Overview

The Southern Police Institute (SPI) is associated with the University of Louisville's Department of Criminal Justice. SPI's purpose is to advance professional development for law enforcement practitioners. The program consists of twelve weeks (480 hours) of in-residence, college level programming focused on equipping police leaders with executive skills.

In 2024, Captain Victoria Weyda attended SPI, where she completed a project focused on reviewing police early intervention systems. Her paper, reformatted, is below.

Introduction

Early intervention systems are not a new concept for police management. In 1981, The US Civil Rights Commission reported that police departments often have a few officers that receive a disproportionate number of complaints and use of force investigations. Since then, this phenomenon has been discovered again and again. The Los Angeles County Sheriff's department discovered that 62 deputies accounted for 500 use of force/harassment complaint investigations (Walker, 2003). Kansas City discovered that 2% of their officers generated 50% of all citizen complaints (Walker, 2003). This phenomenon is still relevant and hits close to home. From doing an informal analysis, I discovered that in Cincinnati between 4/4/23 to 4/4/24, 4% of our officers generated 33% of our citizen complaints. Cincinnati implemented an early intervention system in 2002 as a part of the Collaborative Agreement, which could account for why our numbers are not as disproportionate as other cities. However, no research has been undertaken to prove the efficiency or validity of the current EIS.

While early intervention systems have been used in law enforcement for decades, there is little research evaluating the effectiveness of these systems. There are specific barriers in reviewing and analyzing research on EIS for best practices. First, there is no agreed upon definition of what a successful EIS looks like and how to measure it. Some studies, like Helsby et al (2008) and Lersch et al (2006) focus on the rate of true positives and false positives – how many of the flagged officers went on to have a negative incident. Some studies, like Shjarback (2015) and Briody and Prenzler (2020) evaluated whether implementing the EIS reduced the number of citizen complaints for the entire department. Finally, some studies like Macintyre et al (2008) and Jackson et al (2017) evaluated whether flagged officer's complaints went down following identification and intervention. The second barrier is the lack of rigorous experimental designs. Worden et al (2013) reviewed and analyzed prior EIS studies. He noted that there are alternate explanations for

the change in rates of citizen complaints and use of force other than the EIS intervention. Often when an officer is flagged, it is due to unusually high activity, so a reduction following that could be an expected regression to the mean, as opposed to evidence that the intervention worked. There is a question on whether the metrics being tracked are indicative of an active officer, not necessarily a problem officer. Finally, most studies have not factored in an officer's activity into their pre-post indicators. It is entirely possible and even likely that officers that trigger a flag in the EIS reduce their activity to decrease the likelihood of having additional uses of force and complaints.

The third barrier is the inconsistency in the metrics being tracked to determine if an officer should be flagged. Nearly all police agencies track citizen complaints, even though, on average 90% of citizen complaints are not substantiated and citizen complaint numbers may simply be a measure of officer activity (Lersch, 2001). Lersch et al (2006) also evaluated an EIS based primarily on use of force data. In that agency, an officer was flagged based on having four or more "high" levels of use of force in a quarter. However, when they examined the top 90th percentile of officers using force the most often, none of those officers were flagged by the system; even though they used force in 74% of the opportunities they had to use force. On the other hand, they examined the top 90% percentile of officers having the most opportunity to use force. On average those officers used force in 5% of the opportunities that they had, and 16 of those officers were flagged.

Literature Review – What Works and What Doesn't

I have reviewed seven articles that attempt to quantify whether an EIS system was successful. As stated previously, the methodologies and measurements are dramatically different which makes it difficult to determine what, if anything works. There is, however, evidence of what does not work. This literature review will start by outlining the evidence that an EIS based on an officer exceeding a set threshold in citizen complaints and/or use of force does not predict problem behavior. I will then outline a few articles that did find positive results and what they are doing differently.

Shjarback (2015) did a comprehensive review of new EIS around the country on a broader scale – reviewing whether implementing an EIS reduced citizen complaints for the department. Every police department with over 100 employees that reported implementing an EIS between 2003-2007 to the BJS was included. Shjarback did not make any effort to analyze or group departments based on how their EIS flagged officers. However, Walker published an EIS planning guide in 2003 on behalf of the Department of Justice. This guidebook specifically addressed how to determine thresholds for an EIS, so it is fair to assume that that was the common model from 2003-2007. Across all agency sizes,

they found no significant difference in the rates of citizen complaints before and after implementing an EIS.

James et al (2021) reviewed a threshold-based EIS system for a large municipal department. To evaluate the effectiveness, they created a group of officers "flagged" by the EIS system and a compare group of unflagged officers. The researchers graded 1000 incident reports created between both groups based upon the officer's performance on-scene. They found no significant differences between the officer performance between the groups. They did find the EIS flagged officers performed slightly better in the observe-assess phase and were slightly more likely to use reasonable force. However, the EIS system was unable to identify officers displaying problem behaviors.

Lersch et al (2006) also evaluated an EIS at a large municipal agency. The qualifiers to be flagged included receiving IA complaints, "high" use of force incidents, police shooting incidents, and citizen complaints. While reviewing flagged officers, they found that the only qualifier that resulted in officers being flagged was the "high" use of force. To further examine the implications of officers being flagged based on the number of uses of force, they examined the top 90% percentile of the number of times an officer used force, and the top 90% percentile for the number of opportunities an officer had to use force. They found that in the first group, on average, the officers used force 74% of the time they had the opportunity. This could be taken as an indication that the officers lack de-escalation skills and are more likely to end up in a serious incident. However, none of those officers were flagged by the EIS system. In the second group, they found that officers who had the most opportunities to use force, on average, used force 5% of the time. The EIS flagged 16 of those officers. This suggests that simple numerical threshold of the number of uses of force is not a good indicator of problem behavior. A ratio between uses of force and opportunities or uses of force and arrest would be more valuable.

Briody and Prenzler (2020) evaluated an EIS in New Zealand. While the New Zealand EIS was also triggered based on citizen complaints, they had many more factors that could generate alerts, including: sick leave taken, use of force, vehicle pursuits, employment investigations, etc. The New Zealand EIS is also unique in that after an initial meeting with their supervisor, any other interventions are voluntary. Over a five-year span, approximately 62% of the cases did not move past the initial meeting. While they found a 63% reduction in complaints for officers who participated in the program, it did not change their department's overall complaint rates.

Macintyre et al (2008) explored a novel process being used by the Australian Victoria Police Department. While this system also operates based on thresholds, any officer flagged was

reviewed by the "Research and Risk Unit". This unit did a comprehensive review of the officer's activity and personnel data. The RRU then compiled a report on the officer evaluating their risk and providing recommendations for their supervisor. To evaluate the program, they compared the number of complaints the officer received prior to being profiled by the RRU and afterwards. For the 44 members who were profiled, in the prior 4 years they averaged 15.125 complaints per quarter. Following the EIS intervention, they averaged 4.375 complaints per quarter. This program appears to be extremely effective in reducing citizen complaints; however, it is extremely labor intensive. The RRU profiles took, on average, 4 weeks to compile. For a staff of approx. 15,000 sworn officers, they only profiled 44 officers.

Helsby et al (2018) discussed some of the limitations of threshold-based early intervention systems. Some of these include the fact that threshold-based systems are easy for officers to game – they know when they are approaching the limits and can reduce activity to not exceed them. Also, threshold-based systems are naturally binary with flagged vs not flagged employees. A more robust EIS can evaluate risk factors and assign a weighted ranking to the officers. This study specifically evaluated a machine learning EIS implemented in Charolette and compared it to their existing threshold-based EIS. For the purposes of evaluation, they considered a flagged officer who went on to have an adverse incident within a year of being flagged. By using a machine learning system their true positive rate increased from 34 to 60; and their false negative rate decreased from 323 to 294. While these were statistically significant improvements, the system missed nearly 300 officers that went on to have an adverse incident.

 Jackson et al (2017) created an EIS for a mid-sized Sheriff's agency. To evaluate the effectiveness, they sought out "problem" officers. These officers had been terminated, resigned in lieu of termination, or had serious suspensions. A control officer was paired for each problem officer based on working the same job classification, in the same area, and the same number of years of service. They then retroactively applied their EIS to all included officers to determine the rate of true positives and false positives. For the purposes of the study, they created a "late warning" condition and an "early warning" condition. The late warning condition required an IA investigation with sustained charges. They found that this condition was extremely effective at identifying problem officers, however, doesn't meet the goal of identifying the problem officer *before* the negative incident.   For the "early warning" condition, they flagged any officer that had any IA investigation, and anything marked as "needs improvement" on their performance review OR any officer with 5 or more IA incidents in the last year. The early warning conditions resulted in 65% true positives and 15% false positives.

Conclusions

EIS continues to be recommended by the Justice Department and considered a "best practice", even though very few departments evaluate their EIS, and of those that do, most are not successful at predicting problem officers. The research shows that a simple count of citizen complaints or count of use of force to flag officers fails to identify problem officers. There are systems, however, that use many more types of indicators that appear to be promising. The system in New Zealand evaluated by Briody and Prenzler used many types of data and was successful at lowering complaints for individual officers. The system in Australia evaluated by Macintye et al initially flagged officers based on complaints. However, they followed up with an in-depth review of each officer prior to recommending an intervention. While this was also successful at lowering complaints for individual officers, the in-depth profile is too labor intensive for most agencies.

A new trend in EIS is machine learning and artificial intelligence. This was tested in Charlotte by Helsby et al. They did find that, compared to a traditional threshold system, their true positives increased, and false negatives decreased. However, the system still only identified 17% of the officers who went on to have a negative incident; so, while this may be more effective, it still isn't very effective. Finally, Jackson et al reviewed an EIS at a sheriff's office. This was done retroactively, by starting with the officers that had a discipline issue and determining what factors could have been used to predict it. Unsurprisingly they found that a sustained IA investigation was the best predictor. But they did find that any IA investigation plus needing improvement on a performance review was a good predictor.

What remains clear is that it is important to evaluate an EIS system both forensically and at regular intervals moving forward. We should not focus too heavily on using citizen complaints and use of force as the primary drivers and instead use a variety of personnel activity and data. Finally, it is important to be flexible and willing to change the indicators and thresholds as the data changes.

References

Briody, M., & Prenzler, T. (2020). The New Zealand Police early intervention system: A review of implementation and impact issues. *International Journal of Police Science & Management*, *22*(3), 297-307.

Helsby, J., Carton, S., Joseph, K., Mahmud, A., Park, Y., Navarrete, A., ... & Ghani, R. (2018). Early intervention systems: Predicting adverse interactions between police and the public. *Criminal justice policy review*, *29*(2), 190-209.

Jackson, B. A., Towe, V. L., Wagner, L., Hunt, P., Greathouse, S., & Hollywood, J. S. (2017). Managing officer behavioral risk using early intervention systems: addressing system design challenges for law enforcement and corrections environments. *Policing: A Journal of Policy and Practice*, *11*(1), 103-117.

James, S., James, L., & Dotson, L. (2021). Evaluating the effectiveness of a police department's early intervention system. *Journal of Experimental Criminology*, *17*, 457-471.

Macintyre, S., Prenzler, T., & Chapman, J. (2008). Early intervention to reduce complaints: An Australian Victoria police initiative. *International Journal of Police Science & Management*, *10*(2), 238-250.

Lersch, K. M. (2002). Are citizen complaints just another measure of officer productivity? An analysis of citizen complaints and officer activity measures. *Police Practice and Research*, *3*(2), 135-147.

Lersch, K. M., Bazley, T., & Mieczkowski, T. (2006). Early intervention programs: an effective police accountability tool, or punishment of the productive?. *Policing: An International Journal of Police Strategies & Management*, *29*(1), 58-76.

Shjarback, J. A. (2015). Emerging early intervention systems: An agency-specific pre-post comparison of formal citizen complaints of use of force. *Policing: A Journal of Policy and Practice*, *9*(4), 314-325.

Walker, Samuel. *Early intervention systems for law enforcement agencies: A planning and management guide*. Washington DC: US Department of Justice, Office of Community Oriented Policing Services, 2003.

Worden, R. E., Kim, M., Harris, C. J., Pratte, M. A., Dorn, S. E., & Hyland, S. S. (2013). Intervention with problem officers: An outcome evaluation of an EIS intervention. *Criminal justice and behavior*, *40*(4), 409-437.