Under Pressure? Performance Evaluation of Police Officers as an Incentive to Cheat*

Ekaterina Travova[†](CEBI[‡], EUSP[§])

May 2023

Abstract

This paper examines the use of high-powered performance-based incentives for civil servants, focusing on drug-related cases registered by the Russian police. Using an event study approach and bunching analysis, I show that the incentives arising from the performance evaluation system of police officers can significantly influence their behavior. Specifically, I find evidence suggesting that this impact can result in the manipulation of drug quantities seized by the police, moving offenders from below to above the punishment threshold. Further negative consequences of the strong performance-based incentives are inequality in the enforcement of law, prolonged sentences, and increased probability of pretrial detention. Thus, I determine that police officers are more likely to manipulate the drug quantities seized from men. I also find that the manipulation increases the probability of pretrial detention by 9% and adds one more year of incarceration, which is a 67% increase on the average sentence length without manipulation.

JEL Classification: D73, H11, J45, K14, K42.

Keywords: Drug Crimes, Police Discretion, Performance Evaluation, Incentives.

---

*I am grateful to Andreas Menzel for generous support at all stages of this project. I thank Stepan Jurajda, Randall Filer, Vasily Korovkin, Sergei Mikhalishchev, Libor Dusek, Dmitriy Skougarevskiy, Alexey Knorre, Nikolas Mittag, Jan Palguta, Daniel L. Millimet, and seminar participants at CERGE-EI, EUSP, Hunter College - CUNY and TAMU (online) for helpful discussions and comments. I also thank the two anonymous journal reviewers and the handling editor, Scott Adams, for detailed and insightful feedback. All errors remaining in this text are the responsibility of the author. This study was supported by Charles University, GAUK project No. 702218, the H2020-MSCA-RISE project GEMCLIME-2020 GA No. 681228, and CEBI, Danish National Research Foundation Grant DNRF134. Declarations of interest: none.

[†]E-mail: ekta@econ.ku.dk.

[‡]Center for Economic Behavior and Inequality (CEBI), Department of Economics, University of Copenhagen, Øster Farimagsgade 5, Building 35, 1353 Copenhagen, Denmark

[§]European University at Saint-Petersburg, Shpalernaya st. 1, 191187 Saint Petersburg, Russia.

1 Introduction

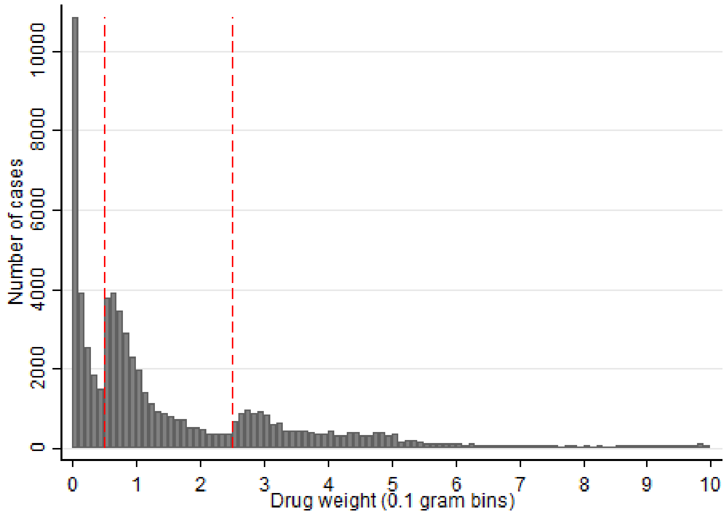
For decades, scholars have been investigating the relationship between incentives and effort. Long-standing theoretical literature emphasizes the importance of sufficient motivation but, at the same time, recognizes that high-powered incentives tailored to easily measurable and quantifiable performance indicators could significantly distort behavior (e.g., Holmstrom and Milgrom 1991, Dixit 1997). The negative effect of these strong incentives is documented, for example, for teachers (Jacob and Levitt 2003, Jacob 2005), nurses and doctors (Banerjee et al. 2008, Alexander 2020), government officials (Fisman and Wang 2017) and law enforcers (Mas 2006, Ash and MacLeod 2015, Makowsky et al. 2019, Acemoglu et al. 2020). In this paper, I study the effect of high-powered performance-based incentives on the behavior of police officers in Russia, and investigate further consequences of these incentives for the offenders.

Russia provides a natural laboratory for investigating the use of incentives in the public sector. Figure 1[1] shows the distribution of heroin cases across drug quantities seized in Russia during the 2013-2014 period. Two dashed lines indicate the threshold drug quantities that define three classes of the seriousness of the drug possession offenses, and accordingly the severity of punishment[2]. The figure reveals a striking pattern suggesting that, at the moment of arrest, many people possess a drug quantity that is just above a threshold beyond which they will be convicted of a more serious offense. In addition, there is a missing mass of cases just below the thresholds. This phenomenon is suggestive of manipulation of the drug quantities seized by the police, moving offenders from below to above the thresholds. In turn, the possible manipulation might be one of the adverse outcomes of implementing strong performance-based incentives in law enforcement organizations.

To study the effect of the high-powered incentives on police officers' behavior, I exploit a specific feature of the Russian institutional context, namely, the existence of two drug control agencies (until 2016), which were similar in many aspects but had different performance evaluation approaches. Both agencies were highly centralized and

[1]Figure 1 replicates Figure 5 in Knorre (2017).
[2]There is also a third threshold (at 500 grams for heroin) that is not depicted on the graph.

Figure 1: Distribution of cases across quantities of heroin seized in Russia during 2013-2014



Note: The baseline sample consists of all heroin-related cases registered in Russia during 2013-2014. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment. This graph replicates Figure 5 in Knorre (2017).

characterized by the lack of systematic accountability of their officers for misconduct. In order to set incentives, one of the agencies compared the performance of its officers within each station over time so that the most recent performance would not be worse than previously. Therefore, the targets - the total number of drug crimes and the number of serious and most serious drug crimes that should be reported by the end of the year - were always known to officers, thus making the incentives to meet the requirements clearer and stronger. In contrast, the other agency used the performance comparison across stations, and officers could only forecast the target that they should reach based on their performance and that of other officers in previous years.

Using an event study approach, I show that evaluating the police officers based on the performance indicators does disturb their behavior during the year. The analysis suggests a 4% increase in the number of (serious and most serious) drug crimes registered in the 30-day period, by the end of which the police station reaches its previous year's performance level. There is also a slight upward pre-trend in the two 30-day periods before the "event". After the "event", the number of cases returns to its average level before the jump. One of the potential mechanisms behind this pattern might be the manipulation of the drug amounts seized by the police, which, in turn, results in the bunching of cases above punishment thresholds observed in Figure 1.

To investigate this potential mechanism, I focus specifically on cases related to a particular drug type - heroin. In the data, only heroin quantities are presented in milligrams (while all other drugs are in grams) which allows me to study discontinuities in the distribution of cases around the thresholds at a granular level. In Subsection 4.1.2.1, I discuss the alternative (to manipulation) explanations for the observed bunching, such as differential treatment by the police (other than moving some offenders around the threshold), the nature of demand and supply in the illicit drug market, and specific features of the criminal justice system. While it is not possible to completely rule out the alternative stories behind the observed discontinuities, I present anecdotal evidence suggesting that manipulations with drug quantities is the most probable explanation. In addition, I argue that these manipulations were more likely in the form of increasing the actual drug weights than merely changing the numbers in documents. While the paper's results and conclusions do not depend on the exact method of manipulation, knowing that the falsification of evidence may occur and may be relatively costless[3] is important for policy-making. Using a standard bunching estimator complemented by a regression analysis, I provide further evidence suggesting that the observed bunching patterns in the distribution of drug-related cases are consistent with them being driven by the performance evaluation system[4].

The use of "targets" and "quotas" resulting in adverse effects is not specific to Russian police and is observed in police performance evaluation in other post-Soviet states as well as in some developed countries, such as the UK, France, Australia and the US (Eterno and Silverman 2017, de Maillard 2018, Ossei-Owusu 2021). Furthermore, in those states of the US, where the use of statistics-based incentives was legally prohibited, this approach is still widely implemented informally by many supervisors of local stations (Ossei-Owusu

[3]As discussed in Subsection 4.1.2.1, the actual manipulation of drug quantities is relatively costless for police officers due to the specific rules applied to weighing the drugs seized: the drug quantity seized is determined not by the weight of the pure drug substance but by the weight of the entire mixture. Therefore, for example, in the case of heroin, an officer could add flour, sugar, or any other white powder in order to arrest an offender for a more serious crime.

[4]It is worth noting that bribery may be another motive for police officers to manipulate the drug quantities, or to threaten offenders with possible manipulation. This is discussed in more detail in Subsection 4.1.2.2

2021).

In the second part of the paper, I identify further negative consequences of the high-powered performance-based incentives: inequality in the enforcement of law, prolonged sentences, and increased probability of pretrial detention. Adopting the novel bunching technique from Diamond and Persson (2016), I investigate the characteristics of victims of possible manipulation of seized drug quantities and estimate the effect of these manipulations on case characteristics and sentence length. The analysis identifies a gender bias in police officers' decisions to manipulate, suggesting that men are more likely to become a victim. I do not find any significant effects in relation to other demographics, socio-economic status or such offender's characteristics as being previously charged with an offense or being under the influence of drugs or alcohol when arrested. Analyzing all felony cases processed by Russian federal district courts during the 2009-2013 period, Volkov (2016) and Chatsverykova (2017) also find a significant gender bias in judges' decisions against men but only for violent crimes or theft and not for drug-related offenses.

The overall estimated effect of possible manipulation on the sentence length of drug users (who constitute almost 70% of all manipulated offenders) is around one additional year of incarceration. This is an almost 67% increase on the average sentence length (1.5 years) without manipulation. I also find that the manipulation increases the probability of pretrial detention by 9%, but it does not affect the probability of pleading guilty or the lengths of investigation and adjudication.

This paper adds to the still scarce but growing literature on performance evaluations and incentive schemes in the public sector. In law enforcement, studying the effect of various policies on the behavior of agents is particularly important because occasionally it becomes a matter of life and death. For example, Acemoglu et al. (2020) investigate the use of high-powered incentives for the military in Colombia and find that rewarding army members for killing guerillas (left-wing insurgents) significantly increases the number of false positives in which innocent civilians are killed and misrepresented as guerillas. This is an extreme example of an adverse response to poorly designed incentives in a highly consequential setting, which is similar to the drug quantity manipulation in Russia investigated in my paper. Nevertheless, these studies broaden the empirical evidence and also provide general insights that are useful for policy-making in other settings.

The manipulation of the drug quantities seized by the Russian police is widely discussed in various media reports[5]. There are also some descriptive studies (Paneyakh 2014, Knorre 2017) and more quantitative analyses of the distribution of drug-related cases (Skougarevskiy 2017, Knorre 2020). However, these papers do not test for the possibility that the incentive structure used by law enforcement agencies induces fraudulent behavior. Thus, this study addresses this gap by presenting a rigorous analysis of possible drug manipulation and the mechanisms behind it. Additionally, in contrast to previous studies, this paper takes a step further and analyzes the consequences of high-powered incentives in law enforcement.

Using Russia as an example, this paper raises a concern about the increasing use of high-powered performance-based incentives in the public sector without taking into account the specifics of the institutional environment. While there are studies that find that strengthening the incentives for public employees could be effective (e.g., Duflo et al. 2012, Singh and Masters 2017), it is important to remember that this does not apply to all settings. For example, the use of these strong incentives has been shown to backfire in the highly centralized environment as in Fisman and Wang (2017), in the environment with a lack of accountability for misconduct as in Acemoglu et al. (2020), and in the environment characterized by both these institutional features as in the case studied in this paper.

Among studies on incentives in law enforcement, there is a scope of papers that focus on incentives arising from various punishment thresholds. The response to these thresholds may come from two sides: offenders who might strategically bunch below the thresholds (Traxler et al. 2018, Lepage 2020), and law enforcers who might want to adjust the punishment for some offenders around the thresholds (Anbarci and Lee 2014, Goncalves and Mello 2021, Bjerk 2005, Ulmer et al. 2007, Rehavi and Starr 2014, Bjerk 2017). I focus on the latter case and, in contrast to most existing studies, analyze a setting in which law enforcers behave in a more repressive way, intentionally increasing the penalty for the offender[6].

[5] For example, see Nadezhdin and Matveeva (2019), Merzlikin (2019), Antonov (2019).

[6] Anbarci and Lee (2014) and Goncalves and Mello (2021) use US data on speeding tickets and find an excess mass at speeds just below the first threshold, above which the fine increases. They take this bunching as evidence of manipulation by police officers, who may wish to avoid onerous punishment for drivers. Bjerk (2005), Ulmer et al. (2007) and Rehavi and Starr (2014) find that some prosecutors

Similar to this study, Tuttle (2019) also finds the bunching of drug offenders above the punishment threshold. However, using US data, he comes to a different conclusion as to why the bunching occurs, which highlights my results in an interesting way. Thus, in contrast to my paper, Tuttle (2019) documents that the observed excess mass of drug offenders is due to prosecutorial discretion: prosecutors may use legal tools to move offenders above the threshold if they believe these offenders deserve a harsher punishment. Additionally, Tuttle (2019) only finds bunching for crack-cocaine traffickers, whereas I document bunching for both heroin users and sellers. In general, the main focus of his paper is on racial discrimination, whereas this study adds to the scarce literature on the negative consequences of strong incentives in the public sector in countries where race and ethnicity play a lesser role in defining social status.

The rest of this paper is organized as follows. Section 2 describes the institutional context and data. In Section 3, I provide the empirical strategy, and in Section 4, I present the results. Section 5 contains concluding remarks.

2 Institutional Context and Data

This section briefly discusses the institutional background, providing information on Russian anti-drug laws and the system of performance evaluation for police officers. Additionally, it describes the dataset used for the empirical analysis.

2.1 Institutional Context

The first independent Russian anti-drug agency was established in 2002. Since then it has been reorganized multiple times, and in 2004 was renamed the Russian Federal Service for Drug Control (FSKN)[7], also known as the "Drug Police". The responsibilities of this agency included controlling legal drugs, combating illicit drug trafficking, and prevention of drug abuse. The FSKN shared jurisdiction with the Public Security Service (Police)

are more likely to charge offenders who were initially arrested for crimes under a mandatory minimum sentencing law with a lesser crime not covered by this law. Bjerk (2017) focuses on drug crimes in the US and finds that first-time drug offenders are likely to avoid prosecution under mandatory minimum law.

[7]Federal'naya sluzhba Rossiiskoi Federacii po kontrolyu za oborotom narkotikov, FSKN.

of the Ministry of Internal Affairs (MVD)[8], but was solely responsible for coordinating and pursuing Russian drug investigations abroad (The Ministry of Internal Affairs of the Russian Federation, n.d.). While the main focus of the FSKN was officially on larger cases (drug trafficking, organized crime, large drug quantities), the MVD mostly dealt with routine low-profile cases, such as drug use and small-quantity drug sales. At the same time, the MVD provided many other public security functions, and drug control was not its only responsibility.

Both agencies were highly centralized and had similar structures with three levels of authority: federal - at the level of the country, regional - at the level of regions[9], and local - at the level of municipal districts (*raions*)[10] and some of the cities. In the case of the MVD, there could be further subdivision of police stations in large municipal districts and cities. In 2013-2014, there were around 750,000 MVD police officers and 36,000 FSKN officers, which reflected the number of drug-related cases registered by these two agencies: almost two thirds of them were initiated by the MVD.

Even though the MVD and FSKN were supposed to have different drug control strategies, in practice, their efforts were duplicative in many aspects (Knorre and Skougarevskiy 2015). Table B1 in Appendix B presents some summary statistics for the MVD and FSKN. Even though almost all differences in parameters are statistically significant, not all of them are economically significant. Thus, the compositions of drug types seized by the agencies were similar with regard to the prevalence of heroin and cannabis group drugs (marijuana, hashish, and others) in both the MVD and FSKN. The average seizure weight in the case of the FSKN was much greater than in the case of the MVD. However, the FSKN very rarely seized large drug quantities, and median seizure sizes for the two agencies were around 5 grams. In addition, the typical offender and the composition of punishment types were also quite similar across the two agencies. In 2016, the FSKN was dissolved, and its functions were transferred to the MVD.

Anti-drug legislation. 95% of all drug crimes registered in Russia in 2013-2014 were prosecuted under articles 228 and 228.1 of its Criminal Code. The severity of a penalty

[8] Ministerstvo vnutrennih del Rossiiskoi Federacii, MVD.

[9] In 2013, there were 83 regions. After the annexation of Crimea in 2014, the number of regions increased to 85.

[10] In 2013, there were 1,815 districts, while in 2014 there were 1,833 districts.

under these articles depends on the type of drug offense (drug use or drug sale) and on the weight of the drug seized, which are classified via threshold quantities as "significant", "large" or "especially large" (Appendix B, Tables B2 and B3). According to the law, the drug quantity seized is determined not by the weight of the pure drug substance but by the weight of the entire mixture. Therefore, if a police officer seizes, for example, one gram of heroin mixed with two grams of sugar, it will be considered to be three grams of heroin.

The punishment for drug possession of a "significant" quantity, with no intention to sell, is a fine of up to approximately $1138[11], corrective labor, restriction of liberty, or imprisonment for up to three years. For "large" and "especially large" quantities, the punishment is imprisonment for three to ten and ten to fifteen years, respectively. In the case of voluntary surrender of drugs to a police officer and active assistance during the investigation, an offender is exempted from criminal liability. If the quantity of drugs seized is less than significant, only administrative penalties of a fine of up to approximately $142[12] or arrest for up to fifteen days can be imposed.

Drug sale is punishable by imprisonment for four to eight years if the quantity is less than "significant", eight to fifteen years for a "significant" quantity, and ten to twenty years for a "large" quantity. "Especially large" quantities carry a fifteen to twenty years or life sentence. Thus, the crime is serious if the quantity of drug seized is less than "significant", and most serious if the quantity is "significant" or higher.

The practice of plea bargaining was introduced in 2001. Pleading guilty significantly simplifies the procedure: a conviction is pronounced without the actual examination of evidence at a court hearing. In addition, a person who accepts a plea bargain waives the right to appeal. In return, by pleading guilty, the offender reduces the maximum punishment by one third. It is important to note that the plea agreement does not lessen the charge and only provides a discount for the highest penalty prescribed for a particular offense. The official criminal procedure involving the plea bargain is as follows. After the investigation is finished and a prosecutor has filed the charges, the accused receives the

[11] All amounts are expressed in US dollars using the average 2013-2014 exchange rate (RUB/USD = 35.16).

[12] All amounts are expressed in US dollars using the average 2013-2014 exchange rate (RUB/USD = 35.16).

case materials to review. At this stage, the police officer who investigated the case should explain to the accused that s/he has a right to apply for a plea bargain. This application has to be approved by the prosecutor first and then decided by a judge in court (Antonov i Partnery 2016). In practice, police officers often offer the plea agreement already at the initial stage of the investigation. If the person accepts it and signs the relevant documents in advance, the police officer saves himself a lot of time and effort on the careful collection of evidence, especially, when the credibility of evidence "collected" could be in doubt, as in the case of drug manipulation (Dombrovickij, n.d., Ajvazyan, n.d.).

According to Titaev and Pozdnyakov (2012), in general, pleading guilty in Russia does not reduce the sentence significantly and even worsens the offender's situation in some cases. Nevertheless in 2013-2014, almost 60% of all cases (30% of drug-related offenses) were processed under a plea agreement. This quite large share could be explained by the legal illiteracy of offenders who simply do not know how the plea bargain may influence their legal situation and accept the agreement at the very beginning of the investigation.

The performance evaluation of anti-drug agency personnel. During the 2013-2014 period, when both the FSKN and the MVD were responsible for enforcing drug laws, each had their own officers' performance evaluation system.

The system used by the FSKN was based on performance indicators that included the total number of drug crimes and the number of serious and most serious drug crimes (per 100 officers). At the end of each year, for each indicator, the FSKN regional offices received a position in cross-region ratings. The final evaluation was determined by the overall rank of the office in relation to other offices based on these ratings. Local subdivisions (stations) at the level of municipal districts or cities within the same region were not competing with each other. However, they could receive unofficial guidance from their regional supervisors regarding the approximate level of performance indicators that should be targeted. This translated further into individual-level instructions for officers within each station.

In contrast to the FSKN, the MVD stations, individually and together as a whole regional office, compared their performance indicators, including the total number of drug crimes and the number of serious and most serious drug crimes (per 100 officers), with their own numbers in the previous year. At the end of each year, they had to show positive dynamics[13] (Novikova 2014). In order to do that, station supervisors assigned monthly, or sometimes even daily, "plans" (targeted levels of performance indicators) to each officer.

If the MVD officer met or surpassed the targets, s/he might receive a monetary reward or promotion (for high-profile cases). While there was no guarantee that the officer would be rewarded for good performance, s/he certainly was reprimanded, warned or even fired in the case of unsatisfactory performance. In addition, the officer could be deprived of monthly bonuses if s/he did not fulfill the plan. The FSKN officers faced a similar rewards and punishment system; however, the definition of "satisfactory" and "unsatisfactory" performance was vaguer than in the MVD. In both agencies, decisions to reward or punish were made by immediate superiors at each level.

Thus, both performance evaluation approaches presented strong incentives for police officers to show the required level of cases solved, and the whole justice system created an enabling environment for fabricating these cases when other tools to meet the targets were not available. It was not only police officers who were evaluated based on easily measurable and quantifiable indicators, but also prosecutors and judges. The evaluation of the prosecutors was linked to the number of convictions, while acquittals were considered "lost" cases and negatively affected the evaluation. Judges were evaluated by the number of appeals and by the "confirmation rate" of their decisions at the higher-instance courts (Schultz et al. 2014). This system incentivized the law enforcers to strategically bargain with each other, which further enabled the police to select the "right" cases to prosecute, falsify evidence, and pressure defendants etc. (Paneyakh 2014). At the same time, police officers in Russia are rarely punished for misconduct. For instance, lawyer Dmitriy Egoshin from the Public Verdict Foundation estimates the share of the police abuse of power cases that does not reach the courts at 70-80% (Torocheshnikova and Sotnikov 2021)[14].

[13]Formally, after the reforms in 2011, the MVD offices had to use an overall score calculated by the complicated formula and to compare their performance across units rather than relative to the previous period (Novikova 2014). However, locally this did not work due to the complexity of the system (Kolegov 2012, Titaev 2019).

[14]Moreover, in 2013-2014, around 60% of the police abuse of power cases that reached the court resulted in conditional sentence and around 15% - in fine, restriction of liberty, or exemption from punishment (Torocheshnikova and Sotnikov 2021). There is no estimation for the share of evidence falsification cases that did not reach the court. However, it was determined that, in 2013-2014, most

Meanwhile, even though the FSKN's performance evaluation system was more transparent, it was more difficult for the FSKN stations to set the "necessary" level of performance indicators, since it had to take into account the performance of other stations in the current period. In contrast, the MVD officers always knew what numbers they should reach. These institutional features could significantly contribute to the difference in the magnitudes of bunching observed for these two agencies, which I investigate in more detail in Subsection 4.1.

2.2 Data

This paper uses a database provided by the Institute for the Rule of Law at the European University at St. Petersburg, Russia[15]. It contains information on almost 300,000 drug crimes reported in Russia during 2013-2014. The information is based on five forms that are created at the different stages of the investigation of a specific case and include the following data:

form 1: identified crime and investigation results;

form 2: socio-economic characteristics of offenders;

form 3: criminal proceedings;

form 4: reparation for damages and the seizure of crime objects;

form 6: trial results[16].

Knorre and Skougarevskiy (2015) and Skougarevskiy (2017) extracted and analyzed all information on primary drug types, weights of drugs seized, offenders' characteristics and court decisions from this database. I follow their approach. Both forms 1 and 4 contain information on weights of drugs seized. Form 1 quantities are determined by a

of the convicted police officers received a conditional sentence (50% of cases), or either fine, restriction of liberty, or exemption from punishment (45% of cases) (Nikonov 2020). In addition, police officers are usually more likely to be acquitted than the general population: for example, in 2013-2014, the share of acquittals for police officers was around 5% while for other offenders it did not even reach 1% (Torocheshnikova and Sotnikov 2021).

[15]Initial data was compiled and prepared at the Institute for the Rule of Law at the European University at St. Petersburg with support from the Russian Science Foundation grant 17-18-01618.

[16]Form 1 is completed by an investigator when s/he decides to initiate criminal proceedings that as to be approved by a prosecutor. During the investigation, forms 2, 3 and 4 are created. All these forms have to be checked by the prosecutor's office before referring the case to the judicial authorities. Form 5 is not in the database since it contains information on victims, while drug crimes are victimless. Form 6 is filled in by a judge. After closing the case, all forms are converted from written to electronic form and submitted to an information center (Shklyaruk and Skougarevskiy 2015).

police officer, who has to weigh the seized drugs, while form 4 is created at a later stage after the prosecutor's approval of case initiation and contains drug quantities measured in a laboratory. Therefore, to investigate the behavior of police officers and the extent to which it could be affected by their performance evaluation, I use form 1 as a primary source of data. In order to identify characteristics of possible victims of manipulation, I merge data from forms 1 and 2. For the investigation of manipulation consequences, I turn to combined data from forms 1, 2, and 6, merged with drug weights from form 4. Weights from both forms are needed for estimating the local average treatment effect (LATE) of possible manipulation on sentence length and case characteristics in accordance with the Diamond and Persson (2016) approach. In this exercise, I restrict the sample to cases related only to drug use for the following reason. In contrast to drug use offenses, the length of imprisonment in drug sale cases might be determined not based on the weight of the drug itself but based on, for example, the amount of money seized in connection to the sale of that drug. Therefore, one needs to analyze the sentence length in these two main groups of offenses separately. At the same time, the sample of drug dealers from the merged dataset based on forms 1, 2, 6, and 4 contains an insufficient number of observations for bunching techniques.

Weights from the two forms coincide for 92.8% of heroin-related cases and for 99.2% of cases related to other drugs. This could be explained by the fact that, in data, only heroin quantities are presented in milligrams, while all other drugs are in grams. Nevertheless, the distributions of heroin cases across drug quantities from the two forms do not differ significantly (Appendix A, Figure A2). Additionally, I run the regression analysis to check whether there is any particular pattern in the weight differences at the form 1 heroin quantities above the thresholds. The results in Table B4, Appendix B suggest that the difference in weights is slightly smaller if the weight from form 1 is above the third threshold. However, this does not affect the paper's analysis because I exclude the long tail of the distribution covering the third threshold for the reasons outlined in Subsection 3.1.

Merging cases across forms decreases the sample for two reasons. First, some of the cases are dismissed at different stages of investigation so not all the cases registered in form 1 have the information in other forms. Second, since the cases do not have a unique identifier across forms, data on them have to be merged based on the combination of the year of reporting, region, agency, police station code, and case number. Due to missing or incorrect data, not all of the cases could be merged across forms. Nevertheless, as Figures A2 and A3 in Appendix A suggest, the distributions of cases and magnitudes of bunching are similar across these gradually merged samples.

The initial dataset based on form 1 contained data on 518,979 drug crimes, including 89,152 heroin cases. 16% of cases related to heroin were excluded from the sample because the quantity of drug seized was missing[17]. Table B5 in Appendix B presents the comparison of means in subsamples with and without missing values for four main samples used in this paper. Missing weight values are likely to be caused either by inaccurate completion of paper forms by police officers or by mistakes during the conversion of these forms into electronic files. Additionally, under some circumstances, a case can be initiated without a drug seizure. For example, as the table suggests, the weights are more likely to be missing for drug sale cases. This could be explained by the fact that one can be accused of selling drugs if police find marked bills in his/her possession. Furthermore, as expected, the documents are more complete for more serious crimes (with longer sentences), when there is conclusive evidence (being arrested under the influence of drugs) or in the case of refusing to plead guilty, which leads to a full investigation, compared to the simplified procedure under the plea bargain.

For the bunching analysis, a serious issue would be if the information on heroin seizures was missing at particular drug weights. However, Figure A4 in Appendix A provides suggestive evidence that this is not likely the case in my data. The graph depicts the distribution of cases across drug weights from form 4 for which weights in form 1 are missing. It suggests that missing quantities from form 1 are likely to be concentrated mainly between 0 and 1 gram but, at the same time, they are distributed similarly to non-missing quantities from both form 1 and form 4. Above 1 gram, the cases with missing form 1 weights are distributed almost uniformly without any particular pattern around the second threshold, which I primarily focus on in this paper[18]. In regression analysis, I am able to control for police station and region-by-month fixed effects, which

[17] The form 4 dataset included information on 236,989 drug crimes out of which 50,782 were related to heroin. Due to missing drug weights, 9% of heroin-related cases were also excluded from the analysis.

[18] See Subsection 3.1 for more details on thresholds, and why I mainly study the second cutoff.

mitigates the potential effect of dropping the observations with missing weight values.

3 Empirical Strategy

3.1 Analyzing Incentives

3.1.1 Event Study

I start the analysis with an event study to investigate whether the expected end-of-year performance evaluation affects the behavior of the police during the year. I run this analysis separately for two main performance indicators: the total number of drug crimes and the number of serious and most serious drug crimes. For each station s in region r, I calculate the total number of crimes (the number of serious and most serious crimes) per day t and, comparing their cumulative value with the 2013 level, determine when the station reached this level. This allows me to define a set of event study dummies $Reached^{m}_{srt}$: 1 for the event "month" (the event day when a police station reached its previous year's "benchmark" and 29 days before that), 5 for pre- and 5 for post- 30-day periods, since, on average, stations reach both levels needed between July and August. The inverse hyperbolic sine transformation (IHST) of either daily total number of crimes or daily number of serious and most serious crimes, $Number_{srt}$, is my main outcome, which I regress on event study dummies and the set of fixed effects:

$$Number_{srt} = \sum_{m=-5}^{5} \beta_m Reached^{s}_{srt} + Station_s + Day_t + Month \text{ x } Region_{rt} + \varepsilon_{srt}. \quad (1)$$

In the regression, I include station fixed effects, $Station_s$, to control for time-invariant station characteristics, day-of-the-month fixed effects, Day_t, to control for monthly time trends, as well as region-by-month fixed effects, $Month$ x $Region_{rt}$, to account for region-specific annual time trends. Standard errors are clustered by station. The identification strategy exploits the variation in the timing of reaching the previous year's performance level for a given station.

There are two types of police stations that could be included in a control group: "not-yet-treated" and "never-treated". Since "never-treated" stations in the given context might exhibit slightly different behavior than "treated" ones, my preferred approach is to restrict the sample to only those stations that surpassed their 2013 "benchmark" (conditional on it not being zero[19]) at some point during the study period (2014). These are around 43% of MVD stations and 32% of FSKN stations that reached or bettered the 2013 target for the total number of drug crimes, and around 38% of MVD stations and 33% of FSKN stations for the number of serious and most serious drug crimes (Table B6, Appendix B). To test the sensitivity of the results to this selection criterion, I repeat the analysis in Appendix B adding stations that did not reach the 2013 performance level[20]. This allows me to compare the results of the event study with the bunching estimates, which are calculated only for the extended sample due to the requirement of a sufficient sample size.

3.1.2 Bunching

Since the event study can provide only general insights about the police officers' behavior but does not tell anything about the observed bunching, I turn to the standard bunching estimator and bunching regression analysis. I focus on heroin seizures because, in the data, only heroin quantities are presented in milligrams, while all other drugs are in grams. This fine scale for heroin allows me to study discontinuities in the distribution of cases around the thresholds at a granular level, choosing the bunching estimator's parameters more precisely. Even more, for some drugs that are measured in grams, it is not possible to run this type of analysis because both of the first two thresholds are between 0 and 1 gram. At the same time, those drugs which are measured in grams but have higher thresholds exhibit similar to heroin discontinuity around the first threshold only (Appendix A, Figure A5). In Subsection 4.1.2.1, I discuss possible reasons for this difference in distributions in more detail.

The standard bunching estimator method (Saez 2010, Chetty et al. 2011, Kleven and Waseem 2013) allows to construct a measure of excess mass of offenders above a

[19]There were no stations that did not register crimes at all, but there were around 18% of MVD stations and 14% of FSKN stations that did not register any severe cases in the previous year and, therefore, had a zero "severe" target in 2014.

[20]There were around 58% of MVD stations and 68% of FSKN stations that did not reach the 2013 tager for the total number of drug crimes, and around 44% of MVD stations and 54% of FSKN stations for the number of serious and most serious drug crimes.

threshold by comparing actual and counterfactual distributions around this threshold. The counterfactual density of seized drug quantities is estimated by fitting a high-order polynomial to the observed distribution, excluding the manipulation region. To estimate bunching at a particular threshold, the parts of the distribution covering other thresholds need to be excluded from the analysis (see Appendix C for further details).

In the bunching analysis, I mainly focus on the second threshold for several reasons. I do not calculate the bunching estimator for the first threshold because, most importantly, the number of weight bins that could be defined below and above the manipulation window around this threshold is insufficient for estimating the counterfactual distribution. Additionally, data on offenses below the first threshold could be incomplete due to police officers' reluctance to deal with cases that do not affect their performance evaluation directly, without being moved above the threshold[21]. Registering a case requires a significant amount of paperwork that could discourage the police officers from doing so if it does not improve the evaluation. Moreover, some officers might show leniency towards minor offenses and not register them. Since calculating the bunching estimator for the first threshold is technically infeasible, I study the probability of bunching above this threshold in the regression analysis; however, the results should be interpreted with caution. Further details are provided below in this subsection.

The third threshold for heroin is at a large quantity - 500 grams. Criminal cases around this threshold have low frequency and, therefore, their distribution is presented in the separate graph in Appendix A, Figure A6. As the graph shows, there is a mass of cases on both sides of the threshold with the peak at the threshold, and the pattern is similar to the distribution of cases around other large round quantities of heroin (100, 200, etc. grams). This pattern is more pronounced for drug sale cases, which suggests that the observed distribution of cases at 500 grams (and other large quantities) might rather follow the distribution of large-scale drug traffickers on the illicit drug market than reflect the police's response to the punishment threshold. At the same time, there are no clear incentives for police officers to have more cases above the third threshold compared

[21] Drug use cases below the first threshold are administrative (non-criminal) offenses which are not taken into account for the performance evaluation. In contrast, drug sale cases below the first threshold are serious offenses and do contribute to the evaluation; however, they could be requalified to drug use cases and lose their value for the police.

to the cases below it. In the case of drug sale, offenses on both sides of the threshold are considered to be most serious, while in the case of drug use, crossing the third threshold increases the seriousness of the crime from serious to most serious. However, in neither of these two cases does moving offenders from below to above the threshold contribute to the overall performance evaluation, because the number of serious and most serious crimes are calculated together. Therefore, I do not investigate the discontinuity at the third threshold, and even exclude the long tail from the analysis, since, qualitatively, it is likely to be uninformative for studying the second threshold. Furthermore, the observed bunching of cases at round drug quantities, especially large ones (Appendix A, Figure A6) could contaminate the estimation of the counterfactual distribution.

To be able to calculate the standard bunching estimator for the second threshold, I need the regularity assumption, i.e., the distribution is "well-behaved" around the threshold in the absence of manipulation. This assumption allows me to use data outside the manipulation window to restore the counterfactual distribution. The estimation suggests a smoothly decreasing shape of the distribution in the counterfactual world. Indirect evidence in support of this shape comes from the distributions of cases related to the other types of drugs (Appendix A, Figure A5), which do not have bunching above the second threshold. If we assume that offenders are rational agents, we could expect a counterfactual distribution with humps just below the thresholds. However, the voluntary bunching of offenders below the threshold is more likely to be observed if punishment increases discontinuously for any quantity exceeding the limit (Traxler et al. 2018, Lepage 2020). According to Russian anti-drug law, in the counterfactual world without manipulation, the punishment should increase smoothly without shifts at the thresholds in the case of drug use or overlap in the case of drug sale. In Subsection 4.1.2.1, I present further arguments explaining why the distribution of cases is unlikely to have bunching above the threshold in the absence of manipulation.

Using the bunching estimator, I start with estimating the magnitude of bunching of offenders above the second threshold in the full sample with the drug weights from form 1. To check that the results are insensitive to the choice of estimation parameters, I repeat the procedure described in Appendix C, using different polynomial orders, values of upper bounds, starting points after the exclusion of the area around the first threshold, and endpoints after the exclusion of the long tail covering the third threshold. At this stage, I choose the baseline parameters of the bunching estimator that will be used in the further analysis of bunching consequences.

Next, I investigate how the magnitude of possible manipulation around the second threshold varies across agencies. Then, I test whether reaching the target linked to the number of serious and most serious crimes affects this magnitude in two agencies differently. For the performance evaluation, the number of serious and most serious crimes is calculated per 100 officers. I do not have information on the size of each regional office; therefore, I cannot exploit the FSKN's cross-region comparison scheme[22]. Instead, assuming that the number of officers at each station is fixed during the 2013-2014 period, I determine the total absolute number of serious and most serious drug crimes solved by each station in 2013. Then, I divide all cases initiated in 2014 (by agency) into two groups: before and after achieving each station's 2013 level, and check whether the bunching varies between these four groups. To calculate the bunching estimator, I need as many observations as possible; therefore, I include both stations that reached their 2013 targets and stations that did not reach these targets during 2014.

Finally, I support the bunching analysis with the regression estimates of the bunching probabilities using the following specification run for each agency separately:

$$Bunch_{isrt} = \beta_0 + \beta_1 Post_{isrt} + X_i + Station_s + Day_t + Month \text{ x } Region_{rt} + \varepsilon_{isrt}, \quad (2)$$

where $Bunch_{isrt}$ equals 1 if the heroin weight associated with case i registered by station s in region r on day t is in the bunching window above the second threshold (between 2.5 and 3.3 grams). The variable of interest is $Post_{isrt}$, which indicates whether case i was registered before or after station s reached the previous year's number of serious and most serious crimes. In addition to controls, X_i, for the type of offense (drug use or drug sale), I also include station fixed effects, $Station_s$, to control for time-invariant station

[22] I conducted an exploratory analysis of the effect of the cross-region comparison scheme, assuming the number of officers to be proportional to the total number of drug crimes solved by each station during each year. First, I identified the FSKN and MVD stations with the highest relative number of serious and most serious drug crimes solved in 2013. Second, for each agency, I split all 2014 cases into two groups: before and after reaching the "best" level of 2013 (established by either the FSKN or MVD station). Finally, I calculated the bunching estimate for four groups of cases. The difference in magnitudes appeared to be insignificant.

characteristics, day-of-the-month fixed effects, Day_t, to control for monthly time trends, as well as region-by-month fixed effects, $Month$ x $Region_{rt}$, to account for region-specific annual time trends. Standard errors are clustered by station. To avoid the contamination of the estimation results by the presence of other thresholds, I cut the sample at 0.8 gram from the left and at 10 grams from the right, as is done in the part with the bunching estimator (see Subsection 4.1.2.2 for details on the bunching estimator's parameters). Similar to the event study, the preferred control group here is "not-yet-treated" stations. Their combination with the stations that never reached the 2013 performance level is used for robustness check.

I also conduct this regression analysis for the first threshold; however, the results should be interpreted with caution due to potentially missing cases below the first threshold. I code $Bunch_{isrt}$ as equal 1 if the heroin weight associated with case i is in the bunching window above the first threshold - between 0.5 and 1 gram. Since the incentives at the first threshold are linked to the different performance indicator, $Post_{isrt}$ is redefined to indicate whether case i was registered before or after station s reached the previous year's total number of drug crimes. The sample is cut by the starting point of the bunching window around the second threshold - at 1.6 grams (see Subsection 4.1.2.2 for details on the bunching estimator's parameters).

3.2 Identifying Victims of Possible Manipulation

In order to recover the characteristics of those who were potentially manipulated by the police, I adopt the technique designed by Diamond and Persson (2016). As in the part with the bunching estimator, I cut the sample at 0.8 gram from the left and at 10 grams from the right (see Subsection 4.1.2.2 for details on the bunching estimator's parameters).

First, I estimate the counterfactual expected values of observable characteristic Y at any drug midpoint R of bin j inside the manipulation area $[r_l, r_u]$ if there was no manipulation, using cases outside of this area:

$$Y_j = \sum_{k=0}^{p} \beta_k R_j^k + \varepsilon_j, \tag{3}$$

where $R_j < \overline{D} - r_l$ or $R_j > \overline{D} + r_u$, $\overline{D}$ is a threshold. Then I calculate the observed average values of characteristic Y for offenders inside the manipulation region below ($\overline{Y}^{never}$) and above ($\overline{Y}^{up}$) the threshold $\overline{D}$:

$$\overline{Y}^{never} = \frac{1}{N^{never}} \sum_i Y_i, \text{ where } \overline{D} - r_l \leq r_i < \overline{D}, \tag{4}$$

$$\overline{Y}^{up} = \frac{1}{N^{up}} \sum_i tY_i, \text{ where } \overline{D} \leq r_i \leq \overline{D} - r_u. \tag{5}$$

Here $\overline{Y}^{never}$ is the average characteristic of those offenders who were arrested with the quantity of drugs just below the threshold and were not selected for manipulation ("never-takers"):

$$\overline{Y}^{never} = \frac{N^{down}}{N^{down} - N^{compliers}} \overline{Y}^{down} - \frac{N^{compliers}}{N^{down} - N^{compliers}} \overline{Y}^{compliers}. \tag{6}$$

Accordingly, $\overline{Y}^{up}$ is the average characteristic of all those offenders who were manipulated ("compliers") and who actually were arrested with a drug quantity just above the threshold ("always-takers"):

$$\overline{Y}^{up} = \frac{N^{always}}{N^{always} - N^{compliers}} \overline{Y}^{always} - \frac{N^{compliers}}{N^{always} - N^{compliers}} \overline{Y}^{compliers}. \tag{7}$$

Using the estimates of the counterfactual values of observable characteristic $\hat{Y}$ and distribution of cases $\hat{C}$, I obtain values of $\overline{Y}^{down}$ and $\overline{Y}^{always}$ in the following way:

$$\overline{Y}^{down} = \frac{\int_{\overline{D}-r_l}^{\overline{D}-\sigma} \hat{Y}_j^R \hat{C}_j dR}{N^{down}} \tag{8}$$

$$\overline{Y}^{always} = \frac{\int_{\overline{D}}^{\overline{D}+r_u} \hat{Y}_j^R \hat{C}_j dR}{N^{always}}. \tag{9}$$

The number of offenders in each part of the manipulation region can be calculated as:

$$N^{never} = N^{down} - N^{compliers}, \text{ where } N^{down} = \int_{\overline{D}-r_l}^{\overline{D}-\sigma} \hat{C}_j dR, \tag{10}$$

$$N^{up} = N^{always} + N^{compliers}, \text{ where } N^{always} = \int_{\overline{D}}^{\overline{D}+r_u} \hat{C}_j dR. \tag{11}$$

Plugging these into (6) and (7) and using estimates from (4), (5), (8), and (9), I solve for the compliers' average value of characteristic Y:

$$\overline{Y}^{compliers} = 0.5(\frac{N^{never}}{N^{never} - N^{down}}\overline{Y}^{never} - \frac{N^{down}}{N^{never} - N^{down}}\overline{Y}^{down})+ \\ + 0.5(\frac{N^{up}}{N^{up} - N^{always}}\overline{Y}^{up} - \frac{N^{always}}{N^{up} - N^{always}}\overline{Y^{always}}). \quad (12)$$

Finally, I compare the mean characteristics of those offenders who were potentially manipulated by the police ("compliers") with the mean characteristics of all offenders who were "eligible" for manipulation but did not receive it ("never-takers"):

$$\Delta Y = \overline{Y}^{never} - \overline{Y}^{compliers}. \quad (13)$$

3.3 Estimating the Effect of Possible Manipulation on Sentence Length

I identify the effect of possible manipulation of drug quantities on sentence length (and on other case characteristics) in two steps, again following Diamond and Persson (2016). As in the part with the bunching estimator, I cut the sample at 0.8 gram from the left and at 10 grams from the right (see Subsection 4.1.2.2 for details on the bunching estimator's parameters).

First, I estimate the relationship between sentence length, S, and the unmanipulated quantity of drugs seized from form 1:

$$S_j = \sum_{k=0}^{p} \beta_k R_j^k + \gamma_R * \mathbb{1}[R_j \geq \overline{D}] + \omega_j, \quad (14)$$

where $R_j < \overline{D} - r_l$ or $R_j > \overline{D} + r_u$ is to ensure that the drug quantity (and related sentence) data used to estimate equation (14) is outside the manipulation region $[r_l, r_u]$ around the threshold $\overline{D}$. This equation yields the expected sentence length at each drug quantity in the counterfactual world without manipulation.

Then, I calculate the counterfactual expected sentence length across the whole set of drug offenders inside the manipulation region, using estimates $\hat{S}_j$ from equation (14) and estimates of the counterfactual distribution of cases $\hat{C}_j$ from equation (C1):

$$\overline{S} = \int_{\overline{D}-r_l}^{\overline{D}+r_u} \hat{S}_j \frac{\hat{C}_j}{\int_{\overline{D}-r_l}^{\overline{D}+r_u} \hat{C}_j} dR. \tag{15}$$

Comparing observed and estimated counterfactual average sentence lengths, I obtain the "intent-to-treat" effect. It shows a difference in the lengths of imprisonment in two worlds, with and without manipulation, for the offenders who have been caught with the quantity of drugs that falls within the manipulation region:

$$ITT = \frac{\sum_{i \in manip\, region} S_i}{N^{manip}} - \overline{S}, \tag{16}$$

where N^{manip} is the number of offenders in the manipulation area.

The procedure described above is repeated with drug quantities from form 4 instead of sentence length. This constitutes the effect of being manipulated on the quantity of drugs seized that is determined officially at the laboratory and then considered by a judge in court. The ratio of ITT from (16) to this first stage effect, in turn, identifies the local average treatment effect (LATE) of being manipulated on the sentence length.

I repeat the procedure described above to determine the effect of possible manipulation on the probability of pleading guilty, length of investigation, length of adjudication, and probability of pretrial detention.

4 Results

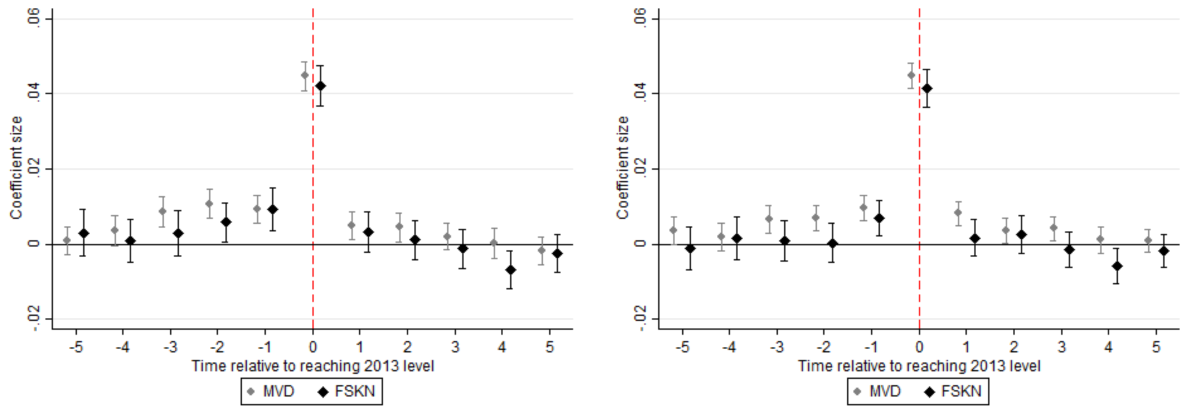
4.1 Performance-Based Incentives for Police Officers

4.1.1 General Effect on Police Behavior: Event Study

Since the previous year's performance presents a direct target in the case of the MVD or a noisier benchmark in the case of the FSKN, reaching this level could significantly affect the behavior of the police during the current year. The regression results for the event study with the sample of stations that reached their previous year's level of performance (conditional on it not being zero) are shown in Figure 2 and Table B8 (columns (1)-(4))

in Appendix B. They suggest a common pattern for both performance indicators for the two agencies: around 4% increase in the number of (serious and most serious) drug crimes registered in the 30-day period, by the end of which the station reaches the 2013 benchmark, and a slight upward pre-trend in the two 30-day periods before it. After the "event", the number of cases returns to its average level before the jump, and this happens more quickly at the FSKN stations. This could be explained by the presence of other performance indicators for the MVD. In contrast to the FSKN, the MVD deals with all types of crimes, not only drug-related ones. Therefore, it is also evaluated based on the total number of *all* crimes and the number of *all* serious and most serious crimes. Drug crimes contribute to these performance indicators that could, to some extent, "disturb" the estimation results for the MVD. Additionally, I find the similar patterns for heroin-related cases only, the distribution of which is characterized by the significant bunching above both thresholds. (Figure A1, Appendix A and Table B9, Appendix B).

Figure 2: The effect of reaching the 2013 level on the total number of drug crimes (left) and the number of serious and most serious drug crimes (right) registered in 2014



Note: The samples include all MVD stations and all FSKN stations that reached the total 2013 number of drug crimes (left) and the 2013 number of serious and most serious drug crimes (right) during 2014. The dependent variable is the inverse hyperbolic sine transformation of either the daily total number of drug crimes or the daily number of serious and most serious drug crimes calculated based on the sample of all drug-related cases from form 1. The specification includes station, day-of-the-month and region-by-month fixed effects. Standard errors are clustered by station. The regression results are reported in Appendix B, Table B8.

While the event study analysis cannot precisely identify the cause of the observed pattern, it is possible that both legitimate increased effort and unlawful moving of offenders from below to above the thresholds by the police are contributing factors. Nonetheless, this analysis provides valuable evidence indicating that the performance evaluation system impacts police behavior, which could be harmful, even in the absence of any deliberate misconduct. Specifically, variations in police officers' effort during the year could lead to differences in the likelihood of crimes being registered, which may result in inequality in the enforcement of law. If police officers do behave dishonestly, for example, manipulate the drug quantities, then the results of the event study are consistent with the idea that manipulation is risky, and when the performance target is still far away, it is unclear whether benefits from manipulation will outweigh its costs. However, the closer the target, the clearer the gain from manipulating drug quantities becomes. Once the target is reached, police officers stop manipulating, also because they do not want to ratchet up the target for the next year. In Subsection 4.1.2, I focus specifically on heroin-related cases to investigate the potential for fraudulent behavior among police officers in response to the performance-based incentives.

Turning to those stations that had zero previous year's level and stations that did not reach their non-zero previous year's levels, I calculate mean monthly numbers of (serious and most serious) drug-related offenses registered by these stations in 2014. The results presented in Figures A7 and A8 in Appendix A suggest that the MVD and FSKN stations both behave in a similar way. However, the motivation behind this behavior could differ in the two agencies due to the difference in the performance evaluation approaches. Thus, the MVD stations with a zero 2013 level registered a low number of severe cases in 2014, because they do not have any incentives to increase the numbers that would become their direct targets for the next year. The MVD stations that did not reach their previous year's performance level registered a slightly higher number of cases per month than the stations that did reach the targets. At the same time, they did not increase effort towards the end of the year because their targets were too far away (Table B6, Appendix B). In the case of the FSKN, stations do not directly target their previous year's performance but rather compete with each other. Therefore, stations with a zero 2013 level did not improve their performance in 2014, probably because they anticipated that, in any case, they would be unable to reach the top of the ranking in 2014. Furthermore, stations that did not reach their own non-zero 2013 performance level might have performed well in 2014 compared to other stations and, therefore, did not need to register as many (severe) cases as they did in the previous year.

In columns (5)-(8) of Table B8 in Appendix B, I present the results of the event study for the extended sample which includes both stations that reached their previous year's performance level and stations that did not reach it. The estimates suggest a similar pattern with a jump in the "event" month. However, the magnitude of the effect is almost twice lower than with the restricted sample. This supports the point outlined in the previous paragraph: stations that did not reach the target probably do not increase their effort towards the end of the year because either they are too far away from the target (for the MVD) or they perform well compared to other stations and do not need to reach the previous year's target (for the FSKN).

4.1.2 Possible Manipulation of Drug Quantities

Combining the information on sanctions for drug-related crimes and the systems of performance evaluation of police officers discussed in Subsection 2.1 suggests the following performance-based incentives for moving offenders across the thresholds. In the case of *drug use*, crossing the first threshold changes the status of an offense from administrative (non-criminal) to criminal, which contributes to the performance evaluation of police officers by increasing the total number of drug crimes solved. Crossing the second threshold increases the seriousness of crime from least serious to serious, which in turn increases the number of serious and most serious drug crimes and positively affects the performance evaluation. The incentive around both thresholds in the case of *drug sale* is ambiguous, since crossing them does not directly contribute to performance indicators. However, police officers may still have the incentive to move offenders above the threshold due to their concern about losing "points" if a drug sale case below the threshold is requalified as a drug use case (e.g., possession without the intent to sell). Nonetheless, if an offender is moved above either threshold, a requalification only reduces the severity of the crime but maintains either the total number of drug crimes or the number of serious and most serious drug crimes at the same level.

4.1.2.1 Bunching Pattern

This subsection discusses the observed pattern in the distribution of heroin cases and potential mechanisms that might explain it.

As the estimation suggests, the excess mass above the second threshold for all heroin-related cases from form 1 is almost six times greater than the average number of cases that would be in the bunching window above this threshold in the counterfactual world (Appendix A, Figure A2). One of the potential mechanisms behind the observed pattern could be manipulation of drug quantities around the threshold. There is anecdotal evidence, various media reports, and personal stories posted on online forums[23] that document that manipulation of drug quantities was (and still is) a widely-used method in Russia to fabricate evidence. For example, in an interview, one of the former policemen (Nadezhdin and Matveeva 2019) describes the following scheme often used by police officers. The police station receives a call from somebody reporting that there are drug users under the influence of drugs in the hall of his building. The police arrive and find unconscious people and a quantity of, for example, heroin. If the drug quantity is below the threshold, the officers could add flour, sugar, or any other white powder in order to arrest the users for a more severe crime[24].

In contrast to merely changing the numbers in documents, increasing the actual drug weight is less likely to be detected. If the police officer puts the wrong number (higher drug weight) on the form at the very first stage when s/he seizes the drugs, then s/he should have an expert at the laboratory who will be able to "confirm" the drug quantity at the next stage of the investigation. However, the laboratory staff does not have the same performance-based incentives as the police, but they can be punished for misconduct. If the police officer decides to change the number after the laboratory has recorded the weight, it could be questioned by the offender who has to be provided with a copy of the protocol filled in with the actual number during the seizure. In addition, an appeal could lead to the re-examination of the drugs seized[25].

Other potential mechanisms behind the observed discontinuity in the distribution might include differential treatment by the police (other than moving some offenders around the threshold), the nature of demand and supply in the illicit drug market, and specific features of the criminal justice system. However, first, there are no incentives for

[23]For example, see Nadezhdin and Matveeva (2019), Merzlikin (2019), Antonov (2019).

[24]This is possible because, according to the law, the drug quantity seized is determined not by the weight of the pure drug substance but by the weight of the entire mixture.

[25]See Shklyaruk and Skougarevskiy (2015) for more details on how various documents (forms) are completed during the investigation.

police officers to focus particularly on offenders who are just above the second threshold and ignore those who are further to the right. The same logic applies to the hole in the distribution directly below the threshold: the offenders just below the second threshold are as "valuable" to police as other offenders who are further to the left (but above the first threshold). The performance evaluation of police officers also depends on the total number of all criminal drug cases solved during the year, not only serious and most serious ones. Therefore, police officers are motivated to "keep" every offender (above the first threshold in the case of drug use) for the end-of-year performance evaluation.

Second, it is unlikely that the observed pattern reflects the distribution of drug quantities sold on the market and consequently possessed by the drug users. As Filchenko and Zhandrov (2022) document, street heroin in Russia was (and still is) packed into 0.3-0.8 gram bags. At the same time, anecdotal evidence based on news about drug seizure cases suggests that heroin was sold in other quantities as well, for example, 1.8[26] and 3.1[27] grams. This could be because there is no strict definition of a dose. The quantity of drug needed to achieve the desired effect depends on the purity of the drug and the past usage of the drug addict. Street heroin is often mixed with some additives, such as sugar, flour or starch, to increase the weight for retail sale and/or with other drugs to increase the effect of the heroin. The composition of the mixture and the weight of the bags are up to drug dealers, and heroin users with different tolerance levels are buying different quantities of drugs to achieve the same effect (Cunha-Oliveira et al. 2013).

Finally, in contrast to some other countries (for example, the US (Tuttle 2019)), law enforcers in Russia do not have legal tools to either increase or decrease the quantity of drugs presented in the court. The drug weights, as in Figure 1, are registered during the seizure and remain unchanged until the case is closed. If an offender accepts a plea bargain, the maximum punishment (usually sentencing) is lowered by one third but the actual drug weight registered remains the same in the documents.

If the observed pattern in the distribution is due to manipulations, then the explanation for discontinuity around the second threshold only in the case of heroin could be that it is less costly and technically easier for police officers to manipulate

[26]See, for example, Belgorod news (https://moe-belgorod.ru/news/incidents/1011942).

[27]See, for example, Saratov news (https://www.4vsar.ru/news/67932.html?ysclid=lakxs9fs7v750990618).

heroin weights[28]. In contrast to other drugs often seized in Russia[29], only cocaine and heroin are usually sold in the form of white powder and, therefore, their quantities can be easily increased by, for example, adding any non-drug white powder. Amphetamine and its derivatives appear as a white powder too; however, it is not their only form: amphetamines are also sold as tablets or crystals, which vary in color from almost clear to white, yellow, and brown (Cunha-Oliveira et al. 2013). Furthermore, both of the first two thresholds for amphetamines are between 0 and 1 gram that together with the quantities being measured in grams could mask any potential discontinuities around these thresholds. In contrast, cocaine, also measured in grams, has the second threshold at 5 grams but there is no observed discontinuity around it. This could be because cocaine users were a “harder” target for the police compared to, for example, heroin users, because cocaine was considered an elite drug at that time due to its high price.

All arguments outlined above in relation to the second threshold are also applied to the first threshold but with one exception. The missing mass below the first cutoff might be partly due to the police officers’ reluctance to deal with cases that do not affect their performance evaluation significantly[30] and (or) due to their leniency towards minor offenses. This suggests that the observed bunching above the first threshold could be driven by both moving offenders from below to above the cutoff and lower probability of cases below the cutoff being registered in the first place.

In the following subsection, I turn to the standard bunching estimator and bunching regression analysis to study the behavior of police officers around the punishment thresholds more closely.

4.1.2.2 Bunching Analysis

The bunching estimator at the second threshold for all heroin cases from form 1 is 6.325[31] (Appendix A, Figure A2). This means that the excess mass above the threshold is almost

[28] Graphs with distributions of other often seized drugs are in Appendix A, Figure A5.

[29] The most common forms of some street drugs are as follows: marijuana - dried plant flowers or hand-rolled cigarettes; hashish - dried plant compressed into blocks, balls or sheets; hashish oil - oily liquid, shatter, wax, or budder; synthetic cannabinoids - oily liquid, or dried shredded plant material sprayed with chemicals; desomorphine - light brown liquid (A DEA Resource Guide 2022).

[30] More details on police incentives around the thresholds are in Subsection 4.1.

[31] The analysis with data from form 4 yields similar results (Appendix A, Figure A2).

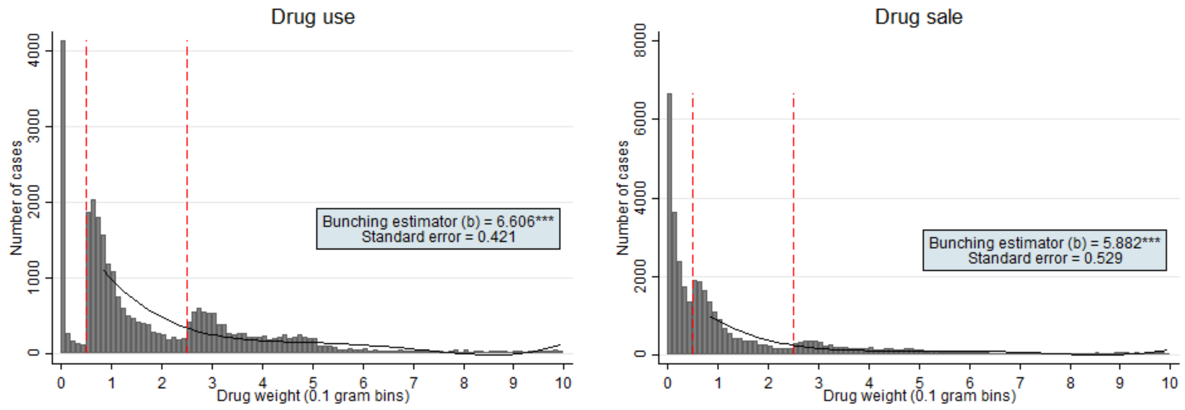
six times greater than the average number of cases that would be in the bunching window above this threshold in the counterfactual world. The effect is slightly stronger in merged samples from forms 1 and 2 and forms 1, 2, 6, and 4 (Appendix A, Figure A3), supporting the result observed in the initial dataset from form 1.

As a baseline endpoint, above which I cut the sample, I choose the 90th percentile - 10 grams. As Table B7 in Appendix B shows, prolonging the tail of the distribution makes the bunching estimator sensitive to variations in parameters. Within 10 grams, the estimated result is robust to varying the starting point, the width of the manipulation window, and the degree of the polynomial I use to fit the counterfactual distribution (Appendix B, Table B7). To avoid the possible overstatement of the effect of manipulation, I choose the main specification yielding the results presented in this paper that gives the smallest possible estimate of bunching.

As noted at the beginning of Subsection 4.1, the incentives to move offenders above the second threshold could be stronger for drug use offenses (in contrast to drug sales) because in this case it directly improves the performance evaluation. This could explain a sharper graph and slightly higher bunching estimate for drug users (left) than for drug dealers (right) in Figure 3. Another reason for the observed difference could be that drug users are the significantly larger group of drug offenders, as well as much easier to locate and, hence, to potentially manipulate.

As the event study results suggest, the performance requirements can significantly affect the behavior of officers during a year. However, this specification does not tell us anything about potential manipulation and how its magnitude (the excess number of cases above the threshold) varies across agencies and over time. To explore the effect of differences in the systems of performance evaluation, I split all heroin-related cases into two groups: those initiated by the MVD and those initiated by the FSKN (Figure 4). The estimation determines a difference in the values of the bunching estimator at the second threshold, significant at the 1% level. The bunching estimate for the MVD cases is 8.254, while for the FSKN cases it is only 3.291. This can be explained by the difference in the two systems of performance evaluation. In the case of the FSKN, final crime statistics are compared with the performance of other police stations and, eventually, other regions. The FSKN officers do not know the exact level that should be reached in order to obtain

Figure 3: Distributions of cases related to drug use (left) and drug sale (right) across quantities of heroin seized

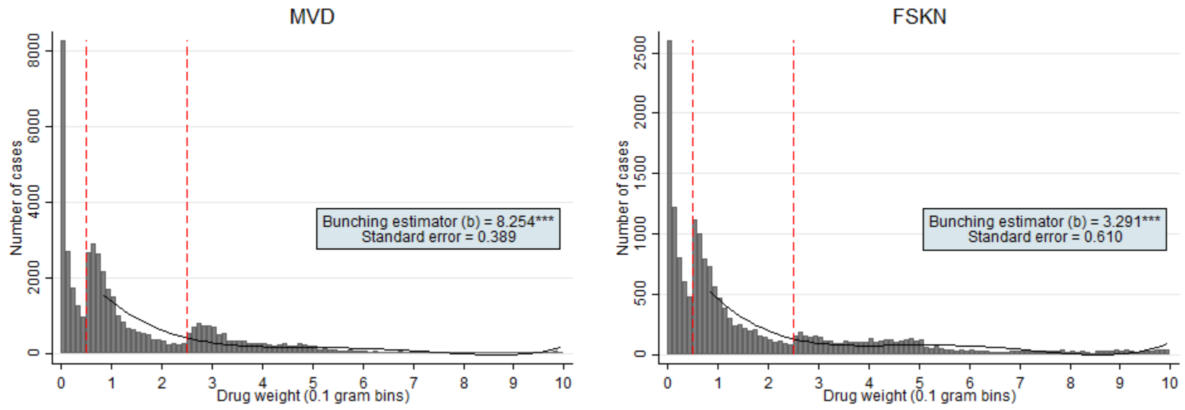


Note: The baseline sample consists of all heroin-related cases from form 1 registered in Russia during 2013-2014. The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of crime and punishment.

a satisfactory performance evaluation. Therefore, the incentives to manipulate in the case of the FSKN are weaker. In turn, the MVD officers compare results with their own performance in the previous period, which is well known to them. Given that the most recent performance should not be worse than previously, the performance evaluation system may incentivize some police officers to behave dishonestly by manipulating drug quantities seized in order to improve their statistics.

Next, I test whether reaching the target differently affects the magnitudes of possible manipulation at the second threshold in two agencies. Figure 5 presents the bunching estimators for the MVD and FSKN stations before and after they reach their previous year's numbers of serious and most serious cases. As expected, the estimation suggests that, in the case of the MVD, the magnitude of possible manipulation is higher when the station had not yet met its previous year's performance level, and that the difference is statistically significant at the 10% level. At the same time, the magnitude of possible manipulation by the FSKN police stations does not significantly depend on reaching, or not reaching the "benchmark". The residual bunching for the MVD stations after reaching the target could be explained by the presence of another performance indicator for this agency - the number of *all* serious and most serious crimes, not only drug-related ones. As discussed in Subsection 4.1.1, drug crimes contribute to this indicator, which

Figure 4: Distributions of cases initiated by the MVD (left) and the FSKN (right) across quantities of heroin seized



Note: The baseline sample consists of all heroin-related cases from form 1 registered in Russia during 2013-2014. The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment.

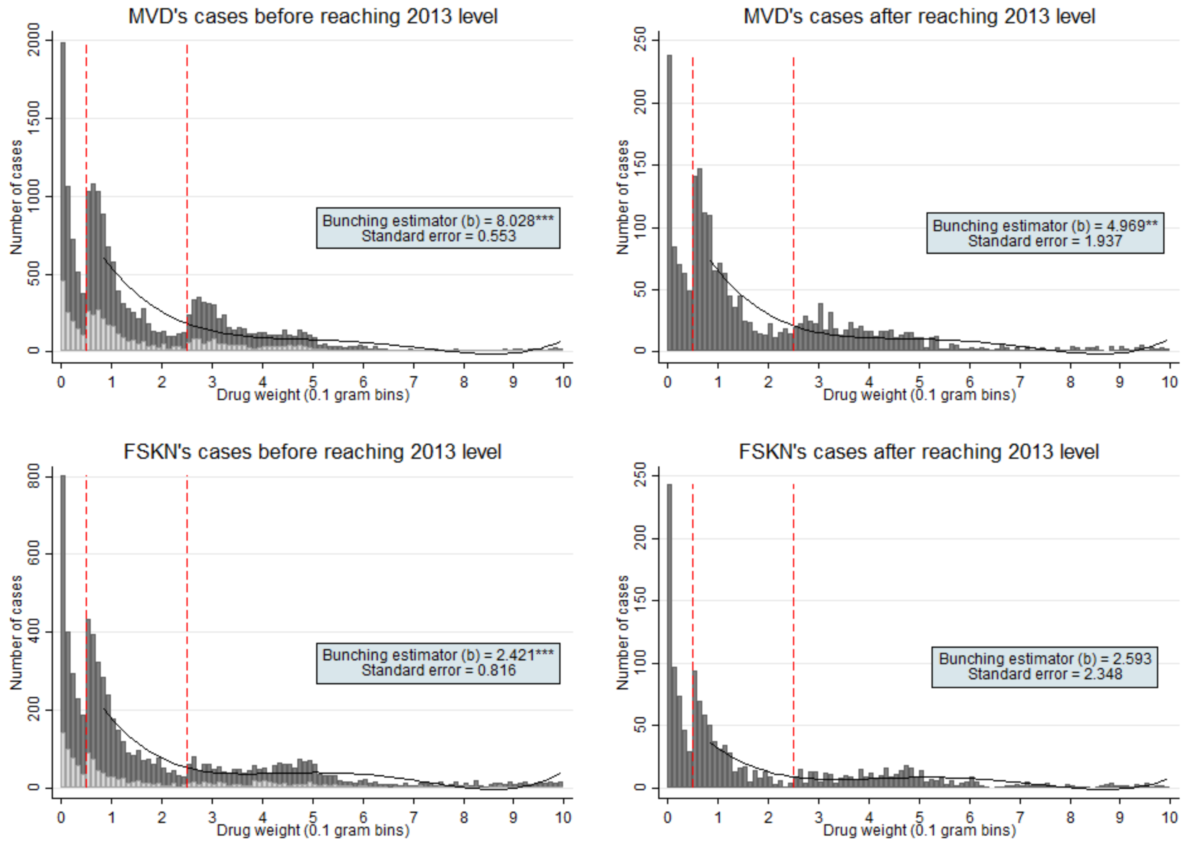
could affect the event study and bunching analysis results.

The sample used for calculating the bunching estimators includes both stations that reached the target and stations that did not reach it during the year (dark grey bars). The restricted sample, which only includes stations that reached the target, is depicted in light gray for the period before the "event" and fully coincides with the extended sample for the period after. These two samples have comparable distributions and produce similar regression estimates of the effect of targeting the previous year's performance level on the probability of bunching above the second threshold. As Table B10 in Appendix B presents, reaching the 2013 number of serious and most serious drug crimes significantly decreases the bunching probability for the MVD stations (columns (3) and (7)); however, there is no effect on the behavior of the FSKN stations (columns (4) and (8)).

I also run the regression analysis for the first threshold to explore the effect of reaching the previous year's total number of drug crimes on bunching probabilities. As columns (1), (2) and (5), (6) show, the estimates are insignificant for both agencies. However, the results should be interpreted with caution due to potentially missing observations below the first threshold, especially for drug use offenses.

All in all, the results in this subsection support the hypothesis that the main driving force for the bunching of drug quantities above the thresholds could be the performance

Figure 5: Distributions of cases across quantities of heroin seized during 2014 by the MVD and FSKN police stations before (left) and after (right) reaching the total number of serious and most serious drug crimes solved in 2013



Note: The baseline sample consists of heroin-related cases from form 1 registered in Russia during 2013-2014. The series shown in bars is a histogram of the observed distribution of cases: light grey bars - cases registered by stations that reached the 2013 level during 2014, dark grey bars - cases registered by both stations that reached the 2013 level and stations that did not reach it during 2014. Light grey and dark grey distributions are overlaid on one another and not stacked in the left graphs and fully coincide in the right graphs. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment.

evaluation system. However, it is worth noting that bribery might be another motive for police officers to potentially manipulate the drug quantities, or to threaten offenders with possible manipulation[32]. However, according to an investigation based on anonymous surveys of 571 victims of extortion from being caught with drugs (Litavrin et al. 2017),

[32]As for other potential motives for manipulation, it is unlikely that the observed bunching could be explained by the officers' taste for stricter punishment or by their concern for drug users. First, manipulation is, nevertheless, risky and does not seem to be implemented due to taste only, without significant benefits for the officers. Second, in Russia, drug addicts do not receive any treatment during incarceration and, without proper therapy, they could resume taking drugs after release (Semioshina 2019). Moreover, sometimes addicts may even have access to drugs in prison (The Federal Penitentiary Service of Russia 2021).

if an offender decides and is able to pay a bribe, in most of the cases s/he does it to buy himself out of prison, not just to decrease the sentence. This means that those individuals are not likely to be in the database at all and these bribery cases are undetectable. In addition, the study suggests that the amount paid in bribes is increasing with the drug quantity; however, it does not find any evidence of the bribe cases bunching at some particular quantities. If, nevertheless, there are cases in which manipulations were triggered by bribery motives but offenders were unable to pay the bribe, it could partly explain the residual bunching for the MVD stations after reaching the target (Figure 5).

4.2 Mean Characteristics of Possible Victims of Manipulation

There are a number of criteria that a police officer can use to select which offenders to push above the threshold. My analysis begins by calculating summary statistics for the whole population of heroin offenders and for those who fall into the manipulation region. Table B11 in Appendix B shows that means are similar across these two samples, suggesting the absence of self-selection into the area around the second threshold. To determine the mean characteristics of possible victims of manipulation, I use the technique described in Subsection 3.2 and present the results in Table 1.

Table 1: Mean characteristics of possible victims of manipulation

	Eligible for manipulation	Manipulated	Diff.	Std. errors
Male	0.809	0.850	-0.041^{***}	0.015
Russian	0.856	0.883	−0.027	0.020
Citizen of Russia	0.958	0.949	0.008	0.010
Resident of region	0.906	0.887	0.023	0.015
At most secondary education	0.595	0.574	0.021	0.023
Unemployed	0.761	0.756	0.005	0.018
Previously charged with offense	0.689	0.695	−0.006	0.021
Under the influence of drugs	0.518	0.521	−0.003	0.018
Under the influence of alcohol	0.013	0.016	−0.003	0.006

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The baseline sample consists of all heroin-related cases from forms 1 and 2 registered in Russia during 2013-2014. Column 1 presents the predicted mean characteristic of all drug offenders who possessed an unmanipulated quantity of drug that fell into the manipulation window below the threshold. Column 2 presents the predicted mean characteristics among the compliers, i.e., the offenders who were actually moved above the threshold. Column 3 tests the difference. To obtain the estimates, I apply the method described in detail in Subsection 3.2.

First of all, I check whether the offender's characteristics, such as gender, nationality, citizenship, and regional residency, affect a police officer's decision to manipulate the seized drug quantities. The only significant difference I find is in mean shares of men among those who were eligible for manipulation but did not receive the "treatment", and those who were pushed above the threshold. The results suggest that men are more likely to have their drug quantities manipulated by police.

Next, I turn to the indicators of an offender's socio-economic status, such as unemployment and education. Individuals from the low socio-economic class are likely to be legally illiterate and not to have any means to protect themselves from police oppression. However, I do not find any evidence that they are more likely to become a victim of manipulation.

In addition, I check whether being previously charged with an administrative[33] or criminal offense and being under the influence of drugs or alcohol when arrested influence a police officer's decision to manipulate. In general, these three features should make offenders more vulnerable targets. Additionally, according to a former police officer, there are almost no random people among the manipulated offenders because it is quite risky to manipulate when one does not know how the person will respond. Usually, police officers target those who have been previously convicted or who they know use drugs (Litavrin 2019). Even though differences in shares between those who stayed below the threshold and those who were moved above it have the expected negative sign, I do not find any statistically significant effect here.

Overall, the analysis suggests that possible manipulations, potentially triggered by the strong incentives, could contribute to the existing gender-based inequality in the enforcement of law. This result is in line with other studies which find that, in general, law enforcers are more lenient towards women: women are less likely to receive speeding tickets, women receive shorter sentences for the same crime, etc. (Makowsky and Stratmann 2009, Butcher et al. 2017, Bindler and Hjalmarsson 2020).

[33] A wrongful action (omission) by a natural person or legal entity which is administratively punishable under The Code of Administrative Offenses of The Russian Federation. This violation of the law is not serious enough to be considered criminal.

4.3 The Effect of Possible Manipulation on Sentence Length

A case-by-case comparison of heroin weights from forms 1 and 4 shows that they coincide in 92.8% of the full sample. Significant deviations seem suspicious and might be the consequence of mistakes made when filling in the card or converting it into an electronic form. At the same time, observations with large discrepancies in weights are randomly distributed and, therefore, could be excluded from the analysis. Table 2 presents the results of an estimation conducted for full and restricted samples, which includes observations with absolute weight differences of less than 14 grams; this being the 95th percentile among absolute nonzero deviations.

Table 2: The effect of possible manipulation on sentence length and case characteristics

	Absolute difference $\leq$ 14		Full sample	
	Coefficient	s.e.	Coefficient	s.e.
	Panel A. Sentence length			
First stage	1.190***	0.005	0.497	2.246
ITT	1.206***	0.045	1.209***	0.052
LATE(sentence, years)	1.013***	0.128	2.434***	0.929
	Panel B. Case characteristics			
LATE(plea)	−0.078	0.093	−0.060	0.158
LATE(investigation, days)	1.749	8.598	1.436**	0.688
LATE(adjudication, days)	−0.799	4.455	−0.701**	0.356
LATE(pretrial detention)	0.092***	0.035	0.075**	0.034

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The baseline sample consists of all heroin-use-related cases from forms 1, 2, 6 and 4 registered in Russia during 2013-2014. See the text for further details defining the subsample of observations with absolute difference in weights of less than 14 grams. Panel A presents estimates of the impact of drug weights from form 1 on drug weights from form 4 (First stage), as well as the ITT effect of manipulation on the sentence length of all individuals in the manipulation region, and the LATE of manipulation on the sentence length of compliers only. Panel B presents the LATE of manipulation on the probability of pleading guilty, length of investigation, length of adjudication, and probability of pretrial detention. To obtain the estimates, I apply the method described in detail in Subsection 3.3.

The first stage effect of possible manipulation of heroin quantities registered by police officers in form 1 on heroin weights recorded in form 4 after the expertise is significant and shows a 1.2 grams increase in the quantities of drugs seized for individuals in the manipulation area. There is also a significant effect of being in the manipulation window on sentence length (ITT). However, in order to see the impact of manipulation on compliers' years of imprisonment, I divide ITT by the first stage effect and obtain LATE(sentence), which suggests a one-year increase in sentence length for individuals who were pushed above the threshold[34] (Panel A, Table 2).

Next, I turn to other case characteristics that could also be affected by the possible manipulation of drug quantities and that could significantly influence the sentence length. First, I check whether manipulation increases the probability of accepting a plea bargain. Pleading guilty significantly simplifies the whole procedure and a conviction is pronounced without actual examination of the evidence at a court hearing. In addition, a person that accepts a plea bargain waives the right to appeal. Therefore, in the case of manipulation, police officers could offer the plea agreement more forcefully because the credibility of evidence collected is in doubt. Since, in manipulation cases, there is no need for an actual investigation and because of the potentially increased probability of pleading guilty, I also check whether there is a significant effect of manipulation on the length of investigation and length of adjudication. Finally, I explore whether manipulation affects the probability of pretrial detention, because pretrial detention not only prevents offenders from fleeing but also allows police officers to keep manipulated offenders under constant pressure. As the analysis in Panel B of Table 2 suggests, the only significant effect here is on the probability of pretrial detention: potentially manipulated offenders are 9% more likely to be detained while they are waiting for a trial.

5 Conclusion

The tradeoff between motivating civil servants and distorting their behavior has always been a central issue of incentives design. Increasingly, the literature documents a negative effect of high-powered performance-related incentives in the public sector. Nevertheless, their use is still a common practice across bureaucracies in many countries. A particularly notable example of such countries is Russia. A recently published report on drug crimes

[34] This estimate is close to that obtained in Skougarevskiy (2017). Applying regression discontinuity design methods to the data on cannabis and heroin cases from Russia, he finds that the length of unconditional incarceration increases by 0.84 years when the drug weight crosses the threshold. My estimate could be higher because I focus solely on heroin cases, which might be considered to be more serious offenses than cannabis-related crimes. In addition, I estimate the effect for compliers, while Skougarevskiy (2017) shows the discontinuity taking into account all offenders in the window above the threshold.

(Knorre 2017) illuminates revealing statistics on the distribution of criminal cases across quantities of heroin seized. These statistics suggest the bunching of offenders who were arrested in possession of a quantity of drugs just above the threshold that is sufficient to be convicted of a more serious crime. At the same time, there is a missing mass of cases just below the threshold. This might be evidence of police manipulating the quantities of drugs seized in order to move offenders from below to above the threshold.

This paper provides an empirical analysis of the mechanism that drives the possible manipulation of quantities of drugs seized using a unique dataset that contains rich information on drug crimes reported in Russia during 2013-2014. Exploiting the specific features of the Russian institutional context, I show how high-powered performance-based incentives could trigger the misbehavior of police officers: around 4% of all heroin offenders in the database were moved above the punishment threshold as a result of manipulation. Additionally, the results suggest that men are more likely to have their drug quantities manipulated by the police. This manipulation increases the probability of pretrial detention by 9% and its overall effect on sentence length is an additional year of incarceration.

The paper shows how strong incentives tied to easily measurable and quantifiable performance indicators could easily distort the behavior of civil servants in an environment where manipulation is ex ante less expected. Forecasting future outcomes itself is a common practice in many public organizations; this provides guidance for the upcoming period. However, the way in which it is implemented could become an issue (Rasul and Rogger 2018, Banerjee et al. 2021), as in the case of drug control in Russia. Therefore, a comprehensive approach is required in order to improve the situation. The first step on the way to efficiency could be decentralizing the performance evaluation system and enabling regional officers to take into account local specifics affecting their performance. In turn, evaluating police officers based on local trends in criminal statistics will smooth the incentives arising from performance indicators. Additionally, increasing the cost of misconduct to police officers could deter their poor behavior and further improve police practices.

References

A DEA Resource Guide (2022). Drugs of Abuse. Technical report, US Department of Justice, Drug Enforcement Administration. `https://www.dea.gov/sites/default/files/2022-12/2022_DOA_eBook_File_Final.pdf`. Accessed January 10, 2023.

Acemoglu, D., L. Fergusson, J. Robinson, D. Romero, and J. F. Vargas (2020). The Perils of High-Powered Incentives: Evidence from Colombia's False Positives. *American Economic Journal: Economic Policy 12*, 1–43.

Ajvazyan, D. (n.d.). *Osobyj Poryadok po Ugolovnomu Delu v Sude.* `https://adv1.ru/articles/1266/`. Accessed November 11, 2022.

Alexander, D. (2020). How do Doctors Respond to Incentives? Unintended Consequences of Paying Doctors to Reduce Costs. *Journal of Political Economy 128*, 4046–4096.

Anbarci, N. and J. Lee (2014). Detecting Racial Bias in Speed Discounting: Evidence from Speeding Tickets in Boston. *International Review of Law and Economics 38*, 11–24.

Antonov, E. (2019). *Kak Peterburzhcy Dokazyvayut, Chto im Podbrosili Narkotiki? Tri Istorii o Zaderzhaniyah, Sudah i Problemah Zakonodatel'stva* . `https://paperpaper.ru/photos/kak-peterburzhcy-dokazyvayut-chto-im-pod/`. Accessed June 28, 2019.

Antonov i Partnery (2016). *Okonchanie Predvaritel'nogo Sledstviya s Sostavleniem Obvinitel'nogo Zaklyucheniya.* `https://pravo163.ru/okonchanie-predvaritelnogo-sledstviya-s-sostavleniem-obvinitelnogo-zaklyucheniya/`. Accessed November 11, 2022.

Ash, E. and W. B. MacLeod (2015). Intrinsic Motivation in Public Service: Theory and Evidence from State Supreme Courts. *The Journal of Law and Economics 58*, 863–913.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2021). Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy, and Training. *American Economic Journal: Economic Policy 13*, 36–66.

Banerjee, A. V., E. Duflo, and R. Glennerster (2008). Putting a Band-Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System. *Journal of the European Economic Association 6*(2-3), 487–500.

Bindler, A. and R. Hjalmarsson (2020). The Persistence of the Criminal Justice Gender Gap: Evidence from 200 Years of Judicial Decisions. *The Journal of Law and Economics 63*(2), 297–339.

Bjerk, D. (2005). Making the Crime Fit the Penalty: The Role of Prosecutorial Discretion Under Mandatory Minimum Sentencing. *Journal of Law and Economics 48*, 591–627.

Bjerk, D. (2017). Mandatory Minimums and the Sentencing of Federal Drug Crimes. IZA Discussion Paper 10544.

Butcher, K. F., K. H. Park, and A. M. Piehl (2017). Comparing Apples to Oranges: Differences in Women's and Men's Incarceration and Sentencing Outcomes. *Journal of Labor Economics 35*(S1), S201–S234.

Chatsverykova, I. (2017). Severity and Leniency in Criminal Sentencing in Russia: The Effects of Gender and Family Ties. *International Journal of Comparative and Applied Criminal Justice 41*(3), 185–209.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records. *Quarterly Journal of Economics 126*, 749–804.

Cunha-Oliveira, T., A. C. Rego, F. Carvalho, and C. R. Oliveira (2013). Medical Toxicology of Drugs of Abuse. *Principles of Addiction-Comprehensive Addictive Behaviors and Disorders 1*, 159–175.

de Maillard, J. (2018). Police Performance Regimes and Police Activity: Compstat in Paris and London Compared. *European Journal of Criminology 15*(5), 589–608.

Diamond, R. and P. Persson (2016). The Long-Term Consequences of Teacher Discretion in Grading of High-Stakes Tests. National Bureau of Economic Research Working Paper 22207.

Dixit, A. (1997). Power of Incentives in Private Versus Public Organizations. *The American Economic Review 87*(2), 378–382.

Dombrovickij, M. (n.d.). *Rassmotrenie Ugolovnogo Dela v Osobom Poryadke — Chto Znachit i Stoit li Soglashat'sya*. `https://legis-group.ru/blog/poleznoe/useful/the-case-in-a-special-manner`. Accessed November 11, 2022.

Duflo, E., R. Hanna, and S. P. Ryan (2012). Incentives Work: Getting Teachers to Come to School. *American Economic Review 102*(4), 1241–78.

Eterno, J. A. and E. B. Silverman (2017). *The Crime Numbers Game: Management by Manipulation*. CRC Press.

Filchenko, A. and V. Zhandrov (2022). The Signs of the Falsification of Evidence and Results of Operational Investigative Activities in the Materials of Criminal Cases on Drug Trafficking. *Volgograd Academy of the Russian Internal Affairs Ministry's Digest 1*(60), 148–155.

Fisman, R. and Y. Wang (2017). The Distortionary Effects of Incentives in Government: Evidence from China's "Death Ceiling" Program. *American Economic Journal: Applied Economics 9*, 202–218.

Goncalves, F. and S. Mello (2021). A Few Bad Apples? Racial Bias in Policing. *American Economic Review 111*(5), 1406–41.

Holmstrom, B. and P. Milgrom (1991). Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization 7*, 24.

Jacob, B. A. (2005). Accountability, Incentives and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools. *Journal of Public Economics 89*, 761–796.

Jacob, B. A. and S. D. Levitt (2003). Rotten Apples: An Investigation of the Prevalence and Predictors of Teacher Cheating. *The Quarterly Journal of Economics 118*, 843–877.

Kleven, H. J. and M. Waseem (2013). Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan. *Quarterly Journal of Economics 128*, 669–723.

Knorre, A. (2017). *Narkoprestupleniya v Rossii: Analiz Sudebnoi i Kriminal'noi Statistiki.* Technical report, The Institute for the Rule of Law at the European University at St. Petersburg. http://enforce.spb.ru/images/Knorre_Drug_crimes_in_Russia.pdf. Accessed October 1, 2017.

Knorre, A. (2020). Do Russian Police Fabricate Drug Offenses? Evidence From Seized Heroin's Weight Distribution. *Journal of Drug Issues*, 378–391.

Knorre, A. and D. Skougarevskiy (2015). *Kak MVD i FSKN Borjutsja s Narkotikami: Sravnitel'nyj Analiz Rezul'tativnosti Dvuh Vedomstv.* Technical report, The Institute for the Rule of Law at the European University at St. Petersburg. http://enforce.spb.ru/images/analit_zapiski/FSKN_MVD_memo_2015_web.pdf. Accessed October 1, 2017.

Kolegov, A. (2012). *Palochnaya Sistema. Ot Plana po Poimke Prestupnikov Ustali Dazhe Policejskie.* https://advokat-ko.ru/2012/06/palochnaja-sistema/. Accessed November 2, 2022.

Lepage, L.-P. (2020). Are Criminals Strategic? Offender Responses to Drug Sentencing Cutoffs. *Labour Economics 66*, 101906.

Litavrin, M. (2019). *Lishnij Ves. Kratkaya Enciklopediya Policejskih Narkoprovokacij.* https://zona.media/article/2019/07/15/encyclopedia-228. Accessed October 28, 2021.

Litavrin, M., A. Sarang, and A. Knorre (2017). *Skol'ko Stoyat Veshchestva, Esli vy — Policejskij. Issledovanie o Vzyatkah za Narkotiki v Rossii.* https://openrussia.org/notes/715620/. Accessed February 2, 2019.

Makowsky, M. D. and T. Stratmann (2009). Political Economy at Any Speed: What Determines Traffic Citations? *American Economic Review 99*(1), 509–27.

Makowsky, M. D., T. Stratmann, and A. Tabarrok (2019). To Serve and Collect: the Fiscal and Racial Determinants of Law Enforcement. *The Journal of Legal Studies 48*, 189–216.

Mas, A. (2006). Pay, Reference Points, and Police Performance. *The Quarterly Journal of Economics 121*(3), 783–821.

Merzlikin, P. (2019). *My Znaem, Chto Pravoohranitel'naya Sistema v Rossii Prognila. No Kak Eto Proizoshlo i Chto s Etim Delat'?* `https://meduza.io/articles`. Accessed June 28, 2019.

Nadezhdin, I. and A. Matveeva (2019). *Oni Tol'ko i Sposobny, Chto Dur' Podkinut'.* `https://lenta.ru/articles/2019/06/13/drugs/`. Accessed June 28, 2019.

Nikonov, M. (2020). *Fal'sifikaciya Dokazatel'stv. Policeyskie Metody.* `https://zonaprava.com/download.php?file=6178`. Accessed February 10, 2021.

Novikova, A. (2014). Sistema Ocenki Policii. Moscow. Public Verdict Foundation. Paper presented at "Human rights and law enforcement agencies" conference. `http://publicverdict.org/articles_images/12001_60506_reform_police.pdf`. Accessed October 1, 2017.

Ossei-Owusu, S. (2021). Police Quotas. *New York University Law Review 96*(2), 529–605.

Paneyakh, E. (2014). Faking Performance Together: Systems of Performance Evaluation in Russian Enforcement Agencies and Production of Bias and Privilege. *Post-Soviet Affairs 30*(2-3), 115–136.

Rasul, I. and D. Rogger (2018). Management of Bureaucrats and Public Service Delivery: Evidence from the Nigerian Civil Service. *The Economic Journal 128*, 413–446.

Rehavi, M. M. and S. B. Starr (2014). Racial Disparity in Federal Criminal Charging and its Sentencing Consequences. *Journal of Political Economy 6*, 1320–1354.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points? *American Economic Journal: Economic Policy 2*, 180–212.

Schultz, A., V. Kozlov, and A. Libman (2014). Judicial Alignment and Criminal Justice: Evidence From Russian Courts. *Post-Soviet Affairs 30*, 137–170.

Semioshina, V. (2019). *Ni Gramma Sozhaleniya.* `https://takiedela.ru/2019/10/ni-gramma-sozhaleniya/`. Accessed October 30, 2021.

Shklyaruk, M. and D. Skougarevskiy (2015). *Kriminal'naya Statistika: Mehanizmy Formirovaniya, Prichiny Iskazheniya, Puti Reformirovaniya.* Technical report, The Institute for the Rule of Law at the European University at St. Petersburg. `http://enforce.spb.ru/images/Products/Crimestat_report_2015_IRL_KGI_web.pdf`. Accessed October 1, 2017.

Singh, P. and W. A. Masters (2017). Impact of Caregiver Incentives on Child Health: Evidence from an Experiment with Anganwadi Workers in India. *Journal of Health Economics 55*, 219–231.

Skougarevskiy, D. (2017). *What Do Graduated Sanctions Tell Us About the Functions of the Law: A Case of Drug Crimes.* Ph. D. thesis, Graduate Institute of International and Development Studies.

The Federal Penitentiary Service of Russia (2021). *Characteristics of Prison Population.* `http://fsin.su/structure/inspector/iao/statistika/Xar-ka%20lic%20sodergahixsya%20v%20IK/`. Accessed October 2, 2021.

The Ministry of Internal Affairs of the Russian Federation (n.d.). *The History of the Ministry of Internal Affairs of the Russian Federation.* `https://xn--b1aew.xn--p1ai/history`. Accessed October 1, 2017.

Titaev, K. (2019). *My Znaem, Chto Pravoohranitel'naya Sistema v Rossii Prognila. No Kak Eto Proizoshlo — i Chto s Etim Delat'?* `https://meduza.io/feature/2019/06/26/my-znaem-chto-pravoohranitelnaya-sistema-v-rossii-prognila-no-kak-eto-proizoshlo-i-`
Accessed November 2, 2022.

Titaev, K. and M. Pozdnyakov (2012). *Poryadok Osobyi - Prigovor Obychnyi: Praktika Primeneniya Osobogo Poryadka Sudebnogo Razbiratel'stva (gl. 40 UPK RF) v Rossiiskih Sudah.* Technical report, The Institute for the Rule of Law at the European University at St. Petersburg. http://www.enforce.spb.ru/images/analit_zapiski/pm_gl_40_UPK_fin.pdf. Accessed January 30, 2019.

Torocheshnikova, T. and D. Sotnikov (2021). *Po Ponyatiyam. Kak v Rossii Sudyat Silovikov.* https://tvrain.ru/teleshow/notes/po_ponjatijam_kak_v_rossii_sudjat_silovikov-533576/. Accessed October 21, 2022.

Traxler, C., F. G. Westermaier, and A. Wohlschlegel (2018). Bunching on the Autobahn? Speeding Responses to a "Notched" Penalty Scheme. *Journal of Public Economics 157*, 78–94.

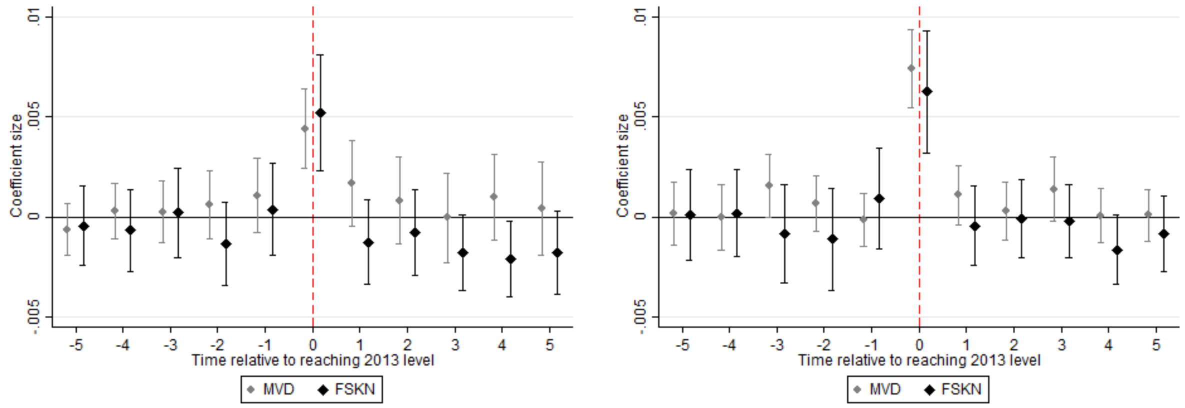
Tuttle, C. (2019). Racial Disparities in Federal Sentencing: Evidence from Drug Mandatory Minimums. *Available at SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3080463. Accessed November 1, 2019.*

Ulmer, J. T., M. C. Kurlychek, and J. H. Kramer (2007). Prosecutorial Discretion and the Imposition of Mandatory Minimum Sentences. *Journal of Research in Crime and Delinquency 44*, 427–458.

Volkov, V. (2016). Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia's Criminal Courts. *Journal of Empirical Legal Studies 13*, 637–665.

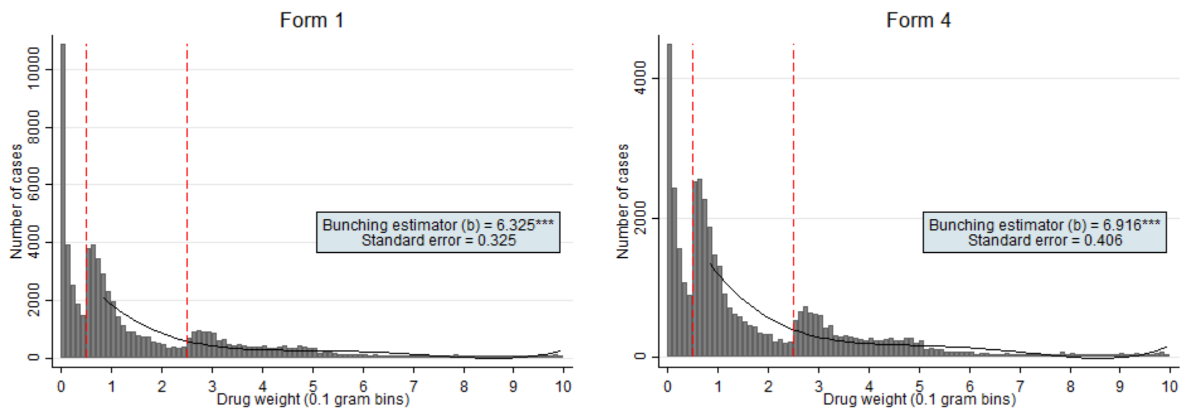
A Supplemental Figures

Figure A1: The effect of reaching the 2013 level on the total number of heroin-related crimes (left) and the number of serious and most serious heroin-related crimes (right) registered in 2014



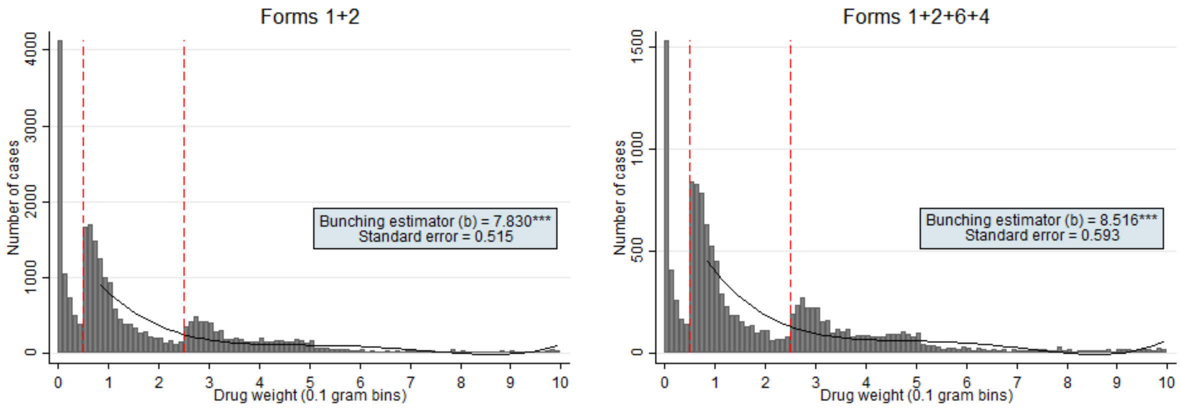
Note: The samples include all MVD stations and all FSKN stations that reached the total 2013 number of drug crimes (left) and the 2013 number of serious and most serious drug crimes (right) during 2014. The dependent variable is the inverse hyperbolic sine transformation of either the daily total number of heroin-related crimes or the daily number of serious and most serious heroin-related crimes calculated based on the sample of all drug-related cases from form 1. The specification includes station, day-of-the-month and region-by-month fixed effects. Standard errors are clustered by station.

Figure A2: Distribution of cases from form 1 (left) and form 4 (right) across quantities of heroin seized



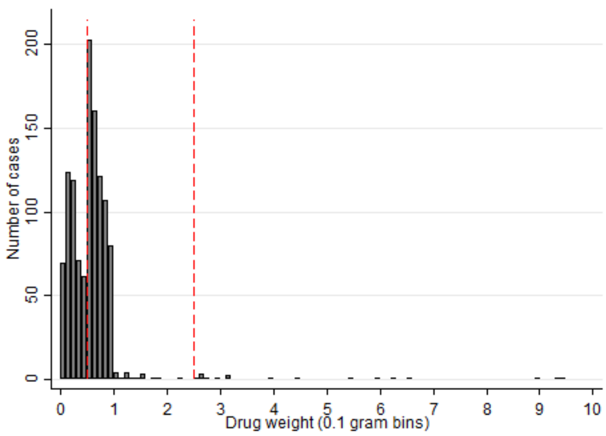
Note: The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment.

Figure A3: Distributions of cases from forms 1, 2 (left) and forms 1, 2, 6, 4 (right) across quantities of heroin seized



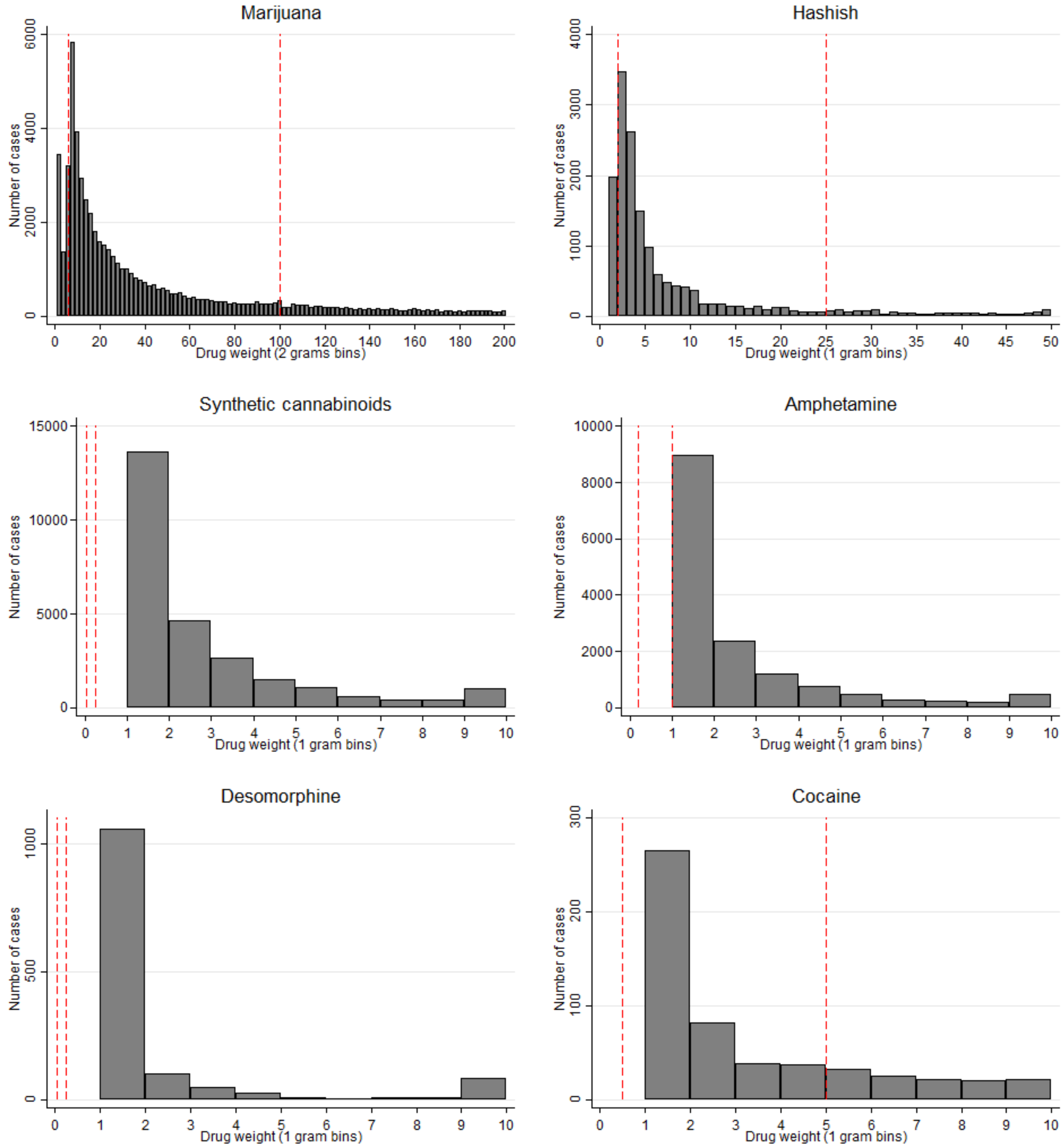
Note: The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment.

Figure A4: Distribution of cases with missing heroin quantities in form 1 across heroin quantities from 4



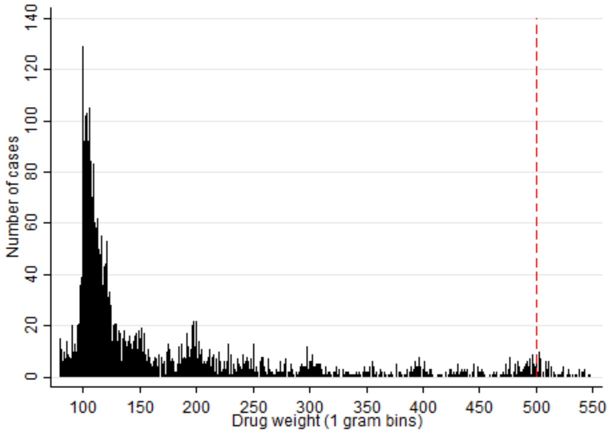
Note: The baseline sample consists of all heroin-related cases for which quantities in form 1 are missing but quantities in form 4 are not missing. The threshold (dashed line) determines the scale of seizure (large and especially large drug quantities), the seriousness of the offense, and the severity of punishment.

Figure A5: Distributions of cases across quantities of drugs seized by drug type



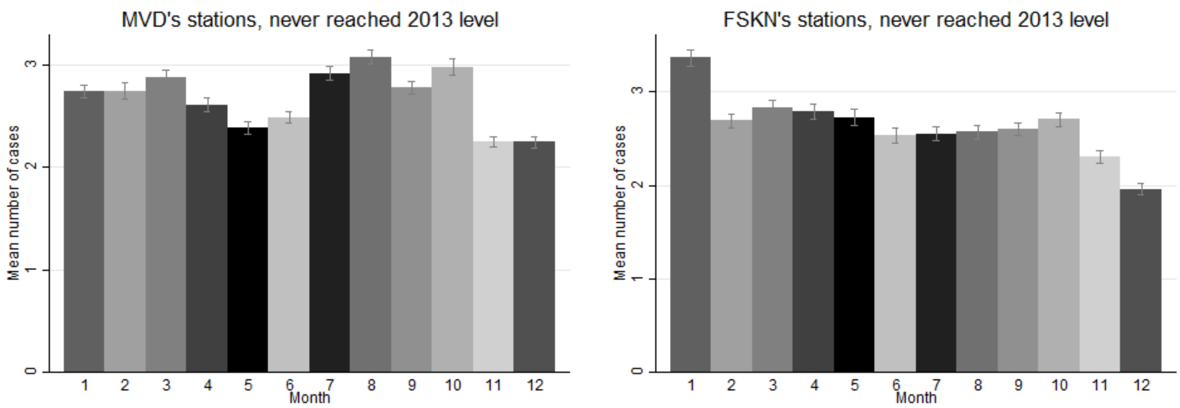
Note: The baseline sample consists of all drug-related cases from form 1 registered in Russia during 2013-2014. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug quantities), the seriousness of the offense, and the severity of punishment.

Figure A6: Distribution of cases across quantities of heroin seized around the third threshold



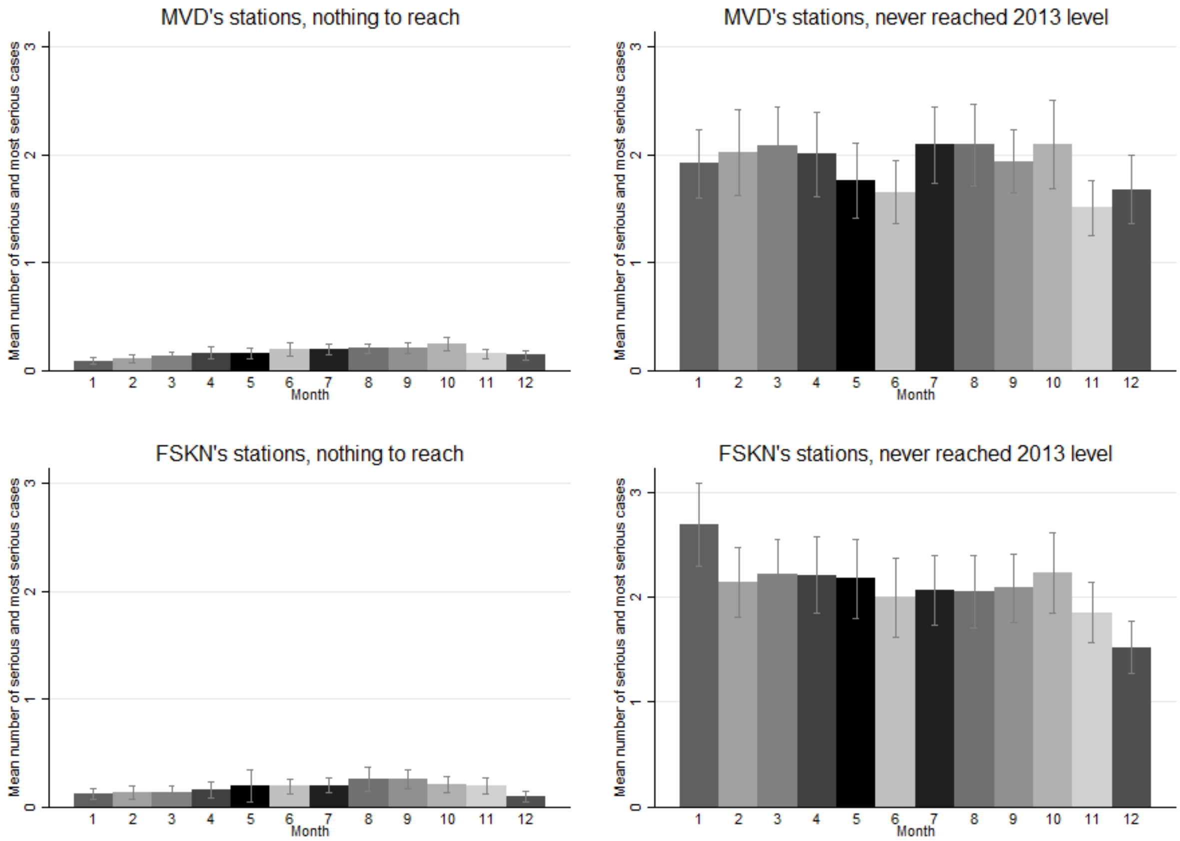
Note: The baseline sample consists of all drug-related cases from form 1 registered in Russia during 2013-2014 where the quantities of heroin seized were between 80 and 550 grams. The threshold (dashed line) determines the scale of seizure (large and especially large drug quantities), the seriousness of the offense, and the severity of punishment.

Figure A7: Mean total numbers of drug-related offenses registered in 2014, by month, agency and group in relation to reaching the 2013 level



Note: The series shown in bars present mean total monthly numbers of drug-related offenses registered in 2014 by the MVD and FSKN stations that did not reach their previous year's performance level during 2014. The graph also shows 95% confidence intervals for each mean.

Figure A8: Mean numbers of serious and most serious drug-related offenses registered in 2014, by month, agency and group in relation to reaching the 2013 level



Note: The series shown in bars present mean monthly numbers of serious and most serious drug-related offenses registered in 2014 by the MVD and FSKN stations of two types: those that had a zero target from previous year (on the left) and those that did not reach their previous year's performance level during 2014 (on the right). The graph also shows 95% confidence intervals for each mean.

B Supplemental Tables

Table B1: The comparison of means and medians for the MVD and FSKN

	Form 1			Form 1+2+6		
	MVD	FSKN	Diff.	MVD	FSKN	Diff.
Drug types:						
Cannabis	0.380	0.322	0.057***	0.471	0.372	0.099***
Heroin	0.242	0.240	0.002	0.198	0.198	-0.001
Amphetamine[35]	0.134	0.147	-0.013***	0.113	0.125	-0.012***
Synthetic cannabinoids	0.118	0.146	-0.028***	0.110	0.161	-0.051***
Papaver, Opiates	0.056	0.061	-0.004***	0.043	0.069	-0.026***
Desomorphine	0.015	0.008	0.007***	0.024	0.016	0.009***
Cocaine	0.002	0.005	-0.002***	0.001	0.003	-0.002***
Other	0.053	0.071	-0.018***	0.040	0.056	-0.016***
Median drug weight, grams	2.34	5	-2.66***	4	6	-2***
Median heroin weight, grams	0.9	1.66	-0.76***	1.03	2.6	-1.57***
Article 228 (use)	0.672	0.435	0.237***	0.910	0.777	0.133***
Article 228.1 (sale)	0.320	0.494	-0.174***	0.086	0.156	-0.070***
Offender characteristics:						
Male				0.924	0.896	0.028***
Russian				0.881	0.861	0.020***
Citizen of Russia				0.972	0.976	-0.004***
Resident of region				0.915	0.966	-0.051***
At most secondary educ.				0.628	0.563	0.065***
Unemployed				0.694	0.631	0.063***
Student				0.001	0.001	0.000
Worker				0.235	0.289	-0.054***
White collar				0.013	0.016	-0.004***
Previously charged with offense				0.589	0.542	0.046***
Under the influence of drug				0.406	0.506	-0.101***
Under the influence of alcohol				0.051	0.016	0.035***
Punishment types:						
Compulsory works				0.055	0.031	0.024***
Corrective labor				0.036	0.023	0.013***
Restriction of liberty				0.022	0.016	0.006***
Suspended sentence				0.409	0.443	-0.034***
Custodial sentence				0.324	0.372	-0.047***
Fine				0.148	0.106	0.043***
Other				0.005	0.010	-0.005***
Sentence length, years				2.722	3.417	-0.695***
Pleaded guilty				0.463	0.199	0.264***
Investigation length, days				55.567	82.115	-26.548***
Adjudication length, days				26.673	32.066	-5.394***
Pretrial detention				0.128	0.130	-0.002
N	222,660	101,686		61,274	17,933	

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table compares means and medians for the MVD and FSKN in two samples: form 1, and form 1 merged with forms 2 and 6. The samples consist of all drug-related cases registered in Russia during 2013-2014.

[35]This group includes amphetamine itself, methamphetamine, mephedrone, metamfepramone and other amphetamine derivatives.

Table B2: Quantities of drugs (grams above) for purposes of articles 228, 228.1 of the Criminal Code of the Russian Federation

	Significant	Large	Especially large
Marijuana	6	100	100000
Hashish	2	25	10000
Hashish oil	0.4	5	1000
Papaver	20	500	100000
Poppy straw	1	5	500
Methadone	0.5	2.5	1000
Heroin	0.5	2.5	500
Amphetamine	0.2	1	200
Desomorphine	0.05	0.25	10
Cocaine	0.5	5	1500
Synthetic cannabinoids	0.05	0.25	50

Table B3: The seriousness of offense and sanctions according to articles 228, 228.1 of the Criminal Code of the Russian Federation

	Article 228 (use)		Article 228.1 (sale)	
Drug quantity	Seriousness	Sentence (years)	Seriousness	Sentence (years)
Less than significant	Administrative offense	Fine/15 days	Serious	4-8
Significant	Least serious	0-3	Most serious	8-15
Large	Serious	3-10	Most serious	10-20
Especially large	Most serious	10-15	Most serious	15-20

Table B4: Absolute differences in heroin weights from forms 1 and 4 across thresholds

	MVD (1)	FSKN (2)
Above first	-0.070	9.304
	(0.427)	(4.526)
Above second	-0.367	4.254
	(0.188)	(4.876)
Above third	-1.208*	-12.891
	(0.099)	(8.491)
Charge use	-0.306	25.777
	(1.025)	(29.259)
Charge sale	0.362	32.786
	(0.669)	(26.520)
Constant	0.785	-23.791
	(0.644)	(26.917)
Station FEs	✓	✓
Day-of-the-month FEs	✓	✓
Region-by-month FEs	✓	✓
Year FEs	✓	✓
R-squared	0.227	0.198
N	33,270	10,949

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents regression estimates for the MVD in column (1) and the FSKN in column (2). The dependent variable is the absolute difference in heroin weights from forms 1 and 4. Variables of interest *Above first*, *Above second*, and *Above third* are dummies indicating form 1 weights in a 1-gram window above the first, second, and third thresholds. The specification includes station, day-of-the-month, region-by-month, and year fixed effects. Standard errors are clustered by station and year.

Table B5: The comparison of means within the missing values analysis

	Form 1			Form 4			Forms 1+2			Forms 1+2+6+4		
	(1)	(0)	Diff.	(1)	(0)	Diff.	(1)	(0)	Diff.	(1)	(0)	Diff.
Initiated by the MVD	0.682	0.643	0.039***	0.749	0.796	−0.047***	0.753	0.817	−0.064***	0.794	0.942	−0.148***
Initiated by the FSKN	0.314	0.355	−0.040***	0.250	0.204	0.046***	0.245	0.182	0.063***	0.205	0.058	0.147***
Initiated by others	0.003	0.002	0.001***	0.001	0.001	0.001	0.002	0.001	0.001*	0.001	0.000	0.001
Article 228 (use)	0.447	0.364	0.083***	0.530	0.440	0.090***	0.727	0.670	0.057***	0.837	0.793	0.045***
Article 228.1 (sale)	0.550	0.592	−0.042***	0.467	0.537	−0.070***	0.269	0.291	−0.022***	0.160	0.187	−0.027***
Male							0.814	0.837	−0.023***	0.822	0.844	−0.022**
Russian							0.851	0.895	−0.044***	0.876	0.920	−0.045***
Citizen of Russia							0.936	0.957	−0.021***	0.945	0.963	−0.018***
Resident of region							0.893	0.931	−0.037***	0.888	0.915	−0.027***
At most secondary educ.							0.604	0.635	−0.031***	0.593	0.663	−0.070***
Unemployed							0.784	0.810	−0.026***	0.772	0.828	−0.056***
Student							0.0001	0.0005	-0.0004	0.0001	0.0007	−0.0006*
Worker							0.182	0.150	0.032***	0.196	0.133	0.063***
White collar							0.014	0.010	0.004*	0.015	0.014	0.001
Previously charged with offense							0.685	0.590	0.095***	0.682	0.571	0.111***
Under the influence of drugs							0.509	0.354	0.155***	0.536	0.358	0.178***
Under the influence of alcohol							0.014	0.014	0.000	0.016	0.016	0.000
Sentence length, years										3.009	2.417	0.592***
Pleaded guilty										0.346	0.604	−0.258***
Investigation length, days										62.863	56.447	6.416***
Adjudication length, days										31.463	29.702	1.761***
Pretrial detention										0.262	0.152	0.111***
N	76,735	12,417		46,593	4,189		30,728	4,268		14,350	1,516	

* p < 0.10, ** p < 0.05, *** p < 0.01

Note: The table compares means in four samples used in the analysis: from form 1, form 4, forms 1 and 2, and forms 1, 2 and 6 merged with weights from form 4. The samples consist of all heroin-related cases registered in Russia during 2013-2014. The (1) columns present means in the subsamples without observations with missing drug weights, the (0) columns present means in the subsamples of observations with missing drug weights, columns Diff. shows differences in means.

Table B6: Number of stations and their 2013 performance levels by group in relation to reaching the 2013 level

	MVD stations			FSKN stations		
	N	%	2013 level	N	%	2013 level
Total number of crimes:						
Reached	1,243	42.5	24	503	32.4	18
Nothing to reach	0	0	0	0	0	0
Never reached	1,287	57.5	51	1,048	67.6	48
Number of serious and most serious crimes:						
Reached	1,107	37.9	17	511	33.0	16
Nothing to reach	529	18.1	0	210	13.5	0
Never reached	1,287	44.0	34	830	53.5	41
Total	2,923			1,551		

Note: The table presents agency specific numbers and shares of stations that reached the 2013 level during 2014, that did not reach the 2013 level in 2014 and for which the 2013 level is equal to zero. The table also shows the average total numbers of drug crimes and the average numbers of serious and most serious drug crimes registered by stations in 2013, which form their performance targets in 2014.

Table B7: Robustness check

Starting point	Endpoint	Polynomial degree k	Manipulation window		Bunching estimator b	s.e.
			Lower bound r_l	Upper bound r_u		
0.7	10	4	1.5	3.3	7.463***	0.305
0.7	10	4	1.5	3.4	8.879***	0.377
0.7	10	4	1.5	3.5	10.828***	0.526
0.7	10	5	1.1	3.3	7.057***	0.415
0.7	10	5	1.1	3.4	8.123***	0.509
0.7	10	5	1.1	3.5	9.515***	0.691
0.8	**10**	**4**	**1.6**	**3.3**	**6.325*****	**0.325**
0.8	10	4	1.6	3.4	7.256***	0.319
0.8	10	4	1.6	3.5	8.453***	0.398
0.8	10	5	1.2	3.3	7.885***	0.579
0.8	10	5	1.2	3.4	9.165***	0.766
0.8	10	5	1.2	3.5	10.851***	0.956
0.9	10	4	1.7	3.3	6.398***	0.246
0.9	10	4	1.7	3.4	6.787***	0.277
0.9	10	4	1.7	3.5	7.421***	0.309
0.9	10	5	1.3	3.3	7.459***	0.473
0.9	10	5	1.3	3.4	8.466***	0.537
0.9	10	5	1.3	3.5	9.751***	0.572
0.7	11	4	1.7	3.3	6.654***	0.267
0.7	11	4	1.7	3.4	7.756***	0.320
0.7	11	4	1.7	3.5	9.220***	0.392
0.7	11	5	1.2	3.3	7.407***	0.506
0.7	11	5	1.2	3.4	8.706***	0.654
0.7	11	5	1.2	3.5	10.449***	0.802
0.8	11	4	1.8	3.3	5.542***	0.255
0.8	11	4	1.8	3.4	6.202***	0.286
0.8	11	4	1.8	3.5	7.055***	0.353
0.8	11	5	1.3	3.3	7.409***	0.472
0.8	11	5	1.3	3.4	8.591***	0.556
0.8	11	5	1.3	3.5	10.151***	0.718
0.9	11	4	1.9	3.3	4.976***	0.258
0.9	11	4	1.9	3.4	5.388***	0.278
0.9	11	4	1.9	3.5	5.912***	0.308
0.9	11	5	1.3	3.3	6.732***	0.453
0.9	11	5	1.3	3.4	7.565***	0.519
0.9	11	5	1.3	3.5	8.624***	0.615

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The baseline sample from form 1 consists of all heroin-related cases registered in Russia during 2013-2014. To obtain the estimates of bunching, I apply the method described in detail in Appendix C.

Table B8: The effect of reaching 2013 level on the total number of drug crimes and the number of serious and most serious drug crimes registered in 2014

	Stations reached				Stations reached & never reached			
	Total		Serious & most serious		Total		Serious & most serious	
	MVD	FSKN	MVD	FSKN	MVD	FSKN	MVD	FSKN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
m=-5	0.001	0.003	0.004**	-0.001	-0.005***	-0.004**	-0.004***	-0.005***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
m=-4	0.004*	0.001	0.002	0.002	-0.000	-0.002	-0.002*	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
m=-3	0.009***	0.003	0.007***	0.001	0.000	0.000	-0.001	-0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
m=-2	0.011***	0.006**	0.007***	0.000	0.001	0.000	-0.002	-0.002
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
m=-1	0.009***	0.009***	0.010***	0.007***	0.003**	0.001	0.004***	0.001
	(0.002)	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
m=0	0.045***	0.042***	0.045***	0.041***	0.024***	0.020***	0.026***	0.021***
	(0.002)	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
m=1	0.005***	0.003	0.008***	0.002	0.002	0.001	0.004***	-0.001
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
m=2	0.005**	0.001	0.004**	0.003	0.001	-0.003	0.000	0.000
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
m=3	0.002	-0.001	0.004**	-0.002	-0.000	-0.003	0.002	-0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
m=4	0.000	-0.007***	0.001	-0.006**	-0.001	-0.008***	-0.001	-0.006**
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
m=5	-0.002	-0.002	0.001	-0.002	-0.003*	-0.003	-0.001	-0.001
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Constant	0.062***	0.048***	0.045***	0.043***	0.066***	0.054***	0.048***	0.050***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
Station FEs	✓	✓	✓	✓	✓	✓	✓	✓
Day-of-the-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
Region-by-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.179	0.105	0.126	0.116	0.216	0.181	0.182	0.157
N	453,695	183,595	404,055	186,515	1,066,895	566,115	873,810	489,465

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents event study estimates by agency for two performance indicators: the total number of drug crimes and the number of serious and most serious drug crimes. In columns (1)-(4), the sample includes only stations that reached the 2013 performance level during 2014. In columns (5)-(8), the sample includes both stations that reached and stations that did not reach the 2013 performance level during 2014. The dependent variable is the inverse hyperbolic sine transformation of either the daily total number of drug crimes or the daily number of serious and most serious drug crimes calculated based on the sample of all drug-related cases from form 1. The specification includes station, day-of-the-month and region-by-month fixed effects. Standard errors are clustered by station.

Table B9: The effect of reaching 2013 level on the total number of heroin-related crimes and the number of serious and most serious heroin-related crimes registered in 2014

	Stations reached				Stations reached & never reached			
	Total		Serious & most serious		Total		Serious & most serious	
	MVD	FSKN	MVD	FSKN	MVD	FSKN	MVD	FSKN
m=-5	-0.015*	-0.021	0.000	0.000	-0.014**	-0.016	-0.001**	-0.001
	(0.008)	(0.018)	(0.001)	(0.001)	(0.006)	(0.013)	(0.001)	(0.001)
m=-4	-0.009	-0.014	-0.000	0.000	-0.006	-0.023*	-0.001	-0.001
	(0.009)	(0.019)	(0.001)	(0.001)	(0.006)	(0.014)	(0.001)	(0.001)
m=-3	-0.017*	0.024	0.002*	-0.001	-0.010	-0.008	0.000	-0.002**
	(0.010)	(0.019)	(0.001)	(0.001)	(0.007)	(0.012)	(0.001)	(0.001)
m=-2	-0.009	-0.001	0.001	-0.001	-0.005	-0.027**	-0.001	-0.002***
	(0.011)	(0.019)	(0.001)	(0.001)	(0.008)	(0.013)	(0.001)	(0.001)
m=-1	-0.005	0.001	-0.000	0.001	0.001	-0.024*	-0.000	-0.001*
	(0.012)	(0.019)	(0.001)	(0.001)	(0.009)	(0.013)	(0.000)	(0.001)
m=0	-0.002	0.004	0.007***	0.006***	0.004	-0.017	0.004***	0.002*
	(0.011)	(0.018)	(0.001)	(0.002)	(0.009)	(0.015)	(0.001)	(0.001)
m=1	-0.002	-0.027	0.001	-0.000	0.008	-0.038**	0.001	-0.001*
	(0.013)	(0.020)	(0.001)	(0.001)	(0.012)	(0.018)	(0.001)	(0.001)
m=2	-0.021	-0.006	0.000	-0.000	-0.009	-0.022	-0.000	-0.001
	(0.014)	(0.024)	(0.001)	(0.001)	(0.013)	(0.022)	(0.001)	(0.001)
m=3	-0.029**	-0.021	0.001*	-0.000	-0.016	-0.034	0.001	-0.001
	(0.015)	(0.027)	(0.001)	(0.001)	(0.014)	(0.024)	(0.001)	(0.001)
m=4	-0.019	-0.058*	0.000	-0.002*	-0.006	-0.060**	-0.000	-0.003***
	(0.017)	(0.029)	(0.001)	(0.001)	(0.016)	(0.030)	(0.001)	(0.001)
m=5	-0.020	-0.055*	0.000	-0.001	-0.004	-0.047	0.000	-0.002*
	(0.017)	(0.031)	(0.001)	(0.001)	(0.016)	(0.029)	(0.001)	(0.001)
Constant	0.167***	0.165***	0.008***	0.008***	0.189***	0.240***	0.011***	0.012***
	(0.006)	(0.011)	(0.000)	(0.001)	(0.004)	(0.005)	(0.000)	(0.000)
Station FEs	✓	✓	✓	✓	✓	✓	✓	✓
Day-of-the-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
Region-by-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.411	0.334	0.068	0.089	0.428	0.354	0.156	0.138
N	30,642	9,378	404,055	186,515	69,757	29,774	873,810	489,465

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents event study estimates by agency for two performance indicators: the total number of drug crimes and the number of serious and most serious drug crimes. In columns (1)-(4), the sample includes only stations that reached the 2013 performance level during 2014. In columns (5)-(8), the sample includes both stations that reached and stations that did not reach the 2013 performance level during 2014. The dependent variable is the inverse hyperbolic sine transformation of either the daily total number of heroin-related crimes or the daily number of serious and most serious heroin-related crimes calculated based on the sample of all drug-related cases from form 1. The specification includes station, day-of-the-month and region-by-month fixed effects. Standard errors are clustered by station.

Table B10: Regression estimates of bunching probabilities in response to reaching the previous year's benchmarks

	Stations reached				Stations reached & never reached			
	First threshold		Second threshold		First threshold		Second threshold	
	MVD	FSKN	MVD	FSKN	MVD	FSKN	MVD	FSKN
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reached, total	-0.054	-0.019			-0.061	0.011		
	(0.083)	(0.047)			(0.051)	(0.032)		
Reached, serious & most serious			-0.142*	0.020			-0.077*	-0.019
			(0.078)	(0.033)			(0.039)	(0.024)
Charge use	0.576**	0.105	0.848***	0.129	0.428***	0.316**	0.126**	0.137*
	(0.249)	(0.200)	(0.175)	(0.181)	(0.059)	(0.135)	(0.057)	(0.082)
Charge sale	0.346	-0.158	0.857***	0.143	0.139**	0.053	0.104*	0.113
	(0.233)	(0.203)	(0.186)	(0.182)	(0.055)	(0.135)	(0.057)	(0.086)
Constant	-0.061	0.440**	-0.685***	0.059	0.182***	0.248*	0.010	0.074
	(0.234)	(0.200)	(0.189)	(0.181)	(0.053)	(0.134)	(0.057)	(0.083)
Station FEs	✓	✓	✓	✓	✓	✓	✓	✓
Day-of-the-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
Region-by-month FEs	✓	✓	✓	✓	✓	✓	✓	✓
R-squared	0.601	0.382	0.546	0.266	0.439	0.360	0.316	0.218
N	3,139	691	3,171	592	9,214	3,663	8,230	3,220

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Note: The table presents regression estimates of bunching probabilities above the first and second thresholds by agency. The samples include heroin-related cases from form 1 registered: by stations that reached the 2013 performance level during 2014 in columns (1)-(4), and by both stations that reached the 2013 level of performance and stations that did not reach it in columns (5)-(8). The sample for the first threshold is restricted to drug quantities between 0 and 1.6 grams (columns (1)-(2), (5)-(6)), the sample for the second threshold is restricted to drug quantities between 0.8 and 10 grams (columns (3)-(4), (7)-(8)). The dependent variable is the binary variable equal to 1 if the drug quantity is in a bunching area (between 0.5 and 1 gram for the first threshold and between 2.5 and 3.3 grams for the second threshold). The specification includes station, day-of-the-month and region-by-month fixed effects. Standard errors are clustered by station.

Table B11: Summary statistics

	Overall		Manipulation region	
	Mean	N	Mean	N
Male	0.814	30,728	0.830	5,008
Russian	0.851	23,025	0.861	3,303
Citizen of Russia	0.936	30,728	0.951	5,008
Resident of region	0.893	28,294	0.895	4,726
At most secondary education	0.604	30,728	0.594	5,008
Unemployed	0.784	30,722	0.755	5,008
Student	0.0001	30,722	0.000	5,008
Worker	0.182	30,722	0.209	5,008
White-collar	0.014	30,722	0.015	5,008
Previously charged with offense	0.685	30,728	0.689	5,008
Under the influence of drugs	0.509	30,728	0.525	5,008
Under the influence of alcohol	0.014	30,728	0.014	5,008

Note: The baseline sample from forms 1 and 2 consists of all heroin-related cases registered in Russia during 2013-2014. See the text for further details defining the subsample in the manipulation region around the threshold.

C Estimation Details

To estimate the magnitude of the response of police officers around the punishment threshold, I adapt the standard method from the bunching literature (Saez 2010, Chetty et al. 2011, Kleven and Waseem 2013).

To obtain the bunching estimator, I estimate the counterfactual density of seized drug quantities by fitting a high-order polynomial to the observed distribution, excluding the region $[r_l, r_u]$ around the threshold $\overline{D}$:

$$C_j = \sum_{k=0}^{p} \beta_k R_j^k + \sum_{r=r_l}^{r_u} \gamma_r * \mathbb{1}[R_j = r] + \nu_j, \tag{C1}$$

where C_j is the number of cases in bin j, p is the order of the polynomial, R_j is the midpoint of bin j. For heroin-related cases, bin size is set to 0.1 gram, which is approximately the smallest dose that can be bought. To obtain the counterfactual distribution, I estimate the predicted values from (C1), omitting the γ_r shifters for smoothing the density around the threshold:

$$\hat{C}_j = \sum_{k=0}^{p} \hat{\beta}_k R_j^k. \tag{C2}$$

The key assumption for the bunching estimator, as well as for any other bunching methodology, is that without manipulation the actual distribution of outcomes in the bunching window would follow the polynomial estimated outside this window.

Comparing the counterfactual and observed distributions, I can estimate the missing mass to the left of the threshold, and the excess bunching mass to the right of the threshold:

$$\hat{M} = \sum_{j=r_l}^{\overline{D}} (\hat{C}_j - C_j) \text{ and } \hat{B} = \sum_{j=\overline{D}}^{r_u} (C_j - \hat{C}_j). \tag{C3}$$

To determine the lower and upper bounds of the excluded interval, I follow Kleven and Waseem (2013). Because the excess bunching above the threshold is quite sharp (compared to the missing mass), the upper bound can be determined visually. With r_u

fixed, I set the lower bound r_l such that $\hat{B} = \hat{M}$.

Finally, I obtain a bunching estimate for the magnitude of manipulation by calculating the ratio of excess mass to the average height of the counterfactual density above the threshold:

$$\hat{b} = \frac{\hat{B}}{\sum_{j=\overline{D}}^{r_u} \hat{C}_j / N}, \tag{C4}$$

where N is the number of bins in the interval $[\overline{D},\ r_u]$.

Since the paper studies the response of the police around the second threshold, I exclude the area around the first threshold, as well as the long tail covering the third threshold.