

March 2023

Seattle EAQ Evaluation Final Report

Prepared for

Seattle Police Department

Loren T. Atherley
Seattle Police Department
P.O. Box 34986
Seattle, WA 98124-4986
Loren.atherley@seattle.gov

Prepared by

Sean E. Wire
Elise P. Kratzer
Heather Prince, PhD
RTI International
3040 E. Cornwallis Road
P.O. Box 12194
Research Triangle Park, NC 27709

RTI Project Number 0218245.000.001



Contents

Executive Summary		1
1. Introduction		3
1.1 Background		3
1.2 Format		4
1.3 Scope.....		5
2. Component 1: Terry Frisk Equity Evaluation		6
2.1 Evaluating the Premise.....		6
2.2 Validating the Methodology		6
2.2.1 Initial Considerations and Current Resolution.....		7
2.2.2 Outstanding Considerations.....		9
2.2.3 Recommendations & Conclusion.....		9
2.3 Summary.....		10
3. Component 2: Location-Based Resource Accountability		11
3.1 Evaluating the Premise.....		11
3.2 Validating the Methodology		11
3.2.1 Initial Considerations and Current Resolution.....		12
3.2.2 Outstanding Considerations.....		13
3.2.3 Recommendations & Conclusion.....		14
3.3 Summary.....		15
4. Component 3: Officer Service Quality Assessment		16
4.1 Evaluating the Premise.....		16
4.2 Validating the Methodology		18
4.2.1 Considerations.....		19
4.2.2 Recommendations & Conclusion.....		21
5. The Seattle Crime Harm Index		23
References		25

Executive Summary

Seattle's Equity, Accountability, and Quality (EAQ) initiative is a holistic risk-management approach that aims to actively manage the balance between crime control and civil liberties and examine the total cost of ownership of public safety. The EAQ model replicates the framework of a police performance management system (CompStat)-style management meeting, using novel metrics developed by the Seattle Police Department to demonstrate compliance with the consent decree and in an attempt to reflect organizational health based on equity, accountability, and quality measures. This report draws the following conclusions about each component in the EAQ.

1. The **post-stop equity component** uses a propensity score-matching approach as an elegant way to isolate racial bias where causal experimentation is not possible. This metric leverages proven methodologies and existing data collection efforts to provide a reasonable measure of racial disparity in officer decision making. We recommend that this component proceed as an integral part of the equity component of EAQ, with additional concurrent efforts to improve data quality and the subsequent accuracy of this measure.
2. The **location-based resource accountability component** satisfies its intended goal of creating a dynamic measure of resident need and police service. The data sources for both resident need (calls for service) and officer location (Automated Vehicular Locator) are both appropriate and available, and the metric itself relies on well-established spatial statistics to identify a novel need. Due to the discussed limitations to geographic specificity, and the reliance on officer location over behavior, we do not view this metric as a definitive measure of over-policing. However, it does serve as a valid start to conversations about why observed disparities occur and how to iteratively improve service equity from there. We recommend this component proceed as a valuable part of the EAQ forum.
3. Due to the recent shift in the interaction quality metric from body-video assessment to **post-interaction surveys**, there are unknowns related to implementation. The use of survey techniques to evaluate perceptions of police performance is a well-established methodology, here facilitated by improved distribution and survey targeting. We believe that, in its current form, the metric's use of Net Promoter Scores provides an effective benchmark for police performance as a whole, with some current technical limitations to evaluating subjective

interaction quality. We can offer only initial recommendations, although its value in the EAQ may improve as future efforts expand survey content and generalizability.

1. Introduction

1.1 Background

In 2012, the City of Seattle entered into a settlement agreement with the U.S. Department of Justice (DOJ) as the result of a DOJ investigation into the Seattle Police Department's (SPD's) use of force practices and concerns about biased policing. Although the investigation did not find that SPD officers engaged in biased policing, it noted concerns regarding racial disparities in outcomes. With these findings, SPD began working with the DOJ Office of the Inspector General to establish strategies and metrics to demonstrate progress toward compliance with post-Federal Consent Decree operations. Through this strategy, SPD is working to extend and sustain progress toward a more equitable delivery of police service established under its consent decree by regularly collecting and reporting Equity, Accountability, and Quality (EAQ) measures that support a focus on continuous improvement. SPD has upgraded its data warehouse and processing infrastructure to provide near real-time patterns of disparity and evolve its general understanding of the collateral harms associated with the delivery of police service. With this, SPD is establishing a CompStat-style forum that will focus on the continuous monitoring of three established EAQ measures and will look at operationalized measures for disparate outcomes, under- and over-policing of communities, and service quality for awareness, mitigation, and continuous improvement. Overall, this process continually maintains focus and progress on the consent decree reforms. SPD has partnered with RTI International to serve as its quality assurance and evaluation partner for implementing the EAQ process.

RTI is a private, nonprofit research organization with the capabilities, infrastructure, and review systems to manage and complete complex projects. RTI's Center for Policing Research and Investigative Science actively partners with law enforcement agencies across the country, with an emphasis on providing rigorous, data-driven results that have direct implications for the field. In September 2021, RTI began documenting the planned EAQ methodologies and technology through a series of meetings with the City of Seattle, SPD, and relevant partners to understand the development and implementation of methodological approaches. For each method, RTI staff reviewed relevant research and literature relating to the foundational hypotheses for the methods and technology presented and consulted internal and external experts on best practices and feasibility of the methodology.

1.2 Format

Broadly, this report serves as a direct follow-up to the pre-implementation report submitted to the SPD in December 2021. During this intervening time, SPD has refined the methodological components of the EAQ, taking into account initial concerns and considerations raised by RTI during the first evaluation. This report will revisit the methodologies of the following three EAQ components, providing an updated evaluation of the proposed application and methodology:

- **Terry Frisk Equity Evaluation:** To measure the level of racial equity in officers' decisions to frisk during a Terry stop, SPD is using Propensity Score Matching (PSM) to determine whether observed disparity in post-stop outcomes can be attributed to the perceived race of the subject, with the expectation that this approach can translate to other areas of service equity.
- **Location-Based Resource Accountability:** To look at areas of over- and under-policing relative to demand, SPD is using spatial analysis to map areas of community need and Automated Vehicular Locator (AVL) data to measure police presence.
- **Officer Service Quality Assessment:** To assess the quality of interactions between officers and community members more frequently, SPD is utilizing SPIDR Tech to obtain feedback from community members following interactions and contribute to a composite rating of interaction quality for the department.

For this report to serve as a comprehensive evaluation, there is some intentional overlap with the pre-implementation report, wherein the project description, logical evaluation and methodological validation are included and updated where necessary. SPD has made concerted efforts to consider RTI's initial criticisms and opportunities for improvement of the proposed methodologies, where possible. We will highlight the resolution of relevant considerations for each component in this report. Where no such clarification or change has been made, we provide a list of outstanding considerations that may be relevant to how the metrics are employed or may be improved in the future. Finally, for each of the three components, we provide a summary judgment of whether and how SPD should proceed in integrating the component into the overall EAQ framework.

1.3 Scope

In this report, we intend to raise all relevant considerations as justified by the existing literature and expert input. However, there are two key restrictions to the scope of this report. First, this report is limited to a logical and methodological assessment of the three components. Regular EAQ meetings were not yet active during the evaluation period, so we are unable to advise on operational parameters, such as the appropriate cadence of the meetings, how to best leverage the metrics, or how these new evaluative criteria might be communicated within and outside the agency.

Second, our assessment of each component of the EAQ is restricted to its intended use as a system-level metric for police performance. Throughout this evaluative process, our inquiries have been focused on evaluating the EAQ outcomes at this high level of aggregation. In practice, however, each of the three metrics has the potential to be used for more detailed explorations of outcomes at the officer level. There are additional methodological and operational considerations about the appropriateness of applying these methodologies to a different unit of analysis. Therefore, we recommend further consideration before proceeding with these metrics at the officer level.

2. Component 1: Terry Frisk Equity Evaluation

2.1 Evaluating the Premise

The primary goal of this EAQ component is to create a dynamic and high-resolution measure of the level of racial equity in officers' decisions to frisk during a Terry stop to determine whether differences in post-stop outcomes can be attributed to the race of the subject. Terry stops were selected as a convenient and accessible sampling point, but the intention is to measure officers' differential perception of dangerousness that can characterize service equity across a range of scenarios. They will use a PSM approach, whereby situational and demographic variables that SPD recorded will serve as controls, so that any remaining difference can, in theory, be attributed to the race of the subject. The sourcing of adequate data and the ability to process those data efficiently will determine the appropriateness of this proposed methodology.

The proposed data source for this component is the universe of recorded Terry-stop contacts to inform the model, supplemented with computer-aided dispatch data. This incident-level metadata provides measures on officer and subject demographics, situational dimensions surrounding each stop, and abstractions of what a reasonable officer might know prior to their decision to frisk. Barring any non-systematic missing data issues, this internal data source is the only viable way to capture the predicates and outcomes of each police stop.

Because true experimentation is not possible in this case, PSM presents a robust quasi-experimental alternative that is logically consistent with the goals of this EAQ metric. As explored in depth in the pre-implementation report, we find that both the availability of the data and the approach to analysis satisfy the research premise, although we caution here against treating this metric as a comprehensive measure of racial disparity.

2.2 Validating the Methodology

This component uses propensity scores to approximate equivalence between groups and isolate race as responsible for any observed disparity. PSM, which is a quasi-experimental method that matches individuals on demographic and situational similarity, is capable of causal inference in theory (Rosenbaum & Rubin, 1983; Dehejia & Wahba, 2002). In reality, however, the inability to account for unknown unknowns prevents balancing on key differences beyond the treatment effect

(Govindasamy, 2016; King & Nielsen, 2019). The approach SPD proposed assigns propensity scores as weights to account for selection assignment differences (Olmos & Govindasamy, 2015).

New York City has successfully used this approach to assess racial disparity in post-stop outcomes (Levchak, 2021). This study echoes previous research that finds these methods reasonably approximate a randomized experimental design to allow for estimating causal effects. The current approach is also modeled on previous efforts in Seattle. In the *Disparity Review: Part 1* (Seattle Police Department, 2019), a similar methodology was used as a proof of concept to capture a citywide measure of disparity in the use of frisking, measuring the differential perception of dangerousness based on the perceived race of the subject of the stop. This component builds on this premise and applies the methodology to specific geographies, units of officers, and time periods to track levels of disparity over time and across levels of aggregation.

2.2.1 Initial Considerations and Current Resolution

1. **Balance.** One of the primary considerations in trying to make this approach dynamic over time is balance. How can the availability of good, matched cases be maintained and support the frequency of the EAQ? Certainly, the number of Terry stops cannot and should not be altered to satisfy this model, so SPD's control is limited to the level of aggregation of the method. Before the cadence for both EAQ and this equity component were set to the monthly level, we had suggested drawing matches from a wider timeframe to create higher quality counterfactuals. Over the development of this methodology, it became clear that the model is stable and sufficiently powered from a month's worth of stops. This allows the approach to provide some temporal insights about post-stop disparities without relying on too few stops. We recommend serious consideration before applying this methodology to smaller units of analysis (shorter timeframe or smaller geographic areas).
2. **Number of events.** When using logistic regression to create the propensity scores, there are limitations to the number of events per variable (EPV) that can be employed. Based on the 119 variables used to explain post-stop differences and the intention to divide the pool of potential matches by time and place, we were initially concerned with overfitting the model and violating the EPV assumption. SPD's Bayesian approach (XGBoost) to the computation of propensity scores can relax this assumption and justifies the inclusion of a substantial number of covariates.

3. **Officer behavior.** One of our early concerns with using post-stop outcomes as the proxy measure for racial disparity is the potential chilling effect that measurement itself may have on officer stop and frisk behavior. This sort of de-policing effect has been observed across agencies subject to the consent decree process (Stone et al., 2009; Chanin & Sheats, 2018). Although we have no reason, at this time, to believe that this has occurred in SPD, it is important to monitor as EAQ progresses and becomes a part of officers' daily lives. Whereas we expect that effective communication about how post-stop outcomes will be monitored at an aggregate level may mitigate this effect, SPD has committed to tracking any de-policing effect. Because the model itself relies on a sufficient number of cases for matching, this method would naturally erode if officer drawback occurred at a substantial level.
4. **Staffing.** We have observed that due to staffing availability, and post-pandemic declines in proactivity, the number of qualifying Terry stops is lower than past levels and may affect match quality or the ability to generate consistent results across every month or among certain officer units. We reiterate that there should be a built-in expectation that this metric can only operate with sufficient stops to inform the model and interruptions in this metric may naturally occur. Based on our observations, SPD's workflow continuously monitors the balance of matches and will not relax model parameters for a potentially faulty value if and when these interruptions do occur.
5. **Missing data.** Along the same lines, missing values persist in the data. An initial look at the historical data shows that 12.1% of Terry stop records have missing or unknown data for the race field. This is currently being monitored by a data governance program (DGAL-192). Since these initial reports, SPD has required the race field to be filled out for all filed contacts where Terry stops are reported. Taken together, this is a methodology where upwards of 15% of stops are not directly pertinent to this question of racial disparity, and another approximately 10% may not have data related to the central question of race. These present a baseline of limitations to the data that may confound initial results but should motivate additional efforts to filter out unwanted cases and improve data quality for race and all covariates.

2.2.2 Outstanding Considerations

Non-Terry stops. Based on initial evaluations of the data, 85% of the stops in the model data are Terry stops, with the remaining 15% including probable-cause stops and post-arrest frisks due to an excess of caution in reporting. The inclusion of non-Terry stops does introduce some unwanted noise in the results; however, this is unlikely to change. These estimates are still reasonably accurate, and although there may be future attempts to flag unwanted stops, there is an expectation that over-reporting of stops will still occur and be included in this analysis.

2.2.3 Recommendations & Conclusion

Overall, this methodology is an elegant approach to isolating the “causal” effect of perceived subject race on officer decision making. The propensity score-based approach, which can only theoretically define causality when all known and unknown covariates are controlled for, is an appropriate and sophisticated substitute for a randomized controlled trial, which is not possible here. The use of post-stop outcomes as a measure for the general impact of race on the outcomes of police interactions is appropriate, as these situations are highly discretionary and offer a rich dataset to control for and isolate the effects of race.

That said, the disparity metric associated with this EAQ component is a proxy measure, a relevant avenue to get at the larger question of racial motivation in officer decision making. At no point should the measure of disparity in post-stop outcomes be viewed as a comprehensive measure of racial disparity across police interactions as a whole. To illustrate, this metric does not measure what is potentially the largest source of disparity in Terry stops: the decision whether to stop someone. Gau and Brunson (2010) and Bandes and colleagues (2019) provide support for the importance of the stops themselves being an impactful source for potential disparity regardless of whether a frisk occurs. We understand that future additions to this EAQ component, including risk-adjusted disparity, may expand the scope, but it is important to frame the current view of post-stop outcomes as only part of the complete picture of disparity.

The utility of this metric is contingent on continuity in stop behavior over time and across the unit divisions to which it is applied. If officers conduct fewer stops or proactivity declines to a point that the models cannot compute a summary disparity score, this does not serve as an indicator that racial disparity has been eradicated, but rather that the source of its measurement is no longer available and must be measured in other ways. Ensuring continuity in the number of qualifying events is in

the best interest of this endeavor. This is not to suggest a mandated increase in the frequency of stops, but rather that effective communication to officers and transparency about how this metric will be used may circumvent any pullback behaviors or a breakdown in this measurement.

There are a few ways this component can be expanded. For example, there is value to conducting targeted interviews or focus groups with officers to better understand their decision making when it comes to Terry stops and post-stop frisking. Any additional knowledge, even if qualitative, may improve future modeling and identify relevant covariates. Any additional explanation for discretion will eat into the variability explained by race, reducing any undue disparity attributed to it.

It may be useful to explore other post-stop outcomes beyond frisking to proffer a more comprehensive look at differential burden by race. Comparisons of the existing comparison groups can be used to examine outcomes such as duration of the stop, justification for the frisk, number of officers present, the likelihood of use of force, and frisk outcomes like identification of a weapon or arrest of the subject. These supplementary analyses likely are not feasible as a dynamic measure but may present aggregated measures of these differences as a measure of disparity.

2.3 Summary

We believe that this EAQ metric leverages proven methodologies and existing data collection efforts to provide a reasonable, if limited, measure of racial disparity in officer decision making. We recommend that this component proceed as an integral part of the equity component of EAQ fora, with additional concurrent efforts to improve data quality and, therefore, the accuracy of this measure.

3. Component 2: Location-Based Resource Accountability

3.1 Evaluating the Premise

The primary goal of this EAQ component is to create a measure of the ratio between community need for police in a specific area and the corresponding level of police presence, to flag areas of over- and under-policing. Identification of these areas with service disparity will allow for the investigation into the causes for this mismatch in service and address behavior or planned resource allocation. Using spatial analysis will plot a known concentrations of community demand for service and using AVL data will identify where officers spend time. Assessing the overlap in these measures can demonstrate a proper dosage of police presence but will also allow for the mapping of areas with misalignment. The sourcing of adequate data and the ability to process those data efficiently will determine the appropriateness of this proposed methodology.

To assess overlap, we must measure both community need and officer presence. The proposed measure of community need is derived from the historical volume of calls for service (CFS) to the police, geocoded to its place of origin. The proposed data source for measuring the levels of police presence is the AVL data, which provide an approximation of where officers spend their time. Both data sources are reasonable proxy measures for community need and police response. The use of spatial analytics to determine areas of concordance and disparity in alignment between need and police service is a reasonable solution, as long as careful attention is paid to call criteria used to predict need and the level of geography at which the metric is aggregated. As explored in depth in the pre-implementation report, we find that both the availability of the data and the approach to analysis satisfy the research premise.

3.2 Validating the Methodology

The use of citizen CFS to approximate need and the use of AVL data to measure police presence are both well-established methodologies. The concentration of crime in a few high-volume locations is an accepted way to define and direct police patrol to these areas of high need (Sherman & Weisburd, 1995; Weisburd, 2015). Seattle's approach—leveraging predictive models of need based on past concentrations—is an extension of this logic, whose value corresponds to the quality of the call inclusion parameters. Likewise, AVL is a commonly used metric for where officers spend time (Weisburd et al., 2015; Wu et al., 2022; Telep et al., 2014).

The innovation of the current approach is examining the geographic relationship between these two metrics. Analogous studies focused on proactive policing (Wu & Lum, 2017) provide a logical foundation for exploring the overlap of police activity at a higher level of specificity. Seattle's current application of AVL and CFS analysis will largely follow initial efforts to explore these concepts in relation to the community's demand for enforcement (Atherley et al., 2022). This work was originally developed as part of a routine SPD research project that applied these insights after the development work was complete. Pending successful hurdling of technological limitations, this component is a feasible measure of the spatial and temporal overlap between demand for service and police adherence to providing that service. The primary contention with this approach is the balance between masking variability between individual streets using density-based clustering to identify larger areas, and the operational need to identify specific areas of interest that may not conform to a traditional street-based approach.

3.2.1 Initial Considerations and Current Resolution

1. **Anticipated need.** We initially pointed out limitations to using a cross-sectional approach in defining public demand for police service. There is some evidence of temporal stability in crime hot spots, but they are often classified into increasing or decreasing CFS trajectories over time. SPD is adopting a dynamic predictive approach, where the call time and location are used to define anticipated need in each of the prediction zones. What the refresh rate for service need may be is unclear, but the ability to iterate on both demand and officer presence is essential, as EAQ carries on for any extended period of time.
2. **Perceptions.** There is an outstanding question of whether community members' perceptions of over- and under-policing match the data. Although the current measure is a primarily a practical measure of accountability, this EAQ component may present a future opportunity to assess equity in service delivery compared to perceived demand and the demand articulated by community perceptions.
3. **Definition of need.** Careful attention must be paid to how need is defined within a community, specifically as it relates to sequestering resident-initiated and officer-initiated calls. Officer-initiated, or on-view, calls may be more indicative of where presence is targeted than of community need. Using existing officer activity to define need may introduce a self-justifying feedback loop, where need is defined as where officers already spend time and

initiate calls. For this reason, we suggest—and SPD agrees—that need is indicated by dispatched CFS to the police.

4. **In-transit data concerns.** Initially, we raised concerns about how AVL transmission while using thoroughfares or traffic corridors may indicate a concentration of officer presence without community need tied to those locations. Whereas it is true there will be no differentiation in AVL data pings between on-scene and in-transit status, newer features of the methodology mitigate the problem presented here. Foremost, the change in geographic clustering methods to a density-based approach (DBScan) results in 796 zones, compared with the nearly 2,500 zones previously predicted using affinity propagation. Operating at higher levels of geographic aggregation means that in-transit data points are likely to be dispersed across zones and be not problematic.
5. **Service levels at locations.** In this methodology, officer presence is assumed to be in service of community need. However, some locations are likely convenient places for administrative tasks, report writing, or meal breaks. As such, AVL location data during these times should not necessarily contribute to a measure of service-levels, although these concentrations should not and will not be disregarded from the overall analysis. These known concentrations may represent potential operational security concerns (ambush risk) and should be known. As part of the EAQ process, these individualized locations are identified and explained in the context of over-policing. Because this is an iterative process, known hot spots can be annotated and subsequently filtered out of EAQ conversations, once they have been addressed.

3.2.2 Outstanding Considerations

1. **Officer behavior vs. presence.** The differentiation between the time spent in an area and the activities conducted while in that area is absent from this EAQ metric. It may be important to capture the activities of the officers beyond when and where the AVL pings their locations. There is a functional difference between an officer driving from point A to point B through a neighborhood and an officer engaging in a 15-minute directed patrol on that block face (Nagin et al., 2015; Koper, 1995). Furthermore, a directed patrol where an officer spends time in the car is different from both community policing efforts or proactive law enforcement, which contribute differently to perceptions of over- and under-policing.

The data here are necessarily limited to officer locations. However, as this EAQ component continues to develop, considering the effects of officer behavior beyond mere presence will be important.

2. **Street segment variability.** This EAQ metric balances having enough granularity to focus on specific locations, while also maintaining a high enough geographic aggregation for predictions to be valid. We believe that SPD's intended approach strikes this balance, but we also contend that the larger polygonal divisions of the city will mask variability that may be relevant to identifying areas of over- or under-policing. The heterogeneity of crime and community need between street segments within a community is well documented (Weisburd et al., 2004; Steenbeek & Weisburd, 2016). First observed in Seattle, streets next to each other, even in places classified as "bad neighborhoods," can have very different needs for police presence, based on the heterogenous distribution of crime at the street level. The current approach, generalizing both need and presence at a meso-geographic level, can miss variability in both within the defined areas. This criticism is not intended to discount the current method, but suggest additional levels of analysis for the future, facilitated by the point level data collection of CFS and AVL pings.

3.2.3 Recommendations & Conclusion

Overall, this methodology satisfies its intended goal of creating a measure of resident need and police service. This success is contingent on functionable AVL data collection and management, and the accurate prediction of need using continuously updated CFS data. The identification of misalignment between these two spatial data layers is a creative solution to identify service disparity. These data sources are both appropriate and available, and the metric itself relies on well-established spatial statistics to identify a novel need.

Beyond its stated goals, there is potential for a positive, unintended consequence of this component. Although they note the importance of the role of leadership and organizational history, de Brito and Ariel (2017) find that the act of monitoring patrol locations can increase fidelity to assigned patrols. However, it is important to define the scope of what this metric can really say. We contend that the residuals indicating over- or under-policing compared to need should serve as conversation starters in the EAQ process, and not a definitive measure of over-policing as experienced by the community. What this metric is best positioned to do is highlight the service areas with the greatest

disparities and serve as an inflection point for considering why that disparity exists and whether any intervention is necessary.

Owed to the meso-level of geographical abstraction and the expectation that officer presence can legitimately go beyond the immediate needs of citizen crime calls, this outcome measure is not necessarily a measure of true inequity but a signal of where it might be found with further investigation. For these reasons, this EAQ metric should be framed as a useful operational tool for better managing police resources and justifying existing police presence, rather than an academic measure of how and when the public experiences these service disparities.

There are a few ways this component can be expanded. Currently, this metric relies on AVL data as a measure for police service in a community; it is the best and most accessible metric available at the timescale required for continuous monitoring. However, AVL can capture more than the strict definition of police service. Likely dispersed across the city, AVL pings during transit, meal breaks, or other administrative tasks that all contribute to our understanding of where police are engaging throughout the city. Referring to the importance of officer behavior as much as presence, as EAQ progresses, the measurement of police service might be refined to include only those officer activities that may contribute to perceptions of over- or under-policing.

Because the source data for both AVL and citizen need (call location) are at the point level, there is a rich potential to explore these concepts at a lower level of geographic aggregation. Although these efforts may not be appropriate for continuous monitoring as part of the EAQ, examination at the street segment level may help to unshroud the masked variability discussed earlier. At this microgeographic level, conversations about streets with service disparities can become a lot more specific. We believe this is worth exploring in conjunction with the current planned measure.

3.3 Summary

This EAQ component leverages well-established and accessible data sources as a reasonable measure for where police are spending time and where they ought to spend time. Due to the discussed limitations to geographic specificity and the inability to distinguish what officers are doing, we do not view this metric as an academic or definitive measure of over-policing. However, the metric does serve as a valid approach to initiating conversations about why observed disparities may occur and iteratively improve on service equity from there. We recommend this component proceed as a valuable part of the EAQ forum

4. Component 3: Officer Service Quality Assessment

The format of Component 3 will differ slightly due to the recent transition from automated body camera transcription to targeted surveys of citizens with recent police contact. This new component was not featured in the pre-implementation report, and this constitutes a first iteration of evaluation and recommendation.

4.1 Evaluating the Premise

The original stated goal of this EAQ component was to create an engaging measure of the quality of interactions between officers and community members. The use of body-worn video and classification models was the most robust methodology to achieve both the frequency and scope of the intended metric. With this methodology no longer in use, the question becomes the suitability of targeted community surveys to satisfy the EAQ goal. The appropriateness of this methodology is predicated on the belief that the collected data are both attainable and believable.

In January 2023, the department implemented a continuous measurement satisfaction survey, administered by an automated platform also used to update those accessing police service. After a community member calls 911, the system sends a set of automated messages (text and/or email) to the contact information they provided. This message confirms their request for service, provides reference information, and some limited instructions preparing them for the response (e.g., documents and materials to have available for an auto theft report), if applicable. After the officer completes service (clears the call), additional automated messages are sent asking if the community member would like to participate in a satisfaction survey. Questions about service satisfaction are presented using the Net Promoter format, whereby the community member is asked whether they would refer a friend or family member dealing with a similar issue to request service from the SPD. Additionally, change in fear of crime questions are asked. The subject is asked if their specific interaction increased or decreased their fear of crime during the day and at night, separately. Some demographic and use type (e.g., resident of the city, works in the city) and unstructured free text response are included.

Satisfaction questions are relative to the resource and phase of service the community member recently interacted with. During the initial response, the Net Promoter question is asked relative to the person the community member spoke with on the phone, the officer, and the department as a whole. If the subject is listed as a victim in a police report, after that report processes through the

Records Management System (RMS), up to 12 hours after the response, an additional victim satisfaction survey is processed. If the report results in a follow-up investigation by a detective, additional satisfaction questions are asked using the same automated process. Finally, once the case is closed (e.g., inactivated, referred for prosecution, declined by the prosecuting attorney's office), a final survey is initiated.

The Net Promoter format was selected to control survey effects and provide immediate comparability across analogue industries. As indicated previously, it is assumed the responses are biased toward those with a motivation to respond. Responses are assumed to reflect those who are extremely satisfied and extremely dissatisfied with the service they received. Given the potential to reactivate the trauma of a person who was recently the victim of a violent crime, all violent crimes against person are removed from the automated survey process. This control may eliminate an overly positive response from someone who is grateful for having had their physical safety protected, directly. The residual emotional effect (midbrain) of a highly stressful, frustrated, or otherwise victimized feeling is moderated by the Net Promoter question format. This question asks the respondent to consider whether, based on part or the totality of their experience, they might recommend a friend or family member take similar action to access services. This referential consideration deploys some additional cognitive processing, reflecting a value judgment made about an external object (a friend or family member) and is commonly employed in customer satisfaction where emotional or impression managed responses are a risk. In addition to allowing for a manageable dimension (complexity and scope of the instrument), increasing response and completion rates,¹ the Net Promoter model does not require a new scale be validated (e.g., test retest reliability, interrater reliability) and provides immediate comparability across industries.

The SPD intends to track movement, as well as cross-industry comparables, for engagement of this metric. Although policing generally suffers from a lack of comparability, the highly emotional and selection-biased nature of satisfaction responses compounds the problem. The cross-industry comparability of this metric allows the SPD and stakeholders to contextualize satisfaction scores in a meaningful way. Additionally, as is the case with disparity measures under the Equity metric (above), trends and patterns provide actionable insights. Whereas a good equity metric can be said to be *as low as reasonably achievable* (ALARA), a common strategy for managing key performance indicators in

¹ Initial operation suggests a sustained response rate of 20% over the first 2 months of operation.

safety and risk, the “good” of a quality metric can be said to be *as high as reasonably achievable* (AHARA). Identifying opportunities to increase and/or optimize quality metrics is the goal of the Quality component of EAQ and is achieved through the use of this method. In this way, it is the movement and relative context (outlier) of the quality measurement that is actionable; the effect of selection bias is effectively mitigated by its intended form of engagement.

Historically, community surveys are substantively robust and infrequent due to the cost and effort of implementation. However, SPD’s leveraging of SPIDR Tech automates the dissemination and collection of surveys and supports the premise that interaction data are widely attainable. There is an inherent loss in specificity of the metric because the data source shifts from an objective record of the interaction to the post facto perception of the surveyed resident. The SPD has given significant consideration to the selection bias inherent in this approach. As this is a convenience sample, and participation is voluntary, without incentive, it is assumed that respondents are motivated to respond. Motivation is both positive and negative: Respondents may be motivated by either an extremely high or extremely low subjective perspective on the service delivered.

However, there is no reason to doubt that the limited scope of questions asked in these surveys is believable. In theory, the use of frequent feedback can provide reasonable estimates for any measure included in the survey, although this is contingent on acquiring a sufficient volume of responses, because participation is neither automatic nor compulsory. In reality, it is important to consider whether the subset of people who do respond and their perceptions are representative of the interactions overall; these specific questions are considered below.

The logic behind this EAQ component does, in theory, present a reasonable data collection and analysis protocol that can address narrow aspects of interaction quality.

4.2 Validating the Methodology

The use of community surveys to evaluate perceptions of the police and police interactions is a well-established methodology and is becoming increasingly common (Rosenbaum et al., 2017; Merenda et al., 2021). Traditionally, these surveys are cross-sectional and provide estimates at the population level. Here, Seattle is leveraging SPIDR Tech to make these surveys more targeted and frequent following any qualifying interaction with the police, enabling a dynamic estimate of perceptions of interaction quality over time. In theory, this works to establish a baseline in the population and then to apply repeated measures to the same population to detect changes.

These approaches—using immediate post-interaction surveys—have not been academically or scientifically evaluated, although they are in use throughout public safety offices and police agencies across the country. Because of the lack of research in this area, we treat the methodology as a logical extension of the larger community surveys that are validated and used to address these same questions.

One novel expansion of this approach in Seattle is the use of Net Promoter Scores (NPS). This evaluation of services, based on whether or not one would recommend this service to a friend or relative in the same situation, is common in marketing research (Fisher & Kordupleski, 2019), but has also been expanding to the medical fields (Krol et al., 2015) and public sector agencies (Luomaaho et al., 2021). In examining the survey content specific to Seattle, we found the potential that focusing on recommendation of services may be (1) tied to either global attitudes about the police beyond the scope of the most recent interaction, or (2) driven more by outcomes than the process and treatment during the interaction (Tankebe, 2013). Given the restrictions on more direct measures of interaction quality, we see potential in this method but raise the following initial considerations as the measurement is further integrated in the EAQ process.

4.2.1 Considerations

1. **Response rate.** It is important to consider the response rate, when extrapolating survey responses into a global measure of citizen satisfaction and interaction quality. Response rate measures how many responses were received out of how many could have been. SPIDR Tech self-reports an average response rate to their post-interaction surveys at about 12.1%. According to initial data, Seattle's survey response rate is between 20% and 25%, which is better than the SPIDR Tech baseline and generally considered within range for NPS scores in other fields. The importance of response rate is contingent on the minimum viable number of respondents and the size of the effect to detect. Distributing surveys to a majority of police interactions daily still results in a large population of surveys from which to draw results. However, the risks of extrapolation increase while drawing on a smaller percentage of the overall population. Here, we risk the assumption that the 20% who respond are behaviorally and substantively the same as the 80% who do not. This raises the next consideration of selection bias.

2. **Selection bias** occurs when the responses received do not represent the population generally. There are two opportunities for selection bias in the administration of these surveys. First, the post-interaction surveys are not universally applied. Whereas the goal of this EAQ metric is to offer a proxy measure for interactions generally, the survey appears to be limited to those who willingly engaged with police in the first place. This excludes victims of violent crime, subjects of proactive police enforcement, traffic stops, and arrestees. Exclusion of this segment of the population ignores the measures of interaction quality in those scenarios where it is potentially most important. The second potential for selection bias comes from the respondent in their decision whether to respond. As indicated by the response rate, there is variability in whether people complete the survey, which poses the question of why these discrepancies exist. As with other opportunities for feedback, only those with the best and worst experiences may be willing to take the time to offer praise or criticism. Understanding and accounting for selection bias is a huge hurdle for the believability of this metric.
3. **Call type.** The cadence of these surveys allows for a highly dynamic measure that can be examined at a higher frequency than the other components, at a weekly or even daily level, to identify or explain outliers. The EAQ metrics are not designed to use individual officers as the unit of analysis; however, this metric is well set up to disaggregate the overall scores by beat or unit, which can serve as part of the incentive structure. One additional unit of analysis may be the call nature related to the survey. Although disaggregation to the call type level may not be possible or supported by the call volume at the same time as resolution, organizing interaction quality by type of call would provide some insights about which situations (for both officers and subjects) may be driving low- or high-quality ratings. This has the operational benefit of targeting additional trainings or interventions to improve ratings and NPS.
4. **Net Promoter Score.** The primary outcome measure in the survey is the NPS focused on whether these services would be recommended to someone else in a similar situation. As discussed, this may be driven by distributive justice outcomes rather than procedural fairness or professionalism. However, it does carry the unique benefit of facilitating cross-industry comparisons, which can serve as an anchoring point for understanding of and conversations about overall performance. The included measure of interaction quality is an open text field

that is not conducive to identifying trends in officer behavior, demeanor, or performance. SPIDR Tech has survey item templates directly related to officer courteousness that more directly mirror the previous attempted measure of quality. This is worth considering as an addition to the current survey.

5. **Perceptions.** In interpreting the survey results and creating the overall metric, it may be difficult to separate out responses driven by the latest interaction with the police, and engrained perceptions due to a history of direct and vicarious interactions with the police. There is evidence that perceptions of the police generally can be affected by recent direct experiences, neighborhood context, vicarious experiences shared by family, and long-held intergenerational beliefs (Harris & Jones, 2020; Wolfe et al., 2017; Fine et al., 2022). In this context, a high NPS may be due to a positive outcome from their most recent interaction, a professional experience high in procedural justice, or a stable belief in the legitimacy of police that manifests regardless of the immediate situation. Whether the reasoning for the score may be teased out is unclear, but adding questions explicitly about the context and outcomes of interest for the most recent interaction may provide additional insight.
6. **Demographics.** Previous evaluations of community perceptions of police satisfaction find that demographics such as age, education, race, and fear of crime can explain some of their ratings, beyond direct experience with the police (Haberman et al., 2016; Weitzer & Tuch, 2005). Seattle's post-interaction surveys contain many demographic questions and, contingent on the completeness of that data, these may be used to explain any observed differences in NPS or satisfaction beyond the context of the most recent interaction.

4.2.2 Recommendations & Conclusion

Because this EAQ component was not evaluated in the same iterative way as the others, the initial considerations comprise our recommendations for organizing and moving forward with this metric. Overall, we believe that the post-interaction survey satisfies the conditions for generating a narrow measure of interaction quality. Compared to the original transcription and analysis of body-worn video, reliance on citizen perceptions introduces an additional degree of subjectivity both in terms of content and the decision to participate. Although this method is an acceptable approach for following up on officer interaction quality, there are some critical changes to the actual content of the surveys that would improve this metric.

As it stands, the content is focused on global measures of police performance and approval. As discussed, these opinions can be colored with experiences far beyond the interaction in question. Either expanding or replacing the survey content to explicitly measure officer professionalism, sentiment, and procedural justice would be valuable additions in line with the goals of this EAQ component to ensure quality across a range of interactions. In extrapolating the results of these surveys, accounting for the limitations of this methodology—including selection bias, exclusion of certain types of interactions, and content that may go beyond the most recent encounter—will be essential. As the methodology develops to consider these recommendations, it may serve as a core quality metric for the EAQ program.

5. The Seattle Crime Harm Index

Rather than focus entirely on raw crime counts, practitioners and researchers have begun to examine the use of harm indexes as a way of analyzing crime. Crime counts do not distinguish between the total number of property thefts versus a robbery.

An index attempts to create consistency across disparate variables. The intent of harm indexes is to create a numerical value for crime that equalizes the type of crime by the amount of harm it generates. The first harm index created used court records as an indicator for determining harm (Ignatans & Pease, 2015). Researchers built on this approach, adding in metrics like traffic accidents and drug offenses (Ratcliffe, 2015), while others focused on victim harm (Greenfield & Paoli, 2013). However, Sherman (2013) was the first to create a crime harm index (CHI) based on sentencing guidelines for first offenses (Sherman 2007, 2013; Sherman et al., 2016).

The SPD wanted to incorporate this approach into their crime analysis process and constructed their own harm index based on Washington State offense codes. SPD chose to follow Sherman's original (2013) Cambridge CHI method and used sentencing data to create their CHI.

The creation of the Seattle Crime Harm Index (SCHI) will allow SPD to compare crime statistics by crime type without losing the variance of the harm some crime causes compared to others. CHI levels the variance between high-volume/low-harm and low-volume/high-harm crime. Sherman (2013) outlined the following steps for creating the Cambridge CHI, using the median number of prison days to calculate crime harm:

1. Count the number of crimes of each type.
2. Multiply the count for each type by the median number of prison days recommended for crimes of that type by first offenders.
3. Call the product of that multiplication (crime count for a crime type X median days in prison) the harm subtotal (HST) of days of prison for that offense type.
4. Repeat steps 1, 2, and 3 for every type of crime recorded for the area or person.
5. Sum up all HSTs to yield the total crime harm (TCH).

SPD emulated the Cambridge CHI steps to develop the SCHI. SPD began the construction of the SCHI by requesting and receiving offense-level data from the Washington State Center for Court

Research, Administrative Office of the Courts. SPD requested data from 2008 to Nov. 21, 2021, to include outcomes (guilty/not guilty) and the resulting sentence for first offenses. SPD calculated the average length in days of sentences for first offenses without sentencing “enhancements” where the verdict was guilty. This became the SCHI value.

Where SPD could not calculate an average for the specific Revised Code of Washington (RCW), SPD consulted the Washington State Adult Sentencing Guidelines Manual. SPD took an average of the highest and lowest of the sentencing range (without “enhancements”) to determine the index value. Monetary fines were converted to the index value by taking the dollar value and dividing it by the minimum wage for the City (\$13.50 in 2020) to determine the number of hours. The result was divided by 8 to create an equivalent to the sentencing days. Finally, to produce a severity score, SPD matched data from the Incident-Offense Data Source and assigned a severity score based on the type of aid response and the Seattle CHI score.

However, SPD was still awaiting information about misdemeanor sentences from the municipal court as of March 2023. Prior to 2020, SPD had a well-established relationship with a data analyst with the municipal court. Unfortunately, that resource left the court; since then reconnecting with municipal court has been difficult. Like all City resources, the court has an overload of public requests and is still recuperating from staffing turnover issues.

The primary goal of SPD was to incorporate the SCHI values into the data warehouse (DAP) Incident-Offense data source. When the SCHI is complete and included in the DAP, it will allow SPD and their research partners to study the concepts of harm in policing and develop alternative deployment strategies. Currently, in the prototype and proof of concept, SPD is only utilizing the SCHI component as an additional identifier in the murder/homicide cases as part of the match verification. SPD expects to have a fully functioning model once they can retrieve sentencing data from the municipal court.

References

Component 1:

- Bandes, S. A., Pryor, M., Kerrison, E., & Goff, P. (2019, February 6). The mismeasure of Terry stops: Assessing the psychological and emotional harms of stop and frisk to individuals and communities. *Behavioral Sciences & the Law*, 37(2), 176–194.
- Chanin, J., & Sheats, B. (2018). Depolicing as dissent shirking: Examining the effects of pattern or practice misconduct reform on police behavior. *Criminal Justice Review*, 43(2), 105-126.
<https://doi.org/10.1177/0734016817704696>
- Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151–161.
<https://doi.org/10.1162/003465302317331982>
- Gau, J. M., & Brunson, R. K. (2010). Procedural justice and order maintenance policing: A study of inner-city young men's perceptions of police legitimacy. *Justice Quarterly*, 27(2), 255-279.
<https://doi.org/10.1080/07418820902763889>
- Govindasamy, P. (2016). *A comparison between propensity score matching, weighting, and stratification in multiple treatment groups: A simulation study* [Doctoral dissertation]. University of Denver.
<https://digitalcommons.du.edu/etd/1173/>
- King, G., & Nielsen, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4), 435–454. <https://doi.org/10.1017/pan.2019.11>
- Levchak, P. J. (2021). Stop-and-frisk in New York City: Estimating racial disparities in post-stop outcomes. *Journal of Criminal Justice*, 73(March-April), 101784.
<https://doi.org/10.1016/j.jcrimjus.2021.101784>
- Olmos, A., & Govindasamy, P. (2015). A practical guide for using propensity score weighting in R. *Practical Assessment, Research, and Evaluation*, 20(13), 1–8.
<http://pareonline.net/getvn.asp?v=20&n=13>

Rosenbaum, P. R., & Rubin, D. B. (1983, April). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55.

<https://doi.org/10.1093/biomet/70.1.41>

Seattle Police Department (SPD). (2019). *Disparity Review – Part 1*. Case 2:12-cv-01282-JLR Document 554-1.

Stone, C., Foglesong, T. S., & Cole, C. M. (2009). *Policing Los Angeles under a consent decree: The dynamics of change at the LAPD*. Cambridge, MA: Program in Criminal Justice Policy and Management, Harvard Kennedy School.

Component 2:

Atherley, L. T., Hickman, M. J., Parkin, W. S., & Helgott, J. B. (2022). Measurement of potential over- and under-policing in communities. *Policing: A Journal of Policy and Practice*, 16(3), 443–461. <https://doi.org/10.1093/police/paac025>

de Brito, C., & Ariel, B. (2017). Does tracking and feedback boost patrol time in hot spots? Two tests. *Cambridge Journal of Evidence-Based Policing*, 1, 244–262. <https://doi.org/10.1007/s41887-017-0018-7>

Koper, C. S. (1995). Just enough police presence: Reducing crime and disorderly behavior by optimizing patrol time in crime hot spots (NCJ Number: 167664). *Justice Quarterly*, 12(4), 649–672. <https://doi.org/10.1080/07418829500096231>

Nagin, D. S., Solow, R. M., & Lum, C. (2015). Deterrence, criminal opportunities, and police. *Criminology*, 53(1), 74–100. <https://doi.org/10.1111/1745-9125.12057>

Sherman, L. W., & Weisburd, D. (1995). General deterrent effects of police patrol in crime “hot spots”: A randomized, controlled trial (NCJ Number: 167663). *Justice Quarterly*, 12(4), 625–648. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/general-deterrent-effects-police-patrol-crime-hot-spots-randomized>

Steenbeek, W., & Weisburd, D. (2016). Where the action is in crime? An examination of variability of crime across different spatial units in The Hague, 2001–2009. *Journal of Quantitative Criminology*, 32, 449–469. <https://doi.org/10.1007/s10940-015-9276-3>

Telep, C. W., Mitchell, R. J., & Weisburd, D. (2014). How much time should the police spend at crime hot spots? Answers from a police agency directed randomized field trial in Sacramento, California (NCJ Number: 248011). *Justice Quarterly*, *31*(5), 905–933.

<https://doi.org/10.1080/07418825.2012.710645>

Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology*, *53*(2), 133–157. <https://doi.org/10.1111/1745-9125.12070>

Weisburd, D., Bushway, S., Lum, C., & Yang, S. M. (2004). Trajectories of crime at places: A longitudinal study of street segments in the city of Seattle. *Criminology*, *42*(2), 283–322.

<https://doi.org/10.1111/j.1745-9125.2004.tb00521.x>

Weisburd, D., Groff, E. R., Jones, G., Cave, B., Amendola, K. L., Yang, S. M., & Emison, R. F. (2015). The Dallas patrol management experiment: Can AVL technologies be used to harness unallocated patrol time for crime prevention?. *Journal of Experimental Criminology*, *11*, 367–391. <https://doi.org/10.1007/s11292-015-9234-y>

Wu, X., & Lum, C. (2017). Measuring the spatial and temporal patterns of police proactivity. *Journal of Quantitative Criminology*, *33*, 915–934. <https://doi.org/10.1007/s10940-016-9318-5>

Wu, X., Koper, C., & Lum, C. (2022). Measuring the impacts of everyday police proactive activities: Tackling the endogeneity problem. *Journal of Quantitative Criminology*, *38*, 242–262.

<https://doi.org/10.1007/s10940-021-09496-8>

Component 3:

Fine, A. D., Padilla, K. E., & Tom, K. E. (2022). Police legitimacy: Identifying developmental trends and whether youths' perceptions can be changed. *Journal of Experimental Criminology*, *18*, 67–

87. <https://doi.org/10.1007/s11292-020-09438-7>

Fisher, N. I., & Kordupleski, R. E. (2019). Good and bad market research: A critical review of Net Promoter Score. *Applied Stochastic Models in Business and Industry*, *35*(1), 138–151.

<https://doi.org/10.1002/asmb.2417>

- Haberman, C. P., Groff, E. R., Ratcliffe, J. H., & Sorg, E. T. (2016). Satisfaction with police in violent crime hot spots: Using community surveys as a guide for selecting hot spots policing tactics (NCJ Number: 249608). *Crime & Delinquency*, 62(4), 525–557.
<https://doi.org/10.1177/0011128713516840>
- Harris, J. W., & Jones, M. S. (2020, March-April). Shaping youths' perceptions and attitudes toward the police: Differences in direct and vicarious encounters with police. *Journal of Criminal Justice*, 67, Article 101674. <https://doi.org/10.1016/j.jcrimjus.2020.101674>
- Krol, M. W., de Boer, D., Delnoij, D. M., & Rademakers, J. J. (2015). The Net Promoter Score—an asset to patient experience surveys? *Health Expectations*, 18(6), 3099–3109.
<https://doi.org/10.1111/hex.12297>
- Luoma-aho, V., Canel, M. J., & Hakola, J. (2021). Public sector reputation and netpromoter score. *International Review on Public and Nonprofit Marketing*, 18, 419–446.
<https://doi.org/10.1007/s12208-021-00280-9>
- Merenda, F., Trent, J., Rinke, C. R., & Buchanan, M. (2021). Understanding citizen satisfaction with the police: Results from a community survey. *Police Practice and Research*, 22(1), 692–710.
<https://doi.org/10.1080/15614263.2020.1772784>
- Rosenbaum, D. P., Maskaly, J., Lawrence, D. S., Escamilla, J. H., Enciso, G., Christoff, T. E., & Posick, C. (2017). The police-community interaction survey: Measuring police performance in new ways. *Policing: An International Journal of Police Strategies & Management*, 40(1), 112–127.
<https://doi.org/10.1108/PIJPSM-07-2016-0119>
- Tankebe, J. (2013). Viewing things differently: The dimensions of public perceptions of police legitimacy. *Criminology*, 51(1), 103–135. <https://doi.org/10.1111/j.1745-9125.2012.00291.x>
- Weitzer, R., & Tuch, S. A. (2005). Determinants of public satisfaction with the police. *Police Quarterly*, 8(3), 279–297. <https://doi.org/10.1177/1098611104271106>
- Wolfe, S. E., McLean, K., & Pratt, T. C. (2017). I learned it by watching you: Legal socialization and the intergenerational transmission of legitimacy attitudes. *British Journal of Criminology*, 57(5), 1123–1143. <https://doi.org/10.1093/bjc/azw038>

Crime Harm Index

- Greenfield, V. A., & Paoli, L. (2013, September). A framework to assess the harms of crimes. *British Journal of Criminology*, 53(5), 864–886. <https://doi.org/10.1093/bjc/azt018>
- Ignatans, D., & Pease, K. (2015). Taking crime seriously: playing the weighting game. *Policing*, 10(3), 184–193. <https://doi.org/10.1093/police/pav029>
- Ratcliffe, J. H. (2015, June). Towards an index for harm-focused policing. *Policing*, 9(2), 164–182. <https://doi.org/10.1093/police/pau032>
- Sherman, L. (2013, May 20). The rise of evidence-based policing: targeting, testing, and tracking. *Crime and Justice*, 42, 377–343. <https://doi.org/10.1086/670819>
- Sherman, L. (2007, November 9). The power few: experimental criminology and the reduction of harm. *Journal of Experimental Criminology*, 3, 299–321. <https://doi.org/10.1007/s11292-007-9044-y>
- Sherman, L., Neyroud, P. W., & Neyroud, E. (2016, April 3). The Cambridge Crime Harm Index: measuring total harm from crime based on sentencing guidelines. *Policing*, 10(3), 171–183. <https://doi.org/10.1093/police/paw003>