

Bystander Intervention in Football and Sports. A Quasi-Experimental Feasibility Study of a Bystander Violence Prevention Program in the United Kingdom

Journal of Interpersonal Violence

1–28

© The Author(s) 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/08862605241239452

journals.sagepub.com/home/jiv



Anastasiia G. Kovalenko¹ 
and Rachel A. Fenton¹

Abstract

In recent years, social campaigns and high-profile cases have brought increased attention to violence against women. Athletes can be role models, shaping both prosocial and antisocial attitudes. Their engagement in violence prevention could be an effective tool to tackle violence against women through bystander intervention. This part of a mixed-method feasibility study reports on the quantitative evaluation of an evidence-led bystander program, Football Onside, implemented at a football club in South West England in June 2018 to February 2020. The study employed a quasi-experimental design with intervention and control groups. Football coaches and club members ($n = 50$) completed measures of rape and domestic abuse myths, bystander intent and efficacy, self-reported bystander behaviors, readiness for change, perceptions of peer helping and myth acceptance, law knowledge, and program evaluation. Fidelity was also assessed. The analysis

¹University of Exeter, Devon, UK

Corresponding Author:

Anastasiia G. Kovalenko, University of Exeter Medical School, St. Luke's Campus, Heavitree Road, Exeter, Devon EX1 2LU, UK.

Email: a.g.kovalenko@exeter.ac.uk

compared between- and within-group differences in mean changes over time using mixed-effects models. Participant ratings of learning outcomes were high, and fidelity was maintained throughout the intervention. Between-group comparison revealed mixed results, with greater improvements in the intervention group for bystander intent and efficacy at post-test and follow-up, domestic abuse myths at post-test, and rape myth acceptance at follow-up. Model contrasts for within intervention group revealed improvements in rape and domestic abuse myth acceptance, bystander intent and efficacy, perceived law knowledge at both time points, and perceived peer myths and helping at post-test. At follow-up, intervention participants reported significantly higher engagement in bystander behaviors. No significant effects were found for perceived importance of legal knowledge. Our research highlights the potential efficacy of a bystander program tailored for football club members. Cluster-randomized control trials are now required to examine bystander attitudes and behavior change processes among professional athletes.

Keywords

domestic violence, sexual assault, prevention, intervention, sexual harassment

The United Kingdom has entered a pivotal moment in history in relation to combatting violence against women and girls (VAWG) in the wake of one particular high-profile abduction and murder in March 2021, which has invigorated the national conversation and consciousness. Understood as a cause and consequence of gender inequality, VAWG is a public health epidemic and a global human rights issue (UN, 1993; WHO, 2010). In the United Kingdom, for example, one in four women experience domestic abuse, and one in five experience any kind of sexual assault in their lifetime (Home Office, 2019). Since the beginning of the COVID-19 pandemic, VAWG, especially domestic abuse, has intensified globally (UN Women, 2021) and has been referred to as “the shadow pandemic.” For example, in April to June 2020, roughly one-fifth (21%, 20%, and 19%) of all offences recorded by the police in the United Kingdom were flagged as domestic abuse, which is a 5%-point increase compared with the same period in the previous year (ONS, 2020). In the year ending March 2022, the police recorded the highest number of sexual offences in England and Wales, with 86% sexual assault victims being female (ONS, 2023a), and it is estimated that only one in five cases is

reported to the police (ONS, 2020). Similarly, in the year ending March 2023, 73.5% of police-recorded domestic abuse victims in England and Wales were female (ONS, 2023b). The elimination of VAWG is a strategic priority both nationally (VAWG Strategy—Home Office) and in terms of the United Kingdom’s international legal obligations (UN, Convention on the Elimination of all Forms of Discrimination Against Women (CEDAW), Council of Europe). Yet while some positive legislative initiatives are being delivered, such as the Domestic Abuse Act 2021, the criminal justice system reveals an ever-increasing inability to provide redress with charging, prosecution, and convictions for sexual offences at a 10-year low (CPS, 2019). The focus on prevention and, in particular, how to engage men in prevention has thus never been more relevant in UK society than in this potentially transformative moment.

One prevention strategy gaining traction in the United Kingdom is bystander intervention. Bystanders can play an important role in the prevention of VAWG as agents of cultural change who intervene to disrupt violence and the social norms that facilitate it (Banyard et al., 2004; Orchowski & Berkowitz, 2022). The historical cultural prominence of sport—and, in the United Kingdom, of football (soccer) in particular—and its vindication of traditional models of masculinity linked to violence (Adams, 2011) render it an “unparalleled platform” (Katz, 2018) for VAWG prevention. There is a paucity of knowledge about bystander approaches and sports settings, with no studies to date evaluating the bystander approach in a professional sport setting in the United Kingdom and only very limited evidence elsewhere. With this gap in mind, this study seeks to add to the literature by presenting findings from a controlled quasi-experimental feasibility study of a bystander program, “Football Onside,” with follow-up at 9 months. “Football Onside” is a bystander intervention for the prevention of sexual and domestic abuse, tailored to a male-dominated professional football setting in the United Kingdom for the first time.

Sports, Masculinity, and VAWG

There is a complex, intersecting, and longstanding association between sport, misogynistic or hegemonic masculinity, and the reproduction of violence-supportive norms and violence perpetration (Dyson & Flood, 2008; Flood, 2011). A body of literature has found that male college athletes have a higher affinity for rape myth acceptance (Boeringer, 1999; Bogen et al., 2020) and are an at-risk group for sexual violence (SV) perpetration (Crosset et al., 1996; McCray, 2015; Young et al., 2017) and victimization (Cheever & Eisenberg, 2022).

Male professional sports settings have been the sites for high-profile incidents (see Katz, 2018), and some surveys have shown an increasing number of reports of rape and domestic abuse cases among professional athletes in contact sports (NBC, 2014; O’Hear, 2001). In the United Kingdom, trials of professional sportsmen for rape have fueled high-profile public debate and controversy within the legal community (McGlynn, 2018). In addition to prominent incidents, very limited research has indicated that domestic abuse rates may increase after high stakes sporting events such as the World Cup and Euro Cup (Kirby et al., 2013; Trendl et al., 2021) and national rival team football matches (Williams et al., 2013). This has, however, been contentious, in particular, due to the heterogeneity of studies, the challenges in measuring, recording, and reporting national domestic violence (DV) data, the lack of evidence outside the United Kingdom and North America, and the limited assessment of the contributing risk factors (Forsdike et al., 2022).

Educators are agreed about the importance of engaging men in prevention efforts (Flood, 2011). As sports define the cultural mainstream with team sports allowing men “to do [] gender in the most culturally esteemed way” (Adams, 2011, p. 580), sports settings simultaneously occupy a critical space and present an opportunity for prevention work. Athletes can be positive societal role models (O’Donohue & Schewe, 2019), promoting prosocial behaviors and healthy ways to express masculinity and disapproval of the social norms, which shore up VAWG (Carlson et al., 2015; Katz et al., 2011).

Coaches can be influential positive nonparental role models (Fraser-Thomas et al., 2008). Further, sports teams may be encoded with strong social bonds, which can be mobilized to promote positive intervention (McMahon & Farmer, 2009). Indeed, social expectations around masculinity are related to both perpetration and men’s willingness and likelihood to intervene as bystanders (Brown & Messman-Moore, 2010; Carlson, 2008; Fabiano et al., 2003; Schwartz & DeKeseredy, 2008). Consequently, there is a well-established history of working with athletes and coaches as “exemplars of traditional masculine success” (Katz, 2018) to prevent VAWG (e.g., Mentors in Violence Prevention (MVP)—Katz, 1995, 2018; Coaching Boys into Men (CBIM)—Miller et al., 2012; Wingman 101—Exner-Cortens & Cummings, 2017).

Bystander Interventions

Sports settings align well with bystander intervention because the bystander approach focuses on prosocial framing and engaging men as allies (Flood, 2011). The increasingly promising evidence base is situated predominantly in school and college settings in the United States (Kettrey & Marx, 2019; Kovalenko et al., 2022; Mujal et al., 2021). In the United Kingdom, the

evidence base is in its infancy, but there is some preliminary evidence in school and university settings (Bovill & White, 2022; Fenton & Mott, 2018; Williams & Neville, 2017) and, more recently, in a general population setting (Gainsbury et al., 2020). Overall, the literature reports improvements across a range of associated measures (for reviews see Kettrey & Marx, 2019; Kovalenko et al., 2022; Mujal et al., 2021). However, there are limitations: only a small number of evaluations include a comparison group, backlash is rarely measured, and the majority have a follow-up period shorter than 7 months, which might explain why few report higher levels of self-reported bystander behaviors (Kovalenko et al., 2022; Mujal et al., 2021). Bystander programs targeting athletes have been predominantly implemented in student populations and report improved bystander intent and efficacy, rape myth acceptance, and perpetration rates (Garrity, 2011; Jaime et al., 2018; Miller et al., 2012; Moynihan et al., 2011b). The findings for behavior change have been mixed, showing small improvements or no significant change (Exner-Cortens & Cummings, 2017; Garrity, 2011). Research evidence on bystander intervention in professional sports, however, is extremely limited (Corboz et al., 2016; Powell, 2011) and lacks rigorous evaluation.

Bystander interventions often intend to increase the knowledge of substantive law, but this is undertheorized in terms of its contribution to change and is rarely reported on (Fenton et al., 2016). Backlash, understood as movement by more than one standard deviation (*SD*) in the undesired direction (Moynihan et al., 2011a), is also rarely reported but is important, as interventions may produce unintended effects.

Community Trusts

Capitalizing on the normative status of football in society and the extent of outreach by football clubs in the United Kingdom (via Community Trusts¹), the CEO of a Community Trust was approached as a potential partner in the development, implementation and evaluation of the “Football Onside” program.

Researchers may encounter institutional and cultural resistance implementing violence prevention programs (Brackenridge, 2002; Parent & Fortier, 2018). Nevertheless, overcoming these barriers is possible with strong support from team leaders, their active engagement in the development and implementation of initiatives, and collaborations with external organizations and communities (Parent & Fortier, 2018). It is therefore not unexpected that the enthusiasm, motivation, and commitment of the CEO were instrumental in the realization of the project, and steering group comprising victim organizations, public health, local council, and other professionals in the area was actively engaged throughout the project. The

intervention and the study was designed as a response to both need and opportunity for a bystander VAWG prevention program specifically targeting professional players, coaches and staff outreaching into the wider community through football.

Football Onside

Football Onside is a feminist, gender-transformative, and social-justice-informed bystander leadership intervention. The program is theoretically and pedagogically underpinned by previous research, documented elsewhere (Fenton & Mott, 2017; 2018; Fenton et al., forthcoming). The theoretical framework is Latané and Darley's (1970) situational model, which explains the bystander decision-making process from noticing an event and identifying it as problematic, to taking responsibility, feeling confident, and possessing the necessary skills to act, through to subsequent intervention. Prochaska and DiClemente's (1986) model of behavior change was utilized (Fenton & Mott, 2018), and Nation et al.'s (2003) principles for effective prevention programs, for both content and delivery including varied teaching methods, dosage, being theory-driven and socioculturally relevant to participants, and delivered by well-trained staff, were followed. In order to ensure that the content was of maximal salience to participants (Nation et al., 2003), a focus group with the Community Trust staff was held to inform content development. Content included specific requests for information on law, including materials on grooming and child sex offences, which was implemented both in the content and in an additional take-home booklet. Further perusal of the baseline data informed program development, for example, concentrating on the rape myths participants most believed in and providing feedback on participants own norm misperceptions. Content was designed to explore performative masculinities and sport, prevalence, impact and myths surrounding SV and DV, consent, and bystander skills-building (Table S1, Supplemental Appendix).

Trainers were recruited from, or recommended by, local victims' organizations. Facilitator training was conducted face to face with the two academics who designed the program, and all materials including detailed facilitator notes were provided in advance. The training consisted of going through the slides and notes over the course of a half day. It is worth noting that there was some resistance to the idea that trainers for victims' organizations could need any further training.

This study sought to explore the effects of Football Onside in a sports setting outside a student population for the first time in the United Kingdom. To address some of the limitations in the evaluation literature, we adopted a nonrandomized design with intervention and control group, with measures at

Table 1. Mixed-Effects Model Results.

Measure	Intervention group Trust 1		Control group Trust 2		Between-group difference ^a		Within-group difference ^a		Effect size ^c
	M	[95% CI]	M	[95% CI]	Time × Group I-C	[95% CI]	T1-T0, T2-T0 [95% CI] ^b	d(g)	
IRMA									
Baseline	1.87	[1.72, 2.05]	1.83	[1.6, 2.05]					
1 month	1.57	[1.41, 1.72]	1.65	[1.42, 1.88]					
9 months	1.49	[1.32, 1.66]	1.75	[1.52, 1.98]	-0.12	[-0.37, 0.12]	-0.3	[-0.44, -0.17]**	
DYMAS									
Baseline	2.74	[2.51, 2.97]	2.57	[2.24, 2.9]					
1 month	2.39	[2.16, 2.61]	2.71	[2.38, 3.06]					
9 months	2.48	[2.23, 2.74]	2.47	[2.13, 2.82]	-0.3	[-0.57, -0.04]*	-0.38	[-0.54, -0.22]**	-0.54(-0.53)
Readiness for change									
Baseline	3.56	[3.38, 3.75]	3.49	[3.23, 3.76]					
1 month	3.61	[3.42, 3.8]	3.29	[3.01, 3.56]					
9 months	3.57	[3.35, 3.79]	3.21	[2.92, 3.5]	0.26	[-0.13, 0.64]	0.05	[-0.17, 0.26]	
Bystander intent									
Baseline	3.7	[3.53, 3.87]	3.58	[3.33, 3.83]					
1 month	3.9	[3.73, 4.07]	3.4	[3.15, 3.65]					
9 months	3.91	[3.72, 4.1]	3.47	[3.21, 3.73]	0.37	[0.09, 0.65]*	0.2	[0.04, 0.35]*	1(0.99)
Bystander efficacy									
Baseline	80.8	[75.67, 85.93]	81.75	[74.22, 89.28]					
1 month	87.07	[81.94, 92.2]	73.73	[66.2, 81.25]	14.29	[6.37, 22.22]**	6.27	[1.85, 10.69]*	0.9(0.89)
9 months	87.36	[81.7, 93.01]	73.93	[66.27, 81.59]	14.38	[5.98, 22.78]**	6.56	[1.55, 11.56]*	0.98(0.97)
Bystander behavior									
Baseline	-2.52	[-4.3, -0.62]	-3.25	[-5.96, -0.54]					
1 month	-0.21	[-2.06, 1.65]	-3.67	[-6.47, -0.88]	2.69	[-1.86, 7.23]	2.26	[-0.27, 4.8]	
9 months	0.69	[-1.56, 2.94]	-2.76	[-05.64, 0.13]	2.7	[-02.12, 7.46]	3.16	[0.31, 6.02]*	

(continued)

Table 1. (continued)

Measure	Intervention group Trust 1		Control group Trust 2		Between-group difference ^a		Within-group difference ^a		Effect size ^c d(g)
	M [95% CI]		M [95% CI]		Time × Group I-C [95% CI]	T1-T0, T2-T0 [95% CI] ^b			
Perceptions of peer helping									
Baseline	3.71 [3.52, 3.91]		3.84 [3.55, 4.12]						
1 month	4.01 [3.82, 4.21]		3.74 [3.42, 4.05]		0.4 [-0.04, 0.84]		0.3 [0.06, 0.54]*		
9 months	3.99 [3.75, 4.23]		3.87 [3.57, 4.18]		0.24 [-0.22, 0.7]		0.28 [-0.001, 0.56]		
Perceptions of peer myths									
Baseline	29.76 [24.79, 34.73]		32.5 [25.08, 39.94]						
1 month	22.05 [17.08, 27.02]		27.89 [20.22, 35.56]		-3.09 [-14.56, 8.38]		-7.71 [-13.93, -1.49]*		
9 months	28.98 [22.9, 35.06]		30.52 [22.93, 35.03]		1.22 [-10.74, 13.17]		-0.78 [-7.95, 6.39]		
The law									
Knowledge of law DV									
Baseline	2.68 [2.4, 2.95]		2.25 [1.85, 2.65]						
1 month	3.43 [3.15, 3.7]		2.62 [2.22, 3.03]		0.37 [-0.17, 0.92]		0.75 [0.44, 1.06]***		
9 months	3.16 [2.84, 3.48]		2.45 [2.04, 2.87]		0.28 [-0.3, 0.86]		0.48 [0.14, 0.83]*		
Knowledge of law SV									
Baseline	2.79 [2.5, 3.09]		2.44 [2.002, 2.87]						
1 month	3.59 [3.29, 3.89]		2.85 [2.4, 3.3]		0.38 [-0.23, 1]		0.79 [0.45, 1.14]***		
9 months	3.39 [3.04, 3.74]		2.79 [2.33, 3.25]		0.24 [-0.41, 0.89]		0.6 [0.21, 0.99]*		

Note. CI = confidence interval; IRMA = Illinois Rape Myth Acceptance Short Form; DVMAS = Domestic Violence Myth Acceptance.

^aEstimates based on linear mixed models.

^bT0 = baseline, T1 = post-test, T2 = follow-up.

^cStandardized calculation based on model-adjusted means in both groups.

* $p < .05$. ** $p < .01$. *** $p < .001$.

baseline, post-test, and 9-month follow-up. Our primary research aims were to evaluate the feasibility of the Football Onside program and to examine the effect of the program between the intervention and control group. We aimed to explore and report the effects within the intervention group in order to map onto extant evaluations, which did not have a comparison. Our secondary research aim was to examine backlash arising from the intervention.

Method

Participants

This study was conducted with two professional Football Club Community Trusts (FCCTs) in South West England from June 2018 to February 2020. Trust 1 received the intervention and Trust 2 acted as control. Of 60 participants initially recruited, nine dropped out after completing the pre-intervention questionnaires due to reasons unrelated to the nature of the study, and one participant, affected by the topic, disclosed during program delivery and was assisted to access professional support. These participants ($n=4$) in the intervention group and ($n=6$) in the control group were excluded from the data analysis. The final sample ($n=50$) consisted of FCCT coaches and managers, football academy students, members of a national personal and social development program, and the regional Football Association.

Procedure

Members of both Trusts were recruited through emails sent by their senior management. Initially, 12 participants from Trust 1 completed the baseline survey prior to participating in a focus group with the program developers. Measures were taken at baseline, post-test at 1-month post-intervention, and follow-up at 9-months post-intervention. Participants provided written informed consent. Football Onside was implemented in Trust 1 across two groups, each of which received three 2-hour sessions, one week apart in February to March 2019, which were delivered by two trained male and female facilitators together. Fidelity of program delivery was observed. The post-test questionnaire at 1-month contained course evaluation feedback questions. Participants received a £15 voucher at post-test, and an additional £10 voucher at follow-up. Anonymity of participants was maintained using generated ID codes enabling baseline and post-questionnaires to be linked. Trust 2 were offered free delivery of the Football Onside program after final data collection. Ethical approval was granted by the University of Exeter College of Social Sciences and International Studies Ethics Committee.

Measures

Self-Reported Learning Outcomes and Fidelity. Intervention group participants rated the extent to which the program met its learning objectives (improved knowledge, attitudes, likelihood to intervene and confidence) on a Likert-type scale from 1 (“Definitely no”) to 5 (“Definitely yes”). Mean scores for each question were calculated. Fidelity was observed in terms of adherence to the curriculum, quality of delivery, and program attendance. The same observer conducted all observations, completing six fidelity observation logs.

The Questionnaire. The survey included validated scales measuring attitudes, intentions, and bystander behaviors around sexual and domestic abuse, designed to map onto the extant literature, as well as questions to gather demographic information. Where necessary language was modified to be relevant to a UK adult audience: “girl” was replaced with “woman,” “campus”/“university” with “my organization” and “groups I work with,” and “community resource” with “professional agency.” Phrases “sexual abuse and intimate partner violence abuse” were replaced with “sexual violence and domestic abuse,” and “911” with “999.”

Demographics. Participants were asked their gender, sexual orientation, age, ethnicity, relationship status, if they knew someone affected by sexual or domestic abuse, and prior participation in a prevention program.

Rape Myth Acceptance. The Illinois Rape Myth Acceptance Short Form (IRMA) scale was used (McMahon & Farmer, 2011). The participants were asked to rate their beliefs about 19 statements such as “Rape happens when a man’s sex drive gets out of control” on a five-item Likert-type scale with higher scores reflecting greater adherence to rape myths. The mean was calculated, and the scale had a good internal consistency ($\alpha = .85$).

Domestic Violence Myth Acceptance. The Domestic Violence Myth Acceptance scale (DVMAS) was used (Peters, 2008). Participants rated their beliefs in 16 domestic abuse myths, for example, “A lot of domestic violence occurs because women keep on arguing about things with their partners,” on a seven-item Likert-type scale where higher mean score indicated greater acceptance of domestic abuse myths. Internal consistency of the scale was $\alpha = .7$.

Readiness for Change. A subset of items from the Readiness for Change scale (Responsibility and Denial subscales) was used (Banyard et al., 2010, 2014). Where questions asked about one type of violence only, it was adapted

to state “domestic abuse or sexual harassment/violence” due to the nature of the study. Item “I think I can do something about sexual violence” was adapted to “I think I can do something about domestic abuse and/or sexual harassment/violence and so I am planning to find out more about what I can do.” The participants answered nine items on a seven-point Likert-type scale, with several items being reverse coded. After reverse coding, the items were averaged so that higher number represented higher readiness for change. Internal consistency was acceptable ($\alpha = .6$).

Bystander Intent. The Intent to Help Friends scale was used (Banyard et al., 2014), along with items from the Bystander Attitude Scale–Revised (BAS-R) (McMahon et al., 2014). Participants rated their likelihood to help in 16 situations, such as “Ring a professional agency to get advice on how to safely challenge a friend, relative, or colleague who is a perpetrator,” on a five-item Likert-type scale where higher mean score indicated higher intent to help. The scale had a high internal consistency ($\alpha = .85$). One additional item was added to explore participants’ intent to “Look up laws on domestic abuse or sexual harassment/violence,” due to program content and the scale maintained internal consistency at $\alpha = .85$ level.

Bystander Efficacy. The Bystander Efficacy Scale (Banyard et al., 2007) was used to assess confidence to intervene. The scale ($n = 17$) had high internal consistency, $\alpha = .89$. Participants were asked to rate their level of confidence to perform certain behaviors, for example, “Confidence to speak up to someone who is making excuses for using physical force in a relationship,” on a scale from 0 to 100. A mean score for each participant was calculated, where a higher number represented greater bystander confidence.

Reported Behavior. To assess actual behavior change the Bystander Behavior Scale (BBS-R) was used (Banyard et al., 2007; McMahon et al., 2011). The participants reported any bystander behavior performed on a 3-item scale “Yes” (coded as “1”), “Wasn’t in the situation” (“0”), and “No” (“-1”). The scores were summed, with higher scores indicating higher engagement, and the scale ($n = 16$) had high internal consistency ($\alpha = .87$). Item “Signaled disapproval at sexist comments or jokes by using body language” was added due to program content and “Verbally challenged sexist comments and jokes” was paraphrased from “Challenge a friend who made a sexist joke.” One item was added to assess the law context: “Looked up the law on domestic abuse or sexual harassment/violence.” The scale maintained a high internal consistency $\alpha = .87$ with the added items. Following McMahon et al. (2017), we assessed the frequency of opportunity (i.e. how many times participants had

the opportunity to intervene) and the frequency of bystander intervention (i.e. how many times participants intervened) separately, descriptively analyzing replies “yes” and “no.”

Perceptions of Peer Helping. A subset of items from Perceptions of Peer Helping scale (Banyard et al., 2014) and BAS-R (McMahon et al., 2014) was used. We replaced “friends” with “friends, relatives and colleagues of the same gender,” consistent with Gainsbury et al. (2020). Participants rated the likelihood of their peers to perform five certain helping behaviors on a 5-point Likert-type scale. For example, “Approach someone they knew if they thought s/he was in an abusive relationship to let them know they were there to help.” Means for each item were calculated with higher number indicating a greater belief that peers would help. Internal consistency of the scale was moderate ($\alpha = .68$).

Perceptions of Peer Myth Acceptance. A subset of items from DVMAS and IRMA was used, adding a question on program content (“Sexist banter is okay if it’s only a joke”). Participants indicated the proportion (0–100%) of their friends, family, and colleagues of the same gender they thought would agree with each of the seven statements. Internal consistency was $\alpha = .6$. Baseline scores on Perceived Peer Helping and Peer Myth Acceptance were subsequently used as a pedagogical tool during the intervention to show participants’ actual and perceived norms.

Perceptions of Law Knowledge. In accordance with program content, two bespoke items were created to measure perception of law knowledge related to rape and to domestic abuse on a five-item Likert-type scale, for example, “How would you rank your knowledge about the law relating to sexual violence?” Higher scores indicated greater perceived law knowledge.

Backlash. We followed Moynihan et al. (2011a), calculating the change in attitude scores over time by at least one standard deviation in the undesired direction. A categorical variable was computed for each participant to see if their score changed or stayed the same.

Analysis Plan

Data were analyzed using Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC. The dataset was checked for normality of distribution of residuals, and for heteroscedasticity. Sensitivity analysis explored

missing items and outliers. We performed a square root transformation for one skewed variable, Bystander Behavior; however, since it did not alter the results, we present untransformed data. Missing data not exceeding 10% of total questions on the scale or subscale were handled using person mean substitution (Shrive et al., 2006); otherwise, the observation was coded as missing for that subscale.

Participants in both groups were compared using descriptive statistics and bivariate tests. We also compared completers with those lost to the 9-month follow-up. Dichotomous variables were analyzed using Chi-square tests, while continuous variables were compared using *t*-tests. Mixed-effects models were conducted to observe changes over time from baseline to post and baseline to follow-up, both between the intervention and control group and within the intervention group. Mixed-effects models are preferable to repeated measures ANOVAs because the approach is more flexible, allows for the analysis of nested cases and predictors (Tabachnick & Fidell, 2013), and handles missing data by estimating parameters based on available individual information (Gueorguieva & Krystal, 2004). Due to small sample size and to reduce the chance of Type I error, we adopted a conservative approach with restricted maximum likelihood and Kenward–Roger approximation (Kenward & Roger, 1997; Luke, 2017). Mixed-effects models included time, group, and the interaction between these variables as fixed effects, and random participant effects to account for between-participant variability. The interaction Time \times Group reflected the effects of the intervention. We applied appropriate contrasts to observe within-group changes in the intervention group. Standardized effect size based on model-adjusted mean comparison was calculated (Lipsey & Wilson, 2001) using Cohen's *d* and then Hedges' *g* to avoid bias due to small sample (Hedges & Olkin, 1985). By way of general guidance, $g < .2$ indicates small effect, $g = .5$ indicates medium effect, and $g > .8$ indicates large effect (Cohen, 1988). Backlash was calculated using Chi-square tests.

Results

Participants

Of the 60 initially recruited, 50 (83%) participants completed baseline and post-intervention surveys, and of those, 37 (74%) participants completed follow-up at 9 months. Attrition rates at follow-up were 32.35% ($n=11$) in the intervention group, and 12.5% ($n=2$) in the control group, $\chi^2(3)=50$, $p < .001$. The majority of participants self-identified as White British (96%), male (84%), and heterosexual (90%), with age ranging between 18 and

58 years ($M=31.44$, $SD=10.945$) (Table S2, Supplemental Appendix). Twenty-one participants (42%) reported knowing someone who had been affected by domestic abuse. Five participants (10%) had participated in a DV or SV program before.

There were no substantial differences between the intervention and control groups in terms of demographic variables and baseline scores. There were no significant differences in baseline scores and demographic variables between those who completed the questionnaires at all time points and those who were lost to follow-up. Similarly, participation in the focus group was not associated with any significant differences.

Primary Research Aims

Perceived Learning Outcomes and Fidelity. Participants consistently reported that the program met its learning objectives with mean scores surpassing 4 on every item ($M=4.32$, 95%CI [4.11, 4.52]) (Table S3 Supplemental Appendix). Mean observer scores across the six sessions ranged from 3.5 to 5.0 ($M=4.67$; $SD=0.61$), with one occasion falling below the 4.0 criterion. It became apparent from observation and participant complaint at the first session with the first group that one facilitator was unable to communicate the materials confidently and knowledgeably in accordance with the program learning objectives. Participant interaction with the facilitators in that session was lower. The trained facilitators were immediately replaced by the female-male program developers (second author) and the subsequent two sessions for group one, and the whole program for group two, was unaffected. Fidelity, engagement and attendance were observed to be subsequently high.

Effects of the Intervention

Between-Group. Table 1 shows model-estimated mean differences and means for between- and within-group changes. We found significant intervention effects at post-test for Domestic Abuse Myth Acceptance, $F(5, 98.22)=3.26$, $p=.009$, and at follow-up for Rape Myth Acceptance, $F(4, 98.22)=6.53$, $p<.001$, with greater improvements in the intervention group, and medium effect sizes on both measures. Significant intervention effects were also found in Bystander Intent, $F(5, 98.29)=3.37$, $p=.008$, and Bystander Efficacy, $F(5, 95.52)=4.34$, $p=.001$, at both time points, with greater improvements in the intervention group, and large effect sizes on these measures. No significant effects were found for other measures.

In terms of frequency of bystander opportunity, descriptive analysis showed that intervention group participants with opportunity to intervene reported engaging in more bystander behaviors at post-test and follow-up

compared with the control group participants who engaged in fewer behaviors over time (Table S4, Supplemental Appendix).

Within-Group. Analysis of the intervention effects within the intervention group revealed significant improvements with medium to large effect sizes on IRMA scores at post-test, $d(g)=-0.68(-0.67)$ and at follow-up, $d(g)=-0.88(-0.88)$, and DVMAS scores at post-test, $d(g)=-0.53(-0.53)$, and at follow-up, $d(g)=-0.4(-0.4)$. Participants significantly improved on their Bystander Intent at post-test, $d(g)=0.39(0.39)$ and at follow-up, $d(g)=0.43(0.43)$, and Bystander Efficacy at both time points, $d(g)=0.42(0.41)$ and $d(g)=0.45(0.45)$, respectively. Law knowledge related to DV and SV was also significantly greater with large effect sizes at post-test, $d(g)=0.93(0.92)$ and $d(g)=0.91(0.9)$ respectively, and at follow-up, $d(g)=0.62(0.61)$ for DV, and $d(g)=0.69(0.69)$ for SV. After one month, participants significantly improved on their Perceptions of Peer Helping, $d(g)=0.52(0.51)$, and Peer Myth Acceptance, $d(g)=-0.53(-0.52)$, but not at follow-up. No significant changes were observed for other measures.

Secondary Outcomes

Backlash. We identified backlash in 9.38% of the intervention group for bystander efficacy post-test ($n=3$) and 4.35% at follow-up ($n=1$). These changes in the undesired direction are outweighed by the proportion of participants whose scores improved by at least 1 *SD* post-test (75%) and at follow-up (86.96%). We also observed backlash in 4.55% of the intervention group in perceptions of peer helping at follow-up ($n=1$), outweighed by a proportion of participants (4.55%) whose scores improved by at least one *SD* on this measure, while there were no differences in the control group ($\chi^2=1.25, p=.53$).

Discussion

The purpose of this study was to examine the extent to which the exposure to the Football Onside program had effects on participants' knowledge, attitudes, and confidence about sexual and domestic violence, as well as their bystander behaviors. We also examined if the program learning objectives were met, and observed fidelity to the program. To our knowledge, this was the first UK-based study exploring the effects of a bystander intervention for the prevention of sexual and domestic violence in a professional sports setting. The results suggest mixed but promising changes and provide preliminary support for both the efficacy of the Football Onside program in UK population settings and the premise that professional sports may be an

appropriate and positive platform for VAWG prevention. The results also provide further evidence for the translatability of bystander interventions from the U.S. context (Fenton & Mott, 2017). These findings should be investigated in larger cluster randomized controlled trials.

The consistently high ratings for self-reported learning outcomes and the observed high engagement with the program suggests that content and mode of delivery were appropriate for the target audience. Further, as our participants were almost exclusively men, a traditionally hard to engage, but critical, group (Casey et al., 2018; Flood, 2011) is important for educators exploring bystander programs in real-world professional settings and adds to the evidence that some men will positively receive prevention training (Rich et al., 2010). This finding attests to both the potential of professional sports settings as critical spaces for prevention (Katz, 2018) and the importance of the bystander approach in positioning men as “social justice allies” (Fabiano et al., 2003). The need to replace facilitators highlights the importance of careful selection and thorough training of even professional facilitators, which will be essential for the delivery, engagement, and sustainability of future programs (Anderson & Whiston, 2005; Fenton & Mott, 2017; Nation et al., 2003).

There were several challenges to program and study implementation. Staff availability during the busy sports season was limited, presenting difficulty in releasing staff to take part in the evaluation and subsequent allocation of sufficient time for research activities. Researchers should be mindful of high staff turnover rates at Community Trusts (Bostock et al., 2021)—a factor beyond our control that was the main reason for study attrition. Securing buy-in to participation in the program in Trust 1, however, was not a challenge itself because it was presented by the CEO as part of staff responsibilities. This active involvement and positive role-modeling by the CEO, and stakeholders, further facilitated participation in the program and study. This aligns with similar findings by Fields et al. (2022), where support from athletic directors and athletes enhanced the sense of community and increased participation in CBIM.

A further facilitator to participant engagement was the cohort of people at which Football Onside’ was aimed. Those working at the Community Trust were generally very community-minded because this is at the core of their activities. Delivering the program to a cohesive peer group with similar job responsibilities and perceived prosocial attitudes underscored the importance of giving space for participants and “the opportunity to gather with like-minded men” to engage with VAWG (McMahon & Dick, 2011).

While the overall effects of the intervention were mixed, significant changes were observed for the measures that in particular correlate with the fundamental theoretical design of the Football Onside program (long-term

improvement for rape myth acceptance, bystander intent and confidence, and short-term improvement for domestic violence myth acceptance), with the exception of the actual helping behavior stage. These measures map well onto the processes of Latané and Darley's (1970) situational model as applied to VAWG, and previous studies (Jouriles et al., 2018; Moynihan et al., 2011a).

The maintained long-term improvement in rape myth acceptance speaks to the importance of meeting participants where they are at (Fenton & Jones, 2017): indeed, baseline RMA data were used to inform program development (Fenton et al., forthcoming). We do not know exactly why the significant improvements in DVMA were not maintained after 9 months but suspect it may be due to the fact that the program focused more explicitly on rape myths, in part because so much more is known from rape myth research. To our knowledge, domestic abuse myths have not yet been measured in evaluations of bystander programs with athletes; hence, further research should investigate change processes for this outcome.

Evidence that Football Onside may increase Bystander Intent and Efficacy is especially promising since these measures are important correlates of pro-social bystander behavior (Banyard, 2008). Moynihan et al. (2011a) observed similar improvements in a student athlete sample, but only in a 2-month follow-up evaluation. Our follow-up period of 9 months, however, is comparatively longer than the vast majority of evaluations (Kovalenko et al., 2022; Mujal et al., 2021), indicating potential for long-lasting change.

The final step in Latané and Darley's (1970) theoretical model is progression to actual bystander behavior. There were no significant differences in behavior between groups, although trends were in the desired direction. Descriptive analysis showed an increase from baseline to follow-up in bystander actions (when in the situation) in the intervention group, when participants have had more substantial time and opportunity to enact interventions. Many other studies have reported nonsignificant effects for bystander behavior (e.g., Jouriles et al., 2018; Moynihan et al., 2011a), including a male-targeted bystander program (Gidycz et al., 2011) and a student athlete sample (Moynihan et al., 2018). The trends within our results suggest that our study may simply have not been sufficiently powered to show significant differences. Behavior is notoriously difficult to change and there are challenges in evaluating bystander behavior. Firstly, at no time point did more than half of either group report being in the described situation, thus limiting the possibility to observe the intervention effects and possibly indicating an under-capture of bystander behaviors. Secondly, despite several attempts to measure opportunity (Cares et al., 2015; McMahon, 2015), this construct is still not fully captured with the existing measures and calculations. Descriptive analysis of the "yes," "no," and "wasn't in the situation" options allows the

assessment of frequency of opportunity and behaviors. However, summing the scores for inferential statistical analysis does not allow for meaningful conclusions about individual behavioral changes to be drawn, especially as some events are rare. Yet if analyzed separately, multiple comparisons of individual behaviors would introduce an increased chance of Type I error. Given the complexity of real-life opportunities to intervene and the diverse range of prosocial behaviors that might be performed, it is unlikely that quantitative methodology can ever prescriptively capture them. Inductive thematic analysis of additional open-ended survey questions (Braun & Clarke, 2006) or interviewing could be a possible solution. Nonetheless, real-life impact at a societal level can be cumulatively achieved by small individual effects (Jouriles et al., 2018).

That participants' Readiness did not significantly improve might be explained by the fact that their prior safeguarding training and job roles meant that they were already ready for change. However, given that our participants also had desirable baseline RMA scores, which did improve significantly after 9 months, desirability at baseline does not necessarily equate with no room for improvement. Further, as other studies have found an association between RMA and denial and taking responsibility (Banyard et al., 2014; Fenton & Mott, 2017), we suggest that further examination is required. We speculate that items on the Readiness scale, such as "I think I should learn more," might not fit the construct at post-intervention when participants have indeed just learned more; they may logically think that they should not learn more because the depth of the intervention means they have learned enough. Further, the presentation to participants of their own misperceptions of norms as a pedagogic device in the intervention might actually serve to increase denial. This is because participants now understand that others are more prosocial and hold fewer problematic views than they originally thought, and thus logically that it is less "of a problem in their organization." We suggest that the items and construct underpinning Readiness should be explored in more detail in further research.

The lack of significant difference between groups on Perceptions of Peer Helping and Perceptions of Peer Myth Acceptance is inconsistent with previous research, which has shown that correcting negative perceptions of peer norms is associated with personal willingness to intervene (Brown & Messman-Moore, 2010; Fabiano et al., 2003). Our finding may well attest to the difficulties in measuring the concept of peers generally. For example, individuals have multiple social identities and identify with multiple social groups, within which different norms may operate (Turner & Reynolds, 2010). The level of social identification with each group may influence their understanding and perception of "peers," and peer norms. We used "friends,

family and colleagues,” which was unvalidated, and acknowledge that this is a wide and possibly confusing comparator group requiring further development and refinement as a measure. However, the other purpose of this measure was to be able to present participants with their own misperceptions of norms as part of the intervention in accordance with social norms theory (Orchowski & Berkowitz, 2022). We note from the observation (and years of facilitator experience) that as a pedagogical device, this strategy results in high interest and engagement among participants. Equally, it may be simply that our sample was underpowered.

We found no significant differences in perceived law knowledge between groups. To our knowledge, no studies have extensively measured law-related knowledge, although it is often part of program curricula (Fenton et al., 2016). Thus, its effect on change processes is unknown. A validated law scale would be a valuable addition to the literature expanding understandings as to how bystander programs work.

Additional changes that were observed only on the within-group level included significant increases in Perceived Law Knowledge related to DV and SV in the intervention group at both time points, and bystander behavior at follow-up. Very few participants evidenced a backlash effect. Worsened scores on the measures of bystander efficacy and perceptions of peer helping were outweighed by a higher proportion of scores improved as a result of participation in Football Onside.

Limitations

Our study has several limitations that should be addressed in future research. Our sample could inhibit generalizability of findings to a broader population for several reasons. First, it was a UK-specific study, and findings may not apply to international contexts where the structure and functions of sports and charity organizations are different from the UK FCCT's. Second, the sample size was small and was also underpowered to reliably detect meaningful differences between groups, and we additionally lost a large number of participants to follow-up due to unanticipated staff turnover. This also means we were unable to look at the interactions of gender and age on the outcomes of interest to produce meaningful results. However, a large cluster randomized trial could address this limitation, as well as treatment assignment bias. If participants are randomly assigned and are blind to condition, this would reduce the chance of social desirability and researcher bias. Third, multiple comparisons in a small sample introduce a higher chance of Type I error. We adopted a conservative approach with Kenward–Roger approximation to minimize this issue;

however, the results should be interpreted with caution. Fourth, our study sample was relatively homogenous with respect to racial/ethnic identity and sexual orientation. However, previous research has shown that White heterosexual men are the bearers of hegemonic masculinity (Carrigan et al., 1985; Donaldson, 1993; Jewkes et al., 2015); hence, we argue that in the current study, it is a strength rather than a limitation. Fifth, reliance on retrospective self-reports can result in underestimates or overestimates in these measures and has been noted by other researchers (Sharot et al., 2007). Participants might not recall certain situations, or the memory could be distorted by emotions. They might also misunderstand and consequently underestimate their bystander involvement. Sixth, we calculated standardized effect sizes based on model-adjusted means between two groups. To date, there have been no guidelines on the calculation of effect sizes from mixed-effects models. More research is needed to advance reporting. Other limitations were introduced by providing a long questionnaire with modified scales. One of the reasons for dropouts at follow-up could be related to the length of the questionnaire. Although our study found acceptable internal consistency for each scale in our sample, the use of replicated items and modified language should be investigated further. Despite the limitations, however, we find it encouraging that the participants reported having a chance to intervene and giving examples of doing so since the beginning of the program.

Conclusions

The current feasibility study evaluated a bystander violence prevention program at FCCTs in South West England. The Football Onside program appears to have a promising impact on Community Trust members' attitudes and confidence to intervene. These findings are among the first to promote bystander intervention in professional football club settings. The two Community Trusts represented an ideal population for the study due to the outreach and impact on young people and communities. Trust coaches as leaders could be positive role models promoting new prosocial norms among their players. Further, this study expands the research base for bystander programs by recruiting a predominantly adult male sample with a mean age of 31. Further research should explore intervention effects through a cluster randomized controlled trial in professional athlete teams.

Acknowledgments

We owe a great deal of gratitude to Dr. Nathan Eisenstadt, Dr. Siobhan O'Dwyer, Dr. Fiona Warren, and Prof. Charles Abraham for their support throughout this research.

We would like to extend our sincere thanks to all the FCCT members involved in this study. Without these individuals, this project would not have been possible.

Declaration of Conflicting Interests

The author(s) declared a potential conflict of interest (e.g., a financial relationship with the commercial organizations or products discussed in this article) as follows: At the time of the research, Dr. Rachel Fenton was the supervisor of the project. She is now the director of a spin-out company set up to provide training on bystander intervention.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: Dr. Anastasiia G. Kovalenko was supported by the University of Exeter International Excellence Scholarship for Postgraduate Research. This research was also supported by the ESRC Impact Accelerator Account awarded to Dr. Rachel Fenton.

ORCID iD

Anastasiia G. Kovalenko  <https://orcid.org/0000-0002-4299-3587>

Supplemental Material

Supplemental material for this article is available online.

Note

1. Community Trusts are not for profit organizations that support health and well-being within their local community through involvement in sports and health projects. CT's are usually recognized among other supporters and nonsupporters in the United Kingdom due to their extensive outreach (Martin et al., 2016; Pringle et al., 2021).

References

- Adams, A. (2011). "Josh Wears Pink Cleats": Inclusive masculinity on the soccer field. *Journal of Homosexuality*, 58(5), 579–596.
- Anderson, L. A., & Whiston, S. C. (2005). Sexual assault education programs: A meta-analytic examination of their effectiveness. *Psychology of Women Quarterly*, 29(4), 374–388.
- Banyard, V. L. (2008). Measurement and correlates of prosocial bystander behavior: The case of interpersonal violence. *Violence and Victims*, 23(1), 83–97.
- Banyard, V. L., Eckstein, R. P., & Moynihan, M. M. (2010). Sexual violence prevention: The role of stages of change. *Journal of Interpersonal Violence*, 25(1), 111–135.
- Banyard, V. L., Moynihan, M. M., Cares, A. C., & Warner, R. (2014). How do we know if it works? Measuring outcomes in bystander-focused abuse prevention on campuses. *Psychology of Violence*, 4(1), 101–115.

- Banyard, V. L., Moynihan, M. M., & Plante, E. G. (2007). Sexual violence prevention through bystander education: An experimental evaluation. *Journal of Community Psychology, 35*(4), 463–481.
- Banyard, V. L., Plante, E. G., & Moynihan, M. M. (2004). Bystander education: Bringing a broader community perspective to sexual violence prevention. *Journal of Community Psychology, 32*(1), 61–79.
- Boeringer, S. B. (1999). Associations of rape-supportive attitudes with fraternal and athletic participation. *Violence Against Women, 5*(1), 81–90.
- Bogen, K. W., Mulla, M. M. M., & Orchowski, L. M. (2020). Gender-equitable attitudes, rape myth acceptance, and perceived peer acceptance of violence among high school students: An examination of gender and athletic involvement. *Journal of Interpersonal Violence, 37*(7-8), 1–17.
- Bostock, J., Bull, M., Woodward, J. & Sibley, J. (2021). Mission vs. Market: Theorizing the tensions within community sport trusts. *The International Journal of Sport and Society, 13*(1), 1–22.
- Bovill, H., & White, P. (2022). Ignorance is not bliss: A U.K. study of sexual and domestic abuse awareness on campus, and correlations with confidence and positive action in a bystander program. *Journal of Interpersonal Violence, 37*(5–6), 2801–2825.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101.
- Brackenridge, C. & Fasting K., Sexual harassment and abuse in sport: The research context. *Journal of Sexual Aggression, 2002. 8*(2): 3–15
- Brown, A. L., & Messman-Moore, T. L. (2010). Personal and perceived peer attitudes supporting sexual aggression as predictors of male college students' willingness to intervene against sexual aggression. *Journal of Interpersonal Violence, 25*(3), 503–517.
- Cares, A. C., Banyard, V. L., Moynihan, M. M., Williams, L. M., Potter, S. J., & Stapleton, J. G. (2015). Changing attitudes about being a bystander to violence: Translating an in-person sexual violence prevention program to a new campus. *Violence Against Women, 21*(2), 165–187.
- Carlson, J., Casey, E. A., Edleson, J. L., Tolman, R. M., Walsh, T. B., & Kimball, E. (2015). Strategies to engage men and boys in violence prevention: A global organizational perspective. *Violence Against Women, 21*(11), 1406–1425.
- Carlson, M. (2008). I'd rather go along and be considered a man: Masculinity and bystander intervention. *The Journal of Men's Studies, 16*(1), 3–17.
- Carrigan, T., Connell, B., & Lee, J. (1985). Toward a new sociology of masculinity. *Theory and Society, 14*(5), 551–604.
- Casey, E. A., Carlson, J., Two Bulls, S., & Yager, A. (2018). Gender transformative approaches to engaging men in gender-based violence prevention: A review and conceptual model. *Trauma, Violence, and Abuse, 19*(2), 231–246.
- Cheever, J., & Eisenberg, M. E. (2022). Team sports and sexual violence: Examining perpetration by and victimization of adolescent males and females. *Journal of interpersonal violence, 37*(1–2), NP400–NP422.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Lawrence Erlbaum.
- Corboz, J., Flood, M., & Dyson, S. (2016). Challenges of bystander intervention in male-dominated professional sport: Lessons from the Australian Football League. *Violence Against Women*, 22(3), 324.
- CPS. (2019). Violence Against Women and Girls Report. <https://www.cps.gov.uk/publication/violence-against-women-and-girls>
- Crosset, T. W., Ptacek, J., McDonald, M. A., & Benedict, J. R. (1996). Male student-athletes and violence against women. *Violence Against Women*, 2(2), 163–179.
- Donaldson, M. (1993). What is hegemonic masculinity? *Theory and Society*, 22(5), 643–657.
- Dyson, S., & Flood, M. (2008). *Building cultures of respect and non-violence: A review of literature concerning adult learning and violence prevention programs with men*. Australian Football League and VicHealth.
- Exner-Cortens, D., & Cummings, N. (2017). Bystander-based sexual violence prevention with college athletes: A pilot randomized trial. *Journal of Interpersonal Violence*, 36(1–2), NP188–NP211.
- Fabiano, P. M., Perkins, H. W., Berkowitz, A. D., Linkenbach, J., & Stark, C. (2003). Engaging men as social justice allies in ending violence against women: Evidence for a social norms approach. *Journal of the American College Health Association*, 52(3), 105–112.
- Fenton, R.A., Eisenstadt, N., and Kovalenko, A.G. (forthcoming). Developing a bystander programme for the prevention of VAWG in a male-dominated sports environment.
- Fenton, R. A., & Jones, C. (2017). An exploratory study on the beliefs about gender-based violence held by incoming undergraduates in England. *Journal of Gender-Based Violence*, 1(2), 147–167.
- Fenton, R. A., & Mott, H. L. (2017). The bystander approach to violence prevention: Considerations for implementation in Europe. *Psychology of Violence*, 7(3), 450–458.
- Fenton, R.A. & Mott H.L. (2018). The Intervention Initiative: theoretical underpinnings, development and implementation, In S. Anitha and R. Lewis, (Eds.), *Gender Based Violence in University Communities: Policy, Prevention and Educational Initiatives in Britain*. Bristol University Press. p. 169–188.
- Fenton, R. A., Mott, H. L., McCartan, K., & Rumney, P. N. (2016). *A review of evidence for bystander intervention to prevent sexual and domestic violence in universities* (Vol. 6, pp. 58–58). Public Health England.
- Fields, A. D., Fenn, M., Ripper, L., Hill, A., Mulbah, P., Kass, N. M., & Miller, E. (2022). A Coach-Delivered Dating Violence Prevention Program: Coach Perspectives on Implementation. *Journal of Interpersonal Violence*, 37(13–14), NP10594–NP10617.
- Flood, M. (2011). Involving men in efforts to end violence against women. *Men and Masculinities*, 14(3), 358–377.
- Forsdike, K, O’Sullivan, G., & Hooker, L. (2022). Major sports events and domestic violence: A systematic review. *Health Soc Care Community*, 30(6), e3670–e3685.

- Fraser-Thomas, J., Côté, J., & Deakin, J. (2008). Examining adolescent sport dropout and prolonged engagement from a developmental perspective. *Journal of Applied Sport Psychology, 20*(3), 318–333.
- Gainsbury, A. N., Fenton, R. A., & Jones, C. A. (2020). From campus to communities: Evaluation of the first UK-based bystander program for the prevention of domestic violence and abuse in general communities. *BMC Public Health, 20*(1), 1–11.
- Garrity, S. E. (2011). Sexual assault prevention programs for college-aged men: A critical evaluation. *Journal of Forensic Nursing, 7*(1), 40–48.
- Gidycz, C. A., Orchowski, L. M., & Berkowitz, A. D. (2011). Preventing sexual aggression among college men: An evaluation of a social norms and bystander intervention program. *Violence Against Women, 17*(6), 720–742.
- Georgieva, R., & Krystal, J. H. (2004). Move over ANOVA. *Archives of General Psychiatry, 61*(3), 310.
- Hedges, L., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Elsevier.
- Home Office. (2019). Violence against Women and Girls and Male Position Factsheets. <https://homeofficemedia.blog.gov.uk/2019/03/07/violence-against-women-and-girls-and-male-position-factsheets/>
- Home Office. (2021). Violence Against Women and Girls (VAWG) strategy 2021 to 2024. <https://www.gov.uk/government/consultations/violence-against-women-and-girls-vawg-call-for-evidence/violence-against-women-and-girls-vawg-strategy-2021-2024-call-for-evidence>
- Jaime, M. C. D., McCauley, H. L., Tancredi, D. J., Decker, M. R., Silverman, J. G., O'Connor, B., & Miller, E. (2018). Implementing a coach-delivered dating violence prevention program with high school athletes. *Prevention Science, 19*(8), 1113–1122.
- Jewkes, R., Morrell, R., Hearn, J., Lundqvist, E., Blackbeard, D., Lindegger, G., Quayle, M., Sikweyiya, Y., & Gottzén, L. (2015). Hegemonic masculinity: Combining theory and practice in gender interventions. *Culture, Health & Sexuality, 17*(Suppl 2), S112–S127.
- Jouriles, E., Krauss, A., Vu, N. L., Banyard, V. L., & McDonald, R. (2018). Bystander programs addressing sexual violence on college campuses: A systematic review and meta-analysis of program outcomes and delivery methods. *Journal of American College Health, 1*–10.
- Katz, J. (1995). Reconstructing masculinity in the locker room: The mentors in violence prevention project. *Harvard Education Review, 65*(2), 163–174.
- Katz, J. (2018). Bystander training as leadership training: Notes on the origins, philosophy, and pedagogy of the mentors in violence prevention model. *Violence Against Women, 24*(15), 1755–1776.
- Katz, J., Heisterkamp, H. A., & Fleming, W. M. (2011). The social justice roots of the mentors in violence prevention model and its application in a high school setting. *Violence Against Women, 17*(6), 684–702.
- Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983.
- Kettrey, H. H., & Marx, R. A. (2019). The effects of bystander programs on the prevention of sexual assault across the college years: A systematic review and meta-analysis. *Journal of Youth and Adolescence, 48*(2), 212–227.

- Kirby, S., Francis, B., & O'Flaherty, R. (2013). Can the FIFA world cup football tournament be associated with an increase in domestic abuse? *Research in Crime and Delinquency, 51*(3), 259–276.
- Kovalenko, A. G., Abraham, C., Graham-Rowe, E., Levine, M., & O'Dwyer, S. (2022). What works in violence prevention among young people?: A systematic review of reviews. *Trauma, Violence, & Abuse, 23*(5), 1388–1404.
- Latané, B., & Darley, J. M. (1970). *The unresponsive bystander: Why doesn't he help?* Appleton Century Crofts.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications, Inc.
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods, 49*(4), 1494–1502.
- Martin, A., Morgan, S., Parnell, D., Philpott, M., Pringle, A., Rigby, M., Taylor, A., & Topham, J. (2016). A perspective from key stakeholders on football and health improvement. *Soccer & Society, 17*(2), 175–182.
- McCray, K. L. (2015). Intercollegiate athletes and sexual violence: A review of literature and recommendations for future study. *Trauma, Violence, & Abuse, 16*(4), 438–443.
- McGlynn, C. (2018). Challenging the law on sexual history evidence: A response to Dent and Paul. *Criminal law review, 3*, 216–228.
- McMahon, S. (2015). Participation in high school sports and bystander intentions, efficacy to intervene, and rape myth beliefs. *Journal of Interpersonal Violence, 30*(17), 2980–2998.
- McMahon, S., Allen, C. T., Postmus, J. L., McMahon, S. M., Andrew Peterson, N., & Hoffman, M. L. (2014). Measuring bystander attitudes and behavior to prevent sexual violence. *Journal of American College Health, 62*(1), 58–66.
- McMahon, S., & Dick, A. (2011). “Being in a Room with Like-Minded Men”: An Exploratory Study of Men’s Participation in a Bystander Intervention Program to Prevent Intimate Partner Violence. *The Journal of Men’s Studies, 19*(1), 3–18.
- McMahon, S., & Farmer, G. L. (2009). The bystander approach: Strengths-based sexual assault prevention with at-risk groups. *Human Behavior in the Social Environment, 19*(8), 1042–1065.
- McMahon, S., & Farmer, G. L. (2011). An updated measure for assessing subtle rape myths. *Social Work Research, 35*(2), 71–81.
- McMahon, S., Palmer, J. E., Banyard, V. L., Murphy, M., & Gidycz, C. A. (2017). Measuring bystander behavior in the context of sexual violence prevention: Lessons learned and new directions. *Journal of Interpersonal Violence, 32*(16), 2396–2418.
- McMahon, S., Postmus, J. L., & Koenick, R. A. (2011). Conceptualizing the engaging bystander approach to sexual violence prevention on college campuses. *Journal of College Student Development, 52*(1), 115–130.
- Miller, E., Tancredi, D. J., McCauley, H. L., Decker, M. R., Virata, M. C. D., Anderson, H. A., Stetkevich, N., Brown, E. W., Moideen, F., & Silverman, J. G. (2012). “Coaching boys into men”: A cluster-randomized controlled trial of a dating violence prevention program. *Journal of Adolescent Health, 51*(5), 431–438.

- Moynihan, M. M., Banyard, V. L., Arnold, J. S., Eckstein, R. P., & Stapleton, J. G. (2011a). Engaging intercollegiate athletes in preventing and intervening in sexual and intimate partner violence. *Journal of American College Health, 59*(3), 197–204.
- Moynihan, M. M., Banyard, V. L., Arnold, J. S., Eckstein, R. P., & Stapleton, J. G. (2011b). Sisterhood may be powerful for reducing sexual and intimate partner violence: An evaluation of the Bringing in the Bystander in-person program with sorority members. *Violence Against Women, 17*(6), 703–719.
- Mujal, G. N., Taylor, M. E., Fry, J. L., Gochez-Kerr, T. H., & Weaver, N. L. (2021). A Systematic Review of Bystander Interventions for the Prevention of Sexual Violence. *Trauma, violence & abuse, 22*(2), 381–396.
- Nation, M., Crusto, C., Wandersman, A., Kumpfer, K. L., Seybolt, D., Morrissey-Kane, E., & Davino, K. (2003). What works in prevention: Principles of effective prevention programs. *American Psychologist, 58*(6–7), 449–456.
- NBC. (2014). Still Playing: 12 NFL Players Have Domestic Violence Arrests. <https://www.nbcnews.com/storyline/nfl-controversy/still-playing-12-nfl-players-have-domestic-violence-arrests-n204831>
- O'Donohue, W. T., & Schewe, P. A. (2019). *Handbook of sexual assault and sexual assault prevention*. Springer Cham.
- O'Hear, M. (2001). Blue-collar crimes/white-collar criminals: Sentencing elite athletes who commit violent crimes. *Marquette Sports Law Review, 12*(1), 427–447.
- ONS. (2020). *Domestic abuse during the coronavirus (COVID-19) pandemic, England and Wales: November 2020*. <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabuseduringthecoronaviruscovid19pandemicenglandandwales/november2020>
- ONS. (2023a). *Sexual offences victim characteristics, England and Wales: Year ending March 2022*. <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/sexualoffencesvictimcharacteristicsenglandandwales/yearendingmarch2022>
- ONS. (2023b). *Domestic abuse victim characteristics, England and Wales: Year ending March 2023*. <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/domesticabusevictimcharacteristicsenglandandwales/yearendingmarch2023>
- Orchowski, L. M., & Berkowitz, A. D. (2022). Chapter 1 – A brief history of the science and practice of engaging boys and men in sexual assault prevention. In L. M. Orchowski & A. D. Berkowitz (Eds.), *Engaging boys and men in sexual assault prevention* (pp. 1–27). Academic Press.
- Parent, S., & Fortier, K. (2018). Comprehensive overview of the problem of violence against athletes in sport. *Journal of Sport & Social Issues, 42*(4), 227–246.
- Peters, J. (2008). Measuring myths about domestic violence: Development and initial validation of the domestic violence myth acceptance scale. *J of Aggression, Maltreatment & Trauma, 16*(1), 1–21.
- Powell, A. (2011). *Review of bystander approaches in support of preventing violence against women*. VicHealth.

- Pringle, A. R., Zwolinsky, S., & Lozano-Sufrategui, L. (2021). Investigating the delivery of health improvement interventions through professional football club community trusts—strengths and challenges. *Public Health in Practice*, 2, 100104–100104.
- Prochaska, J. O., & DiClemente, C. C. (1986). *Toward a comprehensive model of change*. Springer US.
- Rich, M. D., Utley, E. A., Janke, K., & Moldoveanu, M. (2010). “I’d rather be doing something else”: Male resistance to rape prevention programs. *The Journal of Men’s Studies*, 18(3), 268–288.
- Schwartz, M. D., & DeKeseredy, W. S. (2008). The Role of Men. *Journal of Contemporary Criminal Justice*, 24(2), 178–185.
- Sharot, T., Martorella, E. A., Delgado, M. R., & Phelps, E. A. (2007). How personal experience modulates the neural circuitry of memories of September 11. *Proceedings of the National Academy of Sciences*, 104(1), 389–394.
- Shrive, F. M., Stuart, H., Quan, H., & Ghali, W. A. (2006). Dealing with missing data in a multi-question depression scale: A comparison of imputation methods. *BMC Medical Research Methodology*, 6(1), 57.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics: Pearson new international edition* (6th ed.). Pearson.
- Trendl, A., Stewart, N., & Mullett, T. L. (2021). The role of alcohol in the link between national football (soccer) tournaments and domestic abuse – Evidence from England. *Social Science & Medicine*, 268: 113457.
- Turner, J. C., & Reynolds, K. J. (2010). The story of social identity. In T. Postmes & N. Branscombe (Eds.), *Rediscovering Social Identity: Key Readings* (pp. 13–32). Psychology Press.
- Williams, D. J., & Neville, F. G. (2017). Qualitative evaluation of the mentors in violence prevention pilot in Scottish high schools. *Psychology of Violence*, 7(2), 213–223.
- Williams, D. J., Neville, F. G., House, K., & Donnelly, P. D. (2013). Association between Old Firm football matches and reported domestic (violence) incidents in Strathclyde, Scotland. *SAGE Open*, 3(3), 1–7.
- UN. (1993). *Declaration on the Elimination of Violence against Women*.
- UN Women. (2021). *The shadow pandemic: Violence against women during COVID-19*. Available from: <https://data.unwomen.org/publications/vaw-rga>
- WHO. (2010). *Preventing intimate partner and sexual violence against women*. Available from: <https://www.who.int/publications/i/item/9789241564007>
- Young, B. R., Desmarais, S. L., Baldwin, J. A., & Chandler, R. (2017). Sexual coercion practices among undergraduate male recreational athletes, intercollegiate athletes, and non-athletes. *Violence Against Women*, 23(7), 795–812.

Author Biographies

Anastasiia G. Kovalenko, PhD, is a postdoctoral researcher at the University of Bristol (UK). This research was done as part of her PhD at the University of Exeter Medical School (UK). Her research focuses on VAWG prevention and behavior change interventions. She is particularly interested in the bystander approach and

seeks to advance behavior change strategies to prevent violence and promote positive health outcomes.

Rachel A. Fenton, PhD, is an associate professor in law at the University of Exeter Law School (UK). Her research focuses on the prevention of VAWG through public health and bystander approaches, and gender and the law.