ASSESSING THE EFFECTIVENESS OF PRETRIAL SPECIAL CONDITIONS

Full Findings from the Pretrial Justice Collaborative

Chloe Anderson, Erin Valentine, and Daron Holman

JUNE 2023

mdrc
BUILDING KNOWLEDGE TO IMPROVE SOCIAL POLICY

FUNDERS

This research was supported by Arnold Ventures.

Dissemination of MDRC publications is supported by the following organizations and individuals that help finance MDRC's public policy outreach and expanding efforts to communicate the results and implications of our work to policymakers, practitioners, and others: The Annie E. Casey Foundation, Arnold Ventures, Charles and Lynn Schusterman Family Foundation, The Edna McConnell Clark Foundation, Ford Foundation, The George Gund Foundation, Daniel and Corinne Goldman, The Harry and Jeanette Weinberg Foundation, Inc., The JPB Foundation, The Joyce Foundation, The Kresge Foundation, and Sandler Foundation.

In addition, earnings from the MDRC Endowment help sustain our dissemination efforts. Contributors to the MDRC Endowment include Alcoa Foundation, The Ambrose Monell Foundation, Anheuser-Busch Foundation, Bristol-Myers Squibb Foundation, Charles Stewart Mott Foundation, Ford Foundation, The George Gund Foundation, The Grable Foundation, The Lizabeth and Frank Newman Charitable Foundation, The New York Times Company Foundation, Jan Nicholson, Paul H. O'Neill Charitable Foundation, John S. Reed, Sandler Foundation, and The Stupski Family Fund, as well as other individual contributors.

The findings and conclusions in this report do not necessarily represent the official positions or policies of the funders.

For information about MDRC and copies of our publications, see our website: www.mdrc.org.

OVERVIEW

As more jurisdictions across the country are seeking to reduce their jail populations, many view electronic monitoring (EM, the use of an electronic device to monitor a person's movement and location) and sobriety monitoring (regular drug and alcohol testing) as potential alternatives to pretrial detention. In theory, the added layer of supervision that these special conditions provide should encourage people to appear for court dates and avoid activities that could lead to new arrests. Yet most studies of the effectiveness of special conditions have faced methodological limitations and have yielded mixed findings. Furthermore, special conditions such as electronic monitoring and sobriety monitoring carry significant costs—both personal and monetary—for those being monitored and for jurisdictions.

This report contributes cross-jurisdiction evidence on the effects of these special conditions of release using retrospective data from cases initiated between January 2017 and June 2019 in four diverse jurisdictions across the United States: one small and rural, one medium-sized, and two large and urban jurisdictions. The MDRC research team employed a propensity score matching design to test the effectiveness of EM and sobriety monitoring in maintaining clients' court appearance rates and helping them avoid arrest. This method allowed the team to compare court appearance and pretrial rearrest outcomes for individuals released with special conditions with those of statistically comparable individuals who were released without special conditions. The analysis uses a noninferiority approach, which tests whether release without special conditions is at least as effective as (that is, no worse than) release with a special condition.

The analysis found that:

- **Being released on EM or sobriety monitoring did not significantly improve court appearance rates.** The analyses found that the special conditions and non–special conditions groups had similar pretrial court appearance rates. These results were consistent across jurisdictions.
- **Being released on electronic monitoring did not significantly increase the percentage of people who avoided a new arrest during the pretrial period.** In fact, the analysis found that the EM group had a higher pretrial rearrest rate than the non-EM group, a result that was consistent across the two jurisdictions in that analysis. While the factors causing the results are not definitively known, the difference may be a supervision effect: people may be more likely to be arrested if their actions are more closely monitored, compared with others who are less closely monitored. Alternatively, the result may reflect unmeasured differences between the EM and non-EM groups that could not be controlled for in the analysis.
- **Being released on sobriety monitoring did not significantly improve the percentage of people who avoided a new arrest, but there was variation in this effect among jurisdictions.** In two of the four jurisdictions studied, people who were assigned to sobriety monitoring were more likely to avoid new arrests, while in the other two, the result was the opposite.

These findings warrant cautious reflection among policymakers and practitioners on the extent of current electronic and sobriety monitoring use, particularly considering their high personal and financial costs to those directly affected and to jurisdictions. The exploratory findings also highlight a need for additional cross-site studies—in particular, those that employ more rigorous experimental methods—on the effectiveness of special conditions at the pretrial stage. Given the site variation in findings, particularly for sobriety monitoring, more research is also needed to delineate the populations that would benefit from special conditions from those who would not benefit and to illuminate the policies and practices that are associated with the greatest success.

CONTENTS

LIST OF EXHIBITS

TABLE

FIGURE

ACKNOWLEDGMENTS

The Pretrial Justice Collaborative (PJC) is a project led by MDRC's Center for Criminal Justice Research, in partnership with Justice System Partners, and supported by Arnold Ventures. The research team would like to extend thanks to Arnold Ventures for its facilitation of this research, and in particular Kristin Bechtel for her ongoing project guidance and review of draft materials.

The goals of PJC are to help eight participating jurisdictions across the country assess the performance of their pretrial release interventions—with a focus on court appearance and arrest-avoidance outcomes—and to identify the least burdensome pretrial supervision interventions necessary to maintain positive outcomes. The research team is grateful to the PJC jurisdictions for their continued partnership in carrying out this research, and specifically the pretrial service departments and state agencies and county agencies that provided the data that made the analyses presented in this brief possible.

The publication of this brief was made possible with the support of many individuals at MDRC. Sarah Picard, Lily Freedman, Melanie Skemer, Emily Brennan, and Bret Barden reviewed and provided valuable comments on drafts of the brief. Daron Holman, Emily Brennan, Eli Miller, Olivia Williams, and Rachel Sander processed the raw data files that fed into the analyses presented in this brief. We also thank Jose Morales and Jayce Helpley for assistance with report coordination; Joshua Malbin for editing; and Carolyn Thomas for preparing the report for publication.

The Authors

As jurisdictions across the country seek to reduce their jail populations by decreasing the number of people who are held while awaiting trial, many are turning to special conditions of pretrial release such as electronic monitoring (EM) and sobriety monitoring as potential alternatives to pretrial detention.[1] These forms of monitoring are often perceived as offering peace of mind to decision-makers and the public by upholding court appearance rates and protecting public safety. Yet the evaluation literature on the effectiveness of EM and sobriety monitoring is often location-specific, dated, and rarely focused on pretrial applications (particularly for EM), and has generally yielded weak or inconsistent results.[2]

Given the high costs of such monitoring—both to jurisdictions and to those monitored—it is important to understand whether these methods are truly effective in helping individuals meet their pretrial conditions.

This report contributes cross-jurisdiction evidence on the effects of these special conditions of release using retrospective data from cases initiated between January 2017 and June 2019 in four diverse jurisdictions across the United States.[3]

- Site A: a large, urban metropolitan area in the western United States
- Site B: a small, rural county from the same geographic region as Site A
- Site C: a large, urban metropolitan area in the southern United States
- Site D: a medium-sized metropolitan area in the central United States

The MDRC research team employed a propensity score matching design (explained below) to test the effectiveness of EM and sobriety monitoring in maintaining clients' court appearance rates and helping them avoid arrest.[4] The analysis uses a noninferiority approach,

1. In this report, electronic monitoring (EM) refers to the use of an electronic device—often in the form of a bracelet fitted to the ankle or wrist—to monitor an individual's movement and location using GPS or cellular location tracking technology. Sobriety monitoring refers to regularly testing for drug or alcohol use, sometimes with a device one wears.
2. Belur et al. (2020); Hatton and Smith (2020).
3. These four jurisdictions, along with four other jurisdictions not featured in this report, have partnered with MDRC's Center for Criminal Justice Research and Justice System Partners, with support from Arnold Ventures, to form the Pretrial Justice Collaborative (PJC). The goal of the Collaborative is to build and disseminate reliable, usable evidence about the most effective strategies for reducing pretrial detention, minimizing conditions of supervision while cases are adjudicated, and reducing racial and economic disparities, while maintaining court appearance rates and public safety. For more information, see: https://www.mdrc.org/project/pretrial-justice-collaborative#overview.
4. Sensitivity analyses (described in greater detail in the Design and Methods section below) suggest the results described were not sensitive to particular matching decisions. However, it is always possible with propensity score matching that the groups being compared could differ with respect to unmeasured characteristics. Therefore, the results of this analysis should be considered exploratory—that is, they should not be considered as definitive as results from rigorous types of analyses such as randomized controlled trials. Readers should approach the findings and interpretation of findings with caution.

which tests whether release without special conditions is at least as effective as (that is, no worse than) release with a special condition. Appendix A provides more details about the analyses as well as additional tables and figures.

BACKGROUND AND POLICY CONTEXT

EM and sobriety monitoring are used in all U.S. states to monitor individuals on pretrial release, probation, and parole.[5] On any given day there are more than 100,000 people on EM across the country, and this number has probably increased since the COVID-19 pandemic.[6] There is no definitive nationwide estimate of the number of people on electronic and sobriety monitoring during the pretrial period specifically, but one nationwide survey of the nation's largest counties found that 71 percent of counties with some form of pretrial services used pretrial electronic monitoring and 77 percent used pretrial drug testing, indicating extensive use.[7] In fact, individuals on pretrial release are thought to have comprised an increasing share of the EM population since EM's first pretrial application in the 1980s.[8] Recent data from local jurisdictions indicates that pretrial EM use has increased further in recent years as jurisdictions have moved to reduce their pretrial jail populations, including in response to the pandemic.[9] Pretrial sobriety monitoring programs have similarly proliferated widely since their first appearance in the 1970s.[10] The MDRC research team heard anecdotally from a number of jurisdictions that they briefly suspended in-person drug and alcohol testing in response to the COVID-19 pandemic.

Despite the widespread use of these special conditions during the pretrial phase, the evaluation literature is far from definitive as to the effectiveness of pretrial electronic and sobriety monitoring in upholding court appearance rates and helping individuals avoid arrest. The literature on electronic monitoring is mostly focused on its use in probation or parole, with little evaluation literature on its use during the pretrial period.[11] The research that does exist is largely mixed as to EM's effectiveness, suggesting that EM does not guarantee improvements in pretrial outcomes.[12] Furthermore, assigning EM to low-risk individuals has been linked to *higher* rates of failure to appear at court hearings and pretrial rearrest.[13] The evaluation literature on the effectiveness of pretrial sobriety monitoring is similarly mixed,

5. Weisburd et al. (2021).
6. Weisburd et al. (2021); Pew Charitable Trusts (2016).
7. Pretrial Justice Institute (2019).
8. Advancing Pretrial Policy and Research (2021b).
9. Advancing Pretrial Policy and Research (2021b); Weisburd and Virani (2022); Hager (2020).
10. Advancing Pretrial Policy and Research (2021a).
11. Advancing Pretrial Policy and Research (2021b).
12. Advancing Pretrial Policy and Research (2021b); Belur et al. (2020); Cooprider and Kerby (1990); Cadigan (1991); Grommon, Rydberg, and Carter (2017); Hatton and Smith (2020); Sainju et al. (2018); VanNostrand and Keebler (2009); Wolff et al. (2017).
13. VanNostrand and Keebler (2009).

with inconsistent findings across studies.[14] From a theoretical perspective, it is perhaps even less clear why one would expect sobriety and electronic monitoring to improve court appearance and rearrest rates, since they were not designed to do so; rather, in the case of EM these devices were designed to simply track a person's movements and alert the pretrial supervision staff if that person violates any geographic constraints ordered by the court. In the case of sobriety monitoring, the intent was to alert the pretrial supervision staff if a person violated a court order to abstain from drug or alcohol use during the pretrial period.

Furthermore, pretrial electronic monitoring and sobriety monitoring programs come with great costs—both financial and personal—to those being monitored and to jurisdictions.[15] It is not uncommon for EM and alcohol monitoring devices to cost over $5,000 per year for each monitored individual, borne by the jurisdiction, the monitored individual, or both.[16] Drug testing is also expensive—easily costing individuals $60 out of pocket per week and jurisdictions hundreds of thousands of dollars annually.[17] Personal costs are also high and include: cumbersome preapproval requirements to do necessary things such as grocery shopping or picking up children from school, in the case of EM; requirements for charging electronic and alcohol-monitoring devices, which can take hours of a person's time each day; challenges in meeting work and childcare obligations in the case of drug testing, which is announced on short notice and often only available during business hours; privacy concerns; strains on family ties; and an increased chance of being reincarcerated due to technical violations.[18]

ELECTRONIC AND SOBRIETY MONITORING AT THE STUDY SITES

Across all sites, judges used their discretion to assign people to special conditions of pretrial release, including electronic and sobriety monitoring, following an arrest. These special conditions could be assigned with or without other conditions of release, such as pretrial supervision or money bond. All jurisdictions also used a pretrial risk-assessment tool to guide their release decisions; the tool assessed individuals' risks of failing to appear in court and of being rearrested if they were released during the pretrial period. However, there was variation among jurisdictions in how explicitly they used risk scores to assign special conditions (see below). The analyses described in this report on the effectiveness of electronic monitoring were conducted at Sites A and C, and the analyses on the effectiveness

14. Advancing Pretrial Policy and Research (2021a); Britt, Gottfredson, and Goldkamp (1992); Goldkamp and Jones (1992); Hatton and Smith (2020); VanNostrand and Keebler (2009).
15. Arnett (2019); Weisburd et al. (2021).
16. Weisburd et al. (2021); McKnight, Fell, and Auld-Owens (2012). For more details on the cost of transdermal alcohol monitoring devices, see https://www.scramnorthcarolina.com/scram-resources.
17. Cornett (2022); Hatton and Smith (2020).
18. Weisburd et al. (2021); Cornett (2022).

of sobriety monitoring were conducted at all four sites.[19] This section summarizes the use of these special conditions across the relevant sites in the study.

Electronic Monitoring

Table 1 highlights characteristics of each jurisdiction's electronic monitoring practices. Compared with those who were not assigned to electronic monitoring, those assigned to electronic monitoring generally tended to have higher risk scores and more serious charges. At both sites, case managers within each jurisdiction were responsible for monitoring and responding to violations. The Site C jurisdiction additionally engaged a private vendor for some operations.

TABLE 1

Characteristics of Electronic Monitoring (EM), by Site

	Site A	Site C
Information available to judges for EM placement decisions	A recommendation based on the assessed risk score, as well as other information about the case (current charges and criminal history)	Only other existing information about a case: current charges, raw risk score (with no accompanying EM recommendation), and criminal history
Technology	GPS-enabled ankle bracelet	GPS-enabled ankle bracelet

Sobriety Monitoring

Table 2 summarizes features of each jurisdiction's sobriety-monitoring practices. While the jurisdictions in Sites B to D did not use an explicitly risk-based process like Site A to place individuals on sobriety monitoring, many had established practices related to sobriety monitoring based on a person's current charges (for example, driving under the influence or drug charges) and criminal history. Because all four jurisdictions had some type of charge- or risk-based process for assigning people to sobriety monitoring, those assigned to sobriety monitoring generally tended to have more serious charges (that is, felony or violent charges) or higher risk scores than those who were not assigned to sobriety monitoring. As was the case for electronic monitoring practices, case managers in each jurisdiction were responsible

19. Use of EM at Sites B and D was low, and as a result, there was not sufficient sample to assess the effects of EM at those sites.

TABLE 2

Characteristics of Sobriety Monitoring (SM), by Site

	Site A	Site B	Site C	Site D
Information available to judges for SM placement decisions	A recommendation based on the assessed risk score, as well as other information about case (current charges and criminal history)	Only other existing information about a case: current charges, raw risk score (with no accompanying SM recommendation), and criminal history	Only other existing information about a case: current charges, raw risk score (with no accompanying SM recommendation), and criminal history	Only other existing information about a case: current charges, raw risk score (with no accompanying SM recommendation), and criminal history
Technology	Urine analysis Transdermal alcohol-monitoring device	Urine analysis Transdermal alcohol-monitoring device Ignition interlock device	Urine analysis Ignition interlock Portable alcohol-monitoring device Transdermal alcohol-monitoring device	Urine analysis Portable breathalyzer

NOTE: A transdermal alcohol-monitoring device is a device—typically in the form of a bracelet—that continuously monitors an individual's alcohol consumption through perspiration on the surface of the skin. An ignition interlock device is a breathalyzer that is installed in one's car, preventing the driver from starting the car before passing a breathalyzer test.

for monitoring test results and responding to positive test results or other issues. Some jurisdictions also worked with private vendors to help test or fit devices.

Costs

The financial costs of these special conditions were substantial across all sites, and costs were often shared between the jurisdiction and the monitored individual. For example, EM was estimated to cost one jurisdiction roughly $12/day per monitored individual, which could add up to over $80,000 per month for its full EM caseload. In other jurisdictions the monetary costs of EM and sobriety monitoring were often just as high and borne at least partially (if not fully) by the monitored individual. At all sites, assignment to EM or sobriety monitoring often increased the number of times people had to come in contact with supervision staff members, due to device maintenance or follow-up from any violations or positive test results—ratcheting up personal and financial costs even further.

Trends in Use Over Time

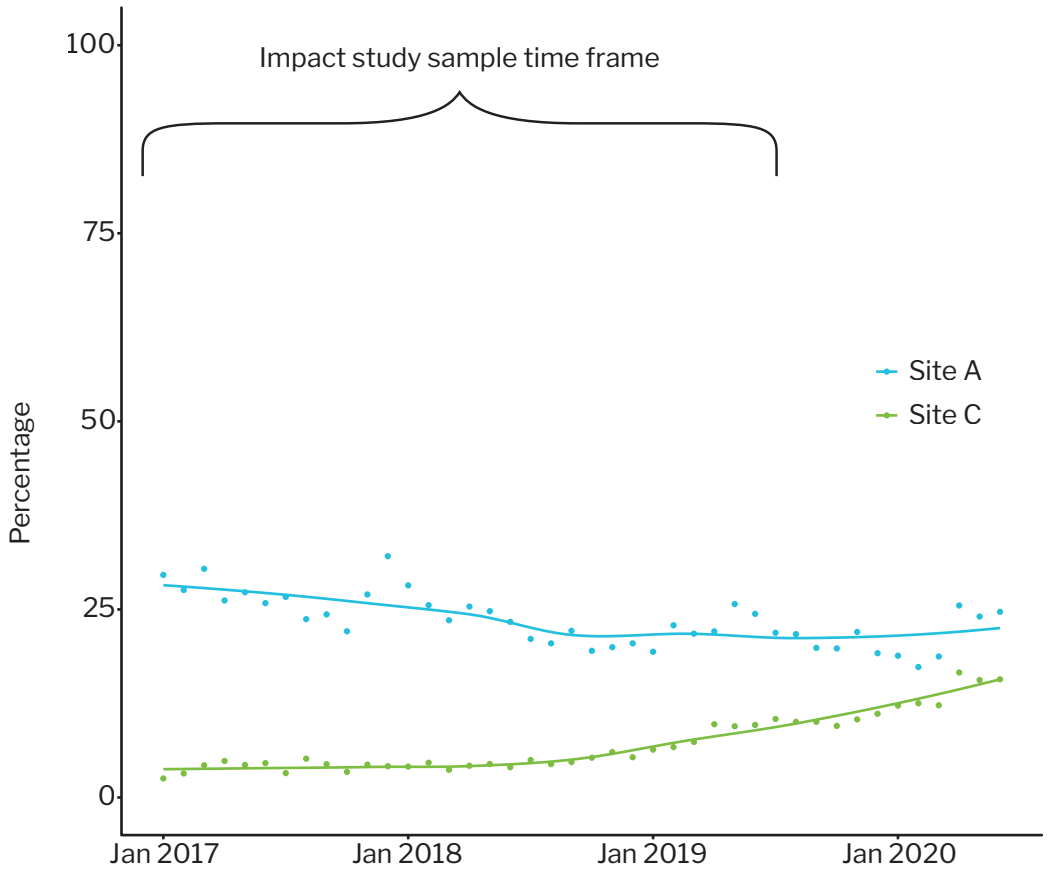
Figure 1 and Figure 2 present trends in the use of electronic and sobriety monitoring for the sites included in each analysis between January 2017 and June 2020.[20] The graphs show the percentage of new cases assigned to special conditions as a share of all supervised cases. With the exception of Site C, which substantially increased its use of both electronic and sobriety monitoring over time, jurisdictions' use of these pretrial special conditions was broadly decreasing or stable across the time period shown. Figure 1 only contains a few months of data after jurisdictions first responded to the COVID-19 pandemic in March 2020, but those months of data generally support what the research team heard from site contacts: that the use of electronic monitoring increased in the months following the pandemic's onset. While jurisdictions continued assigning individuals to sobriety monitoring during this time, as is reflected in the graphs, information that the research team gathered from site contacts indicates that most suspended in-person testing for at least several months at the onset of the pandemic.

Jurisdictions generally perceived electronic and sobriety monitoring as positive alternatives to pretrial detention, yet some also expressed concern over their high costs, unknown efficacy, and potential for "net-widening"—that is, that they might unintentionally lead to more people being controlled by the justice system than would be the case in their absence. All wondered whether court appearance rates and avoidance of arrest could be maintained through even less restrictive means. The sections that follow present findings on the estimated effects of electronic and sobriety monitoring relative to less restrictive alternatives.

20. These graphs include one year more than the time frame for the impact study sample, which goes through June 2019. That sample's timespan is discussed more in the Design and Methods section.

FIGURE 1

Percentage of Supervised Cases on Electronic Monitoring Over Time



SOURCES: Court and pretrial services data from participating sites.

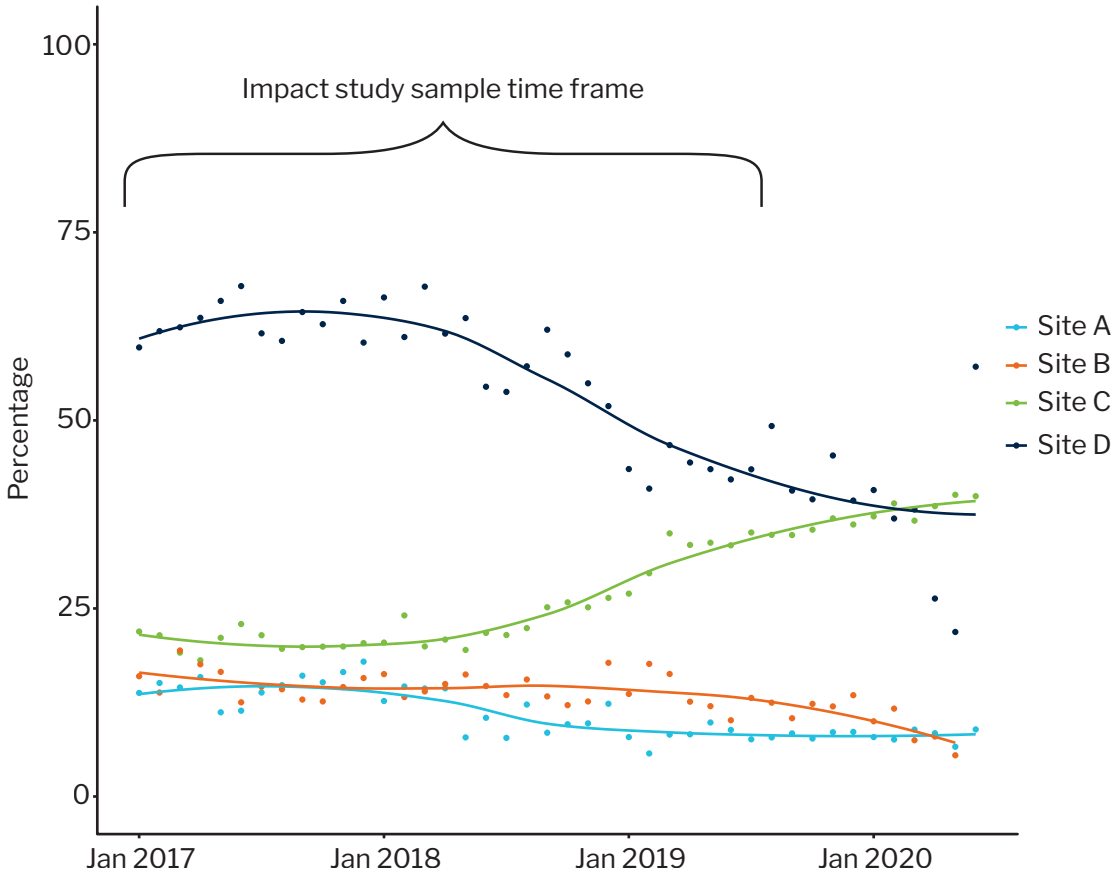
DESIGN AND METHODS

The research team used a propensity score matching design to examine the effectiveness of special conditions in encouraging people to make their court appearances and avoid new arrests. Data from all four sites were used in separate analyses of electronic monitoring and sobriety monitoring. The analyses use retrospective data from cases initiated at the sites between January 2017 and June 2019.[21] The analyses compared outcomes in the six

21. The sample time frame ended in June 2019 in order to allow for at least six months of follow-up data before the start of the COVID-19 pandemic (that is, the follow-up period went through December 2019). Figures 1 and 2 show descriptive trends in electronic and sobriety monitoring through June 2020 for more context.

FIGURE 2

Percentage of Supervised Cases on Sobriety Monitoring Over Time



SOURCES: Court and pretrial services data from participating sites.

months after an arrest for cases that were assigned to EM/sobriety monitoring with those for matched cases that were not assigned to EM/sobriety monitoring. These exploratory analyses were part of a larger set of comparative effectiveness analyses conducted by the MDRC research team.[22]

22. As part of these analyses, the research team also used a regression discontinuity design to assess the effects of electronic monitoring on court appearance and avoidance of arrest; however, that analysis did not provide a strong test of EM, as there was little difference between the EM and non-EM groups in the actual assignment of EM. The results from that analysis are available in full in the technical appendix to Valentine and Picard (2023). For more information about the comparative effectiveness analyses conducted by MDRC as a part of this series, including more technical details about the propensity score matching analyses presented in this report and findings from an analysis of the effectiveness of varying intensities of pretrial supervision, see also Appendix A of this report and the main body text of Valentine and Picard (2023).

In the propensity score matching analyses, individuals who were assigned to EM or to sobriety monitoring were matched, statistically, to individuals who were not assigned to the special condition, but who had very similar characteristics at the time of the initial arrest. Specifically, the research team performed a series of logistic regressions to estimate the likelihood of a case being assigned to the special condition given a set of case characteristics such as age, race, assessed risk score, and severity of charge. The estimated coefficients from this model were then used to create a propensity score (that is, the predicted probability of receiving EM or sobriety monitoring) for each case. Each case from the EM or sobriety monitoring group was then matched to a case from the comparison pool that had a very similar propensity score.[23]

The matching was done separately by special condition and by site. For example, those assigned to electronic monitoring at Site A were matched to cases in a comparison sample from Site A. Separately, those assigned to sobriety monitoring at Site A were matched to cases in a comparison sample from that site, and so on. Once the matched analysis samples were created, the data were pooled across sites to create one EM analysis sample and one sobriety-monitoring analysis sample. Outcomes for the matched EM/sobriety-monitoring groups were then compared with outcomes for their matched comparison groups.[24] The EM analysis sample included about 2,300 cases from Site A for the court appearance analysis and 7,100 cases from Sites A and C for the arrest analysis.[25] The sobriety-monitoring analysis sample included about 33,000 cases from Sites A, B, and D for the court appearance outcome and about 54,000 cases from all four sites for the arrest outcome.

RESULTS

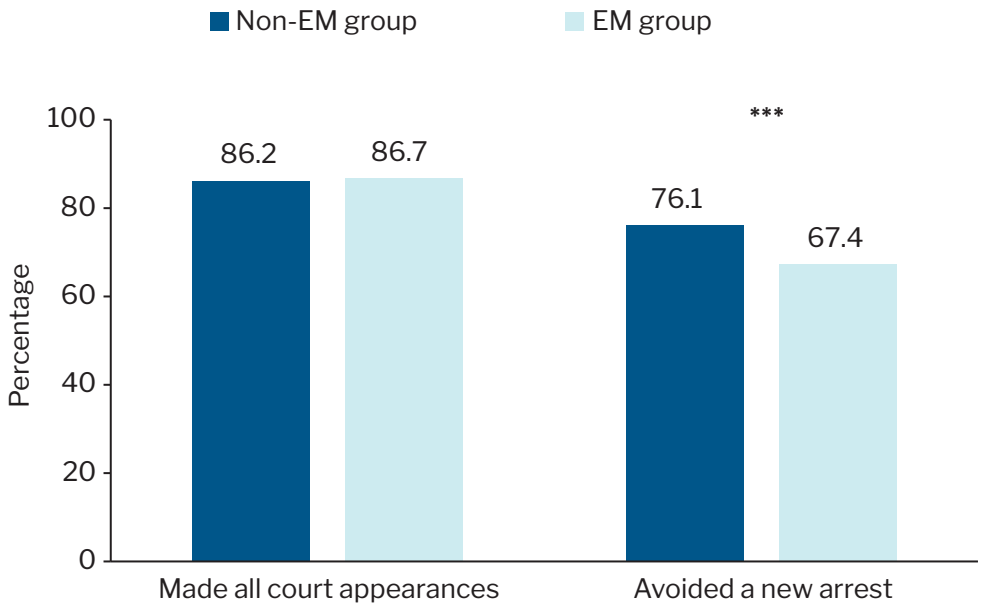
Electronic Monitoring

The results from the electronic monitoring analysis are shown in Figure 3. The results indicate that there was no statistically significant difference between groups in court appearance rates. In other words, electronic monitoring did not improve court appearance rates.

23. Specifically, the research team used one-to-one matching without replacement with a caliper of 0.2.
24. Matching diagnostics suggest that the propensity score matching method yielded statistically similar matched comparison groups. Sensitivity analyses that included variations in the matching model, matching with replacement, different caliper sizes, and equal weighting of sites produced substantively similar findings. This consistency suggests that the results were not sensitive to particular matching decisions. However, as is always the case with propensity score matching, it is possible that the groups differed with respect to characteristics that were not measured and therefore could not be accounted for in the matching process. Therefore, the results of this analysis should be considered exploratory.
25. Data on court appearances were not available from Site C.

FIGURE 3

Electronic Monitoring (EM) Effects



SOURCES: Court and pretrial services data from participating sites.

NOTE: The three asterisks (***) above indicate statistical significance at $p < 0.001$.

The results in Figure 3 also show that **being released on electronic monitoring did not significantly increase the percentage of people who avoided a new arrest during the pretrial period.** In fact, those released *without* EM did better at avoiding arrest, at a rate of about 76 percent, compared with about 67 percent among the EM group. In other words, the EM group was more likely to be rearrested during the pretrial period. This result was consistent across the two sites included in the analysis. It is not clear why the analysis suggests that people on EM were rearrested at a higher rate, but this finding aligns with previous research that has found that EM can sometimes lead to higher rates of arrest (as well as failures to appear in court).[26] One plausible explanation supported by prior research is that the EM group was subject to more restrictive conditions related to their electronic monitoring and thereby had a higher chance of being arrested for technical violations of these conditions.[27] The result also could be due to more intensive surveillance of the EM group resulting in more observed crimes (among those on EM compared with those released without EM). Another possible explanation is that individuals on EM are more likely to be arrested due to unmeasured differences between the EM and comparison samples that could not be accounted for using propensity score matching, such as differences in clinical needs or financial stability.

26. VanNostrand and Keebler (2009).
27. Advancing Pretrial Policy and Research (2021b); Cooprider and Kerby (1990); Sainju et al. (2018); Wolff et al. (2017).

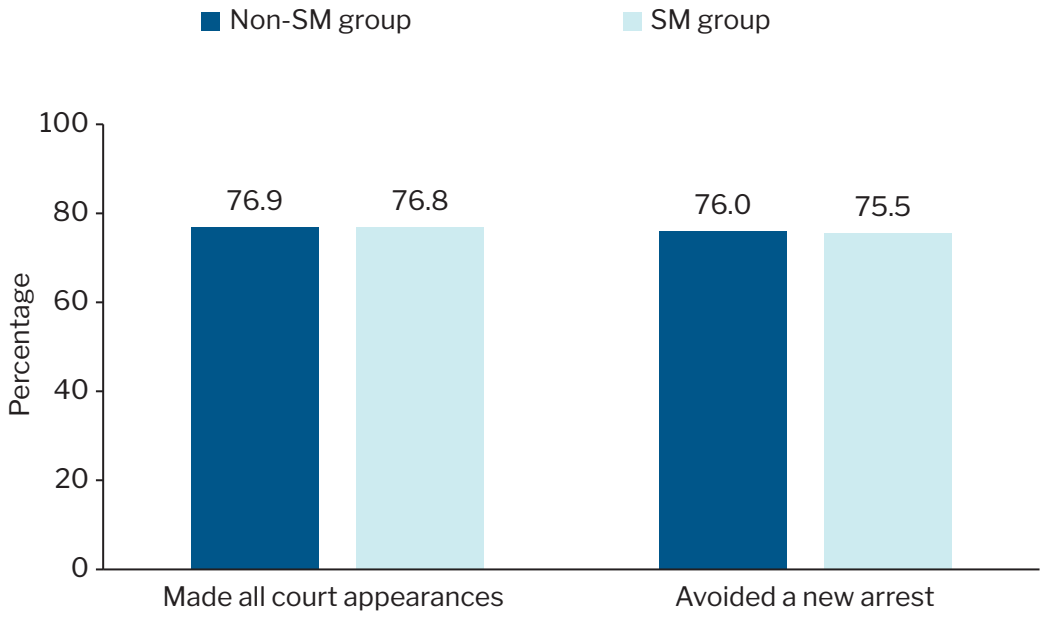
Sobriety Monitoring

The sobriety monitoring analysis indicates that there was no significant effect of sobriety monitoring on either of the outcomes of interest among the pooled sample. The results are shown in Figure 4. About 77 percent of the individuals in both the sobriety-monitoring group and the non-sobriety-monitoring group made all court appearances in the six months after arrest. Similarly, about three-quarters of both groups avoided a new arrest.

Site-specific results (not shown) are consistent when examining the effects of sobriety monitoring on court appearance rates; there was no significant difference between groups with respect to that measure for any of the three sites included in the analysis.

FIGURE 4

Sobriety Monitoring (SM) Effects



SOURCES: Court and pretrial services data from participating sites.

NOTE: None of the differences in the figure are statistically significant at the $p < 0.05$ level.

However, there was some variation across the four sites in the effects of sobriety monitoring on rearrest rates. At two sites, the sobriety-monitoring group was more likely to avoid arrest, while at the other two sites, the result was the opposite. It is not clear why these results occurred, but they are consistent with existing research, which has found both positive and negative effects of sobriety monitoring across different sites, samples,

and research methods.[28] One possibility is that the reach of sobriety monitoring at the sites influenced the estimated effects. The two sites where sobriety monitoring was correlated with higher arrest rates (Sites C and D) used the practice among a relatively large percentage of the people who were released while awaiting trial; the effect was especially large at Site D, which placed the highest percentage of people on pretrial sobriety monitoring (see Figure 2 above). In other words, it is possible that smaller-scope, targeted use of sobriety monitoring is more effective, while expansive use leads to the monitoring of lower-risk clients, for whom it is detrimental. It is also possible that the variation resulted from the propensity score matching methodology, which is unable to account for unmeasured, underlying differences between groups. It could be that at two sites, the sobriety-monitoring and non-sobriety-monitoring groups differed with respect to unmeasured characteristics that resulted in higher arrest rates among the sobriety-monitoring group, but that at the other two sites, the opposite was true.

On the other hand, there was no difference between the two groups in the pooled analysis or at any of the sites when it came to sobriety monitoring and court appearance rates. The results suggest that release without sobriety monitoring was as effective as release with sobriety monitoring in encouraging court appearance. The relationship between sobriety monitoring and rearrest appears to be more complicated, given the variation in results across the sites.

POLICY IMPLICATIONS

These findings warrant cautious reflection among policymakers and practitioners on the extent of current electronic and sobriety monitoring use, particularly considering their high personal and financial costs to those directly affected and to jurisdictions. For example, jurisdictions could consider pilot testing a reduction in their use of electronic and sobriety monitoring—perhaps starting with cases that have low risk-assessment scores—and carefully monitoring outcomes to ensure that clients continue to appear in court and avoid being rearrested. Such a pilot program could be incrementally expanded or paused depending on these outcomes. Policymakers and researchers should also be wary of the assumption that electronic and sobriety monitoring can improve court appearance rates, as they were not designed with that intention. Other, less intrusive methods for improving court appearance rates that have been rigorously evaluated for their effectiveness, such as court-date reminders, should be widely employed for that purpose.[29] The exploratory findings also highlight a need for additional cross-site studies—in particular, those that employ more rigorous experimental methods—on the effectiveness of special conditions at the pretrial stage. Given the site variation in findings, particularly for sobriety monitoring, more research is also needed to delineate the populations that would benefit from special conditions from those that would not benefit, and to illuminate the jurisdiction-specific policies and practices that are associated with the greatest success.

28. Advancing Pretrial Policy and Research (2021a); Hatton and Smith (2020).
29. Zottola, Crozier, Ariturk, and Desmarais (2022).

APPENDIX

A

Technical Supplement

As part of the Pretrial Justice Collaborative (PJC) project, MDRC conducted "comparative effectiveness" analyses that focus on the use of pretrial supervision and other release conditions, and their effects on pretrial outcomes. This appendix provides additional details about the propensity score matching analysis that was used to examine the effectiveness of two pretrial special conditions: electronic monitoring (EM) and sobriety monitoring. It describes the propensity score matching approach, the analysis sample, technical details of the analysis, and results. The analysis was conducted using data from cases initiated between January 1, 2017, and June 30, 2019, at four PJC sites.

- Site A: a large, urban metropolitan area in the western United States
- Site B: a small, rural county from the same geographic region as Site A
- Site C: a large, urban metropolitan area in the southern United States
- Site D: a medium-sized metropolitan area in the central United States

The analyses on the effectiveness of electronic monitoring were conducted for Sites A and C, and the analyses on the effectiveness of sobriety monitoring were conducted for all four sites.[1]

ELECTRONIC MONITORING AND SOBRIETY MONITORING AT THE STUDY SITES

This section describes special-condition practices at the sites from 2017 to 2019. The analysis focuses on this time period because it represents the relatively stable, "normal" conditions that existed before the destabilizing period of the COVID-19 pandemic. According to staff members from these jurisdictions, beginning in March 2020, shutdowns and adjustments related to the pandemic affected pretrial practices (for example, drug testing was suspended at some sites), court processes (for example, in-person hearings had to be postponed or modified), and patterns of crime and policing.

Across all sites, judges used their discretion to assign individuals to special conditions of pretrial release, including electronic and sobriety monitoring, following arrests. The information available to judges at the time of the decision was limited. At all sites, judges had access to existing information about a case and individual, such as current charges, scores from risk assessments predicting failure to appear and new arrests, and criminal history. At Site A, judges were also provided with an EM and a sobriety-monitoring recommendation based on risk assessments predicting domestic violence and drug use. At other sites, there

1. Although Sites B and D also used electronic monitoring, the number of cases assigned to EM at those sites was very small.

were some established practices to base the decision on a person's current charges (for example, driving under the influence or drug charges) and criminal history.

Electronic monitoring at the two relevant sites (A and C) consisted of a GPS-enabled ankle bracelet. Sobriety monitoring at the four sites varied. All four jurisdictions conducted urine analysis. In addition, judges at Sites A, B, and C could order the use of a transdermal alcohol-monitoring device, and judges at Sites B and C could order an ignition interlock device.[2] Other forms of monitoring could include a portable alcohol-monitoring device or a portable breathalyzer.

These special conditions could be assigned with or without other conditions of release such as pretrial supervision or money bond. Case managers in each jurisdiction were responsible for monitoring and responding to EM violations, monitoring drug and alcohol test results, and responding to positive test results or other issues. Some jurisdictions also worked with private vendors to help test or fit devices.

RESEARCH QUESTIONS

The analysis examines whether pretrial release without special conditions is as effective as release with special conditions in helping people make their court appearances and avoid new arrests. People who were released with or without the special conditions could be assigned to other forms of pretrial supervision—for example, weekly or monthly reporting to supervision staff members, in person or over the phone—or to no other forms of supervision. The analysis compares outcomes for people who were not assigned to a special condition (whether or not they received another form of pretrial supervision) with the outcomes of clients who were also assigned to EM or sobriety monitoring (for alcohol, drugs, or both).[3]

The questions addressed by these analyses are exploratory, and findings should not be used to draw causal inferences about the efficacy of the special conditions for producing specific outcomes. The questions the analyses address are:

2. A transdermal alcohol-monitoring device is a device—typically in the form of a bracelet—that continuously monitors an individual's alcohol consumption through perspiration on the surface of the skin. An ignition interlock device is a breathalyzer that is installed in one's car, preventing the driver from starting the car before passing a breathalyzer test.
3. The options for pretrial supervision varied by site. At Site C, most individuals who were not assigned to EM or sobriety monitoring received very little supervision of other kinds. At Sites A and B, some clients were released without any conditions, but most were required to report to the supervision staff at varying frequencies and modes (for example, remotely or in person) depending on their assessed risk. At Site D, all the clients in the sample had some form of reporting requirement, based on assessed risk.

1. Is release without electronic monitoring as effective as release with electronic monitoring in...

    a. supporting clients' appearance in court?

    b. helping clients to avoid new arrests?

2. Is release without sobriety monitoring as effective as release with sobriety monitoring in...

    a. supporting clients' appearance in court?

    b. helping clients to avoid new arrests?

The analysis uses a noninferiority approach, which tests whether release without special conditions is at least as effective as (that is, no worse than) the use of special conditions. The analysis is designed to inform the question of whether jurisdictions could reduce the use of special conditions while maintaining outcomes that are at least as good as those obtained with their current use of special conditions. In other words, is the use of special conditions at current rates at these sites improving outcomes?

OVERVIEW OF PROPENSITY SCORE MATCHING

A propensity score matching design involves comparing outcomes for cases in which individuals experienced a given service or condition (in this case assignment to a special condition) with outcomes for a matched comparison group of cases that had similar characteristics (for example, similar risk scores, demographics, and charges) but did not experience that condition. The two groups are referred to as the program group and the matched comparison group. Propensity score matching is used to balance (match) the program and comparison group with respect to observed characteristics by creating an index (propensity score) that estimates an individual's probability of receiving the program condition based on these characteristics.[4] Program group individuals are then matched to comparison individuals based on propensity scores.

Propensity score matching can only account for characteristics that are measured—that is, it is possible to match cases based on risk scores and other factors that are recorded in a site's data, but the method does not control for any differences between cases in unmeasured characteristics, such as a judge's subjective assessment of whether a person is likely to appear in court. Therefore, there is always some uncertainty about potential bias in the estimated differences in outcomes between matched groups.[5] That is, if the analysis finds that the two groups do have significantly different outcomes, it is possible that this

4. See Rosenbaum and Rubin (1983).
5. For more discussion of these limitations see Bloom, Michalopoulos, Hill, and Lei (2002).

estimated "effect" is actually due to some underlying difference between the groups and that it cannot be attributed to the actual service or condition being studied.

For several reasons, the analysis presented here represents a relatively strong case for the use of propensity score matching. First, the available data are relatively detailed; jurisdictions were able to provide administrative data about individuals, case characteristics, and assessed risks based on criminal histories.

Second, the available data are expected to provide a relatively strong model of the selection process that results in some cases being assigned special conditions and others not. When making decisions about special conditions, judges usually only have a brief moment to observe a particular person, and therefore must make decisions based mostly on the current charge in the case and the individual's criminal history and risk information provided at the hearing. This information is largely available in the data provided by the jurisdictions and the research team was able to match using it. If the propensity score model can largely account for the information that feeds into the real-life decision-making process, then the remaining bias may be small.

Finally, at most sites there was a large pool of potential comparison group cases in the data, which made it likely that good matches could be found for a large proportion of program group cases. This large pool strengthens the analysis, as dropping unmatched cases from a sample can be a source of bias in a propensity score matching analysis.[6]

Despite these strengths, it is of course still possible that there are unmeasured characteristics that differ between the special conditions and matched comparison groups. Therefore, while this analysis provides useful information about the effects of special conditions, uncertainty about potential bias remains. As noted above, the research team considers the results of the analysis to be exploratory. Readers should take the results with a reasonable degree of caution. Regardless of the results, additional research in this area is needed.

SAMPLE FOR ANALYSES

In general, the analysis sample includes:

1. Cases initiated (that is, with arrest dates) between January 1, 2017, and June 30, 2019—the time period (plus six months of follow-up data) before the COVID-19 pandemic began

2. Custodial arrests (that is, not summonses or desk-appearance tickets, which are generally not subject to pretrial supervision)

3. Cases with a completed risk assessment, which includes most custodial cases at the sites

6. King and Nielsen (2019).

4. Cases with charges that do not always result in detention (excluding some murder charges, for example, because no one with that charge will be released while awaiting trial)

5. Cases in which the person was released at some point within the pretrial period and within six months of arrest. Cases in which the individual was detained for the entire pretrial period or for the first six months after arrest (that is, the entire follow-up period) were never subject to special conditions and had no opportunity to appear in court during the pretrial period.

The sample for the electronic monitoring analysis was drawn from Sites A and C only, as the only a small number of cases were assigned to electronic monitoring at Sites B and D during the sample period. In addition, the available measures on court appearance from Site C did not appear to be reliable, so the site was dropped from the analysis of the effect of special conditions on court appearance rates.

MATCHING PROCESS

Matching was conducted separately by special condition and by site. For each site and special condition, the program group (that is, the group of people who were assigned to that special condition) was identified. The comparison pool then consisted of other cases that were initiated during the study period at each site but that were not assigned to the special condition in question.

To model the propensity score, the research team performed a series of logistic regressions to estimate the likelihood of a case being assigned to a special condition given a set of case characteristics (covariates) such as age, race, severity of charge, and assessed risks of rearrest and failure to appear in court. These are represented by the following model:

$logit(SPECIAL_{ij}) = \alpha + \beta_1 X_{ij} + \varepsilon_{ij}$, *where:*

$SPECIAL_{ij}$ = a dichotomous indicator for whether individual *i* with case *j* had a given special condition ordered,

α = the expected mean outcome when all other covariates equal 0,

X_{ij} = a vector of predictive characteristics for individual *i* with case *j* (for example, charge, age at arrest, risk score), and

ε_{ij} = random errors that are distributed independently and identically across cases.

The estimated coefficients from this model represent the relationship between specific characteristics and the likelihood of assignment to a special condition. Once these regressions were run for each site, these coefficients were multiplied by individual case characteristics to create a propensity score for each case.

Once propensity scores were estimated for both program group and comparison pool cases, program group cases were matched to comparison pool cases with similar propensity scores. For example, cases in the program group that had a propensity score of 0.7 (that is, the model estimated that an individual with those characteristics would be assigned to the special condition 70 percent of the time) were matched with comparison pool cases that had a propensity score of 0.7 or very close to it.

Matching was done using one-to-one nearest neighbor matching, meaning that only one record in the program group was matched to exactly one record in the comparison pool and each program group case was matched to the closest unmatched propensity score in the comparison pool. The research team applied a caliper of 0.2. For nearest neighbor matching, the caliper is the maximum tolerated difference between matched subjects. A caliper of 0.2 was chosen based on literature assessing the optimal caliper distance.[7] The research team chose to match without replacement because the comparison pools were large enough for most of the sites, so that matching with replacement (in which multiple program group cases could be matched to the same comparison case) was not necessary. As discussed in more detail below, sensitivity analyses that modified these specifications produced similar results.

Once the matching was done, only the matched cases were then kept for the analysis sample, resulting in a program group of cases for which a comparison case was identified and a comparison group of cases that matched to program group cases. The matched samples were then pooled across sites to create one analysis sample for the electronic monitoring analysis and one analysis sample for the sobriety monitoring analysis. Regression models were then run to estimate the differences in outcomes between the program and comparison groups.

PREDICTORS USED IN MATCHING

As noted above, the matching process was conducted separately for each of the two special conditions at each site. This procedure ensured that program group cases would be matched to comparison cases from the same site. Matching was conducted using as much relevant data as were available for each site. Some factors were available across all sites, while others were site-specific or only available for some sites. Within sites, the same predictors were used for sobriety monitoring and electronic monitoring. The characteristics that were used for the match include the following:

- Demographics: age, race, and sex
- Charge information: class (felony, misdemeanor, other) and category (violent, property, drug, public order)

7. Austin (2011).

- Timing: arrest year

- Scores on risk assessment(s): variable across sites. The jurisdictions at Sites A and B used a risk assessment that produced a single risk score ranging from 0 to 82. The jurisdictions at Sites C and D used a risk assessment that resulted in two separate scores, one predicting failure to appear in court and the other predicting new arrests. In addition, Site A also used specific risk assessments related to domestic violence and substance abuse.[8]

- Criminal history: Site D provided information about prior violent offenses, incarceration, and failures to appear in court. (For other sites, these specific variables were not available, but information on criminal history was reflected in the risk-assessment scores.)

- Release recommendations: recommendations for release, including bond type, recommendations for supervision, etc.

MATCHING DIAGNOSTICS

The research team performed statistical and visual-sensitivity checks to assess the strength of the matching process. First, a good match would result in balance between the two groups with respect to key covariates. In other words, the two research groups should have similar characteristics.

Appendix Table A.1 shows the pooled, cross-site sample means for the matched EM analysis sample. Overall, the matched EM and comparison group samples are very similar, and there are no statistically significant differences between the groups with respect to any of the characteristics shown in the table. Note that much more detailed information, including risk scores, was used to match for each site. The characteristics shown in Appendix Table A.1 include only those variables that were consistently available across sites (risk score variables, for example, varied across sites). As shown in Appendix Tables A.2 and A.3, the matched groups had very similar characteristics (that is, there was good balance) within sites on risk scores and other relevant information, such as recommended release conditions. In addition, both within the site-specific samples and for the pooled analysis, the balance between the two groups was also assessed using a regression model predicting assignment to EM. In all cases, the f-statistic for this regression model was above the statistical significance threshold, indicating that the two groups were not significantly different across those characteristics as a whole.

Appendix Table A.4 shows the pooled, cross-site sample means for the matched sobriety-monitoring analysis sample. In this case, several of the characteristics are significantly different at the 5 percent level. However, in this case, the sample size for the analysis is very

8. For these last two assessments, due to missing data, it was only possible to include a binary variable indicating whether the assessment had been conducted for a particular case. These binary variables were predictive of assignment to special conditions.

large, leading to very small standard errors and very high precision. This precision means that there can be statistically significant differences at the 5 percent threshold even when means are extremely similar. For example, the non-sobriety-monitoring group was 34.0 years old on average, compared with 34.3 years for the sobriety-monitoring group. These means are not meaningfully different but are significantly different at the 5 percent level. For a sample this size, it is reasonable to use a higher threshold to assess significance, since it is possible to pick up on such tiny, nonmeaningful differences.

A clearer picture emerges in Appendix Tables A.5 through A.8, which show more detailed information about the matching characteristics and balance for the individual sites. The tables show that the large sample size and the significant differences between the special condition and comparison groups entirely reflect Site C. For Sites A, B, and D, which have relatively smaller samples, there are very few statistically significant differences between the groups. Only for Site C, where the matched sample size is very large (23,240), are there more significant differences between the two groups, both in individual characteristics and overall, as assessed by a regression model. Despite these significant differences, the research team concluded that the groups were not meaningfully different at a threshold that is more reasonable for such a large sample. Still, in the examination of effects, this issue should be kept in mind, as small differences in mean outcomes could also be estimated as significantly different at relatively low thresholds but not be meaningfully different.

The research team also performed visual checks (not shown), including examining the distribution of propensity scores to determine whether there was sufficient variation in scores and to compare the distribution of scores among the match groups. Overall, these visual checks suggested that there was sufficient variation in propensity scores and that the distribution of propensity scores between the matched groups was similar.

Once the matched groups were determined, the samples were pooled by special condition across sites to create a single EM analysis sample and single sobriety-monitoring analysis sample. Appendix Table A.9 shows the sample sizes of the initial program group and comparison pool samples, as well as the sample sizes for the cases that ultimately matched and were included in the analysis sample. As noted above, the EM analysis was conducted only for Sites A and C. Overall, 81 percent of the cases that were assigned to EM at those sites were matched with a comparison group case, resulting in a sample size of 7,094 for the EM analysis. Eighty-eight percent of cases that were assigned to sobriety monitoring were matched to a comparison group case, resulting in an analysis sample size of 29,320. As noted above, a very large portion of that sample was from Site C.

IMPACT MODEL

The effects of the special conditions were estimated as follows:

$Y_{ij} = \alpha + \beta_1 NOSPECIAL_{ij} + \beta_2 X_{ij} + \varepsilon_{ij}$, *where:*

$NOSPECIAL_{ij}$ = a dichotomous indicator for whether individual *i* with case *j* did not have a given special condition ordered,

Y_{ij} = the outcome measure for individual *i* with case *j*,

α = the expected mean outcome when all other covariates equal 0,

X_{ij} = a vector of predictive characteristics for individual *i* with case *j* (for example, charge, age at arrest, risk score), and

ε_{ij} = random errors that are distributed independently and identically across cases.

The coefficient β_1 represents the estimated effect of *not* being assigned to the special condition versus having it assigned. The analysis clustered standard errors by individual to account for multiple cases per individual during the study period.[9]

RESULTS

This section presents the results from the two main analyses. Each analysis estimates whether being released without the special condition in question was as effective as being released with assignment to that special condition in encouraging people to make all their court appearances and avoid arrest.

Logistic regressions with covariates were performed to estimate the effects of the special conditions on the outcomes. The following covariates were used in estimating effects: age, sex, race, arrest year, charge category, and charge class.[10] Note that the analysis excluded Site C from the estimates of encouraging court appearance since there was not a reliable measure of that outcome for the site.

Electronic Monitoring

Appendix Table A.10 shows the results for electronic monitoring. Release without EM did not have a statistically significant estimated effect on court appearance rates. About 86 percent of those who did not receive EM made their court appearances compared with about 87 percent of those who were on EM.

9. In some instances, individuals had multiple cases within six months of each other, and were assigned to a special condition for one case but not for the other. Those duplicate cases were removed from the sample in order to avoid matching people to themselves.
10. The analyses could only include covariates that all sites had in common.

However, the results suggest that people who were *not* assigned to EM were more likely to avoid arrest than those who were assigned to EM. In other words, those who were assigned to EM were more likely to be rearrested within six months. About 76 percent of those without EM avoided a new arrest during the six-month follow-up period, compared with about 67 percent of those on EM.

There are at least three potential explanations for this result. First, there could be a surveillance effect, in which people in the two groups are involved in crime at similar rates, but being on electronic monitoring means that they are more likely to be caught. Second, there could be a violation effect, meaning that electronic monitoring introduces a new category of arrests that those on EM are not subject to; specifically, individuals on EM could be arrested for violating conditions of EM (for example, by cutting off their EM bracelets). Third, there could be remaining, unmeasured differences between the two groups that were not accounted for in the matching and that are correlated with both EM and arrest; in other words, the difference could be spurious. With this analysis, it is not possible to adjudicate among these explanations. The body of this report provides more context and discussion of this result.

Overall, however, these results suggest that release without EM is at least as effective as release with EM in helping clients to avoid arrest and appear in court, and that EM may actually be associated with an *increase* in rearrests.

Sobriety Monitoring

For the pooled sample, sobriety monitoring did not have any statistically significant estimated effects on either court appearance rates or the avoidance of arrest, as shown in Appendix Table A.11. About 77 percent of groups made all court appearances and about 76 percent of both groups avoided arrest.

Site-specific results (not shown) are consistent when examining the effects of sobriety monitoring on court appearance; there was no significant difference between groups by that measure at any of the three sites included in the analysis. However, there was some variation across sites in the estimated effect of sobriety monitoring on arrests; at two sites, the sobriety-monitoring group was more likely to avoid arrest while at the other two sites, the result was the opposite. The body of this report provides more context and discussion of these results.

Sensitivity Analyses

In order to assess whether the results of the analysis were sensitive to the particular decisions made during the matching process, the research team ran a series of sensitivity analyses using different matching specifications. These analyses included the following:

- Matching with replacement
- Matching using alternative covariate specifications (for example, including additional covariates, including covariate dummies versus continuous variables, etc.)
- Including weights in the pooled analysis such that each site, rather than each case, was weighted equally
- Adjusting the size of the caliper

None of the sensitivity analyses produced results that were meaningfully different from the main results. Appendix Table A.12 shows the range of impact estimates from these analyses, with the lower bound being the lowest estimate from the sensitivity analyses and the upper bound being the highest estimate.

In each case, these bounds are very similar to the estimates from the main analysis. In addition, as noted above, results were generally consistent across the individual sites, with the exception of the estimates of the effects of sobriety monitoring on rearrest. This consistency, especially given that most estimates were not statistically different from zero, provides more support for the argument that the groups were well matched and that the estimates are unbiased. Overall, these additional analyses suggest that the results were not sensitive to specific matching decisions, the weighting of the sites, and the particular sites included.

However, it is important to note that these analyses do not directly address the main limitation of propensity score matching. It is still possible that there are unmeasured differences between the matched groups that were not accounted for in the matching and that could be biasing these results. Readers should take these results with appropriate caution. Additional research continues to be needed to examine the research questions considered in this analysis.

APPENDIX TABLE A.1

Baseline Characteristics of Matched Electronic Monitoring Samples of Sites A and C

Characteristic	No EM Group	EM Group	P-Value
Age	30.2	30.4	0.442
Race (%)			0.680
Black	49.8	48.7	
White	49.0	50.0	
Other	1.2	1.2	
Female (%)	13.6	13.4	0.808
Top charge category[a] (%)			0.218
Violent	30.8	32.3	
Property	25.2	23.5	
Drug	19.8	19.1	
Public order/other	24.2	25.2	
Felony (%)	58.4	56.0	0.647
Year (%)			0.675
2017	22.2	22.1	
2018	51.1	50.6	
2019	26.8	27.3	
Sample size	3,547	3,547	

SOURCE: MDRC calculations based on court and pretrial services data from each site.

NOTES: None of the variables in this table are significantly different by group at the 5 percent level.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.2

Selected Baseline Characteristics of Unmatched and Matched Samples, Electronic Monitoring Analysis, Site A

Characteristic	Initial EM Sample	Comparison Pool		P-Value	Matched EM Group	Matched Comparison Group	P-Value
Age	33.2	35.4	***	0.000	34.6	34.8	0.409
Race (%)			***	0.000			0.977
Black	29.7	22.3			29.5	29.8	
White	68.5	75.9			67.1	67.8	
Asian/other	1.8	1.8			3.4	2.4	
Female (%)	14.6	26.0	***	0.000	17.2	18.9	0.301
Top charge category[a] (%)			***	0.000			0.421
Violent	65.4	14.7			52.9	50.2	
Property	10.3	17.9			13.3	15.4	
Drug	15.6	55.0			22.4	22.2	
Public order/other	8.7	12.5			11.4	12.3	
Top charge class[a] (%)			***	0.000			0.102
Felony	83.4	93.3			86.1	88.4	
Misdemeanor	165.4	6.7			13.8	11.4	
Raw risk score (0-82)	37.5	38.3	*	0.045	37.6	38.4	0.277
Domestic violence assessment conducted (%)	36.8	3.9	***	0.000	25.3	22.1	0.076
Received a recommendation of intensive supervision (%)	51.3	3.1	***	0.000	30.1	31.1	0.585
			***				
Received a recommendation of release without supervision (%)	6.2	47.5	***	0.000	9.8	9.6	0.832
Sample size	1,918	11,875			1,140	1,140	

SOURCE: MDRC calculations based on court and pretrial services data from Site A.

NOTES: In addition to the variables shown, the groups were also matched on additional supervision and release recommendation information and year of arrest.

Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.3

Selected Baseline Characteristics of Unmatched and Matched Samples, Electronic Monitoring Analysis, Site C

Characteristic	Initial EM Sample	Comparison Pool		P-Value	Matched EM Group	Matched Comparison Group	P-Value
Age	28.9	33.9	***	0.000	29.0	28.6	0.112
Race (%)			***	0.000			0.584
Black	58.4	44.2			57.9	59.2	
White	41.0	54.2			41.5	40.1	
Asian/other	0.6	1.6			0.6	0.7	
Female (%)	11.5	27.4	***	0.000	11.6	11.1	0.586
Top charge category[a] (%)			***	0.000			0.322
Violent	22.9	17.2			22.5	21.7	
Property	28.5	22.6			28.3	29.9	
Drug	17.3	18.7			17.6	18.6	
Public order/other	31.3	41.5			31.7	29.9	
Top charge class[a] (%)			***	0.000			0.193
Felony	46.9	26.0			46.1	44.3	
Misdemeanor	53.1	74.0			53.9	55.8	
Risk of FTA (1-6)	2.2	1.9	***	0.000	2.2	2.2	0.328
Risk of arrest for a new crime (1-6)	3.2	2.5	***	0.000	3.2	3.2	0.944
Overall risk level (%)			***	0.000			0.855
Low	30.4	48.9			30.8	31.5	
Medium	55.2	45.6			55.4	55.1	
High	14.4	5.5			13.8	13.4	
Received a release recommendation (%)	41.7	63.0	***	0.000	42.3	44.7	0.092
Sample size	2,442	41,264			2,407	2,407	

SOURCE: MDRC calculations based on court and pretrial services data from Site C.

NOTES: In addition to the variables shown, the groups were also matched on risk of arrest for a new violent crime, additional release-recommendation information, and year of arrest.

Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

FTA = failure to appear in court.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.4

Baseline Characteristics of Matched Sobriety-Monitoring (SM) Samples of All Sites

Characteristic	No SM Group	SM Group		P-Value
Age	34.0	34.3	*	0.036
Race (%)			*	0.044
Black	34.5	35.5		
White	64.2	63.0		
Other	1.3	1.5		
Female (%)	22.6	23.0		0.461
Top charge category[a] (%)			*	0.018
Violent	12.6	13.1		
Property	11.7	11.7		
Drug	30.1	31.3		
Public order/other	45.6	43.9		
Felony (%)	41.1	42.9		0.107
Year (%)			**	0.003
2017	23.5	22.8		
2018	55.3	54.4		
2019	21.2	22.8		
Sample size	11,620	11,620		

SOURCE: MDRC calculations based on court and pretrial services data from each site.

NOTE: Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.5

Selected Baseline Characteristics of Unmatched and Matched Samples, Sobriety Monitoring (SM) Analysis, Site A

Characteristic	Initial SM Sample	Comparison Pool		P-Value	Matched SM Group	Matched Comparison Group	P-Value
Age	34.7	35.1		0.210	34.6	34.8	0.535
Race (%)			***	0.000			0.753
Black	17.2	24.3			17.3	16.6	
White	80.2	73.9			80.1	81.2	
Asian/other	2.6	1.8			2.6	2.2	
Female (%)	24.7	24.0		0.566	25.1	26.4	0.448
Top charge category[a] (%)			***	0.000			0.994
Violent	31.8	21.1			33.1	32.8	
Property	11.0	17.4			11.5	11.5	
Drug	36.4	50.4			37.8	38.2	
Public order/other	20.8	11.1			17.7	17.5	
Top charge class[a] (%)			***	0.000			0.704
Felony	81.8	92.9			84.4	85.0	
Misdemeanor	18.1	7.0			15.5	14.9	
Raw risk score (0-82)	37.9	38.2		0.444	38.1	38.8	0.260
Substance abuse assessment conducted (%)	5.6	0.3	***	0.000	2.1	2.0	0.890
Received a recommendation of intensive supervision (%)	18.3	9.4	***	0.000	18.4	17.0	0.331
Received a recommendation of release without supervision (%)	28.2	42.9	***	0.000	26.7	24.7	0.245
Sample size	1,364	12,499			1,310	1,310	

SOURCE: MDRC calculations based on court and pretrial services data from Site A.

NOTES: In addition to the variables shown, the groups were also matched on additional supervision and release recommendation information and year of arrest.

Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.6

Selected Baseline Characteristics of Unmatched and Matched Samples, Sobriety Monitoring (SM) Analysis, Site B

Characteristic	Initial SM Sample	Comparison Pool		P- Value	Matched SM Group	Matched Comparison Group	P- Value
Age	35.9	35.5		0.204	36.4	35.5	0.062
Race (%)				0.254			0.625
Black	3.5	4.5			3.0	2.8	
White	93.0	91.9			92.9	93.9	
Asian/other	3.5	3.6			4.1	3.3	
Female (%)	33.0	30.1		0.054	32.8	30.1	0.184
Top charge category[a] (%)			***	0.000			0.685
Violent	18.6	31.6			25.8	27.3	
Property	12.0	22.6			16.6	15.9	
Drug	50.5	13.1			31.5	32.4	
Public order/other	18.9	32.7			26.2	24.3	
Top charge class[a] (%)			***	0.000			0.553
Felony	77.3	49.8			68.9	67.7	
Misdemeanor	22.3	48.8			30.6	31.3	
Other	0.4	1.4			0.5	1.0	
Risk category (%)			***	0.000			0.557
Low	44.2	55.8			53.6	51.4	
Medium	36.1	27.1			31.9	32.9	
High	19.7	17.1			14.6	15.8	
Received a recommendation of release without supervision (%)	20.5	47.6	***	0.000	28.1	29.1	0.472
Sample size	1,532	2,823			1,105	1,105	

SOURCE: MDRC calculations based on court and pretrial services data from Site B.

NOTES: In addition to variables shown, samples were also matched on additional release and supervision recommendation and year of arrest.

Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.7

Selected Baseline Characteristics of Unmatched and Matched Samples, Sobriety Monitoring (SM) Analysis, Site C

Characteristic	Initial SM Sample	Comparison Pool		P-Value	Matched SM Group	Matched Comparison Group		P-Value
Age	34.1	33.4	***	0.000	34.1	33.8		0.052
Race (%)			***	0.000				0.097
Black	39.6	47.0			39.0	37.9		
White	59.2	51.3			59.8	61.1		
Asian/other	1.2	1.7			1.2	1.0		
Female (%)	22.1	28.2	***	0.000	22.2	21.7		0.358
Top charge category[a] (%)			***	0.000			***	0.000
Violent	9.6	20.5			9.9	8.9		
Property	11.0	27.5			11.3	11.5		
Drug	32.2	13.5			30.1	28.4		
Public order/other	47.2	38.5			48.7	51.3		
Felony (%)	37.7	23.2	***	0.000	35.7	33.6	***	0.001
Risk of FTA (1-6)	2.0	1.8	***	0.000	2.0	2.0		0.276
Risk of arrest for a new crime (1-6)	2.7	2.5	***	0.000	2.7	2.7		0.079
Overall risk level (%)			***	0.000				0.144
Low	41.6	50.3			42.2	43.0		
Medium	51.6	44.0			50.9	50.6		
High	6.8	5.7			6.9	6.4		
Sample size	11,986	31,720			11,620	11,620		

SOURCE: MDRC calculations based on court and pretrial services data from Site C.

NOTES: In addition to the variables shown, the groups were also matched on risk of new arrest for a violent crime, additional release recommendation information, and year of arrest.

Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

FTA = failure to appear in court

[a]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.8

Selected Baseline Characteristics of Unmatched and Matched Samples, Sobriety Monitoring (SM) Analysis, Site D

Characteristic	Initial SM Sample	Comparison Pool	P-Value	Matched SM Group	Matched Comparison Group	P-Value
Age	34.0	33.9	0.854	33.9	33.8	0.903
Race[a] (%)			0.897			0.837
Black	63.3	64.3		66.6	65.0	
White	36.4	35.5		33.3	34.8	
Female (%)	16.1	21.2 **	0.004	16.3	19.0	0.208
Top charge category[b] (%)		***	0.000			0.162
Violent	3.2	17.5		8.6	12.3	
Property	3.9	12.6		9.9	10.1	
Drug	70.6	34.7		41.0	40.8	
Public order/other	22.3	35.2		40.4	36.8	
Top charge class[b] (%)		*	0.033			0.955
Felony	39.1	43.8		43.0	43.2	
Misdemeanor	28.6	33.2		34.1	33.8	
Other	32.2	23.1		22.9	23.0	
Previous incarceration (%)	64.6	59.4 *	0.015	63.0	60.8	0.415
Previous violent crime (%)	40.7	40.0	0.715	37.6	40.3	0.325
FTA in the past 2 years (%)	52.3	44.4 ***	0.000	50.1	45.9	0.141
Risk score for FTA (%)		***	0.000		*	0.041
1	6.2	6.4		6.2	6.6	
2	11.0	10.9		10.7	10.2	
3	24.7	33.4		25.4	31.7	
4	16.0	18.3		17.6	19.2	
5	23.8	20.8		24.2	21.4	
6	18.3	10.2		15.8	10.8	

(continued)

APPENDIX TABLE A.8 (CONTINUED)

Characteristic	Initial SM Sample	Comparison Pool	P-Value	Matched SM Group	Matched Comparison Group	P-Value
Risk score for a new crime (%)			*** 0.000		*	0.025
1	3.0	2.5		2.1	2.6	
2	13.4	17.0		14.1	16.6	
3	19.6	24.2		21.2	22.4	
4	30.4	35.1		30.4	34.9	
5	17.5	11.8		19.4	13.1	
6	16.1	9.5		12.8	10.4	
Sample size	736	1,743		625	625	

SOURCE: MDRC calculations based on court and pretrial services data from Site D.

NOTES: In addition to variables shown, samples were matched on year of arrest.
Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.
FTA = failure to appear in court.
[a]Nearly all people in the sample were categorized as Black or White; a very small number of clients fell into other categories.
[b]The word "top" refers to the most severe charge category or charge class on a person's case, since a case can have multiple charges attached to it. The order of severity used to determine a case's top charge was (from most to least severe): (1) for charge category: violent, property, drug, public order/other, and (2) for charge class: felony, misdemeanor, other.

APPENDIX TABLE A.9

Sample Sizes of Matched and Unmatched Cases, by Site

Sample	Site A	Site B	Site C	Site D	Total
Electronic monitoring					
Sample assigned to EM					
Matched	1,140	NA	2,407	NA	3,547
Unmatched	778	NA	35	NA	813
Comparison pool					
Matched	1,140	NA	2,407	NA	3,547
Unmatched	10,735	NA	38,857	NA	49,592
Total matched analysis sample	2,280	NA	4,814	NA	7,094
Sobriety monitoring					
Sample assigned to sobriety monitoring					
Matched	1,310	1,105	11,620	625	14,660
Unmatched	54	427	366	1,118	1,965
Comparison pool					
Matched	1,310	1,105	11,620	625	14,660
Unmatched	11,189	1,718	20,100	111	33,118
Total matched analysis sample	2,620	2,210	23,240	1,250	29,320

SOURCE: MDRC calculations based on court and pretrial services data from each site.

APPENDIX TABLE A.10

Sites A and C Effects of Release Without EM Compared With Release With EM

Outcome (%)	No EM Group	EM Group	Estimated Effect	P-Value
Made all court appearances[a]	86.2	86.7	-0.5	0.974
Avoided a new arrest	76.1	67.4	8.7 ***	0.000
Sample size	3,547	3,547		

SOURCE: MDRC calculations based on court and pretrial services data from each site.

NOTES: Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

[a]Estimate includes Site A only, as there was not a reliable measure of court appearance for Site C.

APPENDIX TABLE A.11

Effects of Release Without Sobriety Monitoring Compared With Release With Sobriety Monitoring

Outcome (%)	Non-Sobriety-Montoring Group	Sobriety-Monitoring Group	Estimated Effect	P-Value
Made all court appearances	76.9	76.8	0.1	0.910
Avoided a new arrest	76.0	75.5	0.5	0.476
Sample size	14,660	14,660		

SOURCE: MDRC calculations based on court and pretrial services data from each site.

NOTE: None of the estimates in this table are statistically significant at the 5 percent level.

APPENDIX TABLE A.12

Range of Effect Estimates Produced by Sensitivity Analyses

Outcome (%)	Effect Estimates from the Main Analysis	Lower-Bound Estimates from Sensitivity Analyses	Upper-Bound Estimates from Sensitivity Analyses
Impacts of release without EM compared with EM			
Made all court appearances	-0.5	-1.0	0.6
Avoided a new arrest	8.7 ***	8.5 ***	8.8 ***
Impacts of release without sobriety monitoring compared with sobriety monitoring			
Made all court appearances	0.1	-1.2	-0.1
Avoided a new arrest	0.7	0.1	0.6

SOURCE: MDRC calculations based on court and pretrial services data from each site.

NOTE: Statistical significance levels are indicated as follows: *** = 0.1 percent; ** = 1 percent; * = 5 percent.

REFERENCES

Advancing Pretrial Policy and Research. 2021a. "Pretrial Drug Testing." Website: https://cepp.com/wp-content/uploads/2021/04/Pretrial-Drug-Testing-2020.pdf.

Advancing Pretrial Policy and Research. 2021b. "Pretrial Location Monitoring." Website: https://cepp.com/wp-content/uploads/2021/04/Pretrial-Location-Monitoring-2020.pdf.

Arnett, Chaz. 2019. "From Decarceration to E-carceration." *Cardozo Law Review* 41, 2: 641–720.

Austin, Peter C. 2011. "Optimal Caliper Widths for Propensity-Score Matching When Estimating Differences in Means and Differences in Proportions in Observational Studies." *Pharmaceutical Statistics* 10, 2: 150–161.

Belur, Jyoti, Amy Thornton, Lisa Tompson, Matthew Manning, Aiden Sidebottom, and Kate Bowers. 2020. "A Systematic Review of the Electronic Monitoring of Offenders." *Journal of Criminal Justice* 68, 101686.

Bloom, Howard S., Charles Michalopoulos, Carolyn J. Hill, and Ying Lei. 2002. "Can Nonexperimental Comparison Group Methods Match the Findings from a Random Assignment Evaluation of Mandatory Welfare-to-Work Programs?" Working paper. New York: MDRC.

Britt, Chester L., III, Michael R. Gottfredson, and John S. Goldkamp. 1992. "Drug Testing and Pretrial Misconduct: An Experiment on the Specific Deterrent Effects of Drug Monitoring Defendants on Pretrial Release." *Journal of Research in Crime and Delinquency* 29, 1: 62–78. https://journals.sagepub.com/doi/10.1177/0022427892029001004.

Cadigan, Timothy P. 1991. "Electronic Monitoring in Federal Pretrial Release." *Federal Probation* 55, 1: 26–30. https://www.ncjrs.gov/pdffiles1/Digitization/133410NCJRS.pdf.

Cooprider, Keith W., and Judith Kerby. 1990. "A Practical Application of Electronic Monitoring at the Pretrial Stage." *Federal Probation* 54, 1: 28–35. https://www.ncjrs.gov/pdffiles1/Digitization/123147NCJRS.pdf.

Cornett, Sarah. 2022. "The Fraught and Expensive Cycle of Drug Testing." *Arnold Ventures*. Website: https://www.arnoldventures.org/stories/the-fraught-and-expensive-cycle-of-drug-testing.

Goldkamp, John S., and Peter R. Jones. 1992. "Pretrial Drug-Testing Experiments in Milwaukee and Prince George's County: The Context of Implementation." *Journal of Research in Crime and Delinquency* 29, 4 : 430–465. https://journals.sagepub.com/doi/10.1177/0022427892029004003.

Grommon, Eric, Jason Rydberg, and Jeremy G. Carter. 2017. "Does GPS Supervision of Intimate Partner Violence Defendants Reduce Pretrial Misconduct? Evidence from Quasi-Experimental Study." *Journal of Experimental Criminology* 13, 4: 483–504.

Hager, Eli. 2020. "Where Coronavirus Is Surging—And Electronic Surveillance, Too." *The Marshall Project*, November 22. https://www.themarshallproject.org/2020/11/22/where-coronavirus-is-surging-and-electronic-surveillance-too.

REFERENCES *(CONTINUED)*

Hatton, Ross, and Jessica Smith. 2020. *Research on the Effectiveness of Pretrial Support and Supervision Services: A Guide for Pretrial Services Programs*. Chapel Hill, NC: UNC Criminal Justice Innovation Lab.

King, Gary, and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis* 27, 4: 435–454.

McKnight, A. Scott, James C. Fell, and Amy Auld-Owens. 2012. *Transdermal Alcohol Monitoring: Case Studies*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Pew Charitable Trusts. 2016. "Use of Electronic Offender-Tracking Devices Expands Sharply." Philadelphia: Pew Charitable Trusts.

Pretrial Justice Institute. 2019. *Scan of Pretrial Practices 2019.* Baltimore, MD: Pretrial Justice Institute.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika* 70, 1: 41–55.

Sainju, Karla Dhungana, Stephanie Fahy, Booz Allen Hamilton, Katherine Baggaley, Ashley Baker, Tamar Minassian, and Vanessa Filippelli. 2018. "Electronic Monitoring for Pretrial Release: Assessing the Impact." *Federal Probation* 82, 3: 3–10. https://www.uscourts.gov/sites/default/files/82_3_1.pdf.

Valentine, Erin, and Sarah Picard. 2023. *Assessing the Effectiveness of Varying Intensities of Pretrial Supervision: Full Findings from the Pretrial Justice Collaborative*. New York: MDRC.

VanNostrand, Marie, and Gena Keebler. 2009. "Pretrial Risk Assessments in the Federal Court." *Federal Probation* 73, 2.

Weisburd, Kate, Varun Bahadha, Matthew Clauson, Jeanmarie Elican, Fatima Khan, Kendall Lawrenz, Brooke Pemberton, Rebecca Ringler, Jordan Schaer, Mikayla Sherman, and Sarah Wohlsdorf. 2021. *Electronic Prisons: The Operation of Ankle Monitoring in the Criminal Legal System*. Washington, DC: George Washington University Law School.

Weisburd, Kate, and Alicia Virani. 2022. "Op-Ed: The Monster of Incarceration Quietly Expands Through Ankle Monitors." *Los Angeles Times*, March 15. https://www.latimes.com/opinion/story/2022-03-15/electronic-monitoring-pretrial-bail-mass-incarceration.

Wolff, Kevin T., Christine A. Dozier, Jonathan P. Muller, Margaret Mowry, and Barbara Hutchinson. 2017. "The Impact of Location Monitoring Among U.S. Pretrial Defendants in the District of New Jersey." *Federal Probation* 81, 3: 8–14. https://www.uscourts.gov/sites/default/files/81_3_2_0.pdf.

Zottola, Samantha A., William E. Crozier, Deniz Ariturk, and Sarah L. Desmarais. 2022. "Court Date Reminders Reduce Court Nonappearance: A Meta-Analysis." *Criminology and Public Policy* 22, 1: 97–123.

ABOUT MDRC

MDRC, a nonprofit, nonpartisan social and education policy research organization, is committed to finding solutions to some of the most difficult problems facing the nation. We aim to reduce poverty and bolster economic mobility; improve early child development, public education, and pathways from high school to college completion and careers; and reduce inequities in the criminal justice system. Our partners include public agencies and school systems, nonprofit and community-based organizations, private philanthropies, and others who are creating opportunity for individuals, families, and communities.

Founded in 1974, MDRC builds and applies evidence about changes in policy and practice that can improve the well-being of people who are economically disadvantaged. In service of this goal, we work alongside our programmatic partners and the people they serve to identify and design more effective and equitable approaches. We work with them to strengthen the impact of those approaches. And we work with them to evaluate policies or practices using the highest research standards. Our staff members have an unusual combination of research and organizational experience, with expertise in the latest qualitative and quantitative research methods, data science, behavioral science, culturally responsive practices, and collaborative design and program improvement processes. To disseminate what we learn, we actively engage with policymakers, practitioners, public and private funders, and others to apply the best evidence available to the decisions they are making.

MDRC works in almost every state and all the nation's largest cities, with offices in New York City; Oakland, California; Washington, DC; and Los Angeles.