

Does GPS supervision of intimate partner violence defendants reduce pretrial misconduct? Evidence from a quasi-experimental study

Eric Grommon¹ · Jason Rydberg² ·
Jeremy G. Carter³

Published online: 14 September 2017
© Springer Science+Business Media B.V. 2017

Abstract

Objectives This research examines the effect global positioning system (GPS) technology supervision has on pretrial misconduct for defendants facing intimate partner violence charges.

Methods Drawing on data from one pretrial services division, a retrospective quasi-experimental design was constructed to examine failure to appear to court, failure to appear to meetings with pretrial services, and rearrest outcomes between defendants ordered to pretrial GPS supervision and a comparison group of defendants ordered to pretrial supervision without the use of monitoring technology. Cox regression models were used to assess differences between quasi-experimental conditions. To enhance internal validity and mitigate model dependence, we utilized and compared results across four counterfactual comparison groups (propensity score matching, Mahalanobis

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11292-017-9304-4>) contains supplementary material, which is available to authorized users.

✉ Eric Grommon
egrommon@iupui.edu

Jason Rydberg
jason_rydberg@uml.edu

Jeremy G. Carter
carterjg@iupui.edu

- ¹ School of Public and Environmental Affairs, Indiana University–Purdue University Indianapolis, 801 W. Michigan Street, BS 4067, Indianapolis, IN 46202, USA
- ² School of Criminology and Justice Studies, Center for Program Evaluation, University of Massachusetts Lowell, 113 Wilder Street, HSSB 4th Floor, Lowell, MA 01854, USA
- ³ School of Public and Environmental Affairs, Indiana University–Purdue University Indianapolis, 801 W. Michigan Street, BS 4081, Indianapolis, IN 46202, USA

distance matching, inverse probability of treatment weighting, and marginal mean weighting through stratification).

Results Pretrial GPS supervision was no more or less effective than traditional, non-technology based pretrial supervision in reducing the risk of failure to appear to court or the risk of rearrest. GPS supervision did reduce the risk of failing to appear to meetings with pretrial services staff.

Conclusions The results suggest that GPS supervision may hold untapped case management benefits for pretrial probation officers, a pragmatic focus that may be overshadowed by efforts to mitigate the risk of pretrial misconduct. Further, the results contribute to ongoing discussions on bail reform, pretrial practice, and the movement to reduce local jail populations. Although the cost savings are not entirely clear, relatively higher risk defendants can be managed in the community and produce outcomes that are comparable to other defendants. The results also call into question the ability of matching procedures to construct appropriate counterfactuals in an era where risk assessment informs criminal justice decision-making. Weighting techniques outperformed matching strategies.

Keywords Domestic violence · GPS supervision · Intimate partner violence · Pretrial · Pretrial misconduct · Pretrial supervision

Introduction

The use of global positioning system (GPS) technology to monitor the movements and activities of offender populations has gained significant momentum since the late 1990s. Increasingly, this form of technology is being deployed by criminal justice system administrators to mitigate the risk of repeat intimate partner violence (IPV) victimizations. At least 25 states have passed legislation enabling the use of GPS monitoring for IPV offenders (Gur et al. 2016). One arena where this form of monitoring is particularly relevant is during the pretrial period of criminal justice system processing that occurs after a custodial arrest and ends once a court disposition decision has been rendered (Bales et al. 2010). Victims are most susceptible to further victimization, harassment, intimidation, and associated retaliatory behaviors during this time as defendants attempt to dissuade victims from participating in formal criminal justice system processing (Erez et al. 2012; Han 2003; Sherman and Berk 1984). Pretrial GPS supervision is thought to reduce these behaviors by strengthening protection orders, enabling the near real-time surveillance of defendants, deterring contact with victims, assuring court appearances, and increasing public safety.

Despite the growing deployment of GPS technology to monitor IPV defendants during pretrial, there has been little research available to examine the efficacy of this strategy. One multisite evaluation has been conducted (Erez et al. 2012) and is in need of replication given the pace with which this form of technology continues to evolve. Salient to the current study, Erez et al. (2012) requested new research that is able to integrate matching techniques to create suitable comparison groups and produce unbiased assessments of pretrial outcomes. This study builds upon the important findings gleaned from this foundational study and contributes to the broader knowledge base on the use of GPS supervision in the criminal justice system in several ways.

First, we assess the effect GPS supervision has on failure to appear to court. Although one of the core functions of pretrial service operations is to ensure that defendants attend court hearings (Coopridge 2014), the ability of GPS supervision to affect failure to appear rates has yet to be examined. Second, we build knowledge on technology-assisted pretrial services by incorporating an analysis of failure to appear to scheduled meetings, one of several mechanisms used to assure defendants appear to court (Erez et al. 2012; Goldkamp and White 2006; Ibarra et al. 2014). Third, we integrate recent advances in matching and weighting procedures to generate appropriate counterfactual comparisons to minimize selection bias issues. These three focal aspects of the present study contribute new empirical inputs to a sparse body of disparate evidence regarding the utility of pretrial GPS supervision for IPV defendants, and provide methodological guidance to evaluation research where randomization is impractical and available comparisons provide few means to improve the internal validity of conclusions.

Empirical background

Approximately 1.3 million women and 835,000 men are victims of a physical assault from an intimate partner each year (National Institute of Justice 2015). Though these statistics demonstrate the concerning prevalence of IPV in the USA, estimates suggest that incidents of IPV are underreported by 40–50% (Tjaden and Thoennes 2000). Defensive actions taken by victims to report IPV incidents to law enforcement or simply separate from their abusive partner may expose victims to further abuse and victimization (Block 2003; Erez et al. 2012; Han 2003; Mahoney 1991; Sherman and Berk 1984) and may also escalate incidents, leading to lethal outcomes (Campbell et al. 2003). Providing intervention, assistance, and protection to IPV victims is difficult given the complex dynamics inherent to interpersonal relationships. Offenders possess knowledge of a victim's personal residence and those of friends/family, routine travels and recreation, employer, location of child care or school, and phone and email contacts.

Mandatory protection orders have become the first line of criminal justice system intervention during pretrial (Logan and Walker 2009). Yet, the ability of protection orders to prevent further victimization has been called into question by policymakers, practitioners, academics, and victims themselves (Logan et al. 2006). In a narrative review of the available evidence, Logan et al. (2006) report a wide range of protection order violation rates, ranging in value between 23% and 70% of victim samples examined. Spitzberg (2002) estimated a 40% violation rate across a sample of 32 studies on stalking that examine victim coping strategies.

The integration of technology has been offered as one means to strengthen protection orders and minimize pretrial misconduct (Erez et al. 2012). Despite the proliferation of GPS for monitoring pretrial IPV offenders, only a few empirical studies have accumulated data on the topic, leaving a paucity of evidence regarding its utility (Erez and Ibarra 2007; Erez et al. 2004). Qualitative research has noted that offender monitoring systems utilizing radio frequency signals to create inclusion/exclusion zones for home detention or curfew purposes create an environment where victims perceive greater safety and criminal justice system responsiveness (Erez and Ibarra

2007; Erez et al. 2004). Preliminary trends gleaned from two Midwestern probation departments indicate that very few defendants (i.e., between 1% and 2% of defendants referred to or placed on electronic monitoring) violate exclusion zones surrounding a victim's residence (Erez et al. 2004). Radio frequency technology is limited, however, in that it can only monitor the presence or absence of individuals in designated areas where a receiver has been installed. GPS expands surveillance capabilities to any location an individual may travel and is, thus, perceived to remedy this shortcoming (Brown et al. 2007).

To date, there has only been one quantitative examination of GPS monitoring for pretrial IPV offenders. Erez et al. (2012) produced three independent case studies using retrospective quasi-experimental designs to examine the effect of GPS monitoring on pretrial activities. The researchers concluded that GPS supervision was associated with almost zero contact attempts, fewer pretrial supervision violations, and reduced likelihoods of rearrest during the 1-year follow-up period in two out of three sites. The null effects at one site were attributed to the heterogeneity of GPS defendants and the methods used to construct the samples. More broadly, Erez et al. (2012) suggested that the effect of pretrial GPS supervision on relevant pretrial outcomes will depend, in part, on pretrial services program models. This interpretation contrasts from Spitzberg (2002), who found that protection order violation rates are influenced by defendant and victim characteristics.

The results also raise two important questions. First, Erez et al. (2012) examined pretrial "program violations" while defendants were under supervision, but it is not clear how this variable was operationalized. Therefore, it is not possible to determine whether GPS ensures that defendants attend court hearings or comply with reporting requirements found in most pretrial service program models. These are two of the most critical functions of pretrial operations. Second, treatment and comparison samples constructed by Erez et al. (2012) were generated across similar years within each site and estimated multivariate statistical control models were specific to each site. No additional procedures were used to compensate for selection bias. It is possible that differences in pre-treatment characteristics between the constructed groups confounded the estimated treatment effects (Shadish et al. 2002). The limited results from the three study sites of Erez et al. (2012), and the methodological limitations therein, demonstrate the need for more research on pretrial GPS supervision programs. There remain open questions of whether this form of supervision can affect rates of failure to appear to court or the mechanisms used to administer justice and hold defendants accountable.

Insights about the effect of GPS technology on the supervision of pretrial defendants can be gleaned from three early studies of radio frequency supervision with general offender populations. Most (73%) of the defendants placed on electronic monitoring in Marion County, Indiana were successfully terminated from a 90-day pretrial supervision term without violation (Maxfield and Baumer 1992), and although failure to appear rates could not be directly estimated, less than 13% of the sample absconded and 1% of the sample were arrested within a 90-day pretrial supervision term. In Lake County, Illinois, 219 defendants placed on electronic monitoring were less likely to be successfully discharged from their pretrial supervision in comparison to an unmatched sample ordered to traditional pretrial supervision (Coopridier and Kerby 1990). A further subanalysis found that defendants monitored with technology were more likely to violate a pretrial supervision condition, but were less likely to fail to appear to court

or be rearrested. Among 168 defendants placed on electronic monitoring in 17 federal districts across one fiscal year, 5% failed to appear to court and 6% were rearrested (Cadigan 1991), which was slightly higher than the failure to appear and rearrest rates among unmatched defendants managed without technology.

Most of the evidence on GPS monitoring comes from post-conviction applications. A meta-analysis of three quasi-experimental studies examining GPS or other forms of electronic monitoring found that technology-assisted supervision did not have any effects on future recidivism among higher risk offenders (Renzema and Mayo-Wilson 2005). Relevant to the current inquiry, this research noted that poor quality counterfactuals (i.e., improperly matched comparisons, complete lack of a comparison) were a common feature of the electronic monitoring evaluation literature to date.

More recently, Padgett et al. (2006) examined a sample of 75,661 offenders placed under house arrest in Florida to determine if technology-assisted supervision produced differential technical violation, new offense, and absconding outcomes. Those placed on GPS supervision were slightly more likely than those placed on radio frequency supervision to have their supervision revoked for a technical violation, even though the relative risks associated with new offenses and absconding were similar between these two groups. The researchers concluded that GPS supervision produced similar outcomes to radio frequency supervision, albeit at a higher financial cost to community correctional agencies.

Expanding the scope of this initial examination beyond house arrest and integrating a propensity score-based counterfactual, Bales et al. (2010) compared a sample of medium- and high-risk offenders across a 6-year period in Florida who were placed on electronic monitoring during their community supervision term ($n = 5034$) or were supervised via traditional mechanisms without the assistance of technology ($n = 266,991$) after a felony conviction. The results indicated that GPS supervision produced greater net reductions in revocation or absconding outcomes than radio frequency technology.

Additional studies have examined specific post-conviction offender populations. In comparing samples of high-risk sex and gang offenders on GPS to propensity score matched comparisons in California, Gies and colleagues (2013) found that sex offenders under GPS supervision were less likely to commit new offenses, sex offenses, or to have their supervision revoked in relation to the comparison group. Similarly, gang offenders under GPS were less likely to be rearrested, but had higher rates of technical violations and new offense violations. In San Diego County, California, Turner et al.'s (2015) evaluation of a pilot GPS program found that high-risk sex offenders supervised with GPS and specialized caseloads were less likely to fail to register, abscond, or be found guilty of committing a new offense compared to an unmatched sample.

In all, there is a mixture of evidence about whether the use of GPS technology increases or reduces offender compliance with supervision conditions. The available findings are diverse for both pretrial IPV defendants and samples of post-conviction offenders. Evidence on the role of GPS technology to reduce recidivism is also varied, although recent research with higher risk offenders has begun to build consensus that GPS monitoring can reduce recidivism. To this extent, the study reported here was designed to address two central questions: (1) does pretrial GPS supervision for IPV

defendants improve compliance with supervision conditions and (2) does this form of supervision reduce recidivism?

The current study

Context

The effectiveness of any form of offender monitoring technology is a product of many factors, including federal and state laws, local policies, personnel training, supervision of personnel by administrators, and offender eligibility criteria (Erez et al. 2004, 2012, 2013; Ibarra 2005; Ibarra and Erez 2005; Ibarra et al. 2014). It is critical to understand the unique context in which a form of technology is used and the objectives it is anticipated to produce when assessing outcomes (Salvemini et al. 2015).

The study site sponsoring the research was housed in a pretrial services division of the city and county community corrections department in a large jurisdiction of approximately three million residents in the Western region of the USA. The site was chosen from a snowball sample of jurisdictions identified by subject matter experts and professional organizations, with eligibility based on experience managing pretrial IPV defendants with GPS technology, sizable populations of pretrial defendants being managed with GPS, and data accessibility. The study site had GPS monitoring experience dating back to 2002 and annually averaged roughly 400 defendants ordered to pretrial GPS supervision for IPV offenses.

Determinations of placement to pretrial GPS supervision are a function of two processes. First, face-to-face interviews are conducted by pretrial services with arrestees booked into jail before the arrestee's initial hearing. Interviews are coupled with information compiled from victims' advocates from the police department and prosecution offices, as well as queries to statewide criminal history systems, local court databases, and pretrial services' management information system to score two pretrial risk assessment tools: one for all offenses and one specifically for risk of future domestic assault. Second, a bond advisement report based on offense details and risk assessment scores recommends the level of supervision if pretrial release is granted by the court, with GPS being the most intense level offered.

In addition to monitoring facilitated with technology, defendants must hold physical and telephone check-ins, and attend case management meetings. Supervision-level recommendations are based upon an internal matrix tool that was created by pretrial services and was vetted by presiding judges, prosecutors, and defense counsel representatives. Recommendations are not binding; judges ultimately have the discretion to determine pretrial release and supervision conditions. If pretrial supervision is ordered, defendants are required to report directly to pretrial services for intake. Defendants are informed that their movements will be tracked at all times in real time, and that their history of movement is archived for monitoring and investigative purposes.

The study site employs a one-piece GPS ankle unit and subscribes to an active monitoring plan where GPS points are captured every minute when defendants are not in violation and every 15 s when defendants are in violation. GPS points are monitored by a centralized monitoring center and the supervising pretrial officer. When defendants are in compliance, GPS information is accessed on an as-needed basis. Defendants who

are in violation trigger notifications to an officer's department-issued cellphone, email, and computer. Action must be taken by the officer to recognize and respond to the notification. Officers observe offender movement patterns with proprietary software, which enables sending pre-recorded verbal notifications to defendants and audibly "ping" a device at any time. Defendants must acknowledge these notifications. Pretrial officers at the study site do not have arrest powers, instead relying on local law enforcement to conduct welfare checks and make custodial arrests. Violations are reported to the prosecutor's office, who take discretionary action.

In relation to the pretrial GPS supervision models offered by Ibarra et al. (2014; see also Erez et al. 2012), the study jurisdiction most closely resembles the data-informed due process and punitive model. Inclusion zones, curfew restrictions, and home visits are not used. Victims are informed that GPS monitoring is simply a device and, as a tool, the technology cannot provide protection. In addition to setting an exclusion zone to the terms of the protection order, multiple exclusion zones, up to 1000 ft, can be set by the supervising officer with assistance from the victim or victim advocate. Casework approaches do not seek to fulfill rehabilitative ideals, but they also do not aim to simply detect and punish. Rather, the relationships with defendants are somewhere in between these extremes.

Methods and analytic approach

Research design and participants

A retrospective quasi-experimental design was utilized. The flow of defendants to the study site with IPV offenses was followed for a 1.5-year period to generate a large sample that most accurately reflected pretrial supervision decision-making at the study site. Most importantly, this sampling period enabled the construction of a comparison group of IPV defendants who were ordered and placed on traditional pretrial supervision without GPS technology. Overall, 3480 IPV defendants were processed by the study site across a 1.5-year period. Thirty-two percent ($n = 1116$) did not bond to pretrial supervision and were, therefore, removed from the eligibility pool. Twenty-nine percent ($n = 1000$) were ordered to pretrial GPS supervision and 573 (16% of all IPV defendants, 57% of ordered) bonded and were placed on pretrial GPS supervision. The remainder of the sample (39%, $n = 1351$) were ordered to traditional pretrial supervision. Of those ordered, 910 (26% of all, 67% of ordered) bonded and were placed on traditional pretrial supervision. The final sample consisted of 1483 defendants; 573 defendants under pretrial GPS supervision form the treatment group and an unmatched pool of 910 defendants who bonded to pretrial supervision but were not supervised with GPS technology. Figure 1 presents a CONSORT flow diagram of IPV defendants for this study.

Measures

Two management information systems of the pretrial services division were used. One system maintains records about defendants referred to the court and documents the period between jail booking and the bonding decisions made at the initial hearing. The second system details court and supervision activity. Both systems were used for demographic, criminal history, instant offense, risk assessment, bond recommendation, and bond order information. Misconduct measures were extracted from the second system.

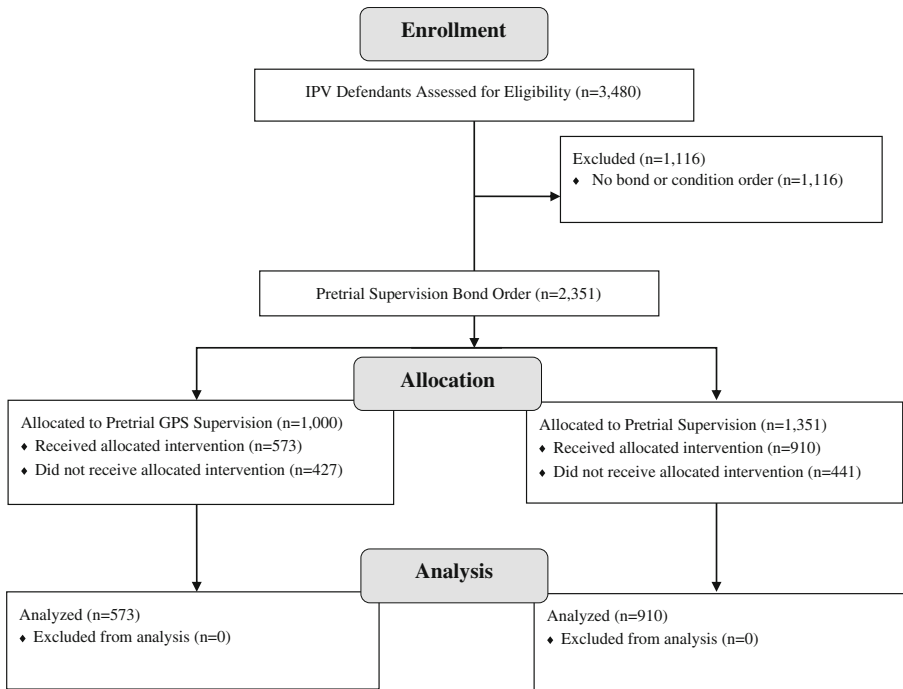


Fig. 1 CONSORT flow diagram of intimate partner violence (IPV) defendants in quasi-experimental conditions

Dependent variables Four dependent variables were used to measure pretrial misconduct. *Failure to Appear to Court* documents whether a defendant has failed to appear to a scheduled court hearing. *Failure to Appear to Meeting* represents whether a defendant has failed to attend scheduled meetings with their assigned officer. Although failing to appear to court is a more serious violation than failing to appear to meetings, the inability to attend meetings can also be viewed as an important signal of unobservable issues that may affect pretrial misconduct and public safety (Ibarra 2005). *Rearrest* is defined as any new arrest for any new offense. *Domestic Rearrest* is defined as any new arrest for a domestic violence offense, including violations of protection or court orders. All of the dependent variables are dichotomous measures. Also integrated are measures of the timing in days to all of the dependent variables, operationalized as the difference between the date a dependent variable occurs and the date on which a defendant was ordered to pretrial supervision.

Independent variable All defendants were involved in one of two conditions across their pretrial supervision term. The treatment group consisted of 573 IPV defendants who bonded and were placed on pretrial GPS supervision per court order. Defendants in this group were mandated to pretrial GPS supervision and were, by default, subjected to more intense surveillance than other defendants managed by pretrial services. This group was also required to maintain more frequent contact with pretrial services through telephone check-ins (held 1–4 times a month) and in-person case management meetings (held 1–4 times a month).

The comparison group consisted of 910 unmatched IPV defendants who bonded and were placed on pretrial supervision. Defendants in this group were not mandated to pretrial GPS supervision or any other form of electronic monitoring at the study site. This group was also required to maintain contact with pretrial services through telephone check-ins (on an as-needed basis or up to 4 times a month) and in-person case management meetings (on an as-needed basis or up to 4 times per month). Beyond these notable differences between the two groups, defendants were similarly managed. Both groups were subject to mandatory protection orders, court reminder calls, were required to meet with their pretrial officer after court appearances, inform pretrial services of law enforcement contact, and were subject to having their bond revoked for violating a protection order or other forms of pretrial misconduct.

Counterfactual estimation strategy

Table 1 displays the demographic characteristics of the sample. Chi-square and two-way analysis of variance (ANOVA) models were used to examine baseline differences between quasi-experimental groups. Many significant differences exist between the two groups, notably the differences in presumed risk that are driven by instant offense charges and classes. Members of the treatment group facing offenses with more serious charges that entail violations of an existing restraining order or other forms of threatening or harassment have higher actuarial risk scores, have had more contact with the criminal justice system, and received higher bond amounts. In all, there is clear evidence that both groups are imbalanced on observable pre-treatment indicators that inform the placement of defendants on pretrial GPS supervision.

Matching procedures To attempt to reduce the confounding effects of selection bias, a variety of propensity score-based matching and weighting strategies were employed to construct a series of alternative comparison groups. The purpose of using different strategies was to mitigate model dependence issues that plague extant research (King and Nielsen 2016). The primary issue here is that estimated treatment effects may be contingent upon the matching strategy (Bales and Piquero 2012; Gaes et al. 2016), and this approach allows consideration of the sensitivity of the results to different means of counterfactual generation.

Propensity scores were estimated from a total of 27 variables reflecting archival data on defendant demographics, risk assessment scores, instant offense details, and criminal history record information. The propensity score was estimated as a covariate balancing propensity score (CBPS), as described by Imai and Ratkovic (2014), and designed to reflect the average treatment effect on the treated units (ATT). As opposed to utilizing logit regression to produce a propensity score maximizing the prediction of treatment assignment, the CBPS score is explicitly calibrated to optimize covariate balance between the treatment and comparison groups (see also Clark and Rydberg 2016). This particular propensity score was chosen to add another layer of mitigation against model dependence by eliminating the need for complex propensity score model specifications (e.g., polynomials and interactions), which can have notable

Table 1 Demographics, instant offense, criminal history, and bond information

Covariate	Treatment group (<i>N</i> = 573)	Comparison group (<i>N</i> = 910)	Total (<i>N</i> = 1483)
	<i>n</i> (%) / <i>m</i> (SD)	<i>n</i> (%) / mean (SD)	<i>n</i> (%) / mean (SD)
Male*	536 (94%)	726 (80%)	1262 (85%)
White*	390 (68%)	662 (73%)	1052 (71%)
Hispanic	226 (39%)	323 (35%)	549 (37%)
English as second language	16 (3%)	37 (4%)	53 (4%)
Age*	35.43 (10.38)	34.34 (10.42)	34.76 (10.41)
Number of charges*	1.84 (0.85)	1.76 (0.75)	1.79 (0.80)
Most serious charge*			
Assault	331 (58%)	654 (72%)	985 (66%)
Vio. restraining order	95 (16%)	64 (7%)	159 (11%)
Threat/harassment	69 (12%)	58 (6%)	127 (9%)
Disturbing the peace	32 (6%)	72 (8%)	104 (7%)
Burglary/theft	23 (4%)	25 (3%)	48 (3%)
Other property	23 (4%)	37 (4%)	60 (4%)
Most serious class*			
Felony	129 (23%)	132 (14%)	261 (18%)
Misdemeanor	213 (37%)	217 (24%)	430 (29%)
Municipal ordinance	231 (40%)	561 (62%)	792 (53%)
Pretrial risk score*	33.26 (16.16)	23.75 (14.55)	27.42 (15.88)
Domestic assault risk score*	4.81 (3.96)	2.00 (2.76)	3.08 (3.55)
Under supervision*	178 (31%)	141 (15%)	319 (21%)
Prior felony arrest*	293 (51%)	237 (26%)	530 (36%)
Prior misd. arrest*	373 (65%)	493 (54%)	866 (58%)
Prior mun. ord. arrest*	123 (21%)	75 (8%)	198 (13%)
Prior DV assault arrest (O)*	332 (58%)	272 (30%)	604 (41%)
Prior prison sentence (P)*	141 (24%)	105 (11%)	246 (17%)
Prior jail sentence (P)*	146 (25%)	156 (17%)	302 (20%)
Prior split sentence	120 (21%)	159 (17%)	279 (19%)
Prior probation sentence	166 (29%)	230 (25%)	396 (27%)
Bond amount*	\$6689.18 (13,052.30)	\$4234.73 (10,063.98)	\$5183.07 (11,371.31)

* $p < 0.05$. A Kruskal–Wallis H test and Mann–Whitney test also indicated significant differences between the two groups on the mean rank and median bond amount [$\chi^2(1) = 89.66, p < 0.001$]

influences on treatment effects (Imai and Ratkovic 2014; Smith and Todd 2005). This estimation was implemented via the CBPS package in R (Fong et al. 2016). Utilizing the CBPS score, a propensity matched comparison group ($n = 573$) was generated using a one-to-one nearest neighbor matching procedure using the logit distance. Matching was performed without replacement or the designation of a distance caliper.

Propensity score matching strategies have been criticized as being unable to balance unobservable covariates exogenous to the treatment condition and, thus, meet the fundamental objective of approximating a completely randomized experiment (King and Nielsen 2016). In its place, the use of the Mahalanobis distance has been advocated (King and Nielsen 2016). This approach is better suited to identifying comparison units that more closely resemble the covariance matrix of treatment units and more adequately identify matches in relation to propensity score strategies (Gu and Rosenbaum 1993). As such, a Mahalanobis distance matched comparison group ($n = 573$) was estimated and included in the analysis.¹

A gamma statistic (Γ) is estimated to accompany analyses using the match-based comparison groups. Gamma is an indicator of the threat of hidden biases that could reverse or nullify the statistically dependable differences between treatment conditions (Loughran et al. 2015). For two identical matched subjects, gamma represents how strongly an unobserved covariate would need to differentially impact treatment assignment before the statistical significance of treatment effects would be reversed (Keele 2011). For instance, when $\Gamma = 2$, then the treatment condition would need to be twice as likely to experience the outcome due to an unobserved covariate in order for the observed significance of treatment effects to be reversed (Loughran et al. 2015). When $\Gamma = 1$, findings only hold when no unobserved covariates operate on treatment assignment. Larger gamma values increase confidence that the results are not sensitive to hidden biases, and $\Gamma = 1.3$ indicates a sensitive treatment effect.

Weighting procedures Although matching strategies discard a considerable proportion of comparison units, we considered two weighting strategies which utilize information from all available comparison units. An inverse weighted comparison group (IPTW) ($n = 910$) was developed following the procedures used by Bales et al. (2010) and Visser et al. (2017). Utilizing the same CBPS score as above, weights for the inverse weighted comparison group are the inverse of one minus the probability of being ordered on GPS supervision, and weights for the treatment group are set to 1. Additionally, we utilized a recent innovation in propensity score methods known as marginal mean weighting through stratification (MMW-S) (Hong and Hong 2009) to produce a marginal mean weighted comparison group ($n = 865$). In this approach, the sample is first stratified on the quintiles of the CBPS score and comparison units are

¹ In addition to the propensity score and Mahalanobis matching techniques, we considered the application of coarsened exact matching (CEM). This technique reduces observed covariates thought to influence selection decisions down to blocks of strata. In turn, these blocks of strata are used to identify comparison units and match these units to a treatment unit (Iacus et al. 2012). Unfortunately, due to the relatively extreme differences between the GPS and comparison groups in the present application, this matching strategy was unable to produce a sufficient number of exact matches. Known as the “tyranny of dimensionality” (Nagin et al. 2009, p. 145), where exact matches on multiple dimensions are simply not feasible as the number of covariates to be matched increases, this is a major limitation of precision-based matching. Indeed, using the same matching specification as the propensity score and Mahalanobis matching techniques, along with reasonable cutpoints for quantitative variables, the CEM procedure was only able to identify two GPS and two comparison units ($n = 4$) as exact matches, resulting in a 99.9% sample loss. In light of this limitation, we did not employ a CEM comparison group in the estimation of treatment effects.

weighted based on the average conditional probability of treatment assignment within a given strata. Approximating random assignment to treatment is generated by eliminating cases with propensity scores that are not represented in both treatment and control conditions ($n = 45$, 4.9%), and then assigning comparison units weights corresponding to the following:

$$MMW-S_{ATTi} = \frac{O_0 \times (O_{1s}/O_1)}{O_{0s}} \quad (1)$$

where the weight for comparison unit i is a function of the total observed number of comparison units O_0 , the total number of observed treated units O_1 , and the relative frequencies of treated and comparison units in propensity score stratum s (i.e., $s \in \{1, 2, 3, 4, 5\}$) (Hong and Hong 2009). In Eq. 1, the numerator represents the expected frequency of control units in a given propensity score strata given the observed distribution for the treatment group, and the denominator represents the observed frequency of control units in each strata.

Analytic strategy To examine the effect of pretrial GPS supervision on failure to appear and rearrest outcomes, bivariate Chi-square associations were first examined between the unmatched groups. Next, Kaplan–Meier survival estimates are presented to determine whether there are significant differences between groups on the timing of outcome measures for both groups. Cox regression models complete the analyses. These models enable comparisons of the relative effects between unmatched and matched quasi-experimental conditions that control for differential time at risk (i.e., length of pretrial supervision) between defendants.² Six different models were estimated for each outcome variable. The first model examines unconditional relationships. The second model enters all control variables presented in Table 1 to form a multivariate regression model with statistical controls. The remaining models employ matched comparison groups generated from the aforementioned matching strategies.

Results

Descriptive and unconditional comparisons

A large majority of the defendants attended all of their court appearances (95%) and scheduled meetings with pretrial service staff (70%). Most of the sample remained arrest-free while under pretrial supervision (84%). Of those defendants who were arrested during pretrial ($n = 244$), 43% were rearrested for a domestic offense. A larger proportion of the treatment group failed to appear

² The average length of pretrial supervision for the treatment group was 141.60 days (SD = 123.98), and it was 117.94 days (SD = 100.75) for the comparison group. The mean difference between the two groups was statistically dependable [$F(1, 1482) = 16.18, p < 0.001$]. A Kruskal–Wallis H test and Mann–Whitney test also indicated significant mean rank and median length of supervision differences between the groups [$\chi^2(1) = 14.02, p < 0.001$].

to at least one pretrial service meeting [$\chi^2(1, 1483) = 31.39, p < 0.001$], were rearrested [$\chi^2(1, 1483) = 97.72, p < 0.001$], and were rearrested for a domestic offense [$\chi^2(1, 1483) = 44.15, p < 0.001$] in relation to the unmatched comparison group. Despite these trends, the rate of failure to appear to court was slightly higher among the unmatched comparison group (7%) than the treatment group (3%) [$\chi^2(1, 1483) = 11.57, p < 0.001$].

Regarding the timing of outcome measures, there were no significant differences between the two groups on the timing to first failure to appear to court [log-rank $\chi^2(1, 1483) = 0.01, p = 0.93$], first arrest [log-rank $\chi^2(1, 1483) = 0.54, p = 0.46$], or first domestic arrest [log-rank $\chi^2(1, 1483) = 0.42, p = 0.52$]. The timing to the first failure to appear to a meeting with pretrial services did differ between those who were and were not ordered to pretrial GPS supervision [log-rank $\chi^2(1, 1483) = 13.40, p < 0.001$]. On average, members of the unmatched comparison group who failed to appear to a pretrial service meeting did so 63 days into their pretrial supervision term ($M_{CG} = 62.75, SE = 5.06$). The treatment group missed their first pretrial service meeting 25 days later than the comparison group ($M_{TG} = 88.23, SE = 5.16$).

As an initial baseline, these results indicate significant differences in pretrial outcomes and the timing to first failure to appear to a scheduled meeting with pretrial services for defendants ordered to GPS supervision and those who bonded to pretrial supervision without GPS or other forms of electronic supervision. Next, we describe Cox regression models and attend to selection bias issues that may be confounding the initial results.

Counterfactual comparisons

Table 2 presents the results of the matching strategies used to construct comparison groups. Prior to matching, large standardized differences between the two groups were observed, with a larger proportion of the treatment group having a criminal history that included a domestic violence assault arrest (standardized difference = 0.83) and a higher average domestic assault risk score (standardized difference = 0.82). After constructing counterfactuals, the weighting strategies outperformed the matching techniques. The average standardized difference for the IPTW and MMW-S techniques were below the threshold of 0.10 that has been offered as a heuristic for determining if covariate imbalance exists (Austin 2009). Despite this average balance score, two pre-treatment covariates in the IPTW procedure remained imbalanced. MMW-S produced the most equivalent comparison to assess the effect of GPS supervision.

Table 3 displays the results of the Cox regressions with models specified by unconditional, conditional, and matched or weighted comparison groups. As all of the available pre-treatment covariates were properly balanced through the MMW-S strategy, the treatment effect should be interpreted through this set of analyses. Overall, the estimated coefficients indicate that the treatment group had lower odds of failing to appear to meetings with pretrial services staff than the comparison group. The hazard of failing to appear to a meeting with

Table 2 Covariate balance across matching/weighting methods

Covariate	Raw data	Propensity score matching	Mahalanobis distance	IPTW	MMW-S
	Std. difference	Std. diff (% Δ)	Std. diff (% Δ)	Std. diff (% Δ)	Std. diff (% Δ)
Male	0.585	0.172 (- 70.6)	0.220 (- 62.4)	0.063 (- 89.2)	0.047 (- 92.0)
White	0.145	0.097 (- 33.1)	0.070 (- 51.7)	0.038 (- 73.8)	0.059 (- 59.3)
Hispanic	0.115	0.040 (- 65.2)	0.051 (- 55.7)	0.016 (- 86.9)	0.003 (- 97.4)
English as second language	0.099	0.165 (+ 66.7)	0.000 (- 100.0)	0.043 (- 56.6)	0.047 (- 52.5)
Age	0.104	0.037 (- 64.4)	0.116 (+ 11.6)	0.077 (- 26.0)	0.111 (+ 6.7)
Number of charges	0.103	0.091 (- 11.7)	0.140 (+ 35.9)	0.051 (- 50.5)	0.087 (- 15.5)
Assault	0.422	0.240 (- 43.1)	0.266 (- 36.7)	0.034 (- 91.9)	0.038 (- 91.0)
Vio. restraining order	0.423	0.262 (- 38.1)	0.238 (- 43.7)	0.062 (- 85.3)	0.031 (- 92.7)
Threat/harassment	0.279	0.189 (- 32.5)	0.137 (- 50.9)	0.013 (- 95.3)	0.047 (- 83.2)
Disturbing the peace	0.131	0.128 (- 2.3)	0.000 (- 100.0)	0.026 (- 80.2)	0.082 (- 37.4)
Burglary/theft	0.099	0.039 (- 60.6)	0.026 (- 73.7)	0.028 (- 71.7)	0.014 (- 85.9)
Other property	0.004	0.025 (+ 525.0)	0.000 (- 100.0)	0.015 (+ 275.0)	0.039 (+ 875.0)
Felony	0.293	0.078 (- 73.4)	0.147 (- 49.8)	0.059 (- 79.9)	0.028 (- 90.4)
Misdemeanor	0.414	0.220 (- 46.9)	0.275 (- 33.6)	0.173 (- 58.9)	0.093 (- 77.5)
Municipal ordinance	0.618	0.274 (- 55.7)	0.379 (- 38.7)	0.215 (- 65.2)	0.066 (- 89.3)
Pretrial risk score	0.618	0.418 (- 32.4)	0.391 (- 36.7)	0.132 (- 78.6)	0.017 (- 97.2)
Domestic aslt. risk score	0.824	0.583 (- 29.2)	0.688 (- 16.5)	0.236 (- 71.4)	0.094 (- 88.6)
Under supervision	0.530	0.329 (- 37.9)	0.329 (- 37.9)	0.058 (- 89.1)	0.103 (- 80.6)
Prior felony arrest	0.754	0.433 (- 42.6)	0.465 (- 38.3)	0.111 (- 85.3)	0.090 (- 88.1)
Prior misd. arrest	0.317	0.198 (- 37.5)	0.123 (- 61.2)	0.115 (- 63.7)	0.053 (- 83.3)
Prior mun. ord. arrest	0.535	0.383 (- 28.4)	0.453 (- 15.3)	0.088 (- 83.6)	0.052 (- 90.3)
Prior DV assault arrest	0.833	0.543 (- 34.8)	0.569 (- 31.7)	0.139 (- 83.3)	0.032 (- 80.3)
Prior prison sentence	0.487	0.295 (- 39.4)	0.276 (- 43.3)	0.090 (- 81.5)	0.046 (- 90.6)
Prior jail sentence	0.289	0.208 (- 28.0)	0.128 (- 55.7)	0.115 (- 60.2)	0.057 (- 80.3)
Prior split sentence	0.125	0.055 (- 56.0)	0.062 (- 50.4)	0.018 (- 85.6)	0.074 (- 40.8)
Prior probation sentence	0.118	0.005 (- 95.8)	0.005 (- 95.8)	0.001 (- 99.2)	0.002 (- 98.3)
Bond amount	0.211	0.094 (- 55.5)	0.097 (- 54.0)	0.047 (- 77.7)	0.007 (- 96.7)
Average std. difference	0.351	0.207 (- 40.8)	0.209 (- 40.4)	0.076 (- 78.2)	0.053 (- 85.02)

Bold indicates covariate imbalanced between treated and untreated units. IPTW = inverse probability of treatment weighting; MMW-S = marginal mean weighting through stratification

pretrial services staff is nearly 1.5 times lower ($1/0.69 = 1.45$) for the treatment group in relation to the comparison group. There were no differential effects between the two groups on the remaining indicators of pretrial misconduct after pre-treatment covariates were balanced.

Table 3 Cox regression of pretrial misconduct outcomes by quasi-experimental conditions

Estimation model	FTA court Exp(B) (SE)	FTA meeting Exp(B) (SE)	Rearrest Exp(B) (SE)	Domestic rearrest Exp(B) (SE)
Unconditional ($N_{TG} = 573$; $N_{CG} = 910$)				
Treatment group	1.02 (0.29)	0.71 (0.10)***	1.10 (0.14)	1.15 (0.21)
Model fit indices				
-2LL	504.80	4486.76	2204.53	765.52
χ^2 (df)	0.01 (1)	13.19 (1)***	0.53 (1)	0.41 (1)
Conditional ($N_{TG} = 573$; $N_{CG} = 910$)				
Treatment group	1.79 (0.47)	0.83 (0.12)	1.21 (0.19)	1.26 (0.33)
Model fit indices				
-2LL	467.29	4365.15	2152.83	734.60
χ^2 (df)	39.62 (28)	139.40 (28)***	51.19 (28)**	33.26 (28)
Propensity matched ($N_{TG} = 573$; $N_{CG} = 573$)				
Treatment group	1.16 (0.30)	0.77 (0.10)**	1.86 (0.15)***	1.09 (0.23)
Gamma	2.02	1.26	2.33	1.94
Model fit indices				
-2LL	410.67	3746.06	2628.60	710.07
χ^2 (df)	0.23 (1)	6.30 (1)**	18.20 (1)***	0.16 (1)
Mahalanobis distance ($N_{TG} = 573$; $N_{CG} = 573$)				
Treatment group	1.13 (0.30)	0.80 (0.11)*	1.10 (0.15)	1.06 (0.24)
Gamma	2.02	1.34	2.28	2.19
Model fit indices				
-2LL	410.62	3676.28	2051.22	691.79
χ^2 (df)	0.17 (1)	4.56 (1)*	0.42 (1)	0.05 (1)
Inverse weighted ($N_{TG} = 573$; $N_{CG} = 910$)				
Treatment group	1.02 (0.29)	0.77 (0.10)**	1.08 (0.14)	1.05 (0.21)
Model fit indices				
-2LL	507.80	4478.49	2203.65	762.83
χ^2 (df)	0.01 (2)	21.03 (2)***	1.40 (2)	3.51 (2)
Marginal mean weighted ($N_{TG} = 523$; $N_{CG} = 865$)				
Treatment group	1.01 (0.31)	0.69 (0.10)***	1.15 (0.15)	1.20 (0.22)
Model fit indices				
-2LL	469.53	4070.70	1996.33	726.86
χ^2 (df)	0.76 (2)	17.12 (2)***	1.07 (2)	2.18 (2)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; FTA = failure to appear; N_{TG} = treatment sample size; N_{CG} = comparison sample size

It is important to note that, substantively, the results from the MMW-S counterfactual are largely consistent with the remaining results of the Cox regressions with imbalanced treatment and comparison groups. That is, the direction of the estimated effects all indicate that the treatment group is less at risk to miss a meeting with pretrial services staff but more at risk for failure to appear to court and rearrest. It is also worthy

to point out that the estimated treatment effect sizes, while variable across models, are relatively small.³

Discussion and conclusion

The most logical deployment of GPS technology to monitor IPV offender movement is thought to be during pretrial (Bales et al. 2010), where real-time monitoring and justice system objectives to deter future contact with a victim are well integrated. Previous research has offered some insight into the efficacy of pretrial offender monitoring technology for IPV defendants (Erez et al. 2004; Erez et al. 2012), but failed to provide a clear answer about whether this strategy will improve fundamental principles of pretrial services: to ensure defendants attend court appearances and protect public safety (see Coopriider 2014). This research examined whether pretrial GPS supervision of IPV defendants reduces failures to appear and recidivism.

The results suggest that pretrial GPS supervision reduces failures to appear to meetings with pretrial services. The intensity of this form of pretrial supervision appears to be most effective for case management purposes. From a client–practitioner rapport perspective (see Blasko et al. 2015; Bonta et al. 2011), there is inherent value to this finding. GPS defendants are, by definition, high-risk populations, who, beyond being in jeopardy for pretrial misconduct, also must attend in-person meetings more often than lower risk defendants. As demonstrated here, defendants under GPS supervision are compliant. This gives pretrial officers more opportunities to individualize care and supervision through adjusting case plans, intervention strategies, and referrals.

At the same time, the results may be disappointing to those seeking to invest in pretrial GPS monitoring programs for their IPV populations. Although failing to appear to a meeting with pretrial services is a form of noncompliance, it may not be indicative of an actionable problem behavior in need of mediation. More concerning are issues involving failures to appear to court and recidivism. GPS defendants in this study were no more or less likely to fail to appear to court than those without pretrial GPS supervision. Additionally, there do not appear to be any reductions in arrests for any type of offense or arrests specifically for a domestic violence offense among defendants ordered to pretrial GPS supervision.

The results of this research overlap with the quasi-experimental case studies of Erez et al. (2012). Erez et al. (2012) found that, across sites, defendants supervised with pretrial GPS monitoring were either less likely or were no more or less likely to violate terms. Yet, our results suggest that how pretrial supervision term violations are operationalized matters. Pretrial GPS supervision may reduce the risk for failure to appear to a meeting with pretrial services, but has null effects if pretrial supervision term violations are defined as failure to

³ Although the estimated treatment effect from the propensity matched, Mahalanobis distance, and IPTW comparison groups must be discounted due to their observed imbalances, supplemental models that entered imbalanced covariates in the Cox regression equation produced results similar to those presented in Table 3 (see Table A1 in the supplementary material). It is important to note that, once the imbalances are controlled for, the estimated treatment effect of pretrial GPS supervision on rearrest observed from the propensity matched group is no longer statistically dependable. This finding indicates that imbalances were responsible for the statistically significant estimated coefficient in Table 3 and further reinforces the importance of attending to selection bias issues through the careful construction of comparison groups.

appear to court. Regarding recidivism, the findings of this research deviate from Erez et al. (2012) in two regards. First, we do not find the risk of recidivism during the pretrial period to be higher for GPS supervised defendants. Second, we do not observe a reduction in recidivism for domestic violence offenses. Differences in the pretrial GPS supervision program models may contribute to the former departure, as Erez et al. (2012) generated their results from a site that emphasized crime control strategies. The latter divergence may be an artifact of research designs as the current study examines pretrial supervision activity, not long-term follow-ups, and is more attentive to selection bias issues.

Limitations

A few limitations need to be acknowledged. First, the policies and practices of the study site may not be representative of other pretrial service jurisdictions. The study site is a dedicated pretrial services division that most closely resembles the data-informed due process and punitive model developed by Ibarra et al. (2014), using multiple risk assessment tools, an active GPS monitoring plan, and partners with law enforcement and prosecutors to respond to violations. Large multisite studies that examine the between-group effects of pretrial service programs are needed to determine which type of program model is the most effective.

Second, the results of this study should not be generalized to all pretrial defendants. Bond setting decisions are made by judges and involve subjective risk determinations (Erez et al. 2012). Not all defendants receive supervision; those viewed as being low risk will bond on their own recognizance or have the ability to post a bond amount. The results should only be interpreted in relation to defendants who bond to pretrial supervision. Because of the non-binding nature of supervision recommendations, it is possible that a defendant who may have been ordered to traditional pretrial supervision is instead ordered to GPS supervision and the reciprocal decision occurs. We examined the relationship between recommendations and supervision orders to explore this threat. Recommendations and orders aligned 72% of the time ($n = 1070$) and there were no significant differences between the two samples on orders that deviated from the supervision recommendation [$\chi^2(1, 1483) = 0.44$ $p = 0.51$]. Although the influence of net-widening cannot be completely ruled out, it is highly unlikely that these decisions obstructed the overall findings.⁴

Third, less expansive outcome measures of domestic violence will need to be integrated into future studies. The measure used in this study is based upon an arrest and blends actions that may be qualitatively different from one another (i.e., violating a protection order versus assaulting an intimate partner). While we are able to capture incidents beyond the violation of court orders, we likely miss those events that are not reported to justice officials. A more refined measure of repeat domestic violence that

⁴ Since GPS supervision is the highest level of supervision available at the study site, all of the mismatches between a pretrial service supervision recommendation and a judicial order for the treatment group are based upon a recommendation for non-GPS supervision. Most (57%; 75/154) of these defendants were recommended to intensive supervision, which is the next most intense form of supervision available, follows the same terms as GPS supervision, but does not include GPS monitoring or other forms of technology. Among the unmatched pool of defendants in the comparison group, 117 defendants (45% of the comparison group subsample of mismatched recommendations and orders; 117/259) were recommended to GPS supervision. Eighty percent of those recommended to GPS supervision were instead ordered to intensive supervision.

captures multiple forms of trauma will help provide a more complete understanding of the role of offender monitoring technology.

Implications for policy, practice, and research

Important implications for policy, practice, and research can be derived from this study. With continued momentum towards reforming pretrial decision-making, modernizing bail policies, and reducing pretrial jail populations, GPS monitoring can be viewed as a diversionary strategy to reduce jail and system costs (Ibarra and Erez 2005). This research demonstrates that high-risk offenders should be a part of the dialog on whom to divert. Whether the focus is on the entire sample or only the treatment group, 95% of the sample attended all of their court appearances and 84% avoided a subsequent arrest while under supervision. Diversions from jail through GPS assignment may enable defendants to maintain social ties, continue working, attend school, or pursue other avenues that reduce risk of future deviance (Erez et al. 2012).

Monetary costs of pretrial supervision are substantially lower than jail placements. Using audit reports and legislative testimony, the study site per diem per defendant is estimated to be \$3.00 for pretrial supervision, \$11.00 for pretrial GPS supervision, and \$59.00 to \$76.00 for pretrial detention. Relative to pretrial detention, these simple per diem direct costs of pretrial supervision suggest significant cost savings. For instance, the average length of pretrial supervision for the sample was 127 days (SD = 110.86). Keeping the entire sample ($N = 1483$) in jail would cost between \$11.11 and \$14.31 million. Placing the entire sample on GPS supervision would total \$2.07 million, while traditional pretrial supervision would cost \$565,000.

Yet, the findings from this study indicate that GPS supervision was no more or less effective than traditional supervision. A more refined calculation of direct costs that includes the cost of arrest by local law enforcement indicates that anticipated net savings from simple per diem cost calculations are overestimated (see Table 4). In fact, the defendants under GPS supervision in this study generated four times the amount of costs as compared to defendants who were not placed on GPS supervision and generated more system costs than would be anticipated if pretrial supervision was not ordered. Of course, the elementary estimation of direct per diem costs in cost-benefit analyses are highly controversial (Frank 2000). These calculations miss indirect costs and collateral consequences that are difficult to quantify. This issue of cost-effectiveness will continue to be revisited as the use of GPS monitoring expands.

Regarding future research, this study contributes to ongoing debates about the value of matching and weighting procedures to construct a statistically equivalent comparison group. We observed that the two weighting procedures produced superior covariate balance to the matching procedures, likely because those ordered to GPS supervision were qualitatively different from those who were ordered to traditional pretrial supervision. That is, there were few comparison units available who could match closely with a corresponding treatment unit.

Despite their continued popularity, propensity score matching strategies have been falling out of favor with social scientists. Much of the distaste involves three issues (King and Nielsen 2016). First is the inability to demonstrate common support, where information about the distribution of propensity scores across groups is either withheld or provided as a summary statistic, as opposed to highlighting significant pre-treatment differences, as we do

Table 4 Estimation of direct per diem cost savings

	<i>N</i>	No. of days supervised	Flat cost per arrest	Per diem cost	Total cost
Treatment group					
In community	410	121	–	\$11.00	\$1331.00
Sub-total	410	–	–	–	\$545,710.00
Arrestees	163	86	–	\$11.00	\$946.00
Arrestees	163	–	\$1414	–	\$230,482.00
Sub-total	163	–	–	–	\$37,722,764.00
Total	573	–	–	–	\$38,268,474.00
Per defendant	573	–	–	–	\$66,786.17
Comparison group					
In community	829	111	–	\$3.00	\$333.00
Sub-total	829	–	–	–	\$276,057.00
Arrestees	81	161	–	\$3.00	\$483.00
Arrestees	81	–	\$1414	–	\$114,534.00
Sub-total	81	–	–	–	\$9,316,377.00
Total	910	–	–	–	\$9,592,434.00
Per defendant	910	–	–	–	\$10,541.14
Net savings					
Grand total					– \$28,676,040.00
Per defendant					– \$56,245.03

here. Second is the potential for unknowingly selecting distal comparison units for treatment units, as matching to a scalar propensity score does not guarantee that matches are as close to exact as possible. Third, additional imbalances may be created with the discarding of unmatched treatment or comparison units—known as the propensity score paradox (King and Nielsen 2016). Consistent with this notion and with previous research (Bales and Piquero 2012), we are able to illustrate that the estimated size of the treatment effect tends to be larger when a counterfactual is generated through propensity score matching in relation to the remaining techniques.

In place of the propensity score, other techniques, such as Mahalanobis distance matching, have been offered as viable alternatives (King and Nielsen 2016). However, we observed here that the extreme differences between the groups resulted in Mahalanobis matching producing a poor quality counterfactual. Another criticism of matching procedures is “match shopping”, where covariates are entered, dropped, or transformed in an effort to generate defensible covariate balance, introducing model dependence. This issue was minimized here by utilizing the CBPS score, which is robust to propensity score model misspecifications (Imai and Ratkovic 2014).

Because the weighting procedures are not limited in selected subsets of the comparison group to form a counterfactual, superior covariate balance was achieved by assigning relatively larger weights to particular units. Although this feature of IPTW and MMW-S is beneficial in this context, this resulted in extreme estimated weights. IPTW weights ranged from 0.02 to 10.41 (median = 0.24) and MMW-S weights ranged from 0.16 to 9.51 (median = 0.36). The cumulative weight distributions are displayed in Fig. B1 in the

supplementary material, suggesting that a relatively small proportion of the units contribute disproportionately to the overall weight. However, the design effects for the IPTW and MMW-S weights are 1.91 and 3.52, respectively, suggesting that the application of the weights did not result in a significant inflation or attenuation of the variance of the treatment effects.

Beyond the construction of suitable comparison groups in situations with highly imbalanced pre-treatment units, future research must begin to create an improved understanding of the behavior of prosecutors in pretrial GPS supervision. Defendants ordered to GPS supervision are subject to more intense surveillance and will generate more information that can be made available to prosecutors than defendants ordered to traditional pretrial supervision. It is possible that the volume of GPS information shared with a local prosecutor has artificially increased the reporting of pretrial misconduct, even if the actual number of violations between defendants who are and are not monitored by GPS supervision is the same. Additionally, it is necessary to produce knowledge on how GPS information is acted upon to file formal supervision violation charges, amend pretrial supervision terms, or petition the court to revoke pretrial supervision. The likelihood that prosecutors take action varies by jurisdiction (Erez et al. 2012; Ibarra et al. 2014). It is possible that the behavior of defendants under GPS supervision will vary in accordance with the ability of prosecutors to enforce violations. Future research should aim to capture the information exchange between pretrial services and local prosecutors to examine discretionary decisions and their effect on pretrial misconduct and outcomes.

Funding This research was supported by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice under grant number 2010-IJ-CX-K023. The opinions, conclusions, and recommendations reflect those of the authors and not any aforementioned agency. This research has been conducted in accordance with the National Institute of Justice's requirements for research independence and integrity; the authors have no vested interests in commercial communication technology products, processes, or services.

References

- Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine*, 28, 3083–3107.
- Bales, W., Mann, K., Blomberg, T., Gaes, G., Barrick, K., Dhungana, K., & McManus, B. (2010). *A quantitative and qualitative assessment of electronic monitoring*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Bales, W. D., & Piquero, A. R. (2012). Assessing the impact of imprisonment on recidivism. *Journal of Experimental Criminology*, 8(1), 71–101.
- Blasko, B. L., Friedmann, P. D., Rhodes, A. G., & Taxman, F. S. (2015). The parolee–parole officer relationship as a mediator of criminal justice outcomes. *Criminal Justice and Behavior*, 42(7), 722–740.
- Block, C. R. (2003). *How can practitioners help an abused woman lower her risk of death?* Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Bonta, J., Bourgon, G., Ruge, T., Scott, T. L., Yessine, A. K., Gutierrez, L., & Li, J. (2011). An experimental demonstration of training probation officers in evidence-based community supervision. *Criminal Justice and Behavior*, 38(11), 1127–1148.
- Brown, T. M. L., McCabe, S. A., & Wellford, C. (2007). *Global positioning system (GPS) technology for community supervision: Lessons learned*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Cadigan, T. P. (1991). Electronic monitoring in federal pretrial release. *Federal Probation*, 55(1), 26–30.

- Campbell, J. C., Webster, D., Koziol-McLain, J., Block, C., Campbell, D., Curry, M. A., et al. (2003). Risk factors for femicide in abusive relationships: Results from a multisite case control study. *American Journal of Public Health, 93*(7), 1089–1097.
- Clark, K., & Rydberg, J. (2016). The effect of institutional educational programming on prisoner misconduct. *Criminal Justice Studies, 29*(4), 325–344. <https://doi.org/10.1080/1478601X.2016.1229770>.
- Coopridger, K. (2014). A descriptive analysis of pretrial services at the single-jurisdictional level. *Federal Probation, 78*(3), 9–15.
- Coopridger, K. W., & Kerby, J. (1990). A practical application of electronic monitoring at the pretrial state. *Federal Probation, 54*(1), 28–35.
- Erez, E., Ibarra, P. R., Bales, W. D., & Gur, O. M. (2012). *GPS monitoring technologies and domestic violence: An evaluation study*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Erez, E., & Ibarra, P. R. (2007). Making your home a shelter: Electronic monitoring and victim re-entry in domestic violence cases. *British Journal of Criminology, 47*(1), 100–120.
- Erez, E., Ibarra, P. R., & Gur, O. M. (2013). Using GPS in domestic violence cases: Lessons from a study of pretrial programs. *Journal of Offender Monitoring, 25*(1), 5–10.
- Erez, E., Ibarra, P. R., & Lurie, N. A. (2004). Electronic monitoring of domestic violence cases—A study of two bilateral programs. *Federal Probation, 68*(1), 15–20.
- Fong, C., Ratkovic, M., Hazlett, C., Yang, X., & Imai, K. (2016). CBPS: Covariate Balancing Propensity Score. R package version 0.11. <https://CRAN.R-project.org/package=CBPS>.
- Frank, R. H. (2000). Why is cost-benefit analysis so controversial? *The Journal of Legal Studies, 29*(S2), 913–930.
- Gaes, G. G., Bales, W. D., & Scaggs, S. J. A. (2016). The effect of imprisonment on recidivism: An analysis using exact, coarsened exact, and radius matching with the propensity score. *Journal of Experimental Criminology, 12*(1), 143–158.
- Gies, S. V., Gainey, R., Cohen, M. I., Healy, E., Yeide, M., Bekelman, A., & Bobnis, A. (2013). *Monitoring high-risk gang offenders with GPS technology: An evaluation of the California supervision program final report*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Goldkamp, J. S., & White, M. D. (2006). Restoring accountability in pretrial release: The Philadelphia pretrial release supervision experiments. *Journal of Experimental Criminology, 2*, 143–181.
- Gu, X. S., & Rosenbaum, P. R. (1993). Comparison of multivariate matching methods: Structures, distances, and algorithms. *Journal of Computational and Graphical Statistics, 2*(4), 405–420.
- Gur, O. M., Ibarra, P. R., & Erez, E. (2016). Specialization and use of the GPS for domestic violence by pretrial programs: Findings from a national survey of U.S. practitioners. *Journal of Technology in Human Services, 34*(1), 32–62.
- Han, E. L. (2003). Mandatory arrest and no-drop policies: Victim empowerment in domestic violence. *Boston College Third World Law Journal, 23*(1), 159–191.
- Hong, G., & Hong, Y. (2009). Reading instruction time and homogeneous grouping in kindergarten: An application of marginal mean weighting through stratification. *Educational Evaluation and Policy Analysis, 31*(1), 54–81.
- Iacus, S. M., King, G., & Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1–24.
- Ibarra, P. R. (2005). Red flags and trigger control: The role of human supervision in an electronic monitoring program. In S. L. Burns (Ed.), *Ethnographies of law and social control (sociology of crime, law and deviance)* (vol. 6, pp. 31–48). Bingley: Emerald.
- Ibarra, P. R., & Erez, E. (2005). Victim-centric diversion? The electronic monitoring of domestic violence cases. *Behavioral Sciences and the Law, 23*(2), 259–276.
- Ibarra, P. R., Gur, O. M., & Erez, E. (2014). Surveillance as casework: Supervising domestic violence defendants with GPS technology. *Crime, Law, and Social Change, 62*(4), 417–444.
- Imai, K., & Ratkovic, M. (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 76*(1), 243–263.
- Keele, L. J. (2011). An overview of rbounds: An R package for Rosenbaum bounds sensitivity analysis with matched data. Ohio State University typescript.
- King, G., & Nielsen, R. (2016). *Why propensity scores should not be used for matching*. Unpublished manuscript, Institute for Quantitative Social Science, Harvard University, Cambridge.
- Logan, T. K., Shannon, L., Walker, R., & Faragher, T. M. (2006). Protective orders: Questions and conundrums. *Trauma, Violence, & Abuse, 7*(3), 175–205.
- Logan, T. K., & Walker, R. (2009). Civil protective order outcomes: Violations and perceptions of effectiveness. *Journal of Interpersonal Violence, 24*(4), 675–692.

- Loughran, T. A., Wilson, T., Nagin, D. S., & Piquero, A. R. (2015). Evolutionary regression? Assessing the problem of hidden biases in criminal justice applications using propensity scores. *Journal of Experimental Criminology*, *11*(4), 631–652.
- Mahoney, M. R. (1991). Legal images of battered women: Redefining the issue of separation. *Michigan Law Review*, *90*(1), 1–94.
- Maxfield, M. G., & Baumer, T. L. (1992). Pretrial home detention with electronic monitoring: A nonexperimental salvage evaluation. *Evaluation Review*, *16*(3), 315–332.
- Nagin, D. S., Cullen, F. T., & Jonson, C. L. (2009). Imprisonment and reoffending. In M. Tonry (Ed.), *Crime and justice: A review of research* (pp. 115–200). Chicago: University of Chicago Press.
- National Institute of Justice. (2015). *Intimate Partner Violence*. U.S. Department of Justice. <https://www.nij.gov/topics/crime/intimate-partner-violence/pages/welcome.aspx>
- Padgett, K. G., Bales, W. D., & Blomberg, T. G. (2006). Under surveillance: An empirical test of the effectiveness and consequences of electronic monitoring. *Criminology & Public Policy*, *5*(1), 61–91.
- Renzema, M., & Mayo-Wilson, E. (2005). Can electronic monitoring reduce crime for moderate to high-risk offenders? *Journal of Experimental Criminology*, *1*(2), 215–237.
- Salvemini, A. V., Piza, E. L., Carter, J. G., Grommon, E. L., & Merritt, N. (2015). Integrating human factors engineering and information processing approaches to facilitate evaluations in criminal justice technology research. *Evaluation Review*, *39*(3), 308–338.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Belmont, CA: Wadsworth.
- Sherman, L. W., & Berk, R. A. (1984). The specific deterrent effects of arrest for domestic assault. *American Sociological Review*, *49*(2), 261–272.
- Smith, J. A., & Todd, P. E. (2005). Does matching overcome LaLonde's critique of nonexperimental estimators? *Journal of Econometrics*, *125*(1–2), 305–353.
- Spitzberg, B. H. (2002). The tactical topography of stalking victimization and management. *Trauma, Violence, & Abuse*, *3*(4), 261–288.
- Tjaden, P., & Thoennes, N. (2000). *Extent, nature, and consequences of intimate partner violence*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Turner, S., Chamberlain, A. W., Jannetta, J., & Hess, J. (2015). Does GPS improve recidivism among high risk sex offenders? Outcomes for California's GPS pilot for high risk sex offender parolees. *Victims & Offenders*, *10*(1), 1–28. <https://doi.org/10.1080/15564886.2014.953236>.
- Visher, C. A., Lattimore, P. K., Barrick, K., & Tueller, S. (2017). Evaluating the long-term effects of prisoner reentry services on recidivism: What types of services matter? *Justice Quarterly*, *34*(1), 136–165. <https://doi.org/10.1080/07418825.2015.1115539>.

Eric Grommon is an Assistant Professor within the School of Public and Environmental Affairs at Indiana University–Purdue University Indianapolis. His research interests include corrections, prisoner reentry, and the evaluation of criminal justice programs, policies, and operations. His research has been published in *Criminology and Public Policy*, *Evaluation Review*, *Journal of Experimental Criminology*, *Journal of Quantitative Criminology*, and *Justice Quarterly*.

Jason Rydberg is an Assistant Professor in the School of Criminology and Criminal Justice Studies at the University of Massachusetts Lowell, where he is also an associate with the Center for Program Evaluation. He obtained his Ph.D. in Criminal Justice from Michigan State University. His research interests include prisoner reentry, sex offender policy, and the evaluation of criminal justice programming. His research has recently appeared in *Criminology and Public Policy*, *Journal of Criminal Justice*, *Journal of Quantitative Criminology*, and *Justice Quarterly*.

Jeremy G. Carter is an Associate Professor within the School of Public and Environmental Affairs at Indiana University–Purdue University Indianapolis. His research areas include policing, policy evaluation, and justice technologies. His research has appeared in outlets such as *Criminal Justice and Behavior*, *Journal of Criminal Justice*, *Journal of Interpersonal Violence*, *Homicide Studies*, *Evaluation Review*, *Police Quarterly*, *Policing and Society*, *Journal of Crime and Justice*, *Criminal Justice Policy Review*, *Security Journal*, and *Police Practice and Research*.