Under Pressure? Performance Evaluation of Police Officers as an Incentive to Cheat: Evidence from Drug Crimes in Russia*

Ekaterina Travova†(CERGE-EI‡, EUSP§)

May 2020

*I am grateful to Andreas Menzel for generous support at all stages of this project. I also thank Stepan Jurajda, Randall Filer, Vasily Korovkin, Sergei Mikhalishchev, Dmitriy Skougarevskiy, Alexey Knorre, Nikolas Mittag, Jan Palguta and seminar participants at CERGE-EI, EUSP and The City University of New York for helpful discussions and comments. All errors remaining in this text are the responsibility of the author. This study was supported by Charles University, GAUK project No. 702218 and the H2020-MSCA-RISE project GEMCLIME-2020 GA No. 681228.

†E-mail: ekaterina.travova@cerge-ei.cz. Address: Politickych veznu 7, 111 21 Prague, Czech Republic.

‡CERGE-EI, a joint workplace of Charles University and the Economics Institute of the Czech Academy of Sciences, Politickych veznu 7, 111 21 Prague, Czech Republic.

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

Abstract

This paper provides an empirical analysis of manipulations of seized drug amounts by police officers, based on a unique dataset that contains full information on drug crimes in Russia reported during 2013-2014. First, using a bunching estimator, I document a significant excess mass of heroin cases above the punishment cliff. The mass is 6.325 times greater than the average number of cases in a counterfactual scenario without manipulation. Next, I employ an event study approach to investigate the incentives for police officers to manipulate and find that the motivation arises from the officers' performance evaluation system. One of the main indicators applied for evaluation is the number of serious and most serious drug crimes, which can be easily increased by moving offenders from below to above the threshold. Exploring the dynamics of this indicator during a calendar year, I document that it increases by 23% in the month when a police station is close to reaching the previous year's level of performance, current target. Comparing the performance evaluation systems of two separate drug control agencies, I find further evidence of this response to performance requirements. Finally, applying a novel bunching technique, I determine that police officers are more likely to manipulate the drug amounts seized from repeat offenders. The overall effect of manipulation on the sentence length of drug users is an additional year of incarceration, which is a 67% increase, compared to the average sentence length without manipulation.

JEL Classification: H11, H76, K14, K42.

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

1 Introduction

For decades, global anti-drug policies have been based on general principals of eliminating the production, sale and consumption of any illegal psychoactive compounds, and have involved harsh law enforcement and even militarization. Even though the likely failure of a war on drugs has been widely acknowledged, these policies are still in place in many countries, leading to insufficient budget spending and unequal treatment of different groups of drug offenders, with a strong focus on drug users (International Drug Policy Consortium 2018). While the harmful consequences of poorly designed anti-drug laws are well-studied, there is little evidence as to how this war creates incentives that affect the behavior of police, prosecutors and judges. In this paper, I provide evidence of the importance of incentives in the law enforcement agencies, analyzing possible manipulation of amounts of drugs seized by police officers in Russia during 2013-2014.

Russia is a notable example in this context, as according to The Federal Penitentiary Service of Russia (2017), at the end of 2016 almost a quarter of all prisoners in the country were convicted of drug related crimes. Moreover, 70% of these drug offenders were imprisoned for using, not for producing or selling drugs[1]. At the same time, the Russian law for drug possession for personal use carries penalties of up to 15 years of incarceration. In addition, the performance evaluation of police officers was and is still based on easily measurable and quantifiable indicators relative to past performance, which establishes a strong motivation to present as many prosecutions of serious and most serious crimes as possible.

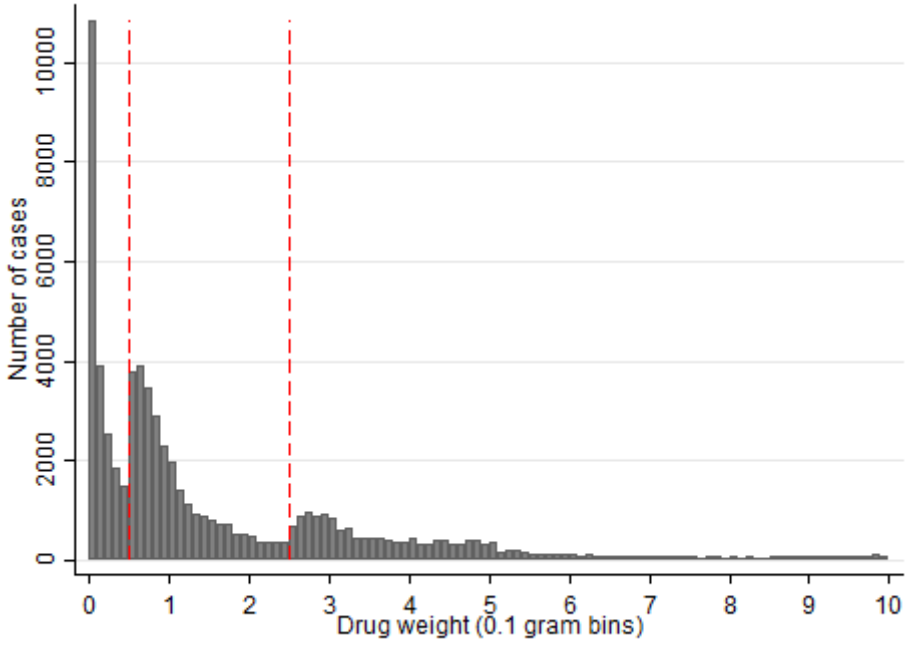
Figure 1 (Knorre 2017) shows the distribution of heroin cases across drug quantities seized in Russia during 2013-2014. Two thresholds, indicated by dashed lines, determine the scales of seizures (less than significant, significant and large drug amounts) that define the severity of crime and punishment[2]. It reveals a striking pattern suggesting that, at the moment of arrest, many people possess a drug

[1] In comparison, worldwide, 1 in 5 prisoners are incarcerated for drug offences. 83% of them are convicted of drug possession for personal use (International Drug Policy Consortium 2018).

[2] There is also a third threshold (at 500 grams for heroin) that is not depicted on the graph.

amount just above a threshold beyond which they will be convicted of a more serious crime. In addition, there is a missing mass of cases just below the thresholds. This phenomenon is suggestive of manipulation of drug quantities seized by the police driven by a rational response of officers to performance evaluation requirements, which so far has only been alleged by various media reports[3]. Presenting a rigorous analysis of possible drug manipulation, my paper fills the gap and provides novel insights for policy making.

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 amounts), the severity of crime and punishment. This graph replicates Figure 5 in Knorre (2017).

First, using a standard bunching estimator[4], I document a significant excess mass

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

[4]The approach was initially developed by Saez (2010), Chetty et al. (2011) and Kleven and Waseem (2013) to study the response to tax regulation. However, increasingly it is applied in many other settings, for example education (Brehm et al. 2017), pensions (Manoli and Weber 2016), social insurance (Le Barbanchon 2016), car speed regulation (Goncalves and Mello 2017), welfare programs (Camacho and Conover 2011), procurement (Palguta and Pertold 2017) and others.

of heroin cases just above the punishment cliff. This mass is 6.325 times greater than the average number of cases in manipulation window above the threshold in a counterfactual world without manipulation. I also present evidence in support of the hypothesis that the observed discontinuities in the distribution are the result of moving people from below to above the threshold, not just due to a self-selection of offenders or differential enforcement around the cutoff.

Next, I employ an event study approach to investigate the response of police officers to the incentives to manipulate seized drug amounts and identify that the pattern observed is consistent with the officers' performance evaluation system. One of the main indicators applied for evaluation is the number of serious and most serious drug crimes, which can be easily increased by moving offenders from below to above the threshold. At the same time, previous year's level of this indicator is used either as a direct target for a current period or as a basis for forecast to set the incentives. Exploring how the number of serious and most serious drug crimes evolves over the calendar year, I document that it increases by 23% in the month when a police station reaches the previous year's level of performance. Exploiting changes in incentives across thresholds and differences in the evaluation approaches of the two law enforcement agencies responsible for drug control in 2013-2014, I find further evidence of this behavioral response to performance requirements. While one of the agencies compared the performance of regional offices cross-sectionally in order to set incentives, the other agency used a comparison within each office over time. As the results show, it leads to a higher magnitude of manipulation in the second case, when the performance target is known to police officers and, hence, incentives are stronger.

It is worth noting that bribery might be another motive for police officers to manipulate the drug amounts or to threaten offenders with possible manipulation. However, while there is no direct evidence, it seems logical to assume that if an offender decides to pay a bribe he does it to buy himself out of prison, not just to decrease the sentence. This means that those individuals are most likely not in the

database at all and bribery cases are undetectable. At the same time, if bribery cases are not distributed uniformly across seized drug amounts, their omission might affect the manipulation window estimation and, hence, the bunching estimate. However, as the sensitivity analysis shows, the change of parameters does not significantly influence the result.

This study adds to the growing research on the performance evaluation and incentive schemes in the public sector. As was highlighted in the seminal paper by Holmstrom and Milgrom (1991), strong incentives could be inappropriate in government jobs, resulting in a negative effect, given that many civil servant's jobs are characterized by multitasking. At the same time, some objectives that the civil servant has to attend to are more easily measured than others. In this context, strong incentives could detract attention away from tasks that are not easily measured, or even induce fraudulent behavior. For example, Mas (2006) studies the effect of arbitration decisions on the performance of police officers. He finds that, in the case of favorable outcomes, crime reports rise and arrest rates and average sentence length decline. In contrast, Prendergast (2001) shows that increased external oversight leads police officers to reduce crime-fighting activities to avoid possible investigation. Banerjee et al. (2012) run a sequence of experiments in India and find that reduced autonomy of police station managers reduces police effectiveness[5]. My paper broadens the empirical evidence needed for better understanding of accountability issue in public sector. It provides useful insights about optimal incentive structure that could be also applied to other than law enforcement setups where inappropriate motivation might trigger the dishonest behavior[6].

In the second part of the paper, adopting the novel bunching technique developed by Diamond and Persson (2016), I identify the characteristics of victims

[5]For a review of field experiments on selection, incentives and monitoring in public sector see Finan et al. (2017)

[6]See Zitzewitz (2012) for a review. For instance, the seminal papers by Burgstahler and Dichev (1997) and Degeorge et al. (1999) study the distribution of earnings reported by firms and find a bunching just above zero earnings or analysts' forecast. This so called earnings management (sometimes referred to as manipulation) could be driven by the response of senior executives to the incentives - implicit and explicit dependence of their rewards on the firm earnings. I find the similar pattern in the behavior of police officers who target their own previous year performance.

of manipulation of seized drug amounts, and the effect of the manipulation on sentence length. The results suggest that while the demographics and socio-economic status of offenders do not have a significant effect on a police officer's decision to manipulate[7], having a criminal history increases the probability of becoming the victim of such manipulation. This adds to the discussion on ricidivizm as well as, more generally, on discretion and discrimination in law enforcement. The overall estimated effect of the manipulation on the sentence length of drug users is around one additional year of incarceration (compared to an average sentence length without manipulation is 1.5 years), and the magnitude of this effect is not dependent on a guilty plea. In contrast to the most existing studies (Anbarci and Lee 2014, Goncalves and Mello 2017, Bjerk 2005, Ulmer et al. 2007, Rehavi and Starr 2014, Bjerk 2017)[8], I analyze a novel setting where low enforcers behave in more repressive way, intentionally increasing, not decreasing, the penalty for the offender. Another study that also finds bunching of drug offenders above the punishment cliff is Tuttle (2019). However, the main focus of this paper is on racial discrimination as the main cause of observed sharp increase in the fraction of crack-cocaine cases above the mandatory minimum threshold. Additionally, in contrast to my study, Tuttle (2019) finds the effect for drug traffickers, which is specifically due to prosecutorial discretion.

The total cost of drug manipulation to society is difficult to calculate precisely. However, undoubtedly, the welfare loss exceeds any benefits from keeping drug users off the streets. Leaving aside ethical issues, one can clearly see the negative effect of

[7]In contrast, Volkov (2016) analyzes all felony cases processed by federal district courts during 2009-2013, and finds a significant bias in judges' decisions against entrepreneurs, and offenders of low socio-economic status. Kurmangaliyeva (2017) determines that the Russian judicial system is more lenient to wealthier defendants.

[8]Anbarci and Lee (2014) and Goncalves and Mello (2017) 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 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 a mandatory minimum law.

multiple manipulations widely discussed in the media: they significantly lower public trust in the police increasing the level of perceived insecurity. This, in turn, decreases the effectiveness of law enforcement and the legitimacy of police actions. In addition, each year the government spends an enormous amount of money on the Penitentiary Service[9], but drug addicts do not receive any treatment during incarceration. After release, most of them start taking drugs again becoming repeat offenders. According to this study, the probability of being manipulated and receive longer sentence is higher for these offenders. This leads to additional budget spending that is far from optimal.

The rest of the paper is organized as follows. Institutional context and data are described in Section 2. In Section 3, I provide the empirical strategy presenting the results in Section 4. 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)[10], also known as the Drug Police. The responsibilities of this agency included control of legal and illicit drug trafficking, prevention of drug abuse, drafting of state policy, and legal regulation. The FSKN shared concurrent jurisdiction with the Public Security Service (Police) of the

[9]The annual budget of the Penitentiary Service of Russia was constantly growing since the establishment and reached $5 billion in 2015 that is comparable with the budget of some European countries: for example, Albania ($4.5 billion) or Moldavia ($2 billion).

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

Ministry of Internal Affairs (MVD)[11], 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 had to be on significant cases (drug trafficking, organized crime, large drug amounts), the MVD mostly dealt with routine low-profile cases, such as drug use and small-quantity drug sales. Almost two thirds of all drug related cases registered during 2013-2014 were initiated by the MVD. At the same time, the MVD provided many other public security functions, and drug control was not its only responsibility. This could explain geographical differences in the numbers of cases initiated by the two agencies. According to Knorre and Skougarevskiy (2015), the FSKN more often operated in less populated localities, taking a targeted approach to searching for drug offenders, while the MVD more often seized drugs in densely populated regions. 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 the Criminal Code of the Russian Federation. The severity of a penalty under these articles depends on the amounts of drug seized, which are classified as "significant", "large" or "especially large" (Appendix, Table B1).

Article 228 imposes criminal responsibility for the illegal acquisition, storage, transportation, manufacture or processing of drugs. The punishment for drug possession on a significant scale, with no intention to sell, (the least serious crime) is imprisonment for up to three years. On large and especially large scales (serious and most serious crimes), the punishment is imprisonment for three to ten and ten to fifteen years, respectively (Appendix, Table B2). 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 amount of drug seized is less than significant, the person can only be brought to administrative responsibility punished with a fine or administrative arrest for up to 15 days.

[11]Ministerstvo vnutrennih del Rossiiskoi Federacii, MVD.

Article 228.1 outlaws the illegal manufacture, sale or dispatch of drugs, which is punishable by imprisonment for four to eight years if the amount is less than significant, eight to fifteen years for a significant amount, and ten to twenty years for a large amount. Especially large amounts carry a fifteen to twenty years, or life, sentence (Appendix, Table B2). In this case, the crime is serious if the amount of drug seized is less than significant, and most serious if the amount is significant or higher (Knorre 2017).

The practice of plea bargaining was introduced in 2001 and became quite common in Russia. During 2013-2014 around 60% of all cases (30% of drug related offences) were adjudicated in relation to plea bargaining. Pleading guilty significantly simplifies the procedure: a conviction is pronounced without the actual examination of evidence at a court hearing. In addition, a person that accepts a plea bargain waives the right to appeal. In return, by pleading guilty the offender lowers the upper bound of the sentencing range by one third.

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 evaluation system. The system used by the FSKN was based on performance indicators that included the number of drug crimes solved, the number of serious and most serious drug crimes solved and the total amount of drugs seized (per 100 officers). 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 the sum of these positions.

On the other hand, the system of performance evaluation of the MVD was based on an overall score for each regional office, which was calculated as the weighted average of values of non-departmental and departmental assessments. Among parameters that characterized a police office's work were the number of crimes solved, the number of serious and most serious crimes solved and the number of serious and most serious drug crimes solved (per 100 officers). These statistics belonged to the highest-weighted group of parameters in the overall score. Crucially,

in contrast to the FSKN, the MVD stations compared performance with their own evaluation in the previous year[12] (Novikova 2014).

Regarding the consequences for police officers, there were disciplinary and financial measures that could be applied to personnel based on their performance. If the officer met or surpassed the targets, he might receive a monetary bonus to his monthly salary or promotion (for high-profile cases). While there was no guarantee that the officer would be rewarded for good performance, he certainly was reprimanded, warned or even fired in the case of unsatisfactory performance. In addition, the officer could be deprived of monthly bonuses, in addition to the fixed salary, if he did not fulfill the plan.

Thus, the system of performance evaluation presented strong incentives for police officers to show the required level of cases and prosecutions. At the same time, even though the FSKN's system was more transparent, it was more difficult for the FSKN stations to set the "necessary" amount of manipulations, 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 manipulation by these two agencies, which I investigate in more detail in Section 5.2.

In addition, not only police officers were evaluated based on easily measurable and quantifiable indicators, but also prosecutors and judges. The evaluation of the prosecutors was tighten 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 incentivizes prosecutors and judges to behave in a repressive way[13] and created an enabling environment for

[12]Formally, after the reforms in 2011, the MVD offices had to compare their performance across units rather than relative to the previous period. However, locally this did not work due to the complexity of the system.

[13]In 2018, the rate of acquittal reached its historic minimum in post-Soviet Russia - 0.24% compared to 0.3-0.4% in the 1990s (Sokolov 2019).

fabricating cases at the lower level.

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[14]. 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 offender;

form 3: criminal proceedings;

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

form 6: trial results[15].

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, which coincide only for 92.8% of cases. However, the distributions of cases across drug amounts do not differ significantly. Form 1 quantaties are determined by a police officer, who has to weight the drug seized, while form 4 is created at a later stage after the prosecutor's approval of case initiation and contains drug amounts measured in the laboratory. Therefore, to estimate the magnitude of possible manipulation and investigate to what extent it varies by drug type, article and agency, I use data from form 1. In order to identify characteristics of victims of manipulation, I merge data from forms

[14]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.

[15]Form 1 is completed by an investigator when he or she decides to initiate criminal proceedings that should be approved by a prosecutor. During the investigation, forms 2, 3 and 4 are created. 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 should contain information on victims, while drug crimes are victimless. Form 6 is filled in by a judge. After closing the case, all forms should be converted from written to electronic form and submitted to an information center (Shklyaruk and Skougarevskiy 2015).

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. I restrict the sample to cases related only to drug use for two reasons. First, separation by article is needed due to the existing specifics of different types of crime. Second, the drug dealers sample from merged dataset based on forms 1, 2, 6 and 4 contains an insufficient for bunching techniques number of observations.

The initial dataset based on form 1 contained data on 518,979 drug crimes including 89,152 heroin related cases. 14% of cases related to heroin were excluded from the sample because the amount of drug seized was missing[16]. Missing values are likely to be caused either by inaccurate completion of forms by police officers or by mistakes during the conversion of the forms into electronic files. Additionally, under some circumstances, a case can be initiated without drug seizure. See Table B3 in the Appendix for more information on the samples discussed in this paper and missing values. While differences in means of the working sample and the set of observations with missing weights are statistically significant for almost all factors, their values themselves are small in most cases. 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 drug) or in the case of refusing to plead guilty, which leads to a full investigation, compared to the simplified procedure under the plea bargain.

3 Empirical Strategy

To study the differences in magnitudes of manipulation and how they could be related to the incentive structure, I apply the standard bunching estimator (Saez 2010, Chetty et al. 2011, Kleven and Waseem 2013). This method allows to construct a measure of excess mass of offenders above the threshold by comparing actual and

[16]The form 4 dataset included information on 236,989 drug crimes out of which 50,782 were related to heroin. Due to missing drug weight, 8% of heroin related cases were also excluded from the analysis.

counterfactual distributions around this threshold.

To check that the results are insensitive to the choice of estimation parameters, I repeat the procedure described in Appendix C.1, using different polynomial orders k, values of upper bound r_u and starting points after the exclusion of the area around the first threshold. I also vary the upper point for drug weight where I cut the sample since the long tail with few observations does not contain much information. This robustness check also allows me to choose the estimation parameters that are applied in further analysis.

Additionally, in order to analyze the effect of performance requirements on the behavior of police officers in more detail, I use an event study framework. The identification strategy exploits the variation in the timing of reaching the previous year number of serious and most serious crimes within stations (controlling for month effects). This approach restricts the sample to those police stations that during the study period (2014) surpassed their 2013 "benchmark" (conditional on it not being zero). For each station i I calculate the total number of serious and most serious drug crimes per month and, comparing these values with the 2013 level, determine when the station reached this level. This allows me to define a set of event study dummies: each dummy equals 1 if the current month r is t periods before/after the event - reaching the "benchmark". In total, I have 13 dummies: 1 for the event month, 6 for pre- and 6 for post-periods, since, on average, stations reach the level needed after 6 months. The logarithm of monthly number of serious and most serious drug crimes Y_{ir} is my main outcome which I regress on event study dummies and station and month fixed effects:

$$Y_{ir} = \sum_{t=-6}^{6} \alpha_t \mathbb{1}[T_{ir} = t] + \gamma_i + \delta_r + \varepsilon_{ir}. \tag{1}$$

To test for a break in the level and decline of the number of serious and most serious drug crimes I apply a more parametric specification. First, controlling for station and month effects, I regress my main outcome on dummy $Event_{ir}$ that equals

1 for the month when station reaches the 2013 level and dummy $Post\text{-}reaching_{ir}$ that equals 1 for all periods after reaching the 2013 level.

$$Y_{ir} = \beta_1 Event_{ir} + \beta_2 Post\text{-}reaching_{ir} + \gamma_i + \delta_r + \varepsilon_{ir}. \tag{2}$$

Then, I add continuous variable $Event\ time_{ir}$ that takes value of 0 in event month with negative values before and positive values after $t = 0$, and its interaction with $Post\text{-}reaching_{ir}$:

$$\begin{aligned} Y_{ir} = \beta_1 Event_{ir} + \beta_2 Post\text{-}reaching_{ir} + \beta_3 Event\ time_{ir} + \\ + \beta_4 Post\text{-}reaching_{ir} Event\ time_{ir} + \gamma_i + \delta_r + \varepsilon_{ir}. \end{aligned} \tag{3}$$

Finally, I check for a difference in the effects of reaching the previous year level for two drug control agencies.

To estimate the mean characteristics of manipulation victims, and the impact of manipulation on sentence length and the probability of pleading guilty, I follow Diamond and Persson (2016). Their technique is based on comparison of the observed and estimated counterfactual distributions of a parameter studied. This allows me to recover the average summary statistics of offenders moved above the threshold and the causual effect of this movement on outcome variables. Appendix C.2 and C.3 provides additional technical details. The standard errors are calculated using bootstrap.

4 Results

4.1 Manipulation of Seized Drug Amounts

Among all drugs in my data, I only find significant bunching around second threshold in the case of heroin. Graphs with distributions of some of the other most often seized drugs are in the Appendix, Figure A1. The bunching estimator for all heroin cases

from form 1 is 6.325 (Appendix, Figure A2). This means that the excess mass above the 2.5 grams threshold is almost six times greater than the average number of cases that would be in the manipulation window above the threshold in the counterfactual world without manipulation. The effect is slightly stronger in merged samples from forms 1 and 2 and forms 1, 2, 6, and 4 (Appendix, Figure A3), supporting the result observed in the initial dataset from form 1.

The result is robust to variations in the specification I use (Appendix, Table B4). For all estimations of the mean victim characteristics and the effect of manipulation I set the upper bound of the manipulation window equal to 3.3 and the polynomial degree equal to 4. In addition to good fit, these parameters give the smallest possible estimate of bunching that does not allow overstatement of the effect of manipulation.

A simple explanation for the bunching only in the case of heroin could be that it is easier for a police officer to manipulate this drug, given the small amounts needed to cross the threshold. In addition, heroin was one of the most popular and potent drugs in Russia at that time. A large share of heroin users were from the lowest socio-economic class, which made the manipulation even easier for police officers. For example, according to the interview with a former policeman (Nadezhdin and Matveeva 2019), in many cases, the police officers used the following scheme. They receive a call from somebody reporting that at the hall of his building there are drug users, under the influence of drugs. The police arrives and finds unconscious people and an amount of heroin. If the drug quantity is below the threshold, the officers could add flour, sugar or any other white powder and arrest the users, leading to their conviction for a serious crime. The rules applied to weighting the drugs seized make the manipulation a relatively costless way to meet the performance requirements. Thus, to move a person above the threshold, a police officer does not even need to have the additional amount of heroin itself, since, according to the law, the drug quantity seized is determined not by the weight of the pure substance but by the weight of the entire mixture.

Another point that should be clarified is the choice of the counterfactual

distribution's form. If we assume that offenders are rational agents we would obtain a counterfactual distribution with humps just below the thresholds. Since these humps could not be estimated, I make the assumption of the counterfactual distribution with a smoothly decreasing shape. In this case, the bunching estimator yields a lower bound. Indirect evidence in support of a smoothly decreasing shape comes from the distributions of cases related to the other types of drugs (Appendix, Figure A1), which do not have bunching (at least around the second threshold).

Eyeballing the distributions of seized amounts of heroin from different Russian regions shows that manipulation magnitudes vary across country. However, rigorous estimation is unfeasible in this case, since splitting the sample into 83 subsamples (as many as there are regions) significantly reduces statistical power when estimating the region-specific extent of manipulation. Therefore, I divide all regions into only two groups, on or away from the main drug-trafficking routes[17]. Figure A4 in the Appendix shows that the magnitude of manipulation in regions along the routes is more than twice as high as that in regions away from the routes. This could be explained by the following factors. First, in regions which are on the drug-trafficking routes, the share of population that could potentially be manipulated (drug users, drug dealers) is greater. Second, police officers in these regions might be more experienced in dealing with drug related crimes. Additionally, there might be differences in incentive structures in the two drug control agencies, which is investigated in more detail in the following subsection.

4.2 Incentives for Manipulation of Seized Drug Amounts

The significant bunching found above raises the question of what causes the police officers' response to the threshold. According to Paneyakh (2014), the main driving force for dishonest behavior is the system of performance evaluation of police officers. To investigate this, I exploit changes in punishment across thresholds, and differences

[17]Information on drug-trafficking routes is taken from the website of Russia's international news agency https://ria.ru/20100603/242406939.html. Accessed on December 1, 2018.

in the evaluation approaches of the two drug control agencies.

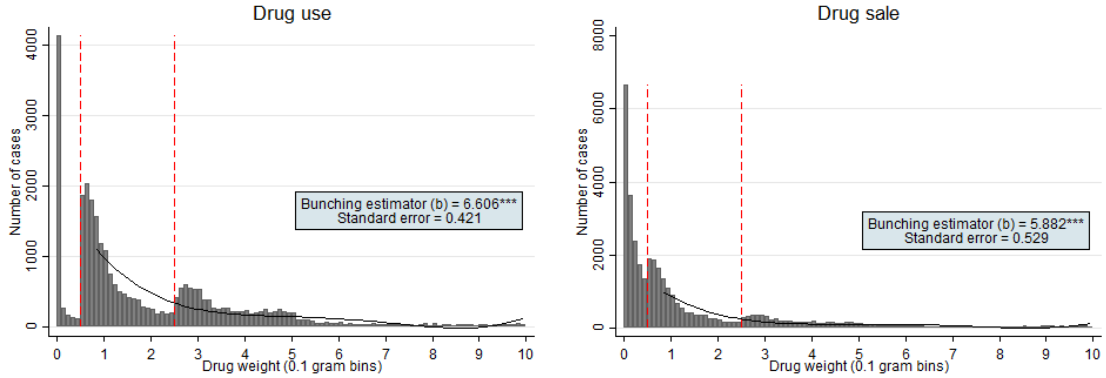
Combining the information on sanctions for drug related crimes and the systems of performance evaluation of police officers discussed in Section 3.1. suggests the following incentives for moving offenders from below a threshold to above it. In the case of drug use (article 228), crossing the first threshold changes the status of offence from administrative to criminal[18], which improves the performance statistics, since in the MVD's system the weight for the number of (drug) crimes solved is much higher than for the number of administrative offences registered. In the FSKN system, administrative offences are not taken into account at all. In the case of drug sale (article 228.1), crossing the first threshold increases the severity of crime from serious to most serious, but does not contribute to the overall evaluation because the number of serious and most serious (drug) crimes are calculated together. I do not analyze possible manipulations around the first threshold for two reasons. First, data on offences below the first threshold could be incomplete due to police officers' reluctance to deal with cases that do not affect their performance evaluation significantly. In addition, some officers might show leniency towards minor offences and not register them. Second, the number of weight bins that could be defined below the threshold is insufficient for estimating the counterfactual distribution.

These issues do not arise in the analysis of police officers' responses to the second threshold. In the case of drug use, crossing the second threshold increases the severity of crime from least serious to serious, which in turn positively affects the evaluation. The incentive for moving offenders from below to above the threshold in the case of drug sale is ambiguous, since manipulation does not directly contribute to performance indicators. However, it could be explained by police officers' concern about losing "points" if a drug sale case is requalified to a drug use case (for example, storage without the purpose of sale). At the same time, if the drug amount seized is large (above the second threshold), a requalification only decreases the severity of the crime (from most serious to serious). However, that does not change the number

[18]The difference between administrative and criminal offences is similar to the difference between misdemeanor and felony.

of serious and most serious (drug) crimes solved by the police and, hence, does not worsen the performance statistics. Figure 2 presents a sharper graph and slightly higher bunching estimate for drug users (left) than for drug dealers (right), which could be explained by different incentives at the threshold. In addition, drug users are the significantly larger group of drug offenders, as well as much easier to locate and, hence, manipulate.

Figure 2: Distributions of cases related to drug use (left) and drug sale (right) across quantities of heroin seized



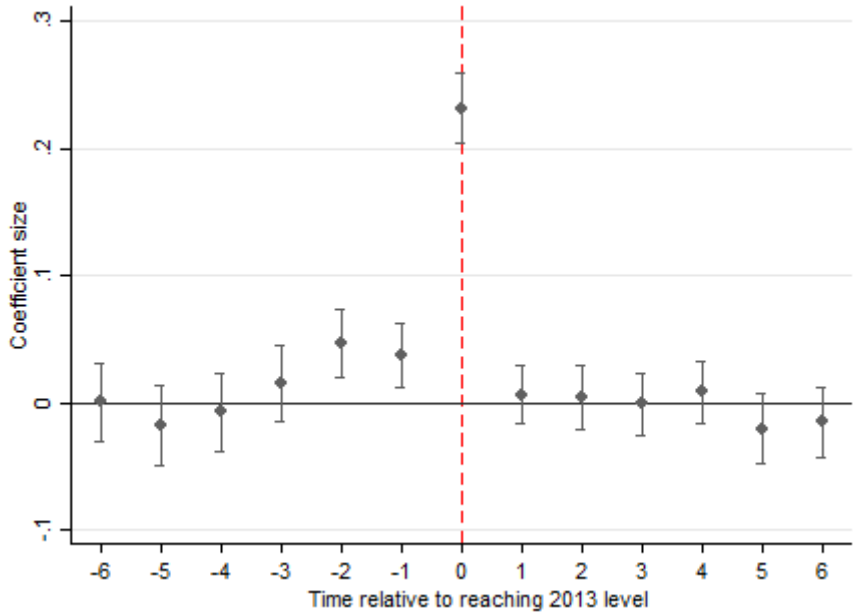
Note: The baseline sample consists of all heroin related cases from form 1 registered in Russia during 2013-2014. The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines)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

At the third threshold (500 grams for heroin), which is not presented in graphs, bunching is not observed, probably due to weak incentives and (or) insufficient number of observations. Therefore, I do not explore the police officers' responses to this threshold, and even exclude the long tail from the analysis, since it does not affect the counterfactual distribution around the relevant (second) cut-off and estimates.

As mentioned above, moving offenders from below the second threshold to above it increases the number of serious and most serious drug crimes solved, which improves the chances of police officers meeting the requirements. Since previous year's performance presents a direct target in the case of the MVD or might be

used for forecasting in the case of the FSKN, reaching this "benchmark" level could significantly affect the behavior of the police during the current year. In order to analyze the possible influence, I use an event study framework. The regression results are shown in Figure 3 and Table A5 in the Appendix.

Figure 3: The effect of reaching 2013 level on the number of serious and most serious drug crimes registered 2014



Note: The sample includes all MVD's and FSKN's stations that reached the total 2013 number of serious and most serious drug crimes during the period studied (January - December 2014). The regression results are reported in the Appendix, Table A5. Standard errors are clustered by station.

The results suggest a significant increase of the number of serious and most serious drug crimes registered in the month when the station reaches the 2013 "benchmark" and then a sudden drop in estimates. This supports the idea of the officers targeting the previous year's performance to meet the evaluation requirements.

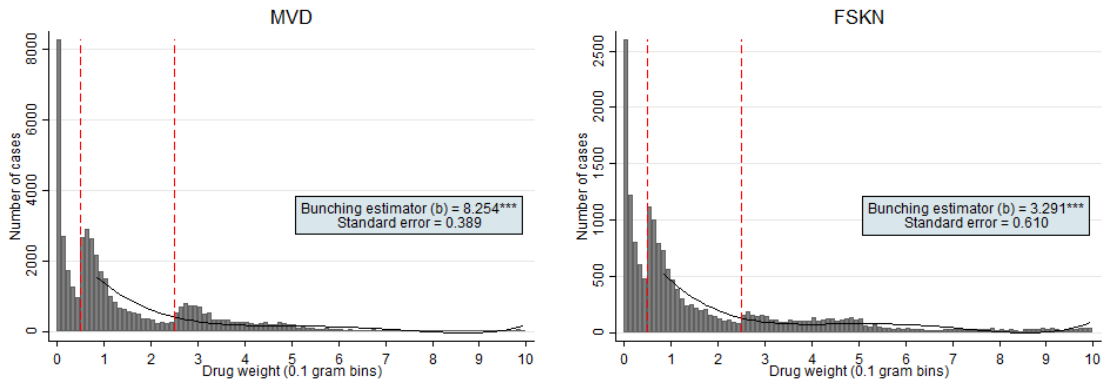
To test for a break in the level and decline of the number of serious and most serious drug crimes I apply a more parametric specification, discussed in Section 4. The estimation results presented in the Appendix, Table A6 shows that both the level drop and trend break in the number of serious and most serious drug crimes are

negative and significant (columns 1, 2). Moreover, even though the trend break in the number of cases could be a departure from an apparent upwards prior trend, the net trend after reaching the 2013 level remains negative. Adding interactions with a dummy for the MVD stations makes the negative trend insignificant (column 3). At the same time, the estimation results suggest that the drop in the number of cases is the same for both the MVD and the FSKN, while the preceding increase is smaller for the MVD. This result is unexpected but could be explained by the difference in goals of two agencies: while the main focus of the FSKN was on significant drug cases (trafficking and organized crime), the MVD mostly dealt with routine low-profile cases and was also responsible for preventing and solving other crimes that depending on the severity could also increase the overall number of serious and most serious crimes.

Turning to differences in the systems of performance evaluation, I break 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, 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 a satisfactory performance evaluation. Even though they can predict this level to some extent, the incentives to manipulate in the case of the FSKN are, therefore, weaker. In turn, the MVD offices 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, manipulating drug amounts seized in order to improve their statistics.

As the event study results suggest (Table A6 in the Appendix), the performance requirements can differently affect the behavior of officers in the two agencies during

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 amounts), the severity of crime and punishment.

a year. However, this specification shows the difference in the numbers of all serious and most serious drug crimes and does not tell us how the magnitude of manipulation (the number of cases above the threshold) varies over time. For example, the FSKN officers might manipulate the seized drug amounts more before submitting the final report, if the regional office did not reach the number of serious and most serious drug crimes of the best performing office in the previous year (assuming that these previous results are taken into account for forecasting). In contrast, the MVD officers should manipulate more if they did not reach their own previous year's level.

For the evaluation, the number of serious and most serious drug crimes is calculated per 100 officers. I do not have information on the size of each regional office, and therefore I cannot exploit the FSKN's cross-region comparison scheme[19]. Instead, assuming that the number of officers at each station is fixed during the

[19]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's and MVD's 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's or MVD's station). Finally, I calculated the bunching estimate for four groups of cases. The difference in magnitudes appeared to be insignificant.

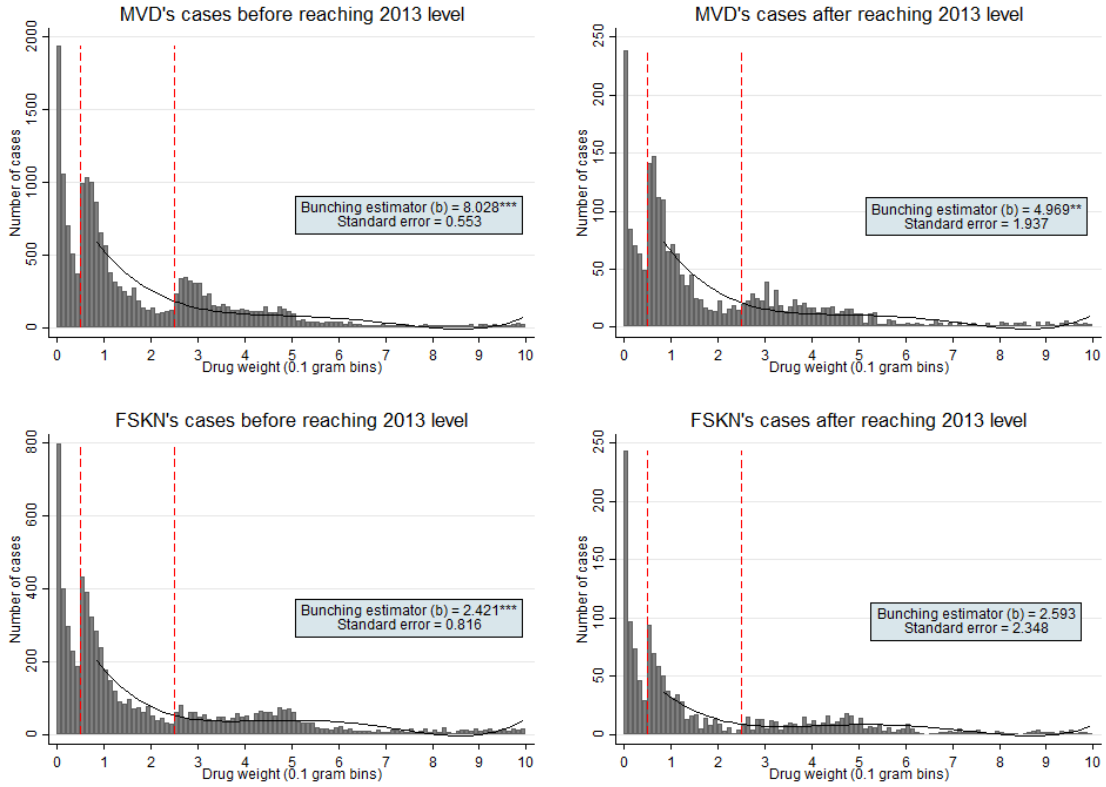
2013-2014 period, I determine the total absolute number of serious and most serious drug crimes solved by each station in 2013. Then, I divide all cases initiated in 2014 (by agency) into two groups: before and after achieving each station's 2013 level. Finally, I check whether the bunching varies between these four groups. As expected, the estimation shows that in the case of the MVD, the magnitude of 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 manipulation by the FSKN police stations does not significantly depend on reaching, or not reaching the "benchmark" (Figure 4). Thus, these results support the hypothesis that the driving force for manipulation of drug amounts is the performance evaluation system.

It is worth noting that bribery might be another motive for police officers to manipulate the drug amounts or for moving people in the opposite direction - from above to below the threshold. However, while there is no direct evidence, it seems logical to assume that if an offender decides to pay a bribe he does it to buy himself out of prison, not just to decrease the sentence. This means that those individuals are most likely not in the database at all and bribery cases are undetectable. At the same time, if bribery cases are not distributed uniformly across seized drug amounts, their omission might affect the manipulation window estimation and, hence, the bunching estimate. However, as the sensitivity analysis shows, the change of parameters does not significantly influence the result.

4.3 Mean Characteristics of Possible Victims of Manipulation

There are a number of criteria that a police officer can use to select 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 B7 in the Appendix 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 victims of manipulation, I use

Figure 5: Distributions of cases across quantities of heroin seized during 2014 by the MVD's and FSKN's 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 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 amounts), the severity of crime and punishment.

the technique described in Appendix C.2 and present results in Table 1.

First of all, I check whether such demographics as gender and nationality affect a police officer's decision to manipulate the seized drug amounts. I find a difference in the 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. However, this difference is only weakly significant. At the same time, there is no effect related to the offender being Russian.

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

Table 1: Mean characteristics of possible victims of manipulation

	Eligible for manipulation	Manipulated	Difference	s.e.
Male	0.809	0.858	-0.049^{*}	0.025
Russian	0.856	0.869	-0.013	0.023
At least college	0.386	0.400	-0.014	0.029
Unemployed	0.761	0.756	0.005	0.018
Repeat offender	0.670	0.733	-0.063^{**}	0.030
Administrative offence[20]	0.076	0.059	0.017^{*}	0.010
Under the influence of drug	0.518	0.467	0.051	0.033
Under the influence of alcohol	0.013	0.016	-0.003	0.008

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

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 amount of drug that fell into the manipulation window below the threshold. Column 2 presents the predicted mean characteristic 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 Appendix C.2.

Turning to the indicators of offender's socio-economic status, such as employment status or education level, I find that the data do not support the hypothesis of the drug amounts of low status individuals being manipulated to a greater extent. This could partly be explained by a possible corruption motive, when a police officer expects either to extract a bribe from the offender or improve his performance statistic. As the analysis shows, there is also no significant effect related to an offender being under the influence of drugs or alcohol at the moment of arrest, which, in general, should make this group of offenders a more vulnerable target. The only factor that appeared to be significant in a police officer's decision making is the offender having a previous criminal history: repeat offenders are more likely to be pushed above the threshold. This could be explained by it being easier to manipulate a person whose socio-economic characteristics are known. All in all, the results support the idea proposed in some media reports that almost any average drug offender could become the victim of manipulation.

4.4 The Effect of 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 only; the weight difference for the rest of the data varies from −3586 to 1604 grams[21]. 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 manipulation on sentence length and probability of pleading guilty

	Absolute difference ≤ 14		Full sample	
	Coefficient	s.e.	Coefficient	s.e.
	Panel A. Sentence length			
First stage	1.071***	0.009	0.366	2.461
ITT	1.089***	0.052	1.091***	0.054
LATE(sentence)	1.016***	0.217	2.978***	0.695
	Panel B. Pleading guilty			
LATE(plea)	−0.167	0.818	−0.484	0.430
LATE(sentence)^{plead}	0.942***	0.060	0.942***	0.065
LATE(sentence)$^{not\ plead}$	0.924***	0.131	−0.780	1.104

* 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 ITT effect of manipulation on the sentence length of all individuals in the manipulation region, and LATE of manipulation on the sentence length of compliers only. Panel B presents LATE of manipulation on the probability of pleading guilty and on the sentence length of those drug offenders who did and did not plea guilty. To obtain the estimates, I apply the method described in detail in Appendix C.3

The first stage effect of manipulation of heroin amounts registered by police officers in form 1 on heroin weights recorded in form 4 after the expertise is

[21] Average nonzero weight difference is −1.231 gram.

significant and shows the 1.1 grams increase of drug seized amount for individuals in the manipulation area. This implies that police officers work and manipulate in collaboration with laboratory experts. 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[22].

Turning to the possible heterogeneity of the effect of manipulation, I estimate LATE for those who pleaded guilty and who did not accept a plea bargain. 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 offences) 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. Additionally, police officers could offer the agreement more forcefully if the credibility of evidence collected is in doubt as in the case of, for example, manipulation of drug amounts. In turn, a plea bargain leads to a conviction without the actual examination of evidence at a court hearing. However, my analysis shows that crossing the threshold does not increase the probability of pleading guilty (LATE(plea) in Table 2). At the same time, the difference in the effects of manipulation on the sentence length of those who accepted the plea bargain and those who did not plead guilty (LATE(sentence)^{plead} and LATE(sentence)$^{not\ plead}$) is statistically insignificant. The reason could be the quite common practice of pinning the unsolved crime on a person who is already convicted of something, and then to push this person to plead guilty to both crimes.

[22]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 offences 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.

The total cost of drug manipulation to society is difficult to calculate precisely. However, undoubtedly, the welfare loss exceeds any benefits from keeping drug users off the streets. Each year the government spends the enormous amount of money on the Penitentiary Service, but drug addicts do not receive any treatment during incarceration. After release, most of them start taking drugs again and could be convicted for a second time. Those who decide to go back to normal life face significant difficulties, and longer incarceration exacerbates their situation, strengthening barriers to reintegration and increasing the probability to commit a "real" crime. Even more importantly, multiple manipulations widely discussed in the media lower public trust in the police increasing the level of perceived insecurity. This, in turn, decreases the effectiveness of law enforcement and the legitimacy of police actions.

5 Conclusion

The likely failure of a war on drugs has been widely acknowledged, yet in many countries anti-drug policies are still based on harsh law enforcement and even militarization. This leads to inefficient budget spending and unequal treatment of different groups of drug offenders, with a strong focus on drug users. Russia is a particularly notable example. A recently published report on drug crimes in Russia (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 with an amount of drugs just above the threshold 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 manipulation of drugs quantities seized by the police, which so far has only been alleged by various media reports.

This paper provides an empirical analysis of possible manipulation of amounts 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 the importance of incentives from performance evaluation in driving the manipulations by police officers. Additionally, the results suggest that individuals with a criminal history are more likely to have their drug amounts manipulated by the police. The overall effect of this manipulation on sentence length is an additional year of incarceration, which is not dependent on a guilty plea. Finally, I calculate the additional pressure on the overall country budget of around $13.5 million. However, this is a lower bound of the total social cost induced by manipulations, since it does not take into account that longer incarceration strengthens barriers to reintegration after release, increases the probability of recidivism and amplifies the spillover effect.

This paper clearly shows the inefficiency of the existing performance evaluation system and motivation scheme and raises a question on optimal incentive structure. Forecasting expected results itself is a common practice in many public organizations; this provides guidance for the upcoming period. However, the way it is implemented could become an issue, as in the case of drug control in Russia. The fact that police officers are punished for not achieving their targets without taking into account any reasons for this failure may trigger their dishonest behavior and manipulation of drug amounts. Thus, one step on the way to efficiency could be to decentralize the performance evaluation system and give regional offices some flexibility in setting the performance requirements.

References

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.

Banerjee, A., R. Chattopadhyay, E. Duflo, D. Keniston, and N. Singh (2012). Improving Police Performance in Rajasthan, India: Experimental Evidence on Incentives, Managerial Autonomy and Training. National Bureau of Economic Research Working Paper 17912.

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.

Brehm, M., S. A. Imberman, and M. F. Lovenheim (2017). Achievement Effects of Individual Performance Incentives in a Teacher Merit Pay Tournament. *Labour Economics 44*, 133–150.

Burgstahler, D. and I. Dichev (1997). Earnings Management to Avoid Earnings Decreases and Losses. *Journal of accounting and economics 24*, 99–126.

Camacho, A. and E. Conover (2011). Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy 3*, 41–65.

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.

Degeorge, F., J. Patel, and R. Zeckhauser (1999). Earnings Management to Exceed Thresholds. *The Journal of Business 72*, 1–33.

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.

Finan, F., B. A. Olken, and R. Pande (2017). The Personnel Economics of the Developing State. In *Handbook of Economic Field Experiments*, Volume 2, pp. 467–514. Elsevier.

Goncalves, F. and S. Mello (2017). A Few Bad Apples? Racial Bias in Policing. Princeton University Working Paper 608.

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

International Drug Policy Consortium (2018). *A Civil Society Shadow Report "Taking Stock: A Decade of Drug Policy"*. `http://fileserver.idpc.net/library/Shadow_Report_FINAL_ENGLISH.pdf`. Accessed January 30, 2019.

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. 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.

Kurmangaliyeva, M. (2017). Criminal Justice and Wealth Inequality: How Much Freedom Can Money Buy in Russia? Available at SSRN https://ssrn.com/abstract=2816363. Accessed October 1, 2017.

Le Barbanchon, T. (2016). Optimal Partial Unemployment Insurance: Evidence From Bunching in the US. Bocconi University Working Paper.

Manoli, D. and A. Weber (2016). Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions. *American Economic Journal: Economic Policy 8*, 160–182.

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.

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

Palguta, J. and F. Pertold (2017). Manipulation of Procurement Contracts:

Evidence from the Introduction of Discretionary Thresholds. *American Economic Journal: Economic Policy 9*, 293–315.

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.

Prendergast, C. (2001). Selection and Oversight in the Public Sector, With the Los Angeles Police Department as an Example. National Bureau of Economic Research Working Paper 8664.

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.

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.

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.

Sokolov, A. (2019). *Goskorporaciya "Pravosudie". Chast' Pervaya. Issledovanie o Tom, Mozhno li Dokazat' Nevinovnost' v Rossiiskom Sude.* `https://www.proekt.media/research/opravdatelny-prigovor/`. Accessed May 21, 2019.

The Federal Penitentiary Service of Russia (2017). *Characteristics of Prison Population*. http://fsin.su/structure/inspector/iao/statistika/Xar-ka%20lic%20sodergahixsya%20v%20IK/. Accessed October 1, 2017.

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. 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.

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*.

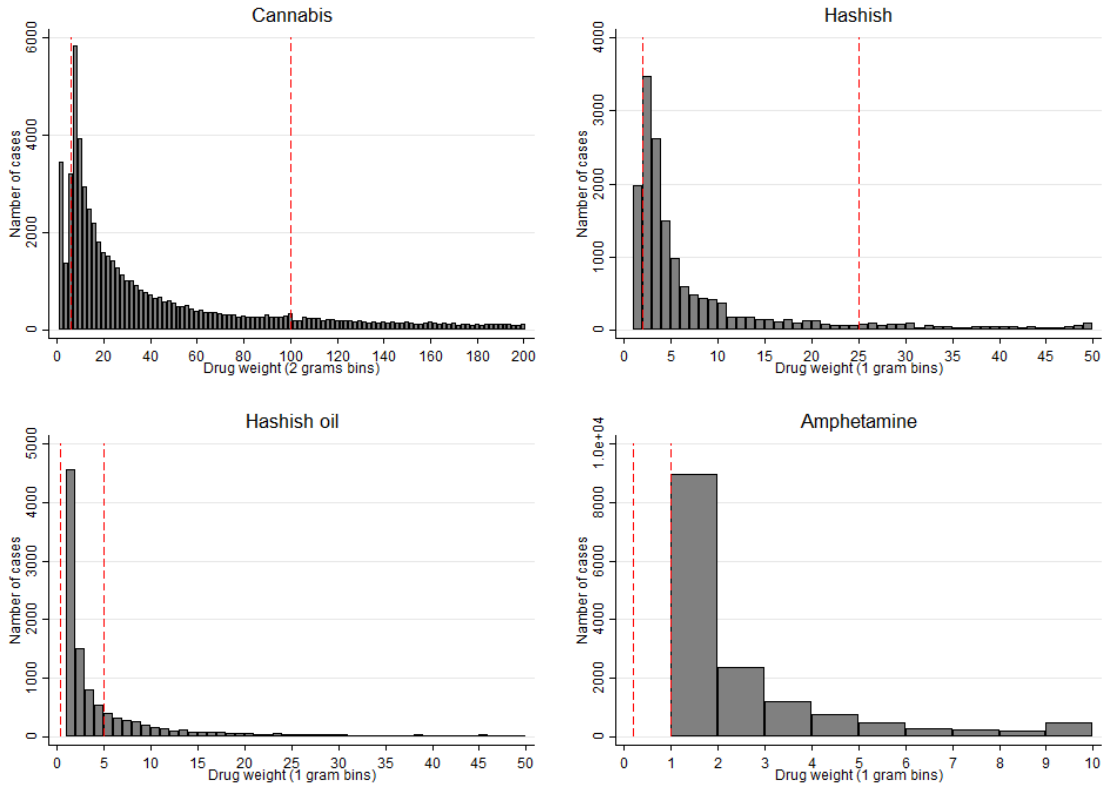
Ulmer, J. T., M. C. Kurlychek, and J. H. Kramer (2007). Prosecutorial Discretion and the Imposition of Mandatory Minimum Sentences. *Journal of Research in Crime and Delinquency 44*, 427–458.

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

Zitzewitz, E. (2012). Forensic Economics. *Journal of Economic Literature 50*, 731–769.

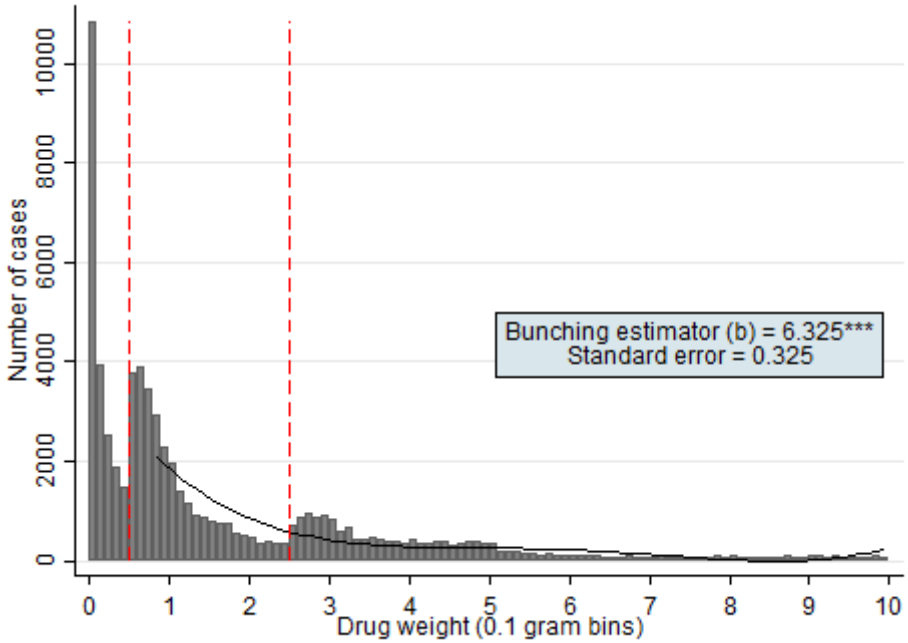
A Supplemental Figures

Figure A1: Distributions of cases across quantities of drugs seized by drug type



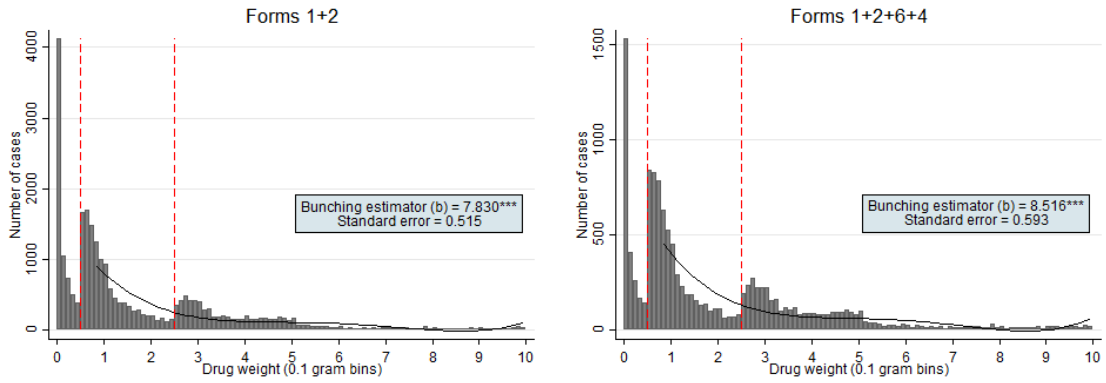
Note: The baseline sample consists of all drug related cases from form 1 registered in Russia during 2013-2014. The thresholds (dashed lines) determine the scale of seizure (less than significant, significant and large drug amounts), the severity of crime and punishment.

Figure A2: Distribution of cases across quantities of heroin seized



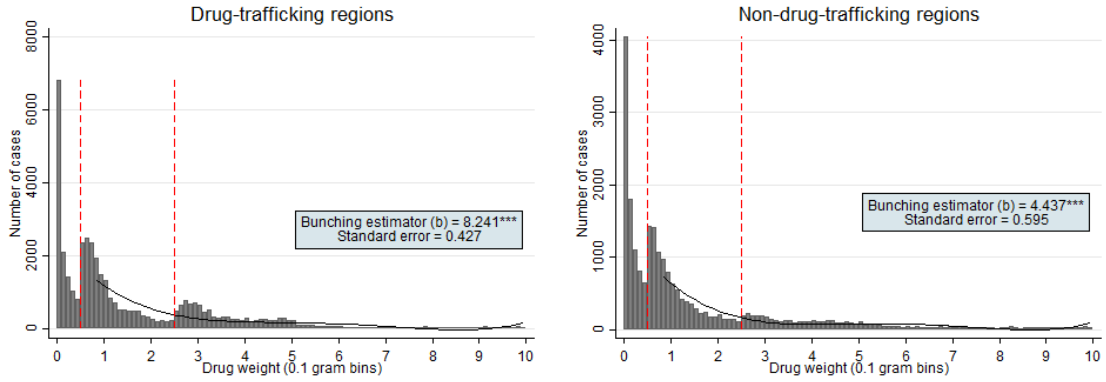
Note: The baseline sample consists of all heroin related cases from form 1 registered in Russia during 2013-2014. The series shown in bars is a histogram of the observed distribution of cases. The solid line is a fourth-degree polynomial fitted to the empirical distribution. The thresholds (dashed lines)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

Figure A3: Distributions of cases from forms 1, 2 (left) and forms 1, 2, 6, 4 (right) across quantities of heroin seized in Russia during 2013-2014



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 amounts), the severity of crime and punishment.

Figure A4: Distributions of cases across quantities of heroin seized in regions which are along (left) or away from (right) the main drug-trafficking routes



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 amounts), the severity of crime and punishment.

B Supplemental Tables

Table B1: Amounts of drugs (grams above) for purposes of articles 228, 228.1 of the Criminal Code of Russian Federation

	Significant	Large	Especially large
Cannabis	6	100	100000
Heroin	0.5	2.5	500
Amphetamine	0.2	1	200
Papaver	20	500	100000
Desomorphine	0.05	0.25	10
Hashish	2	25	10000
Cocaine	0.5	5	1500

Table B2: The severity of offence and sanctions according to articles 228, 228.1 of the Criminal Code of Russian Federation

Drug amount	Article 228 (use)		Article 228.1 (sale)	
	Severity	Sentence (years)	Severity	Sentence (years)
Less than significant	Administrative offence	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 B3: The comparison of means within the missing values analysis

	Form 1			Form 4			Forms 1+2			Forms 1+2+6+4		
	(1)	(0)	(1) - (0)	(1)	(0)	(1) - (0)	(1)	(0)	(1) - (0)	(1)	(0)	(1) - (0)
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***
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***
At least college							0.381	0.350	0.031***	0.392	0.325	0.068***
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
Repeat offender							0.672	0.584	0.088***	0.669	0.565	0.103***
Administrative offence							0.067	0.025	0.042***	0.070	0.026	0.045***
Under the influence of drug							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										3.009	2.417	0.592***
Pleaded guilty										0.346	0.604	−0.258***
Observations	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. Columns (1) present means in the subsamples without observations with missing drug weight, columns (0) present means in the subsamples of observations with missing drug weight, columns (1) - (0) shows differences in means.

Table B4: Robustness check

Starting point	Polynomial degree k	Manipulation window		Bunching estimator b	s.e.
		Lower bound r_l	Upper bound r_u		
0.7	4	1.5	3.3	7.463***	0.305
0.7	4	1.5	3.4	8.879***	0.377
0.7	4	1.5	3.5	10.828***	0.526
0.7	5	1.1	3.3	7.057***	0.415
0.7	5	1.1	3.4	8.123***	0.509
0.7	5	1.1	3.5	9.515***	0.691
0.8	**4**	**1.6**	**3.3**	**6.325*** **	**0.325**
0.8	4	1.6	3.4	7.256***	0.319
0.8	4	1.6	3.5	8.453***	0.398
0.8	5	1.2	3.3	7.885***	0.579
0.8	5	1.2	3.4	9.165***	0.766
0.8	5	1.2	3.5	10.851***	0.956
0.9	4	1.7	3.3	6.398***	0.246
0.9	4	1.7	3.4	6.787***	0.277
0.9	4	1.7	3.5	7.421***	0.309
0.9	4	1.3	3.3	7.459***	0.473
0.9	4	1.3	3.4	8.466***	0.537
0.9	4	1.3	3.5	9.751***	0.572

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

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.1.

Table B5: The effect of reaching 2013 level on the number of serious and most serious drug crimes registered in 2014

	(1)
t=-6	0.0003
	(0.015)
t=-5	−0.018
	(0.016)
t=-4	−0.008
	(0.015)
t=-3	0.015
	(0.015)
t=-2	0.047***
	(0.014)
t=-1	0.038***
	(0.013)
t=0	0.232***
	(0.014)
t=1	0.006
	(0.012)
t=2	0.004
	(0.013)
t=3	−0.0007
	(0.013)
t=4	0.008
	(0.013)
t=5	−0.020
	(0.014)
t=6	−0.015
	(0.014)
Year fixed effects	Yes
Station fixed effects	Yes
Observations	24,060
R-squared	0.649

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

Note: The sample includes all MVD's and FSKN's stations that reached the total 2013 number of serious and most serious drug crimes during the period studied (January - December 2014). The dependent variable is the logarithm of the number of serious and most serious drug crimes per month calculated based on the sample of all drug related cases from form 1. Standard errors are clustered by station.

Table B6: The effect of reaching 2013 level on the number of serious and most serious drug crimes: Trend break specification

	(1)	(2)	(3)
Event	0.209***	0.193***	0.293***
	(0.013)	(0.014)	(0.028)
Post-reaching	−0.033***	−0.034***	−0.057**
	(0.012)	(0.012)	(0.023)
Event time		0.003	−0.001
		(0.002)	(0.004)
Post-reaching x Event time		−0.007***	−0.007
		(0.003)	(0.005)
MVD x Event			−0.137***
			(0.032)
MVD x Post-reaching			0.031
			(0.027)
MVD x Event time			0.007
			(0.005)
MVD x Post-reaching x Event time			−0.002
			(0.006)
Year fixed effects	Yes	Yes	Yes
Station fixed effects	Yes	Yes	Yes
Observations	24,060	24,060	24,060
R-squared	0.648	0.648	0.649

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

Note: The sample includes all MVD's and FSKN's stations that reached the total 2013 number of serious and most serious drug crimes during the period studied (January - December 2014). The dependent variable is the logarithm of the number of serious and most serious drug crimes per month calculated based on the sample of all drug related cases from form 1. Standard errors are clustered by station.

Table B7: Summary statistics

	Overall	Manipulation region
Male	0.814	0.826
Russian	0.851	0.859
At least college	0.381	0.400
Unemployed	0.784	0.748
Student	0.0001	0.000
Worker	0.182	0.210
White-collar	0.014	0.019
Repeat offender	0.672	0.713
Administrative offence	0.067	0.060
Under the influence of drugs	0.509	0.541
Under the influence of alcohol	0.014	0.017
Observations	30,728	5,026

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 around the threshold.

C Estimation Details

C.1 Detecting Manipulation of Seized Drug Amounts

First, I estimate the magnitude of the response of police officers around the crime severity threshold. I quantify this effect, adapting 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 amounts by fitting a high-order polynomial to the observed distribution, excluding the region$[r_l, r_u]$ around the threshold $\overline{D}$:

$$C_j = \sum_{k=0}^{p} \beta_k R_j^k + \sum_{r=r_l}^{r_u} \gamma_r * \mathbb{1}[R_j = r] + \nu_j, \tag{4}$$

where C_j is the number of cases i in bin j, k 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 (4), omitting the γ_r shifters for smoothing the density around the threshold:

$$\hat{C}_j = \sum_{k=0}^{p} \hat{\beta}_k R_j^k. \tag{5}$$

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:

$$\hat{M} = \sum_{j=r_l}^{\overline{D}} (\hat{C}_j - C_j) \text{ and } \hat{B} = \sum_{j=\overline{D}}^{r_u} (C_j - \hat{C}_j). \tag{6}$$

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

$$\hat{b} = \frac{\hat{B}}{\sum_{j=\overline{D}}^{r_u} \hat{C}_j / N}, \tag{7}$$

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

Since the paper studies the rational response of the police only around the second threshold, I exclude the area around first threshold from estimation.

C.2 Identifying Victims of Manipulation

In order to recover the characteristics of those who were manipulated by the police, I adopt the technique designed by Diamond and Persson (2016).

First, I estimate the counterfactual expected values of observable characteristic Y at any drug quantity bin R inside the manipulation area, using offenders outside of this area, if there was no manipulation:

$$Y_j = \sum_{k=0}^{p} \beta_k R_j^k + \varepsilon_j, \tag{8}$$

where $R_j < \overline{D} - r_l$ or $R_j > \overline{D} + r_u$. Then I can 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 $\overline{D}$:

$$\overline{Y}^{never} = \frac{1}{N^{never}} \sum_i Y_i, \text{ where } \overline{D} - r_l \leq r_i < \overline{D}, \tag{9}$$

$$\overline{Y}^{up} = \frac{1}{N^{up}} \sum_i tY_i, \text{ where } \overline{D} \leq r_i \leq \overline{D} - r_u. \tag{10}$$

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

$$\overline{Y}^{never} = \frac{N^{down}}{N^{down} - N^{compliers}}\overline{Y}^{down} - \frac{N^{compliers}}{N^{down} - N^{compliers}}\overline{Y}^{compliers}. \tag{11}$$

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

$$\overline{Y}^{up} = \frac{N^{always}}{N^{always} - N^{compliers}}\overline{Y}^{always} - \frac{N^{compliers}}{N^{always} - N^{compliers}}\overline{Y}^{compliers}. \tag{12}$$

Using estimates from (5) and (8), I can obtain values of $\overline{Y}^{down}$ and $\overline{Y}^{always}$ in the following way:

$$\overline{Y}^{down} = \frac{\int_{\overline{D}-r_l}^{\overline{D}-\sigma} \hat{Y}_j{}^R \hat{C}_j{}^R dR}{N^{down}} \tag{13}$$

$$\overline{Y}^{always} = \frac{\int_{\overline{D}}^{\overline{D}+r_u} \hat{Y}_j{}^R \hat{C}_j{}^R dR}{N^{always}}. \tag{14}$$

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

$$N^{never} = N^{down} - N^{compliers}, \text{ where } N^{down} = \int_{\overline{D}-r_l}^{\overline{D}-\sigma} \hat{C}_j{}^R dR, \tag{15}$$

$$N^{up} = N^{always} + N^{compliers}, \text{ where } N^{always} = \int_{\overline{D}}^{\overline{D}+r_u} \hat{C}_j{}^R dR. \tag{16}$$

Plugging these into (11) and (12) and using estimates from (9), (10), (11) and (12), I solve for the compliers' average value of characteristic Y:

$$\begin{aligned}\overline{Y}^{compliers} = 0.5(&\frac{N^{never}}{N^{never} - N^{down}}\overline{Y}^{never} - \frac{N^{down}}{N^{never} - N^{down}}\overline{Y}^{down})+ \\ &+ 0.5(\frac{N^{up}}{N^{up} - N^{always}}\overline{Y}^{up} - \frac{N^{always}}{N^{up} - N^{always}}\overline{Y^{always}}).\end{aligned} \tag{17}$$

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

$$\Delta Y = \overline{Y}^{never} - \overline{Y}^{compliers}. \tag{18}$$

C.3 Estimating the Effect of Manipulation on Sentence Length

I identify the effect of manipulation of drug quantities on sentence length (and on the probability of pleading guilty) in two steps, following Diamond and Persson (2016).

First, I estimate the relationship between sentence length S and the amount of drug seized from form 1:

$$S_j = \sum_{k=0}^{p} \beta_k R_j^k + \gamma_R * \mathbb{1}[R_j \geq \overline{D}] + \omega_j, \tag{19}$$

where $R_j < \overline{D} - r_l$ or $R_j > \overline{D} + r_u$. Equation (19) gives the expected length of sentence at each drug amount inside the manipulation region in the counterfactual world where no offender is manipulated.

Then, I calculate the counterfactual expected sentence length across the whole set of drug offenders inside the manipulation region:

$$\overline{S} = \int_{\overline{D}-r_l}^{\overline{D}+r_u} \hat{S}_j \frac{\hat{C}_j^{\,R}}{\int_{\overline{D}-r_l}^{\overline{D}+r_u} \hat{C}_j^{\,R}} dR. \tag{20}$$

Comparing observed and estimated counterfactual average sentence lengths, I obtain the "intent-to-treat" effect, which shows a change in the length of imprisonment due to the offender having been caught with the actual amount of drug that falls within the manipulation region:

$$ITT = \frac{\sum_{i \in manip\, region} S_i}{N^{manip}} - \overline{S}, \tag{21}$$

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

The procedure described above can be repeated with drug quantities from form 4 instead of sentence length. This constitutes the effect of being manipulated on the amount of drug seized that is determined officially at the laboratory and then considered by judge at court. The ratio of ITT from equation (21) to this effect, in turn, identifies the local average treatment effect (LATE) of being manipulated on the length of imprisonment.