Routledge
Taylor & Francis Group

a OPEN ACCESS | Check for updates

Police lethal force errors and stress physiology during video and live evaluation simulations

Paula M. Di Nota [ID][a], Juha-Matti Huhta[b,c], Evelyn C. Boychuk [ID][a] and Judith P. Andersen [ID][a]

[a]Health Adaptation Research on Trauma (HART) Lab, Department of Psychology, University of Toronto, Mississauga, Ontario, Canada; [b]Police University College of Finland, Tampere, Finland; [c]Faculty of Education and Culture, University of Tampere, Tampere, Finland

ABSTRACT

Police officers are regularly evaluated for their competency in a variety of skills related to use of force (UOF), including lethal force decision-making, which is usually tested using stressful reality-based scenarios in virtual or live formats. The current observational study fills a literature gap by examining performance (i.e., shoot/no-shoot errors) and stress physiology among 187 police officers during virtual (i.e., video-based) and live UOF scenarios as part of their agency's annual requalification assessment. While moderately low rates of lethal force errors we\re observed overall, there were significantly fewer errors in live (0.81%) versus video scenarios (5.92%). Both conditions elicited significant stress physiology, as measured by heart rate (HR) relative to rest, with higher maximum heart rate in live scenarios. Based on emerging empirical literature and the current findings, we contribute to the discussion on the practical benefits and limitations of video and live simulation approaches in policing. We also provide evidence-based recommendations on how each approach may be most effectively employed for the purpose of evaluating police officers' UOF skills.

1 Introduction

Limited research examines the methods used by police agencies to evaluate officers during use of force (UOF) requalification assessment. Police evaluation lacks industry-wide standards and varies in content and mode of delivery, including the use of live and virtual simulations (CCJ, 2021; Kleygrewe et al., 2022). Requalification assessment, often held annually, is a standard procedure that recertifies officers to maintain their occupational authorities, including employing UOF tools and tactics (DiNota, Chan, et al., 2021). Most applied research on the physical and cognitive (i.e., decision-making) skills related to UOF focuses on training rather than evaluation (for reviews see Bennell et al., 2007; DiNota & Huhta, 2019; DiNota, Andersen, et al., 2021; Jenkins et al., 2020). Given the life-or-death consequences associated with the authority to use force, the current study addresses an important gap in the literature by examining UOF *requalification* procedures in contemporary police practice at a large Canadian municipal agency.

Reality-based simulations test the effective use of complex combinations of UOF-related skills during dynamic encounters typically faced by officers in the field (DiNota, Andersen, et al., 2021). Police agencies utilize reality-based simulations, also called scenarios, to elicit pre-determined

behavioural outcomes depending on the specific skill that is intended for evaluation (e.g., lethal force decision-making while engaging an armed hostile suspect; verbal de-escalation of an individual in crisis) (Koedijk et al., 2021). Scenarios can be presented in live or virtual formats. Live simulations involve in-person interactions between an officer responding to trained UOF instructors or actors delivering scripted scenarios in staged environments (e.g., apartment dwelling, school). Virtual simulations include a variety of formats, ranging from basic video-based configurations to high-fidelity virtual (VR) or augmented reality (AR) systems. While the latter virtual equipment can measure shot accuracy using simulated duty weapons and dynamically modify situational factors (e.g., suspect actions or responses) based on an officer's commands or behaviours in real-time, many small to medium sized (i.e., less than 1000 officers) police agencies do not (yet) have the resources to purchase or maintain these sophisticated systems. Therefore, the current study observed a relatively more basic virtual configuration whereby pre-recorded video scenarios were projected onto a wall in a training facility (see Materials and Methods and Supplementary Materials for details). Despite widespread use of these various evaluation methods in current police practice, there is a lack of empirical research investigating the evaluation of UOF decision-making outcomes in both live and virtual modalities in the same sample.

An integral component of designing simulations that reflect reality – in live or virtual formats – is to mimic the stress that an officer would face in the field, which is known to effect UOF decision-making (Baldwin et al., 2022). A common indicator of stress physiology in police research is heart rate (HR) reactivity (i.e., increases relative to resting HR), which can be non-invasively measured during live (Andersen & Gustafsberg, 2016; Andersen et al., 2018; Baldwin et al., 2022) and virtual UOF scenarios (Kleygrewe et al., 2023; Michela et al., 2022; Nieuwenhuys et al., 2015; Nieuwenhuys, Savelsbergh, et al., 2012). The use of HR is the most valid non-invasive biomarker of autonomic arousal given that it is more robust to respiration and movement artifacts that render other cardiovascular (e.g., heart rate variability) and neuroendocrine biomarkers (e.g., cortisol) unreliable during live-action events (Chan et al., 2022; Billman, 2013; Hayano & Yuda, 2019; Thayer & Sternberg, 2006). The current study measured HR reactivity to confirm that scenarios elicited stress physiology comparable to field operations and evaluation studies (Anderson et al., 2002; Baldwin et al., 2019, 2022). Further, the use of both subjective (e.g., behavioural ratings) and objective (e.g., physiological biomarkers) key variables of interest is a recommended standard in evidence-based policing research, rather than reliance on only one type of outcome (DiNota, Chan, et al., 2021; Koedijk et al., 2021).

The growing interest in virtual simulation approaches for police UOF evaluation (and training) has culminated in several reports (Bennell & Jones, 2004; Bennell et al., 2007; Giessing, 2021; Michela et al., 2019; Murtinger et al., 2021; Zechner et al., 2023) but only a single experimental study that directly compares UOF training responses in virtual and live simulation conditions. Kleygrewe et al. (2023) found greater autonomic arousal (i.e., maximum HR) and physical activity in a single extended live scenario and greater mental effort in three shorter virtual scenarios, reflecting differential physiological effects and implementation strategies (i.e., number and duration of individual scenarios) for police UOF training scenarios. The aim and unique contribution of the current study is to fill a remaining gap in literature by comparing UOF performance outcomes (i.e., lethal force errors) in live and virtual (i.e., video-based) simulations during a police agency's in-house recertification assessment (i.e., not a training exercise).

2 Materials and Methods

The current study methods have been reported in detail in the Supplementary Materials and elsewhere (see Andersen et al., 2021; Di Nota et al., 2021), with key information reported below. Importantly, none of the scenarios were experimentally manipulated and are therefore representative of the professional UOF evaluation approaches, outcomes, and evaluation criteria currently available in the literature as reported by North American and European police agencies (DiNota, Chan, et al., 2021; Kleygrewe et al., 2022, 2023).

2.1 Participants

Data were collected from a total of 187 (31 female, 2 no sex provided, M_{Age} = 33.3 ± 6.2, $M_{Experience}$ = 7.48 ± 5.3 years) active-duty frontline police officers at a large urban police service () during the agency's annual requalification assessments. Participants were informed that their behavioral and heart rate (HR) data were being collected for research purposes only and would not impact their evaluations. All procedures were approved by the Human Participants Research Ethics Board at the University of Toronto, and all subjects gave written consent prior to their participation.

2.2 Procedure

At the start of each annual requalification assessment day, willing volunteers were fitted with HR monitors before completing two UOF simulation blocks (video, live) comprised of five total individual scenarios: two 'shoot' (one live, one video) and three 'no-shoot' scenarios (two video, one live) that were representative of calls officers typically respond to in that region. All scenarios began with an audible dispatch describing the scene, instruction to use all tactics and techniques that would be used in the field (e.g., communication, tools, skills), and were performed consecutively in the same order with no break in between except for facilitator debriefs and feedback. A total of 110 officers performed the live simulation block first and 77 officers performed the video simulation block first. Potential order effects (i.e., making more errors in the first or second block) are not a possible confound as only one participant made an error in more than one scenario (one live, one video).

Live simulations took place at the police agency in rooms furnished to resemble real-world indoor environments (e.g., apartment, rooming house) with approximately 12 m^2 of functional space. The average duration of the live simulation block, including dispatch and feedback, was 13 mins 19 sec (n = 186, min = 5:38, max = 24:28). Detailed procedures for the **video simulations** can be found in the Supplementary Materials and (Andersen et al., 2021; Di Nota et al., 2021), which included three pre-recorded video scenarios filmed at real indoor and outdoor locations (e.g., park, warehouse, office building). Videos were projected onto a screen in a matted training room at the agency with an area of functional space similar to the live simulations, minimizing the confounding effects of differences in range of motion between conditions. The average duration of the video block, including dispatch and feedback, was 11 mins 52 sec (n = 174, min = 3:37, max = 20:02).

2.3 Measures & Analyses

Lethal force decision-making errors were scored by certified UOF instructors as a binary correct or incorrect response; officers received a score of 1 for each decision-making error (i.e., shooting during a no-shoot scenario or failing to use appropriate force when required to). Error rates were computed for each simulation condition (live, video) and decision type (shoot, no-shoot) by dividing the total number of errors for each scenario by the total number of decision-making opportunities (i.e., the total number of observations for each scenario) and converting to a percentage.

Autonomic arousal was recorded using Bodyguard 2 monitors (FirstBeat Technologies Ltd., Jyväskylä, FI) worn under the officer's clothing during all evaluations. Using time-stamped notes, two cardiovascular measures were derived: 1) average resting HR (HRRest) during 5 mins of seated rest in the morning prior to any assessments and 2) average maximum HR (HRMax), which was calculated by averaging the 5 sec before and after the peak HR value achieved during the scenarios (see Andersen et al., 2021). Due to time-keeping limitations, start and end times for individual video scenarios were not recorded. Thus, HRMax in the video condition is calculated by averaging the 5 sec before and after the peak HR value achieved between the start of the first video scenario to the end of the last video scenario. The higher HRMax value from the two live scenarios was analysed for

Table 1. **Video and live simulation error rates**. All 187 participants performed the video and live simulations as part of their agency's annual use of force requalification day. Eight behavioral data points were missing due to facilitators not recording a score. The number of observations included in error rate calculations are indicated for each simulation condition (video, live) and individual scenario.

	Video Simulations			Live Simulations	
	Scenario 1 Knife in park	Scenario 2 Warehouse	Scenario 3 Insecure premises	Scenario 1 Rooming house	Scenario 2 Mental health call
Decision type	No-shoot	Shoot	No-shoot	Shoot	No-shoot
n	185	186	186	186	184
Number of errors	14	5	14	3	0
Error rate (%)	7.57	2.69	7.53	1.61	0.00
	Video error rate: 33/557 = **5.92%**			**Live error rate**: 3/370 = **0.81%**	

each participant. Due to technology failure, three officers had no HR data, six officers are missing HRRest values, two officers are missing all HRMax values but have HRRest, and two officers are missing HRMax for video simulations only. All available participant data were included in the analyses. Outlier analyses reveal no HRMax values >3 SD from the mean during video or live simulations.

Shapiro Wilk tests revealed that none of the study outcomes were normally distributed (ps <0.03). Accordingly, non-parametric within-subject analyses were conducted using Wilcoxon Signed-Rank tests to compare lethal force errors between video and live simulations, HRRest to HRMax during video and live simulations, and HRMax during video and live simulations. Significance criteria was set to $p = 0.05/8 = 0.006$ to account for multiple comparisons. Statistical analyses were conducted using SPSS (Version 28, IBM Corp.).

3 Results

3.1 Lethal Force Decision-Making

Lethal force error rates were significantly higher in the video simulations (5.92%) compared to live simulations (0.81%, Table 1) ($z = -3.860$, $p < 0.001$) (Figure 1). According to decision type, there
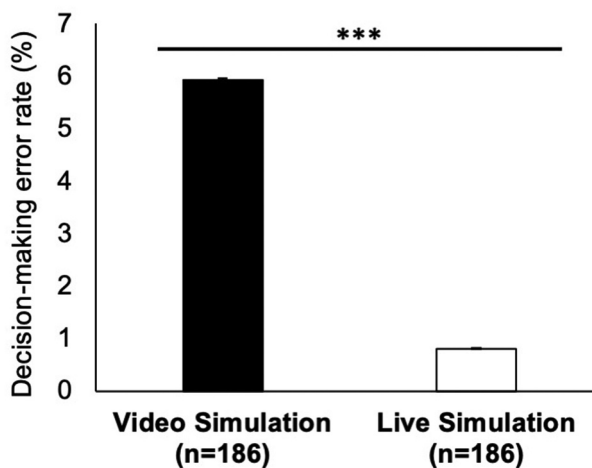

Figure 1. Police lethal force error rates during video and live simulations. When averaging across multiple scenarios and use of force decision types (i.e., shoot/no-shoot), officers made significantly more decision-making errors during video simulations relative to live simulations. Error bars show SEM.
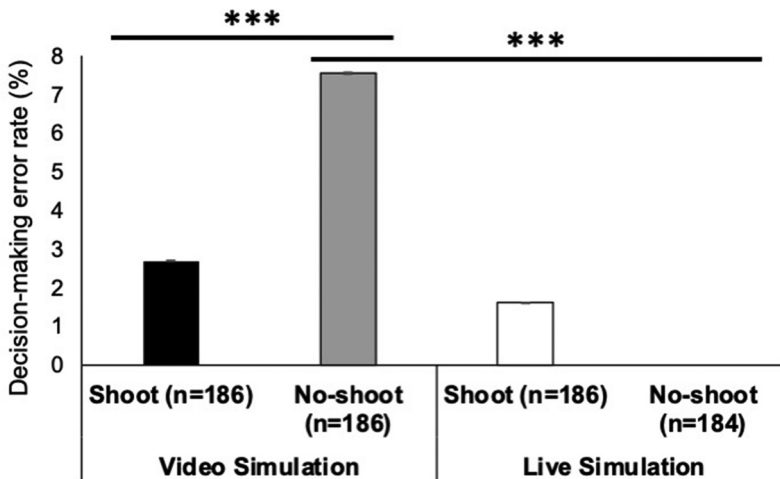Note: *** $p < 0.001$

Figure 2. Police lethal force error rates by decision type during video and live simulations. within the video simulation condition, officers made significantly more decision-making errors during no-shoot scenarios (grey bar) relative to shoot scenarios (black bar). Lethal force error rates did not differ between live simulation decision types, with no errors made by any officers in a no-shoot scenario responding to an individual in mental crisis. Error bars are extremely small, reflecting low variability in error rates, and are plotted to show SEM.
Note: *** $p < 0.001$

were significantly more errors during the two no-shoot video scenarios (7.55%) compared to the shoot video scenario (2.69%, $z = -4.116$, $p < 0.001$, Figure 2). No significant differences were observed between video and live shoot scenarios ($z = -0.707$, $p = 0.480$) or between the two decision types in the live simulations ($z = -1.732$, $p = 0.083$), with no errors observed during the no-shoot scenario and an error rate of 1.61% during the live shoot scenario. Accordingly, lethal force errors were significantly higher during the video no-shoot scenarios compared to the live no-shoot scenario ($z = -4.914$, $p < 0.001$).

3.2 Autonomic Arousal

Relative to average resting HR ($M = 78.5$, $SD = 12.66$, $n = 178$), both video ($M = 114.6$, $SD = 17.67$, $n = 180$, $z = -11.299$, $p < 0.001$), and live simulations ($M = 126.8$, $SD = 20.80$, $n = 182$, $z = -11.476$, $p < 0.001$) elicited significant autonomic arousal. HRMax during the live simulations was significantly higher than during the video simulations ($z = -9.270$, $p < 0.001$, $n = 180$, Figure 3).

4 Discussion

The current observational study revealed significant autonomic arousal, as measured by heart rate, in both video and live scenarios, indicating levels of stress that are consistent with prior literature. Despite greater increases in autonomic arousal during live scenarios, officers committed fewer lethal UOF errors compared to video simulations during their agency's annual requalification evaluation (Figure 1). Although errors were relatively low overall, it is important to note that any errors made during an evaluation may have direct professional consequences (e.g., loss of authorities) for the officers. Given the lack of studies in this area, we present these findings to engage in an empirical conversation about how scenario design and delivery may optimize performance while meeting evaluation objectives.
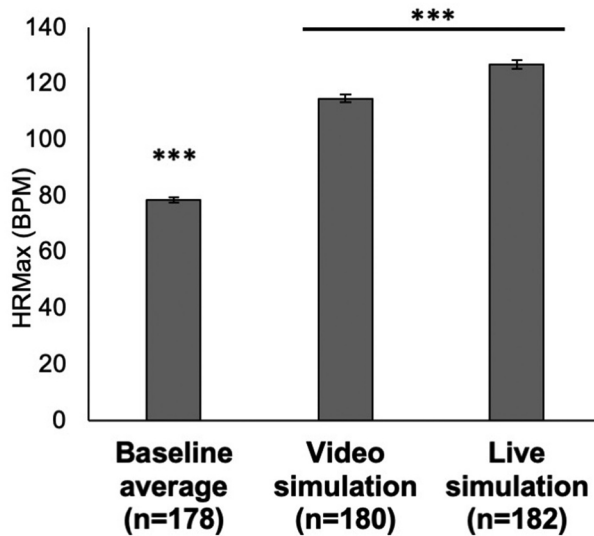
Figure 3. Autonomic arousal during video and live simulations. Both video and live simulation conditions elicited significant autonomic arousal relative to rest, with live simulations eliciting higher HRMax compared to video simulations. Error bars show SEM. Note: *** $p < 0.001$

4.1 Designing effective police evaluations

As recently highlighted elsewhere, the most important consideration when designing police UOF evaluation (and training) exercises is to ensure that the methodological approach (i.e., virtual or live) fulfills intended goals (DiNota, Andersen, et al., 2021; Jenkins et al., 2020; Kleygrewe et al., 2023; Koedijk et al., 2021). Especially for evaluations that bear professional consequences (i.e., qualification or requalification), the chosen modality should elicit specified and pre-determined skills and approximate the operational conditions under which the skills will be used (Koerner & Staller, 2020). Consider an exercise where an officer encounters an armed person and must distance themselves and change positions for better protection. If the officer failed to demonstrate appropriate tactics due to increased mental load, poor sense of presence, and/or perceptual/ technical limitations of a virtual system (discussed below), they could face significant negative evaluative and professional implications. Increased mental load also limits available resources for other psychological processes like critical decision-making and may increase officers' reliance on more severe force options (O'Neill et al., 2019), possibly accounting for the observed increase in video UOF errors (Figure 1). However, the relationships between mental load (increased in virtual simulations), autonomic arousal (increased in live simulations), and performance errors (increased in virtual simulations) in critical police contexts remain unclear. The specific UOF-related skills that are better generalized to the field when evaluated (or trained) using live or virtual/video simulation approaches remain an additional untested, but important, topic for future applied research.

By using a variety of physical settings and actors (i.e., suspects, bystanders), live and virtual simulations evaluate officers' 'non-visible' psychological competencies including threat/risk assessment, situational awareness, and stress modulation (Andersen & Gustafsberg, 2016; Andersen et al., 2018; Michela et al., 2022). The duration of both simulation types can be manipulated to elicit fast decision-making in unpredictable encounters or can be 'paused' to reflect on officers' internal cognitive skills and thought processes (DiNota & Huhta, 2019). Favourable performance outcomes build officers' confidence in meeting challenging situational demands, as UOF (especially lethal force) encounters are typically rare in an officer's career (Baldwin et al., 2018; Jenkins et al., 2020). Accordingly, errors that are not attributed to the officer's (lack of) UOF skills but to technical

aspects of the simulation approach (as exemplified above) should be minimized to avoid 'training scars' that may degrade officer's confidence and self-efficacy (DiNota, Andersen, et al., 2021). While simulations build officers' repertoire of successfully navigated incidents, overrepresentation of lethal encounters may condition increased weapon use (Giessing, 2021), underscoring the need to evaluate a broad variety of UOF-related skills including situational awareness/assessment and social-verbal skills.

In line with extant research, the current findings suggest that live simulations are more effective for evaluating UOF competencies that involve: 1) relatively more physical activity (e.g., chasing, apprehension, arrest, self-defense) utilizing field-compatible UOF equipment (modified for safety) in order to elicit and encode correct motor strategies in representative (i.e., stressful, 3-dimensional) environments; and 2) non-, para-, and verbal communication skills (i.e., de-escalation, professionalism), including interactions with members of vulnerable populations (DiNota, Andersen, et al., 2021; Huey et al., 2021). Live simulation training with individuals in mental crisis improves proficiency in critical decision-making, confidence in communication, and reduces social stigma among police (Krameddine et al., 2013) and social workers (Regehr & Birze, 2020).

Virtual simulations are optimal for smaller 'drills' of discrete, component skills that need to be performed in mentally and physiologically demanding contexts (e.g., transitioning between UOF options, stress management) (DiNota, Andersen, et al., 2021; Michela et al., 2022). Where virtual systems allow, measures of shot or movement accuracy can also effectively evaluate marksmanship skills. Virtual simulations also control for variability in individual scenario presentation and maximize resource efficiency (i.e., instructors' energy, time) and investment potential (Bennell et al., 2007; Giessing, 2021; Zechner et al., 2023). Especially for larger agencies, saving 1–2 mins per officer and including an additional video scenario as observed in the current study can equate to hours saved in annual evaluation and training exercises. While virtual simulations may be useful for supplementing or complimenting live approaches, they should not *replace* live simulations without first being empirically validated for: a. eliciting desired evaluation/learning outcomes, and b. minimizing potential unintended consequences (i.e., reduced sense of presence, increased UOF errors relative to other approaches) (Kleygrewe et al., 2023; Murtinger et al., 2021).

4.2 Perceptual limitations and considerations on downstream cognition and decision-making

Many virtual simulation configurations used by police agencies (including the one used in the current study) are relatively 'basic' compared to advanced VR and AR systems and may influence performance outcomes by presenting physical and perceptual limitations that do not exist in live simulations or real-world interactions. For instance, if evaluating officers' tactical ability to enter and clear a space with corners, withdraw, or take initiative, a virtual simulation cannot provide the realistic visual and spatial cues necessary to evaluate these skills and behavioural dimensions (Huhta et al., 2021). Depending on the size of the video screen and how far officers stand before it, the visual angle of projected images may also be unrealistic (e.g., person standing close by will appear disproportionately large). A critical physical limitation of virtual approaches involves the use of incompatible simulation equipment (Düking et al., 2018). Especially if the evaluation objective is to assess duty weapon skills (e.g., marksmanship, accessing during an encounter) simulated equipment must be equivalent (Zechner et al., 2023).

The human brain is finely tuned to detect asynchrony between senses, like when a film's audio and video tracks are off by milliseconds (Spence & Squire, 2003). When visual cues like direction and speed of the projected image are incongruous or not linked to the officer's movement in the simulation environment, the scenario may feel 'unnatural' or even induce negative physical effects including cybersickness (Giessing, 2021; Kleygrewe et al., 2023). Neuroscientific research also shows that subtle facial muscle expressions, movements, and speech patterns are processed better

during live versus virtual interactions (Redcay et al., 2010; Tomasello et al., 2005). Especially when digital avatars are used instead of recorded human actors, compromised visual perception can lead to downstream errors in officers' social evaluation, judgement, communication, and decision-making skills. Despite efforts at designing highly immersive and engaging environments, artificial sensory cues presented by virtual simulations can limit an officer's sense of presence, undermining scenario realism and evaluation/training objectives (Michela et al., 2019; Zechner et al., 2023), and possibly account for the higher rate of lethal force errors in virtual no-shoot conditions compared to live scenarios (Table 1 , Figure 2).

Limitations of the current work include an imbalance and lack of reproducibility in the number, duration, and content of video and live scenarios. An optimal experimental study design would present identical scenarios to matched officers in both modalities. This would control for scenario complexity and better identify what it is about the nature of virtual or live simulations that may contribute to more errors in one condition over the other, as observed in the current study. However, real-world policing does not operate under or facilitate such stringent experimental conditions. As exemplified by Kleygrewe et al. (2023), applied research in police UOF decision-making must take a more pragmatic approach. The current study limitations reflect the numerous (and often conflicting) realities of contemporary police practice, including available resources (i.e., time, available trainers, infrastructure, VR equipment) and professional requirements (i.e., individual agency evaluation/training standards). Future research should also investigate the implications of combined UOF evaluation/training paradigms on performance outcomes, which are commonly employed by police agencies to conserve resources despite serving different goals (i.e., assessment versus learning). The low frequency of performance errors precludes a determination of whether mistakes in earlier scenarios influence the likelihood of subsequent errors, which remains an important, albeit difficult, question for future research.

The definitive purpose of police UOF evaluation (and training) is to ensure the safety of officers and the public. Given the dire and widespread consequences of police UOF errors, it is essential that the behavioural outcomes are elicited, and neurophysiological mechanisms activated by virtual and live simulation environments are considered. Based on the current findings and relevant multi-disciplinary literature reviewed above, we caution against implementing possibly incompatible simulation approaches that undermine the purpose of UOF evaluation.

Notes on contributors

Paula M. Di Nota, Ph.D., is a postdoctoral research fellow at the University of Toronto Mississauga in the Department of Psychology. Her research investigates the relationships between stress physiology, training, performance, and health in police and other public safety personnel.

Chief Inspector Juha-Matti Huhta, M.Ed., is a use of force instructor at the Police University College of Finland and a doctoral candidate in the Faculty of Education and Culture at Tampere University. He has been a police officer for over 20 years and has experience in special units including K-9 and Regional Special Response Teams. His primary research interest focuses on understanding and developing situational awareness in policing for the purpose of improving training and education.

Evelyn C. Boychuk, M.Sc., was a research assistant in the Health Adaptation Research on Trauma (HART) Lab at the University of Toronto Mississauga at the time of this study.

Judith P. Andersen is an Associate Professor at the Department of Psychology and Affiliated Faculty of Medicine at the University of Toronto. She is a health psychologist who studies the health and performance outcomes associated with stress and trauma. Dr. Andersen's recent work focuses on developing interventions to improve the health and performance among public safety personnel

ORCID

Paula M. Di Nota 🔟 http://orcid.org/0000-0003-4282-6129
Evelyn C. Boychuk 🔟 http://orcid.org/0000-0002-1234-9862
Judith P. Andersen 🔟 http://orcid.org/0000-0001-8477-1432

References

Andersen, J. P., DiNota, P. M., Beston, B., Boychuk, E. C., Gustafsberg, H., Poplawski, S., Arpaia J. (2018). Reducing lethal force errors by modulating police physiology. *Journal of Occupational and Environmental Medicine*, *60*(10), 867–874. https://doi.org/10.1097/JOM.0000000000001401

Andersen, J. P., DiNota, P. M., Boychuk, E. C., Schimmack, U., & Collins, P. I. (2021). Racial bias and lethal force errors among Canadian police officers. *Canadian Journal of Behavioural Science/Revue Canadienne des Sciences du Comportement*. Advance online publication. https://doi.org/10.1037/cbs0000296

Andersen, J. P., & Gustafsberg, H. (2016). A training method to improve police use of force decision making: A randomized controlled trial. *SAGE Open*, *6*(2), 1–13. https://doi.org/10.1177/2158244016638708

Anderson, G. S., Litzenberger, R., & Plecas, D. (2002). Physical evidence of police officer stress. *Policing: An International Journal of Police Strategies & Management*, *25*(2), 399–420. https://doi.org/10.1108/13639510210429437

Baldwin, S., Bennell, C., Andersen, J. P., Semple, T., & Jenkins, B. (2019). Stress-activity mapping: Physiological responses during general duty police encounters. *Frontiers in Psychology*, *10*, 2216. https://doi.org/10.3389/fpsyg.2019.02216

Baldwin, S., Bennell, C., Blaskovits, B., Brown, A., Jenkins, B., Lawrence, C., McGale, H., Semple, T., & Andersen, J. P. (2022). A reasonable officer: Examining the relationships among stress, training, and performance in a highly realistic lethal force scenario. *Frontiers in Psychology*, *5629*. https://doi.org/10.3389/fpsyg.2021.759132

Baldwin, S., Hall, C., Blaskovits, B., Bennell, C., Lawrence, C., & Semple, T. (2018). Excited delirium syndrome (ExDS): Situational factors and risks to officer safety in non-fatal use of force encounters. *International Journal of Law and Psychiatry*, *60*, 26–34. https://doi.org/10.1016/j.ijlp.2018.06.011

Bennell, C., & Jones, N. J. (2004). The effectiveness of use of force simulation training: Final report. (Public Safety Canada). https://www.researchgate.net/publication/232568382_Does_use-of-force_simulation_training_in_Canadian_police_agencies_incorporate_principles_of_effective_training [Retrieved June 22, 2021].

Bennell, C., Jones, N. J., & Corey, S. (2007). Does use-of-force simulation training in Canadian police agencies incorporate principles of effective training? *Psychology, Public Policy, & Law*, *13*(1), 35. https://doi.org/10.1037/1076-8971.13.1.35

Billman, G. E. (2013). The LF/HF ratio does not accurately measure cardiac sympatho-vagal balance. *Frontiers in Physiology*, *4*(26). https://doi.org/10.3389/fphys.2013.00026

CCJ. (2021). *Effectiveness of Police Training: Policy Assessment*. Council on Criminal Justice Task Force on Policing). Retrieved June 1, 2021 https://counciloncj.foleon.com/policing/assessing-the-evidence/policy-assessments/

Chan, J. F., DiNota, P. M., Planche, K., Borthakur, D., & Andersen, J. P. (2022). Associations between police lethal force errors, measures of diurnal and reactive cortisol, and mental health. *Psychoneuroendocrinology*, *142*(2022). https://doi.org/10.1016/j.psyneuen.2022.105789

DiNota, P. M., Andersen, J. P., Huhta, J.-M., & Gustafsberg, H. (2021). Evidence-based instruction of police use of force: Practical methods and pedagogical principles. In E. Arble & B. Arnetz (Eds.), *Interventions, Training, and Technologies for Improved Police Well-Being and Performance* (pp. 72–101). (IGI Global). 10.4018/978-1-7998-6820-0.ch005

DiNota, P. M., Arpaia, J., Boychuk, E. C., Collins, P. I., & Andersen, J. P. (2021). Testing the efficacy of a 1-day police decision-making and autonomic modulation intervention: A quasi-random pragmatic controlled trial. *Frontiers in Psychology*, *12*, 719046. https://doi.org/10.3389/fpsyg.2021.719046

DiNota, P. M., Chan, J., Huhta, J.-M., & Andersen, J. P. (2021). Objective and subjective evaluation of police use of force behaviour. *International Journal of Environmental Research and Public Health*, *18*(5351), 1–12. https://doi.org/10.3390/ijerph18105351

DiNota, P. M., & Huhta, J. M. (2019). Complex motor learning and police training: Applied, cognitive, and clinical Pperspectives. *Frontiers in Psychology*, *10*, 1797. https://doi.org/10.3389/fpsyg.2019.01797

Duking, P., Holmberg, C.-H., & Sperlich, B. (2018). The potential usefulness of virtual reality systems for athletes: A short SWOT analysis. *Frontiers in Physiology*, *9*, 128. https://doi.org/10.3389/fphys.2018.00128

Giessing, L. (2021). The Potential of Virtual Reality for Police Training Under Stress: A SWOT Analysis. In E. Arble & B. Arnetz (Eds.), *Interventions, Training, and Technologies for Improved Police Well-Being and Performance* (pp. 102–124). (IGI Global). 10.4018/978-1-7998-6820-0.ch006

Hayano, J., & Yuda, E. (2019). Pitfalls of assessment of autonomic function by heart rate variability. *Journal of Physiological Anthropology*, *38*(1), 1–8. https://doi.org/10.1186/s40101-019-0193-2

Huey, L., Andersen, J., Bennell, C., Campbell, M. A., Koziarski, J., & Vaughan, A. D. (2021). *Caught in the currents: Evaluating the evidence for common downstream police response interventions in calls involving persons with mental illness*. Royal Society of Canada). Retrieved July 15, 2021 https://rsc-src.ca/sites/default/files/MH%26P%20PB_EN_5.pdf

Huhta, J.-M., DiNota, P. M., Nyman, M., Ropo, E., & Pietilä, E. (2021). Universal police behaviours during critical incidents and their connection to personality: A preliminary study. *The Nordic Journal of Studies in Policing*, *2*, 1–18. https://doi.org/10.18261/issn.2703-7045-2021-01-04

Jenkins, B., Semple, T., & Bennell, C. (2020). An evidence-based approach to critical incident scenario development. *Policing: An International Journal*, *44*(3), 437–454. https://doi.org/10.1108/PIJPSM-02-2020-0017

Kleygrewe, L., Hutter, R. V., Koedijk, M., & Oudejans, R. R. (2023). Virtual reality training for police officers: A comparison of training responses in VR and real-life training. *Police Practice & Research*, 1–20. https://doi.org/10.1080/15614263.2023.2176307

Kleygrewe, L., Oudejans, R. R., Koedijk, M., & Hutter, R. I. (2022). Police Training in Practice: Organization and Delivery According to European Law Enforcement Agencies. *Frontiers in Psychology*, *12*, 798067. https://doi.org/10.3389/fpsyg.2021.798067

Koedijk, M., Renden, P. G., Oudejans, R. R., Kleygrewe, L., & Hutter, R. I. (2021). Observational behavior assessment for psychological competencies in police officers: A proposed methodology for instrument development. *Frontiers in Psychology*, *12*, 1–14. https://doi.org/10.3389/fpsyg.2021.589258

Koerner, S., & Staller, M. Police Training Revisited—Meeting the Demands of Conflict Training in Police with an Alternative Pedagogical Approach. (2020). *Policing: A Journal of Policy and Practice*, *15*(2), 927–938. paaa080. https://doi.org/10.1093/police/paaa080

Krameddine, Y. I., DeMarco, D., Hassel, R., & Silverstone, P. H. (2013). A novel training program for police officers that improves interactions with mentally ill individuals and is cost-effective. *Frontiers in Psychiatry*, *4*(9), 1–10. https://doi.org/10.3389/fpsyt.2013.00009

Michela, A., van Peer, J. M., Brammer, J. C., Nies, A., van Rooij, M. M., Oostenveld, R., Dorrestijn W, Smit AS, Roelofs K, Klumpers F, Granic, I. (2022). Deep-breathing biofeedback trainability in a virtual-reality action game: A single-case design study with police trainers. *Frontiers in Psychology*, *13*, 29. https://doi.org/10.3389/fpsyg.2022.806163

Michela, A., van Rooij, M. M., Klumpers, F., van Peer, J. M., Roelofs, K., & Granic, I. (2019). Reducing the noise of reality. *Psychological Inquiry*, *30*(4), 203–210. https://doi.org/10.1080/1047840X.2019.1693872

Murtinger, M., Jaspaert, E., Schrom-Feiertag, H., & Egger-Lampl, S. (2021). Cbrne training in virtual environments: SWOT analysis & practical guidelines. *International Journal of Safety and Security Engineering*, *11*(4), 295–303. https://doi.org/10.18280/ijsse.110402

Nieuwenhuys, A., Savelsbergh, G. J., & Oudejans, R. R. (2012). Shoot or don't shoot? Why police officers are more inclined to shoot when they are anxious. *Emotion*, *12*(4), 827–833. https://doi.org/10.1037/a0025699

Nieuwenhuys, A., Savelsbergh, G. J., & Oudejans, R. R. D. (2015). Persistence of threat-induced errors in police officers' shooting decisions. *Applied Ergonomics*, *48*, 263–272. https://doi.org/10.1016/j.apergo.2014.12.006

O'Neill, J., O'Neill, D. A., Weed, K., Hartman, M. E., Spence, W., & Lewinski, W. J. (2019). Police Academy Training, Performance, and Learning. *Behavior Analysis in Practice*, *12*(2), 353–372. https://doi.org/10.1007/s40617-018-00317-2

Redcay, E., Dodell-Feder, D., Pearrow, M. J., Mavros, P. L., Kleiner, M., Gabrieli, J. D. E., & Saxe, R. (2010). Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. *Neuroimage*, *50*(4), 1639–1647. https://doi.org/10.1016/j.neuroimage.2010.01.052

Regehr, C., & Birze, A. (2020). Use of Simulation Methods in Social Work Research on Clinical Decision-Making. *Clinical Social Work Journal*, *49*(2), 1–12. https://doi.org/10.1007/s10615-020-00778-5

Spence, C., & Squire, S. (2003). Multisensory integration: Maintaining the perception of Synchrony. *Current Biology*, *13*(13), R519–R521. https://doi.org/10.1016/S0960-9822(03)00445-7

Thayer, J. F., & Sternberg, E. (2006). Beyond heart rate variability: Vagal regulation of allostatic systems. *Annals of the New York Academy of Sciences*, *1088*(1), 361–372. https://doi.org/10.1196/annals.1366.014

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, *28*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Zechner, O., Kleygrewe, L., Jaspaert, E., Schrom-Feiertag, H., Hutter, R. I. V., & Tscheligi, M. (2023). Enhancing Operational Police Training in High Stress Situations with Virtual Reality: Experiences, Tools and Guidelines. *Multimodal Technologies and Interaction*, *7*(2), 14. https://doi.org/10.3390/mti7020014