

**OUTLINE OF PROPOSED PILOT STUDY FOR COURT-ORDERED PILOTS ON
DOCUMENTING POLICE-CITIZEN LEVEL 1 AND LEVEL 2 ENCOUNTERS AND
ACTIVATION OF BWCs FOR LEVEL-1 ENCOUNTERS**

November 9, 2018

This document outlines a design for a pilot program that would address the court's orders for an evaluation of the Facilitator's recommendations to (a) document police encounters at *DeBour* Level 1 and Level 2, and (b) require officers to activate BWCs during Level 1 encounters. The proposed pilot would use systematic social observations—in other words, trained observers would ride along with officers and collect data. It is anticipated that the pilot would begin in April 2019. Overall, the proposed pilot is intended to inform the court on the impact of implementing one or both of these changes.

This pilot will be conducted in a limited number of precincts and PSAs. It will provide information that can help guide the court's decisions about what a department-wide set of policies regarding documentation and mandatory BWC activation should be. The information generated by this study should increase the chances that the department-wide policies, when implemented, will be most effective, feasible, and well-executed.

The study would examine four major impacts:

1. **Inappropriately initiated police encounters.** To what extent are police initiating L1, L2, and L3 encounters without appropriate levels of knowledge or suspicion? How does this vary according to whether documentation is required for all levels of police encounters—condition (a) above, or whether officers are required to activate BWCs during L1 encounters—condition (b) above?
2. **Undocumented L3 encounters.** To what extent are police not documenting L3 encounters? How does this vary according to conditions a and b above?
3. **Police administrative workload.** What is the police administrative workload? How does this vary according to conditions (a) and (b) above? An assessment of workload would also examine any opportunity costs in terms of potential impact upon public safety.
4. **Fourth and Fourteenth Amendment assessment.** Would documentation or BWC recording provide useful data in assessing Fourth and Fourteenth Amendment violations?

In addition to the above impacts, the following questions raised by the court orders will be considered:

5. How frequently do encounters occur at each level?
6. How frequently do L1 and L2 encounters escalate to higher levels?
7. Are there encounters for which recording would not provide benefits (e.g., encounters not involving an investigative or law enforcement purpose), such that privacy concerns might outweigh any justification for recording?
8. How, if at all, does the video recording of citizens in L1 encounters affect citizens' orientation toward the police?

To make for a manageable study that can be conducted in a reasonable time frame with reasonable resources, the monitor's team proposes an evaluation design that considers four possible conditions, which are summarized below and in the following table.

Condition #1

- Follow Patrol Guide procedures; no changes in policy
- Observer ride-along with members of the service

Condition #2

- Follow Patrol Guide procedures **EXCEPT** members of the service will activate BWCs for Level 1 *DeBour* encounters in addition to the situations where BWC activation is currently required
- Observer ride-along with members of the service

Condition #3

- Follow Patrol Guide procedures **EXCEPT** members of the service will document Level 1 and Level 2 encounters electronically by preparing an Investigative Encounters Worksheet and submitting it at the end of each tour. Negative reports are required
- Observer ride-along with members of the service

Condition #4

- Follow Patrol Guide procedures **EXCEPT** members of the service will activate BWCs for Level 1 *DeBour* encounters in addition to the situations where BWC activation is currently required
- Members of service will document Level 1 and Level 2 encounters electronically by preparing an Investigative Encounters Worksheet and submitting it at the end of each tour. Negative reports are required
- Observer ride-along with members of the service

	No BWC change	Change in BWC
No Documentation Change	1 Follow current Patrol Guide Procedures; no changes in policy. Observer ride-alongs with members of the service.	2 Follow Patrol Guide procedures except members of the service will activate BWCs for L1 <i>DeBour</i> encounters in addition to situations when BWC activation is currently required. Observer ride-alongs with members of the service.
Change in Documentation	3 Follow Patrol Guide procedures for BWC activation. Members of the service will document L1 and L2 encounters electronically.	4 Follow BWC Patrol Guide procedures except members of the service will activate BWCs for L1 <i>DeBour</i> encounters in addition to the situations in which BWC activation is currently required; and

	Members of the service prepare an Investigative Encounters Worksheet to submit at end of each tour. Negative reports required.	members of the service will document L1 and L2 encounters electronically.
	Observer ride-alongs with members of the service.	Members of the service prepare an Investigative Encounters Worksheet to submit at the end of each tour. Negative reports required.
		Observer ride-alongs with members of the service.

The proposed pilot study would observe officers operating under the four different conditions identified in the above table and measure outcomes under each condition. Comparison of the four conditions would enable the formation of judgments about the likely consequences of adopting each possible combination of documentation/BWC option under consideration.

Observations of the police encounters with the public would occur in a range of police units that frequently engage with the public, the same set of units for each of the four conditions:

- Precinct officers assigned to RMP duties
- Precinct anti-crime teams (plainclothes)
- Police Service Area (PSA) Housing officers performing interior vertical patrols

Precincts will be stratified by level of activity and randomly assigned to each of the four conditions.

Precinct two-officer uniformed patrol. Three precincts would be assigned to each condition outlined in the above table: one precinct with a lower than average level of police-citizen contact, one with an average level of police-citizen contact, and one with a high level of police-citizen contact.¹ In each of the three selected precincts, field observers would conduct eight-hour shift observations with a randomly selected two-officer uniformed patrol unit on each of three work shifts (first, second, and third platoons). There would be two observation sessions per work shift. Thus, for each of the four conditions, 18 uniformed patrol-unit observation sessions would be conducted, for a total of 72 sessions across all four treatment conditions

Precinct anti-crime teams. These plainclothes units typically operate during two shift periods: 10am–6pm and 6pm–2am. They are expected to engage in more frequent Level 2 and 3 encounters than would officers on uniformed patrol. Observations of anti-crime officers would be conducted in the same precincts selected for observation of uniformed patrol units (three precincts per treatment condition). In each precinct, field observers would conduct two eight-hour shift observations on each of the two work shifts. In the event that more than one anti-crime team is operating (separately), the observed team will be randomly selected. Thus, for each of the

¹ This could be determined from department tallies of workload indicators, such as calls for service and staffing levels (i.e., calls for service per full-time patrol officer).

four conditions, 12 full-shift anti-crime unit observations would be conducted, for a total of 48 sessions across all four treatment conditions.

PSA officers performing interior patrols. These units operate under nine Police Service Area commands, not precinct commands. Each PSA covers multiple precincts. Given the greater similarity of Public Housing patrol in terms of workload and service conditions for interior patrol, PSAs will not be stratified by workload. One PSA will be selected and assigned to each of the four service conditions. The housing units selected for observation will be matched as closely as possible for level of workload and then randomly assigned to treatment conditions (probably mid-range in workload level). Because conditions may vary according to work shift, units assigned to perform vertical patrols will be observed on each work shift (first, second, and third platoons). To observe a sufficient number of citizen encounters, observers will conduct four observation sessions per shift (three shifts) in each of the four PSAs selected. Thus, for each of the four conditions, 12 full observation sessions will be conducted in each of the selected PSAs, a total of 48 sessions across all four treatment conditions.

The total number of field observation sessions conducted is thus broken down as follows:

- Precinct two-officer uniformed patrol: three precincts (by low, medium, and high levels of activity) x six work shifts (two observations per platoon) x four treatment conditions = 72 observation sessions
- Precinct anti-crime teams: three precincts (same as uniformed patrol) x four work shifts (two shifts observed twice) x four treatment conditions = 48 observation sessions
- Public Service Area interior patrols: three work shifts x four observations per shift x four treatment conditions (one PSA for each condition) = 48 observation sessions
- Total number of observation sessions: 168²

Numbers of Encounters Observed

It is difficult to predict how many police-citizen contacts of interest (all levels) will be generated. Based on consultation with NYPD, experience suggests that the numbers fluctuate considerably by geographic area and time of day, the average ranging perhaps from as many as 10-20 contacts per eight-hour shift. Based on prior systematic observational studies in other police departments (Parks et al. 1998:2-25), the average across all eight-hour shifts may be around 12-14 contacts (excluding on-scene, citizen-initiated contacts). A reasonable estimate would predict an average of 12 citizen contacts per shift, but to be even more conservative, we will assume only nine contacts per shift across all observed police units. A rate of nine citizen contacts per shift across all observed units means that 168 eight-hour shifts would yield 1,512 observed contacts. On average there would be 378 citizen contacts per treatment condition, although this could vary, especially if the treatment condition affected the proclivity of officers to initiate encounters.

Based on a sample size of 1,512 (378 per treatment condition), the study would have 80% power with a significance level of 0.05 to detect small effect-size differences between the four conditions (differences of 0.08 standard deviations) based on Cohen's effect-size measure F

² The distribution of the 168 observation sessions across these department units does not reflect the actual distribution of officers in the department. Statistical weighting can be used in the analysis to make appropriate adjustments where department-wide estimates are desired.

(Cohen, 1988). Power calculations were conducted using STATA version 15.0.3.³ The power will increase to detect smaller effect sizes by including adjustments for clustering observations within the same precincts and shifts.

If an extraordinary event were to affect significantly the unit selected for observation on a given date/shift (e.g., building explosion, chemical spill, large traffic accident, weather catastrophe, etc.), then that observation session would need to be redone. In the event that the 168 observation sessions generate significantly fewer encounters for analysis than the target 1,512, the project will increase the number of observation sessions until the target number of encounters is achieved.

ELECTRONIC DOCUMENTATION

For conditions 3 and 4, officers will need to document Level 1 and Level 2 encounters in some way. The NYPD has agreed to create a new “app” or electronic form on officers’ phones to gather this information. The new form could include a few key pieces of information: event type, ICAD number, call type, location, initial level of suspicion (L1-L4), final level of suspicion (L1-L4), civilian gender, race, ethnicity, age, and whether there is BWC footage of the encounter. In addition, the form could have a field where the officer states the reason for initiating the encounter. NYPD could provide these data to the research team for the observation sessions conducted by observers in the field, and the two data sets could be merged. A draft worksheet that illustrates the type of data that an officer could record after encounters on a tour is attached as Appendix 2. This worksheet would not be handwritten by the officers, but would aggregate the data they recorded electronically. The final list of fields in the L1/L2 form on the officers’ phones will be subject to preliminary field testing and discussions with the parties as the phone “app” is being developed.

MEASUREMENT ISSUES

This section describes important measurement issues. Refer to Appendix 1 for an overview of Systematic Social Observation (SSO) as a methodology and key data items to be gathered through this method.

Some of the study efforts will yield data from which statistically significant conclusions might be drawn. Such conclusions may not be possible for other data because there may be too few observations, because the sample size may be too small to detect statistically meaningful differences between conditions, or because the data will be too difficult (perhaps impossible or too onerous) to collect. That is why some aspects of the proposed pilot will yield descriptive information only. However, such descriptive information – a not uncommon aspect of pilots – will provide useful information about whether and how to implement proposed changes, or perhaps whether to continue the pilot or continue it in an altered form.

Inappropriately initiated police encounters

³ Power calculation from G Power 3.1 (with fixed effects for five precincts and two covariates and an N of 1,512) yields .98 power to detect a small effect size of .10.

The central question here is whether more documentation of police activity and more BWC filming of that activity will lead to lower levels of inappropriate police-initiated activity – in particular, unconstitutional *Terry* stops, but also inappropriately conducted L1 and L2 encounters. Making judgments about whether more documentation or BWC recording will reduce unconstitutional stops requires observing events that are not now being documented, either because NYPD policy does not require documentation or because officers are failing to document or record encounters that policy requires to be documented or recorded. Systematic observation of police-citizen encounters (assigning trained field observers to accompany and observe police interaction with the public) can generate data that are independent of the functioning of the police documentation system.

Different types of information are needed to make judgments about the appropriateness of police-initiated encounters at all levels: (a) the circumstances of the encounter (e.g., a radio run to a shots fired incident); (b) the observed behavior or condition of the citizen (e.g., bulge in jacket, consent to search given); (c) behavior of the officer (how the citizen was approached, whether the questioning was accusatory, whether consent to search was requested, whether the officer detained the citizen, etc.); and (d) the knowledge possessed by the police that would justify initiating an encounter of the sort the officer engaged in (e.g., officer knows of specific crime patterns in the precinct; citizen known by officer to be a drug dealer; officer knows an area is a drug dealing hotspot). Items (a), (b), and (c) can be observed directly by the observer as well as the officer. Item (d) can be known to the observer only if the officer volunteers it in the course of the event or following it.

But the only way to capture item (d) reliably in the field is by the observer soliciting this information. Asking officers repeatedly to justify their actions could stimulate changes in the officer's behavior in the presence of the observer that would otherwise not have occurred. Asking these questions repeatedly could be particularly problematic for maintaining the integrity of the control condition (no changes in documentation from the current condition), because that in effect is an informal way of documenting the events.

One solution would be for observers to routinely “debrief” officers more generally after each encounter about factors that affected their decision-making. The officer is asked to recount his/her thought processes about what was going on as the situation unfolded. This has reasonable prospects of making clear the officers' perceptions and reasoning for whatever actions were taken – with the advantage of not focusing narrowly on legal matters.⁴ However, this sort of debriefing also could impact the officers' activities and change the control conditions (see comments in Appendix 1 about “Concerns regarding reactivity”).

The field observers will generate the data that describe the observed encounter. A separate team of experts will evaluate the propriety of police actions during these encounters. The monitor team's experts will review observers' narratives of the encounters and post-encounter debriefings.

⁴ This “debriefing” approach has been used in generating information to make judgments about the constitutionality of searches based on accounts of those events and the debriefings that followed (Gould and Mastrofski 2004).

Failure to document L3 encounters

The level of an encounter is readily observable by field researchers, as it does not require a judgment as to the appropriateness of the officer's actions. The observer must merely note whether the conditions defining an L3 encounter occurred. The research team can independently compare the observer's description of the encounter and notation of an L3 encounter with what the officer reported (e.g., whether a stop report was filed). If the Department can provide this in a timely manner (within eight hours of the end of the shift), then the observer can identify and enter a code for all observed L3 stops that were/were not properly documented at the time he/she coded the observational data for the shift. Otherwise this must be done separately.

Police administrative workload

The question here is: What is the additional burden placed upon the police by requiring additional documentation of the citizen encounters? It is difficult to measure this using systematic field observation because police officers may document these events in bits and pieces throughout the shift (a minute here, a minute there). It may be difficult for an observer sitting in the back seat of a sector car to know when one of the officers is doing this. Further, officers may not do all of their data entry documenting the events during the same shift that field observers are assigned to be in the field observing the officers. Officers may perform this task at the end of the tour (after the observer has left), or at the beginning of the next shift that the officer works (and not observed by the researchers).

Given the challenges of directly observing documentation and timing it, an alternative is for the Department to have officers at the end of each shift report (on a paper report form or electronically) how much time they spent documenting encounters on that shift.

There is also another aspect of the administrative workload that may be difficult to measure. Requiring additional documentation of encounters may discourage officers from engaging in encounters that they might otherwise undertake. In addition, taking the time to document an encounter may mean that an officer will not be back "in service" (i.e., available for an assignment) until he or she has completed the documentation, and thus will not be available to respond to a radio call. Alternatively, an officer may elect to report back in service before completing his or her documentation. While these circumstances are difficult to measure quantitatively, the debriefings and the observers' notes can provide a descriptive analysis that will be useful for the court.

Fourteenth Amendment assessment

As noted at the beginning of this document, one question the pilot would attempt to answer is whether documentation of Level 1 and 2 encounters provides useful data in assessing Fourteenth Amendment violations. Aggregate racial data on such encounters would be collected. Some benchmark would need to be used, however, to determine whether there are racial disparities and whether those disparities might reflect Fourteenth Amendment violations. The challenge of

getting meaningful benchmark data (for traffic and pedestrian stops) is covered in the recent National Academies report, *Proactive Policing: Effects on Crime and Communities* (Weisburd and Majmundar, 2017, Ch. 7).

Other measurement issues

In addition to the major measurement issues described above, the court orders raise other questions to address.

How frequently do encounters occur at each level?

The NYPD currently has no reliable indicator of the frequency of all L1 and L2 encounters, and the low number of L3s documented seems to suggest that current *Terry* stop documentation undercounts the actual number of such stops. Systematic field observation can note the frequency of these events, the nature of the police service/law enforcement issue at hand (e.g., investigating a specific offense, checking out a suspicious situation, or providing non-enforcement assistance to a citizen), and whether/when certain actions were taken (e.g., police request for consent to search, frisk, etc.). In addition, researchers can review and code a sample of BWC videos (under conditions 2 and 4 only as presented in the table), permitting a more focused examination of these events.

How frequently do L1 and L2 encounters escalate to higher levels?

Knowing how frequently these lower-level encounters escalate will help to decide whether to require them to be included in a documentation system, and if required, whether to include all of them or only some types of lower-level encounters in the ultimate post-pilot policy that is established for documentation and BWC recording. For example, data gathered by systematic field observation will enable researchers to estimate the “yield” of L3 encounters that escalated from lower level encounters. The higher the yield, the stronger the case for documentation of L1 and L2 encounters. And it may be useful to identify the features of encounters that are most likely to escalate. The analysis of SSO data to address this issue can also be supplemented by the coding of BWC footage (see prior section).

How, if at all, does the video recording of citizens in L1 encounters affect citizens’ orientation toward the police?

The prospects of conducting a feasible survey of respondents who have had recent Level 1 contacts with an acceptable response rate are not good. However, the second-wave surveys for the current BWC pilot being conducted after the BWCs have been in place for a year, ask those who have been stopped in a car or on the street whether the officer was wearing a camera. If yes, the survey asks whether the fact that police officers were wearing video cameras to record the interaction made the experience more satisfactory, less satisfactory, or did not really affect the person’s satisfaction with the experience either way. This survey will not provide data on any potential BWC chilling effect in Level 1 encounters, as opposed to Level 3 encounters, but it may provide some general information on how citizens view BWCs.

Other privacy questions

The NYPD determined that the current BWC policy should not mandate officers to record all witness interviews. In its response to public and officer input gathered by the NYU Policing Project, the NYPD stated that “[c]ertain witnesses may feel uncomfortable being recorded, such as sex crimes victims, confidential informants, child victims, or witnesses who simply feel too fearful to have their statements recorded and ultimately available to the accused as required by criminal law procedures.” NYPD Response, p. 12. The pilot can collect data on the volume of victims and witnesses encountered.

Coding BWC footage

BWC footage for L1, L2, and L3 encounters will be gathered under conditions 2 and 4 to enable analysis of three questions (only under conditions 2 and 4 will the BWC recording of L1 encounters be required):

- How frequently do encounters occur at each level?
- How frequently do L1 and L2 encounters escalate to higher levels?
- What impact does the presence of SSO observers have on officers’ initiation of encounters?

Coding a sample of BWC footage offers an alternative methodology to in-person SSO for addressing the first two questions, which call for a descriptive analysis. BWCs record more focused information than is typical of a field researcher, but they can be replayed multiple times to capture information that is difficult to gather when the field researcher is observing and capturing features of the encounter in real time. In addition, sampling of BWC videos may be done to make possible a comparison of officer behavior and performance when accompanied by an SSO observer and when not accompanied.

The sampling plan will have the following features. The study will draw a random sample of BWC videos during SSO observation sessions that have been completed under conditions 2 and 4. The study will also review a set of “matching” videos from the same officer for the nearest previous same-day-of-the-week. So, if Officer X is observed on third platoon on June 1 (a Saturday), Officer X’s BWC footage for that shift will be reviewed along with footage from the first prior Saturday that the officer worked the third platoon (for example, May 25).

The coding of BWC footage will be conducted after the conclusion of SSO field observations.⁵ Inasmuch as possible, the same coding protocols will be used as with field observations. The study anticipates using a cadre of SSO observers who would switch to this work upon completion of their SSO duties – after a short training period for adjustments required for coding BWC footage.

⁵ It will be important for NYPD to preserve the videos for all SSO observation sessions and the matching rides associated with those sessions.

STAFFING SYSTEMATIC OBSERVATION⁶

Systematic observation requires two types of tasks: (a) conducting the observational ride-along while gathering data in the field, and (b) entering the coded data into an electronic format that can be analyzed. In addition, field observers will be required to provide a narrative account of each police-citizen encounter, which assists with quality control of coding and often helps with interpreting the quantitative coded data, making it possible to render judgments about the appropriateness of police actions. The post-ride-along data entry task (performed after the ride-along observation session has concluded) requires care and usually consumes as much time as the actual observation does.

This sort of systematic observation research will yield the highest quality observations and coding by assigning the work to a small team of well-trained, well-paid, and motivated research staff. A full-time observer could average two ride-alongs per week (including observation time in the field, coding time after field work, and time correcting coding errors). This would require 11 weeks for data collection and data entry by eight full-time observers (168 ride-alongs divided by two/week divided by eight observers). Add an additional 15% time for unforeseen problems (illness, family emergencies, problems with ride-along), and a staff of eight full-time observers could be expected to complete the 168 rides in 13 weeks.

Full-time observational staff is ideal, but it may be difficult to find good people who are available on a full-time basis for a relatively short data collection project. Alternatively, systematic observation research has relied heavily on the use of currently enrolled university students, whose schedules make it possible for them to do part-time work when not attending classes (especially advanced graduate students who may have completed their course-work). Twelve such students, each completing one ride-along per week could complete 168 ride-alongs in 14 weeks. However, part-time workers attending school do not have as much schedule flexibility to meet the sampling requirements, so it is reasonable to add 25% time, totaling 18 weeks.

In addition to the observers, two full-time quality-control supervisors need to review the observers' work and give timely feedback to correct errors. Someone also needs to perform liaison with the NYPD, explaining the study to key people in the chain of command in selected precincts/units, coordinating ride schedules, and dealing with issues as they arise. Another staff position to be filled is a person to create a data entry system for the field observers. Qualtrics, <https://www.qualtrics.com/>, has been used in recent studies employing systematic social observation. This person would also need to format the data to be used for data analysis (e.g., SPSS, SAS, R). Data management staffing may be required to integrate department-supplied data (e.g., official department reports of L3 stops) with SSO data.

TIMELINE FOR DEVELOPING AND EXECUTING SSO AND BWC CODING FOR PROPOSED PILOT STUDY

⁶ The staffing estimate here does not include research staff required for other research activities that might be used, including any coding of BWC footage and comparing Department documents to field observations.

Research instrument development and field testing: 4.0 months

This involves conducting preliminary ride-alongs with each type of unit to be observed, developing an observation protocol and instruments, putting the instruments into a computer-compatible data entry format (e.g., Qualtrics), field testing and modifying the instruments. This also includes developing protocols for the monitor team's experts to judge the propriety of police actions during police-initiated encounters, and obtaining any necessary Institutional Review Board clearance for conducting research on human subjects.

Recruitment and hiring of observation staff: 2.0 months [will occur during period of instrument development and field testing above]

SSO staff can be recruited from graduate programs in criminal justice, law schools, and other disciplines in NYC and nearby universities. This staff includes observers and team leaders for quality control. Ideally, the team leaders would have prior experience with SSO, but the prospects of this seem uncertain. Applicants should be individually interviewed. Hiring a number of multilingual field observers (especially those fluent in Spanish) will be desirable. Also during this period, the monitor's team should work closely with key NYPD staff to ensure appropriate commanders and staff are well-informed about the project and ready to receive it.

More observers should be hired to train than are required to do the work, to allow for the likelihood that some will prove unsatisfactory during the training session or will elect not to continue after training.

Training SSO observers: 2.0 months*

On-site training will not only involve some classroom time, but also a significant amount of time in the field practicing SSO and then coding results, checking the work of the observers, and correcting errors. The training delivered to field observers would come from several sources, including people experienced in SSO (prior projects) and the monitor team. It is essential that the trainers include people who are familiar with NYPD procedures. Both the NYPD and the plaintiffs will be able to observe the training.

Note that training will include instruction by key Department staff, prosecutors, and monitor team members on standards for proper conduct of L1-L4 encounters.

Conduct SSO field observations: 3.25-4.5 months*

During this period the data will be collected, coded, and edited. Amount of time depends upon full/part-time staffing of SSO positions.

*NOTE: These are the months when field researchers would actually be in the field observing members of the service in ride-along sessions (training and data collection).

Train for and conduct BWC video coding: 1.5-2.5 months

During this period training will be given for coding BWC videos (.25 months). BWC videos are selected and coded (1.25-2.25 months). The amount of time required will depend upon full/part-time staffing of BWC video coding.

Data analysis and report writing: 4.0 months

During this period, the data will be analyzed and a report will be written. Progress reports will be made during the course of the pilot study as part of the periodic reports by the monitor.

Total time for SSO: 14.75-17 months

Timeline for NYPD Tasks

Drafting Operations Order for Pilot Precincts and PSAs, specifying the four conditions

Training Officers in the Pilot Precincts and PSAs

Developing Level 1 and 2 Documentation Form

Adjustment period for officers and system to correct problems and officers to become familiar with new procedures.

APPENDIX 1. OVERVIEW OF SYSTEMATIC SOCIAL OBSERVATION OF POLICE

Since the mid-1960s systematic social observation (SSO) of police has proven to be a useful alternative to other forms of observation and data collection. SSO combines features of ethnographic field observation and survey research. Like an ethnography, it places observers in the “natural setting” where police interact with the public, but like survey research, it requires those observers to follow a set of detailed protocols about what specifically is to be recorded and how to record it. This makes it possible to classify observations into standardized and quantifiable categories. For example, SSO protocols could direct field researchers to determine whether an officer acted disrespectfully towards a citizen she encountered, in what ways, and for how long (momentarily or repeatedly). Detailed instructions establish what constitutes officer disrespect and how to apply the classification codes provided. This standardization of field work and coding makes it possible to replicate these results across different observers, different officers, and different situations. It allows researchers to perform quantitative analyses of field work. In contrast to survey research, the informant about the officer’s degree of disrespect might be the citizen encountered by the officer, whose impressions and opinions would be solicited after the fact. The criteria used by citizens to make these judgments may vary greatly, but SSO yields observations according to criteria that are standardized *a priori*. As with survey research, the selection of specific officers for observation can be performed according to probability sampling, allowing statistical inferences to be made (Mastrofski et al. 1998:vii; Reiss 1971; Worden and McLean 2014:471).

SSO studies of policing have focused primarily on patrol officers. A recent review of this methodology counted five large-scale SSO data collection projects, beginning with the President’s Commission study by Albert J. Reiss in 1966 and five smaller projects (Worden and Mclean 2014). Each project has generated numerous publications, and additional projects have been conducted since the overview (e.g., Mastrofski et al. 2016; Worden and McLean 2017). The scope of topics has been broad, but SSO’s contributions to knowledge have been most frequently used to describe and explain variations in the exercise of police discretion, especially the use and abuse of authority (Mastrofski et al. 2010; Worden and McLean 2014).

Concerns regarding reactivity

The most frequently mentioned concern about SSO as a methodology is that officers react to the presence of an observer or the characteristics/behavior of the observer – thus rendering a distorted picture of how the officer would have behaved if not accompanied by an observer. The common concern is that in the presence of an observer, officers will be more professional, law-abiding, and more considerate of citizens than they otherwise would have been.

The threat of reactivity with SSO cannot be dismissed out of hand, but several things are worth noting. First, it may not be as great as many skeptics anticipate (Mastrofski and Parks 1990; Worden 1989). Reiss (1971) noted that it is naïve to presume that many circumstances and aspects of police behavior readily lend themselves to police manipulation motivated by awareness of being observed. Officers are more influenced by the particulars of the circumstances, as well as their own skills and habits, even with regard to the use of physical force. One of the more consistent findings is that the presence of an observer tends to increase

the likelihood of officer proactivity, especially if the observer is female, and that the more “helpful” the observer is to the officer, the greater the officer’s tendency to use force. Notably, however, levels of reactivity diminish over the passage of time during field work (Worden and McLean 2014:485). Ultimately, it is important to keep in mind that data collection alternatives to SSO are not without their own reactivity issues. Survey respondents are susceptible to the desire to respond in ways that present themselves in the best light. Department records that rely upon officers to self-report events in which they were engaged are clearly highly reactive, since officers know that this information is used to hold them to account (Mastrofski et al. 2010:236). A police agency’s use of body-worn cameras (BWC) may well dwarf any reactivity effects of a ride-along observer, since video footage offers what, for many, constitutes a detailed and reliable (if not necessarily complete) record of events. And for many, a fundamental function of the BWC is in fact to deter police from engaging in inappropriate behavior.

Ways to minimize the reactivity of SSO include: (a) observers avoiding the expression of opinions to officers about what they observe, (b) training observers in ways to establish rapport with officers, and (c) making and following pledges of confidentiality. After the fact, researchers can conduct analyses to detect the nature and extent of reactivity, as well as perform sensitivity analyses to determine its effect on findings (Mastrofski et al. 2010; Worden and McLean 2014).

To assess the extent to which the presence of observers may change the behavior of the officers being observed, the pilot will compare data on the activity of officers during their ride-alongs with data on the activity of those same officers when they do not have an observer with them. This will be done in two ways. First, BWC footage will be compared for a sample of officers with SSO observers and a matching set of BWC videos for the same officers but without observers (see section of this proposal on coding BWC footage). Second, officer behavior will be tracked based on records already maintained by the department on officer activity. For example, if an officer is being observed on the third platoon on a Thursday, data can be collected on the officer’s activities during the four prior third platoon Thursdays and the four subsequent third platoon Thursdays that the officer works. This analysis will be limited to activities that the NYPD routinely gathers (encounters initiated by the officer, citations issued, arrests made), but it will enable researchers to determine whether there is a distinct difference in officers’ proactivity and enforcement levels when they are observed from when they are not observed. Ideally the data will be conducted at the unit level (that is, the actions of all officers observed in the selected unit). The pilot will also compare data on the activities of officers who have been observed with officers in the same unit who were not observed.

From a practical perspective, SSO presents several challenges. It is time-consuming to set up research instruments and observational protocols, hire and train observers, and maintain data quality control. It is not a cost-effective way to gather data on rare events (e.g., police use of lethal force).⁷ Most importantly, its success relies on police officers in the field allowing observers to be sufficiently close to police-citizen encounters so that they can see and hear what is happening. Also important is the willingness of officers to participate in the debriefings

⁷ In 7,200 hours of observation in one large SSO study, a police firearm was drawn in only 53 encounters, and only once was it discharged (at a snake) (Worden and McLean 2014:486).

following the completion of encounters. Under these circumstances, it would be ideal for researchers to be able to provide officers assurances of confidentiality that are backed by assurances from the court and the monitor that they would not be subject to legal process.

Key SSO items for this study

Below is an outline of the sorts of data that SSO field observers will be collecting. It is not exhaustive and is subject to change/refinement as the study develops. It is also anticipated that with respect to legal questions, the narrative accounts and information collected and coded by the observers would be reviewed by a panel of experts drawn from members of the monitor team, and that these experts would make the determinations regarding police legal compliance.

Identifying information

- Precinct/PSA
- Type of unit (patrol, plainclothes anticrime, housing)
- Officer IDs (both officers in two-officer units)
- Observer ID
- Date and start time of shift
- Time observation began

Encounter characteristics

- How officer was mobilized for this encounter (e.g., dispatched, citizen on-scene, officer-initiated without citizen request or department direction)
- Time dispatch received, encounter began, encounter ended
- Location (type of location, GPS coordinates)
- Department record number (e.g., ICAD number, stop report number) – obtained from officers
- Type of problem (most serious) police expected to deal with (taken from an extensive list of hundreds of problems, each with a unique code)
 - As described by dispatch (if relevant)
 - As it appears initially at scene
 - As it appears at end of encounter

Citizen-contact characteristics and actions

In two-officer units, the observer will report actions of both officers when it is possible for the observer to see/hear both. When not possible, the observer will select the officer likely to be engaged the most with citizens and focus observations on that officer.

- How police-citizen contact was initiated at scene (officer, citizen, unable to determine)
- Length of police-citizen interaction
- BWC activated for this citizen? (none of it, part of it, all of it)
- Citizen characteristics: gender, race/ethnicity, age
- Actions by officer(s)
 - How police approached/spoke to citizen initially (general/nonthreatening, command, or threat-focused)

- Citizen response to initial approach (acceded, ignored/declined, left scene)
- Police communicated citizen not free to leave?
- Police requested consent to search citizen/property?
- Police searched citizen/property
 - Evidence available to justify search (e.g., bulge under coat)
- Police made arrest?
- Was this encounter documented electronically? (yes, no, officer indicated intention to file report in future, unable to determine)
- Was a stop report filed for this encounter (yes, no, officer indicated intention to file report in future, unable to determine)
- Encounter level at beginning of contact with this citizen (Levels 1-4)
- Encounter escalated to higher level after initial contact? (what level)
- Police force used toward citizen (if any)? (type of force)
- Citizen resistance/demeanor toward police
- Citizen demeanor toward other citizens (if present)

Supervisor oversight

- Was a police supervisor present during this encounter? (not at all, part of encounter, all of encounter)
- Did supervisor have a discussion with the officers involved in the encounter?

Post-encounter debriefing of officer

Following the completion of an encounter, the observer will as soon as feasible debrief the officers on the prior encounter, asking the officers to describe what they had observed or knew about the situation and the people involved as they were deciding what to do. This information will be documented by the observer in a narrative account and be used to make judgments about the legality of the encounter and actions taken during the encounter.

“We’re interested in getting your reading of this situation so we can better understand the decisions that were made. Could you please talk us through this encounter, focusing on the things that influenced your decisions about what to do and how to do it? To start off, what led you to initiate contact with the citizen(s) in this situation? And what were the key factors that led you to handle the situation as you did?”

APPENDIX 2. INVESTIGATIVE ENCOUNTERS WORKSHEET

Officer Name	Smith
Officer Tax	959026
Officer	
Command	13 Pct
Assignment	Sector A
Day	Monday
Date	20-Aug
Platoon	3rd

[illegible]

REFERENCES

- Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum.
- Gould, Jon B. and Mastrofski, Stephen D. 2004. Suspect searches: Assessing police behavior under the U.S. Constitution. *Criminology & public policy* 3(3):315-362.
- Mastrofski, Stephen D., Tal Jonathan-Zamir, Shomron Moyal, and James J. Willis. 2016. "Predicting Procedural Justice in Police-Citizen Encounters." *Criminal Justice and Behavior*.43(1):119-139.
- Mastrofski, Stephen D. and Roger B. Parks. 1990. Improving observational studies of police. *Criminology* 28(3):475-496.
- Mastrofski, S. D., Parks, R. B., and McCluskey, J. D. 2010. Systematic Social Observation in Criminology. In A. Piquero and D. Weisburd (eds.), *Handbook of Quantitative Criminology*, pp. 225-247. New York: Springer-Verlag, pp. 225-247.
- Mastrofski, S., R. B. Parks, A. J. Reiss, Jr., R. E. Worden, C. DeJong, J. B. Snipes, and W. Terrill. 1998. *Systematic Observation of Public Police: Applying Field Research Methods to Policy Issues*. NCJ 172859. Washington, DC: National Institute of Justice.
- Parks, Roger B., Mastrofski, Stephen D., Reiss, Albert J., Jr., et al. 1998. *St. Petersburg Project on Policing Neighborhoods: A study of the Police and the Community*. East Lansing, MI: Michigan State University.
- Reiss A. J., Jr. 1971. Systematic observation of natural social phenomena. In Costner, H. L. (ed.), *Sociological methodology*. San Francisco: Jossey-Bass, Inc., pp. 3-33.
- Weisburd, D. and Majmundar, M. K. 2017. *Proactive Policing: Effects on Crime and Communities*. Washington, DC: The National Academies Press.
- Worden, R. E. 1989. Situational and attitudinal explanations of police behavior: A theoretical reappraisal and empirical assessment. *Law and Society Review* 23:667-711.
- Worden, R. E. and McLean, S. J. 2014. Systematic social observation of the police. In Reisig, M. D. and Kane, R. J. (eds.), *The Oxford Handbook of Police and Policing*. New York: Oxford University Press, pp. 471-496.
- Worden, R. E. and McLean, S. J. 2017. *Mirage of Police Reform: Procedural Justice and Police Legitimacy*. Oakland, CA: University of California Press.