY

Mike Hough and Mike Maxfield, Editors

Crime Prevention Studies, volume 22

s first sweep in 1982, the British Crime Survey — and its counterparts in the U.S. and ations — have become invaluable sources of data for research and policy development. ers by a distinguished international group of scholars describe key findings of national urveys in a variety of research and policy areas, including: international comparisons of ctimization; covariation of victimization and offending; the measurement of police rmance; the impact of crime in different types of communities; attitudes to crime and justice; fear of crime; and the unequal distribution of risk.

national crime surveys have made substantial contributions to knowledge, according to hors these surveys must adapt to changing circumstances if they are to continue to be ue. Future directions include continuing to incorporate new technology in samples and ey designs; broadening the focus beyond "normal" crimes and individual victims; and ting better measures of crimes such as fraud, organized crime, corruption, and Internet-facilitated crime.

Recent volumes in the Crime Prevention Studies series

– **Imagination for Crime Prevention** Edited by Graham Farrell, Kate Bowers, ohnson and Mike Townsley (2007).

– **Putting Theory to Work** Edited by Johannes Knutsson nald V. Clarke (2006).

– **Situational Prevention of Child Sexual Abuse** by Richard Wortley and Stephen Smallbone (2006).

– **Designing Out Crime from Products and Systems** by Ronald V. Clarke and Graeme R. Newman (2005).

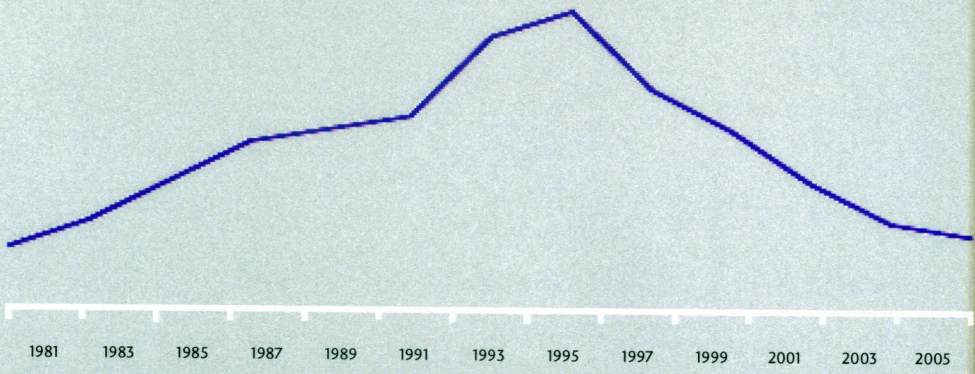
l Justice Press
x 249
, NY 10952 U.S.A.

Publishing
tt House, Uffculme, Cullompton

ISBN 10: 1-881798-75-5
ISBN 13: 978-1-881798-75-0

HOUGH & MAXFIELD CRIME PREVENTION STUDIES VOLUME 22

SURVEYING CRIME IN THE 21ST CENTUR



Mike Hough
Mike Maxfield
Editors

CRIME PREVENTION STUDIES
VOLUME 22