

Evaluating Performance of Government Inspection Bodies: A Possible Approach

Elena Dobrolyubova

Abstract

Ensuring the public safety and limiting administrative barriers have been the two core objectives of most reforms in the area of business inspections and regulatory enforcement for the past decade. However, measuring these efforts has proved quite challenging both in OECD countries and in Russia. Lack of attention to the results achieved or misinterpretation of progress may affect the success of the reform efforts.

The objective of this paper is to develop a framework for defining and evaluating both the effectiveness and the efficiency of regulatory enforcement and to apply this framework to several areas of inspection activities (such as occupational and food safety) in Russia. The proposed framework advances the existing international approaches to measuring the performance of government inspection bodies and accounts for perspectives of citizens, businesses and governments as crucial stakeholders of inspection reforms. The paper is based on both qualitative and quantitative methods (sociological surveys and statistical research).

The results suggest that evaluating the performance of business inspections should include all aspects of minimizing risks and losses in the controlled area – from prevention of violations (reducing risks) to ensuring the reimbursement, should the risk event happen. Various levels of results, including final outcomes (impact), intermediate outcomes and outputs, are needed as they allow for detecting important inconsistencies and gaps in the performance of government inspection bodies. The use of various sources of performance data, independent from inspection bodies, is critical for the successful implementation of the proposed framework.

Keywords:

Administrative reform, business inspections, effectiveness, efficiency, inspection bodies, performance evaluation, regulatory performance

1. Introduction

Reducing administrative barriers and fostering entrepreneurship have become important issues on the agendas of both developed and developing economies. While significant efforts are devoted to improving business regulations and introducing new forms of communication between governments and businesses, reducing the pressure of business inspections has also become a key component of recent administrative reforms (Blanc 2012).

On the other hand, the growing concerns about the risks which the business operations may pose to the public safety call for more effective inspection services to be put in place (OECD 2013). Moreover, a mere increase in inspection numbers and scope cannot reduce the probability of risk events; rather, the targeting, methods and scope of inspections are to be changed to provide for achieving another important objective for business inspections reforms – reducing risks and protecting the interests of the public (World Bank 2011).

Lastly, since many countries face severe budget constraints, there is a need for improving the cost-effectiveness of the public administration in all areas, including government inspection. Since inspection staff often represents a significant part of the total civil service, improving the performance of this part of the public administration is another objective pursued under the inspection reforms (Blanc 2012).

In 2014, OECD experts summarized the experience of the inspection reforms and formulated the best-practice principles for regulatory enforcement and inspections (OECD 2014). The implementation of these principles calls for developing a system of measuring both the effectiveness and the efficiency of enforcement and inspection activities. However, unlike in some other areas relevant to the interaction of businesses with governments (i.e. opening a new business, registering property, obtaining construction permits, etc.) (World Bank 2016; WEF 2016; etc.), there is no readily available cross-country data on measuring the effectiveness and efficiency of business inspections. Except in specific areas, such as revenue administrations, there is no uniform internationally recognized framework for measuring the efficiency and effectiveness of business inspections; while country examples are quite diverse, both in terms of the types of indicators used and in terms of the way the evaluation procedure is organized (Monk 2012). Even though inspections and enforcement are seen as an important component of the regulatory cycle (Coglianese 2012), to date the international cross-country comparisons have been mostly focusing on other aspects of regulatory policy development and implementation, such as regulatory impact analysis and stakeholder consultations (Arndt et al. 2016).

In Russia, the reform of regulatory enforcement was approved in 2016 as one of the 11 priority national projects¹; however, the performance framework for

1 Approved by the Council on Strategic Development and Priority Projects under the President of the Russian Federation in 2016.

the reform is still being developed. This paper is intended to contribute to the ongoing debate on measuring the administrative performance of the government inspection bodies.

The objectives of the paper are (i) to develop a framework for defining and evaluating both the effectiveness and the efficiency of regulatory enforcement and inspections and (ii) to apply this framework to several areas of inspection activities in Russia.

2. Measuring public administration performance: Key approaches, issues and the case of inspection and enforcement agencies

Performance management, involving measuring outcomes and outputs of public administration activities and using monitoring and evaluation results for decision-making purposes, has been among the top issues on the administrative reform agenda for the past few decades (Bouckaert and Halligan 2008). Though the extent of implementing performance management varies among countries with different legal and administrative traditions (Hammerschmid and Löffler 2015), the interest in implementing performance management is high, with the recent economic crisis being an additional incentive for introducing and further developing performance-management practices (Grossi et al. 2016).

Despite the broad application of performance-management approaches, so far there is no consistent terminology used in this area; even such a concept as “performance measurement” is still used differently by various authors (Alach 2017). Traditionally the performance-measurement concept was based on using qualitative performance indicators for measuring organizational activities and achievements (Choong 2013). Recently some scholars have pointed to the need of including some qualitative social metrics to avoid political influence on performance evaluation (Salais 2010) and account for the complexity of the society (Virtanen and Vakkuri 2015). Other studies pointed out that the high social impact in the public sector should be seen as an incentive for rather than an obstacle to using performance information (Moynihan et al. 2012). At the organizational level, another important aspect of performance measurement is regularity – performance information should be collected regularly and timely so that it could be used for informing management (including budgeting) decisions (Ter Bogt et al. 2015). This aspect limits the range of evaluation methods which can be applied to measuring organizational performance.

There are many authors who contributed to general evaluation theory (Christie and Alkin 2008); however, the performance-measurement systems implemented in the public sector to date are mostly based on one of the following three approaches: (i) the logic model (or logical framework model); (ii) the balanced scorecard; or

(iii) Moore's strategic value triangle. The logic model involving three main components (inputs, outputs, and outcomes), where each of the components contributes to the achievement of the next component (i.e. inputs ensuring the achievement of outputs, outputs resulting in the achievement of outcomes), originated in the US (Williams 2014) and has been extensively used both by governments and international organizations worldwide. Another approach is based on the balanced scorecard, which combines a set of measures for evaluating a set of perspectives (i.e. resources, client orientation, internal processes, innovations) of organizational development (Kaplan and Norton 1992). Finally, the strategic-value triangle proposed by Moore is based on measuring performance based on the public value, legitimacy and support, as well as the operational capacity to produce the public value (Moore 2000) and emphasizes the importance of stakeholder engagement in evaluation.

Performance evaluation in the public administration usually involves the comparison of actual performance with the performance targets (Alach 2017), while benchmarking approaches are also used (Magd and Curry 2003). Other methods, such as cost-benefit analysis and social return on investment, are less common for regular evaluation of organizational performance and are used mostly for project and program evaluations (Cordes 2017).

Both theoretical and empirical studies demonstrate mixed results in terms of linking performance-measurement (and performance-management) efforts with actual performance outcomes. In some cases, such reforms seem to have brought positive results (Dan and Pollitt 2015), in others the impact is not evident (Gao 2015) or ambiguous (Greiling 2006; Christensen and Laegreid 2015).

The reasons for such differences are often related to the quality of performance-measurement systems used in the public sector and their potential to inform rather than misinform managerial decisions (Kravchuk and Schack 1996), the quality of performance information used (Schwarz and Mayne 2005), as well as other, i.e. institutional factors (Reichborn-Kjennerud and Vabo 2017; Barbato and Turri 2017). The recent research by Speklé and Verbeeten suggests that the effectiveness of the performance-measurement systems depends both on the quality of the system itself and on the way the system is used by the managers (Speklé and Verbeeten 2014). Given the broad application of performance-management practices and the significant role they play in governance accountability, there is a need for further developing the existing approaches to constructing and implementing performance-measurement systems used in the public sector (Lewis 2015).

One of the possible approaches to that is linking performance-measurement systems to the types of government functions. The prevailing approaches described above were initially intended for measuring public-service delivery rather than performing other types of functions, such as regulation, inspection or enforcement. For instance, the logic model assumes that an increase in inputs should result in improved outputs and, ultimately, outcomes, which is not the case for government

inspections, where the increase in the number of inspection visits does not automatically lead to a higher extent of safety but results in an additional administrative burden on businesses (Hampton 2005). The application of the balanced scorecard methodology requires a definition of the client (Northcott and Ma'amora Taulapapa 2012), which is controversial in the case of enforcement agencies having to deal with a broad range of stakeholders. The ultimate public value produced by the government inspection bodies is public safety. However, this public value is influenced by other factors, which are not controlled by these bodies (OECD 2014); hence, the possibility of the direct application of the strategic-value triangle is also limited.

So far, the specific approaches to measuring performance in inspection and enforcement bodies received insufficient attention, with the important exception of the framework proposed for measuring the performance of the best value inspectorate in the UK local government (Boyne et al. 2002). Some possible approaches for measuring inspection reforms were developed by the World Bank (World Bank 2006) and OECD (Blanc 2012), as described in sub-section 3 of this article. However, to date, there is no uniform approach to measuring and benchmarking the performance of government-inspection bodies (Charlebois and Hielm 2014), while the empirical research and reviews are still mostly limited to selected country or sector examples (Doroshenko 2012; Ma et al. 2015; Dobrolyubova 2016; Safe Work Australia 2013).

3. Methodology

To address the existing gap, the article first reviews the existing empirical approaches to measuring the performance of government inspection bodies in OECD countries. The analysis of the cases allows for identifying the objectives of such type of bodies and systemizing the indicators used for measuring performance and the data sources.

Based on the results of this empirical analysis, the paper then adapts the logic-model approach to the case of government inspection bodies through identifying (i) the levels of performance measurement needed; (ii) performance dimensions which should be measured at each of the levels; and (iii) the algorithm of evaluation which aims at addressing the limitations of the direct application of the existing theoretical approaches to the case of government-inspection bodies.

The proposed framework is then applied to selected areas of regulatory enforcement and business inspections in Russia with both statistical analysis and a sociological survey data being used. The results illustrate both possibilities and some limitations of the proposed methodological approach.

4. Measuring the performance of government inspection bodies: International practice

The first attempts to systemize the approaches to measuring the efficiency and effectiveness of business inspections were undertaken by the World Bank in 2006, when a framework of performance indicators for measuring inspection reforms was developed. This framework aimed at measuring the achievement of 4 key inspection-reform objectives: improving the compliance with transparent regulatory requirements; reducing the uncertainty for businesses; preventing corruption; and minimizing administrative costs for businesses and optimizing public authorities (World Bank 2006). The proposed indicators were based both on business-perception surveys and on the data collected by inspection bodies; however, some of the indicators included into the framework were quite formal (i.e. the percentage of the inspections conducted out of the total number of planned inspections).

More recent research proposes to measure the progress of inspection reforms based on 5 main dimensions: (i) overall administrative burden of business inspections; (ii) targeting business inspections; (iii) using the new tools to conduct inspections; (iv) information-sharing and supporting compliance; and (v) resource allocation (Blanc 2012). The data used for measuring the performance of government-inspection bodies is usually either collected by the inspection body itself or is based on the official statistics; there are also examples of using sociological surveys (Blanc 2009) and applying a standard cost model for measuring administrative costs of government inspections. The formats for performance evaluation vary significantly: from annual performance reports (Finland, Switzerland), evaluation of performance dynamics (Sweden, Estonia) to organizational and functional reviews (Chile), annual audit (Norway), ISO performance audit (Belgium) and sociological surveys with outcome evaluation at the national level (New Zealand) (Monk 2012). Examples of the performance indicators used to measure the effectiveness and (in some cases) also efficiency of government inspections are presented in Table 1.

Table 1
Indicators used for measuring regulatory enforcement and inspections

Dimension	Possible performance indicators	Data source
1. Administrative burden on businesses	<ul style="list-style-type: none"> • share of businesses covered by inspections; • inspection frequency and duration; • total time for 1 inspection; • direct administrative costs 	Sociological surveys (Netherlands, UK, Italy, Denmark, Lithuania, Belarus, Kyrgyz Republic, Mongolia) and data provided by inspection bodies
2. Targeting inspections	<ul style="list-style-type: none"> • allocation of inspections among the groups of businesses with different risk profiles 	Data provided by inspection bodies (UK)

Dimension	Possible performance indicators	Data source
3. Using <i>new tools</i> for conducting inspections	<ul style="list-style-type: none"> percentage of inspections using new tools (check-lists, inspector identification, advance notices, etc.), out of total number of inspections 	Sociological surveys (Italy, Lithuania)
4. Information and compliance support	<ul style="list-style-type: none"> percentage of businesses participating in information activities; business satisfaction with information provided by inspection bodies; perceptions of businesses in terms of clarity and transparency of regulations 	Sociological surveys (Italy, Lithuania, UK)
5. Resource allocation	Allocation of financial and human resources on different tasks: <ul style="list-style-type: none"> information (consultation); analysis and planning; inspection and enforcement 	Data provided by inspection bodies (i.e. UK). Often the budget classification does not allow retrieving the data easily.

Source: compiled by the author based on the dimensions proposed in Blanc (2012)

The performance indicators used by the UK Health and Safety Executive (HSE 2016) illustrate the application of this framework. While the inspection tracks the progress in terms of outcomes in the sector (i.e. injuries, working-days loss; professional-diseases rates, economic effects), the indicators used for measuring the inspection's performance include either intermediate outcomes or output-level measures, such as:

- percentage of regulations which were removed or improved (this indicator illustrates reducing the administrative-burden dimension in the above table);
- percentage of top-category major hazard installations visited (targeting-inspection dimension);
- number of visits to HSE's website and user satisfaction rate (information and compliance-support dimension).

However, in addition to the above performance indicators, HSE also uses some other indicators which do not correspond to the dimensions presented in Table 1. For example, the indicator "percentage of dutyholders who had taken action as a result of an inspector's visit" helps to measure the influence of inspection visits on actions taken by businesses to improve occupational safety. Other performance indicators such as "timeliness in investigating fatal accidents, non-fatal accidents" and "percentage of convictions as a result of legal proceedings initiated" demonstrate how the inspectorate's activity influences the compensation of the damage incurred as a result of breaching occupational safety regulations. Overall, the system of HSE performance indicators helps measuring the achieve-

ments against the overarching organizational goal, which is “to make Britain one of the safest places to work, saving thousands of lives, preventing many more injuries at work, improving people’s health and reducing the economic and social costs of health and safety failures” (HSE 2016).

Another interesting case is related to the US Food Safety and Inspection Service (FSIS). In accordance with its Strategic Plan for 2017–2021, FSIS has three goals: (i) preventing foodborne illness and protecting public health; (ii) modernizing inspection systems, policies and the use of scientific approaches; and (iii) achieving operational excellence (FSIS 2016).

The high-level goals are measured based on outcome-level indicators, such as microbiological contamination rates or numbers of illnesses (i.e. Salmonella, Campylobacter, etc.) from FSIS-regulated products. The objective of modernizing inspection procedures is measured based on the following indicators:

Intermediate outcome:

- % of establishments whose non-compliance rate decreases 120 days after receiving an Early Warning Alert;

Compliance rates

- % of slaughter establishments that are compliant with all livestock restraint and/or stunning requirements;

Information accessibility

- % of analysts able to access, analyze and visualize FSIS data;
- % of employees with online access to FSIS-approved systems;
- number of establishment-specific and other FSIS datasets made publicly available.

The cases presented above demonstrate that the performance indicators used for inspection bodies aim at balancing different functions performed with a special emphasis on information accessibility and promoting voluntary compliance. Neither inspection uses the data on the number of visits or number of sanctions imposed as indicators of performance.

Overall, the analysis of the above cases suggests that both inspectorates analyzed (HSE and FSIS) aim at using at least some performance measures capturing the influence of their activities (both in terms of information and in terms of inspection) on improved safety in the regulated area. This intermediate measure helps to link the outputs produced (i.e. information provided, alerts issued) to the overarching goals of maintaining public safety and minimizing possible risks in the selected area. Thus, the spectrum of performance dimensions for government-inspection bodies proposed by Blanc and presented in Table 1 should be broadened.

5. The proposed framework for evaluating the performance of government-inspection bodies

The proposed framework adapts the logic-model approach discussed in section 1 to measuring and evaluating the effectiveness and efficiency of government-inspection bodies. To formulate the outcomes of the government-inspection bodies' performance, we use the concept of the risk-management cycle, which includes the stages of (i) risk prevention, (ii) minimizing the adverse effects (or damage) and (iii) compensating the damage. The novelty of the proposed approach is related to including the compensation stage into measuring the performance of government-inspection bodies. This stage helps to balance the interests of the businesses and the public (or the third parties), while traditionally more attention is paid to the earlier stages. Also, accounting for compensation (or reimbursement) sets a proper context for evaluating the effectiveness of government inspections as opposed to possible alternative control mechanisms, which could be used to minimize risks in a given area.

Thus, the regulatory-enforcement *effectiveness* reflects the achieved level of protection of public values from risks characterized by a decrease in the frequency and scope of accidents and damage, as well as the extent to which such damage has been reimbursed (compensated). The *efficiency* of regulatory enforcement and inspections means that the high effectiveness is being achieved with simultaneous minimization of inspection-related costs incurred by the state. Hence, in case the regulatory-enforcement effectiveness is low (i.e. the damage is growing), the efficiency of the business inspections cannot be rated as high. This is an important modification of the standard logic-model approach, which assumes that increased inputs correspond to better outcomes and outputs. In the case of enforcement and inspection, given the overall objective of protecting the public values from risks, there is a need to limit the possibility of increasing efficiency at the expense of decreasing effectiveness and, hence, higher risks to the public.

While the OECD Best Practice Principles call for measuring regulatory-enforcement effectiveness based on outcomes (OECD 2014), it is broadly recognized that there are other factors influencing outcomes (such as mortality or disability rates, damage to assets, pollution of environment etc.) rather than the activities of enforcement and inspection bodies. The damage may occur both in cases when regulatory norms were violated and in cases when no breach of current norms took place. Hence, there is a need for modifying the standard logic model and measuring *intermediate outcomes*, reflecting the extent to which the inspection bodies influence the "*final*" outcomes. Such intermediate outcomes are supported by outputs dependable on the performance of inspection bodies, which include (i) the extent of awareness about the regulatory requirements among the businesses; (ii) compliance rates; (iii) the extent of severe violations and how they were addressed; (iv) compensation of damage; (v) respecting the law during the control activities; and

(vi) reducing business administrative costs (Figure 1). A specific list of indicators is to be developed for each major inspection area (i.e. food safety, occupational safety, environmental safety, etc.).

The proposed evaluation framework allows measuring the performance of government-inspection bodies from the perspective of all key stakeholders: citizens (interested in public safety and compensation of damage), businesses (interested in the reduction of administrative costs and compliance with the legal requirements on the part of the inspection bodies) and government at large (interested in the efficient use of resources on inspections and enforcement).

Figure 1
Effectiveness and efficiency of regulatory enforcement and inspections:
Evaluation framework

Final Outcomes	<ul style="list-style-type: none">- Reducing the level of risk for the public values (life, health, property, environment, etc.)- Minimizing the instances of damage to the public values- Compensating the damage incurred		
Intermediate Outcomes	<ul style="list-style-type: none">- Reducing the level of risk by ensuring that mandatory legal requirements are followed		
Outputs	Information (awareness)	Compliance rate	Addressing severe violations
	Compensating the damage	Respecting the law during inspections	Reducing business-administrative costs
Inputs	Budget expenditures on enforcement & inspection, human resources		

Source: author

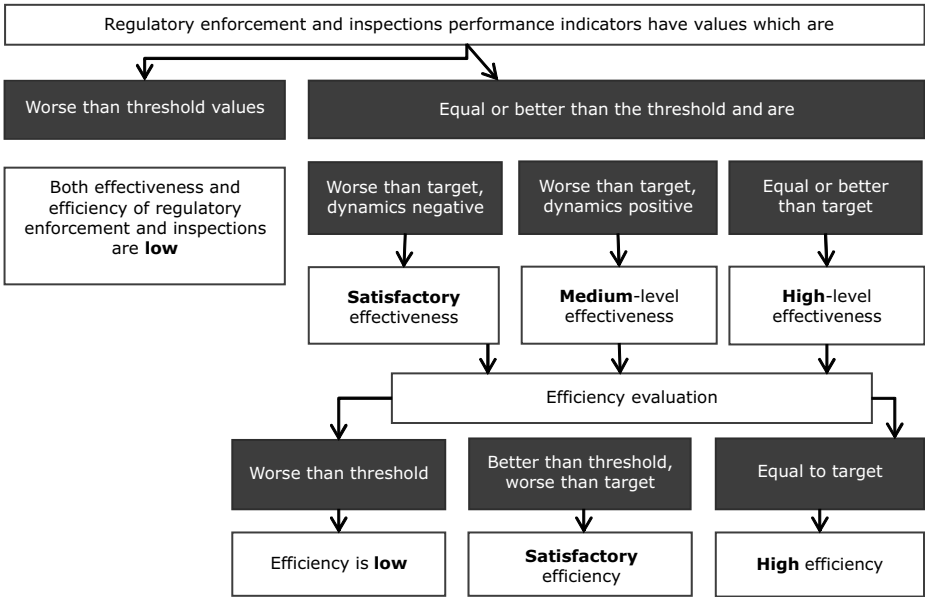
Evaluating the effectiveness of regulatory enforcement and inspection calls for choosing the right benchmark and to balance the evaluation. To this end, three types of approaches are proposed: (i) target-based evaluation, based on comparing actual and target values² of performance indicators; (ii) dynamics-based evaluation, based on comparing the actual value of the performance indicator with its value in the preceding year; and (iii) threshold-based evaluation, based on comparing the actual value to the threshold value representing the minimally acceptable level of risk (it can be established or estimated based on the average performance for the past few years). Notably, these approaches complement each other; this combina-

2 Target values may be set in strategic documents and (or) estimated based on international comparisons.

tion of metrics allows mapping the actual performance to the acceptable interval and to factor in the dynamics in this area.

The general evaluation algorithm which helps to interpret the evaluation results is presented in Figure 2. As discussed above, in case the effectiveness is deemed low (i.e. the values of indicators characterizing final and intermediate outcomes and outputs are below the threshold), the efficiency of regulatory enforcement and inspections is also automatically low. In case the indicator value is equal or better³ than the established threshold, the effectiveness may be rated satisfactory, medium or high, depending on the indicator dynamics and the target value.

Figure 2
Algorithm for evaluating effectiveness and efficiency of regulatory enforcement and inspections



Source: author

Since both the efficiency and the effectiveness of regulatory enforcement and inspection activity cannot be evaluated based on only one indicator (for the aspects measured, refer to Figure 1), a system of weights should be applied. The weights are assigned based on expert assessments (i.e. using the Delphi method and the like) or,

³ The terms “better” and “worse” are used because in regulatory enforcement and inspections often a lower value of an indicator (i.e. number of accidents or mortality rates) represents a better situation than a higher value. The substance of the indicator is to be accounted for in the calculations.

in some cases, based on historical data of incidents and associated costs per type of damage incurred.

6. Evaluating performance of regulatory enforcement and inspections in selected sectors in Russia

6.1 Evaluating the performance of business inspections of occupational safety

Based on the general framework presented in Section 4, a system of performance indicators measuring final outcomes, intermediate outcomes and outputs of occupational safety inspections was developed (Table 2). The indicators used for evaluation come from various sources, including (i) the official statistics on occupational injuries (outcome-level indicators), (ii) the administrative data collected by the regulatory enforcement and inspection body (Rostrud) and (iii) the data collected through a business survey⁴ (intermediate outcome and output level indicators).

Since for most indicators used for the evaluation no target values have been established, dynamics-based evaluation and threshold-based evaluation was carried out. As the latest available statistical data is for 2015, the administrative data for the same period was used. For statistical and administrative data, the thresholds are estimated as average values for the past 5 years. For survey data, the threshold values are deemed equal to the average data for all types of inspections based on survey outcomes. Since no comparable enterprise survey was carried out in 2013–2015, the dynamics-based evaluation is deemed equal to 100 %. The weights for the indicators were assigned based on expert assessment.

In accordance with the law, the compensations to the injured and to their families are paid from the Federal Social Security Fund. There is no reliable data on whether such compensation of damage is sufficient, so for the purposes of further research it is advisable to conduct regular polls among the persons who suffered occupational injuries.

The results of the evaluation are presented in Table 2. As demonstrated by the calculations, both dynamics-based evaluation and threshold-based evaluation exceed 100 %, which means that the effectiveness of occupational safety inspection is at least medium (in accordance with the algorithm presented in Section 4); there is significant improvement in terms of the reduction in injury frequency and the negative economic impact of occupational injuries.

4 The business survey was conducted by RANEPa in March 2017 in 28 Russian regions. The sample included 1,000 enterprises and self-entrepreneurs, mostly small businesses, representing all major sectors of economy. For survey methodology and results see Dobrolyubova et al. (2017).

Table 2
Effectiveness of business inspections of occupational safety in Russia

Indicator	Weight	2014	2015	Threshold value	Dynamics-based evaluation	Threshold-based evaluation
Final outcomes	0.5				107.7	135.8
Occupational injuries, per 1000 employed	0.3	1.4	1.3	1.9	107.7	143.1
Fatal occupational injuries, per 1000 employed	0.4	0.067	0.062	0.082	108.1	132.6
Work days lost due to disability, mln.	0.3	1.5	1.4	1.9	107.1	132.9
Intermediate outcomes	0.3				108.7	109.0
Number of identified accidents	0.4	9666	8643	10343	111.8	119.7
Ratio of prevented and non-prevented damage (ratio in the number of legal entities with violations which pose a threat to occupational safety to the number of legal entities with violations which were the cause of accidents)	0.4	1.43	1.57	1.55	109.9	101.1
Percentage of businesses which claim that occupational safety has improved due to inspection visit	0.2	–	37.3	36.1	100.0	103.3
Outputs	0.2				106.0	109.8
Percentage of businesses which evaluate their awareness about regulatory requirements in occupational safety as high and very high	0.1	–	57.9	60.3	100.0	96.0
Average administrative costs for business per one inspection (including informal payments)	0.1	–	145.6	105.3	100.0	72.3
Compliance rate (percentage of businesses for which no major violations were detected based on inspection)	0.3	87.82	82.73	86.45	94.2	95.7

Indicator	Weight	2014	2015	Threshold value	Dynamics-based evaluation	Threshold-based evaluation
Percentage of businesses which claim they have eliminated all violations (out of the total number of enterprises with violations)	0.2	–	67.3	75.7	88.9	88.9
Percentage of businesses which agree with the outcome of inspection visit	0.1	–	85.3	87.3	100.0	97.7
Percentage of inspections which were cancelled (by the management, prosecutor office or court)	0.1	0.02	0.01	0.03	200.0	266.7
Percentage of cancelled inspections which resulted in administrative measures on the inspectors for violations	0.1	100	100	100	100.0	100.0
Evaluation of Effectiveness					107.7	122.6

Sources: calculated by the author based on Rosstat data (www.gks.ru), Rostrud data (www.rostrud.ru), and RANEPa survey (Dobrolyubova et al. 2017).

Intermediate outcomes were calculated based on enterprise survey data (percentage of businesses claiming that the occupational safety has improved as a result of the inspection) and on the data collected and published by the inspectorate (number of accidents and the ratio of prevented and non-prevented damage). The intermediate outcomes also demonstrate largely positive trends. The inspections tend to be proactive, as the number of severe violations that pose a threat to occupational safety exceeds the number of accidents, while the total number of accidents is decreasing. Some 37.3 percent of businesses which were examined by labor inspection claim that the inspection visit resulted in improved occupational safety (on average, the improvements of safety of production is noted by 36.1 percent of respondents).

To evaluate outputs of the inspections, 7 performance indicators were used, based on the statistics published by the inspection body and on the enterprise survey data (Table 2). While overall assessment of outputs is positive, there are some negative trends to be noted. Firstly, occupational safety inspections result in high administrative costs. The average administrative cost born by businesses (which in-

cludes material costs related to the inspection, labor costs and informal payments) is estimated to account for 145.6 thousand RUR (compared to 105.3 thousand RUR for all types of business inspections). Secondly, the compliance rate (the proportion of businesses without major violations) is decreasing. This trend may be partially explained by better targeting of inspections (when riskier businesses are inspected more often, violations are also detected more often). At the same time, the survey outcomes demonstrate that only 67.3 percent of businesses where violations were found during the inspection fully eliminated these violations; some 85.3 percent of businesses agreed with the inspection outcomes.

The example of occupational safety demonstrates that while final outcomes and intermediate outcomes have been improving, the progress on outputs is not so evident. Hence, there is a gap between outputs of inspection activities and the societal impacts in this area. This trend, if sustained for an extended period, is likely to result in the deterioration of outcome-level results. This example also shows that government inspections are one but not the single factor influencing risks in occupational safety.

The positive trends in occupational safety should also be compared to the existing international benchmarks. According to Eurostat data, the average occupational injury rate accounts for 14 cases per 1000 employed. This indicator exceeds the data for Russia about 10 times. However, the difference in indicators can be at least partially attributed to specifics of data collection. For example, in the UK occupational injuries are registered both by businesses (and this data shows 3 cases per 1000 employed) and based on labor-market data which shows 2.5 times higher results.⁵ The impact of the data collection methods is somewhat supported by the fact that the lowest frequency in occupational safety is noted in Eastern European countries which do not invest much in occupational safety (i.e. Lithuania 2.4 cases per 1,000 employed; Latvia 2 cases per 1,000 employed and Bulgaria 0.7 cases per 1000 employed). It should also be noted that the rate of fatal occupational injuries in the EU is significantly lower than in Russia (0.016 cases per 1000 employed in 2013 as compared to the current level in Russia of 0.062 per 1000). The total number of work days lost due to disability in Russia accounts for 1.4 million or 65.7 days per 1000 employed, which is 6.5 times higher than the EU average (10.87 days per 1000 employed). Hence, the economic impact of insufficient occupational safety in Russia is much higher than in the EU.

The budget costs related to occupational safety inspections have been decreasing for the past few years. In 2015 the budget expenditures accounted for about 2 bln RUR. The average expenditures per one inspection decreased from 15.94 thousand RUR in 2013 to 14.60 thousand RUR in 2015. The average labor costs per inspection also dropped from some 29 man-hours to 27.7 man-hours. Hence, in terms of efficiency, the trend is positive.

5 URL: <http://www.hse.gov.uk/STATISTICS/overall/hssh1415.pdf>

Overall, the analysis suggests that there are positive trends in terms of the rising effectiveness and efficiency of regulatory enforcement and inspections in occupational safety; however, Russia is still significantly lagging behind the EU countries in terms of fatal injuries frequency and economic losses due to breaches of occupational safety.

6.2 Evaluating the performance of business inspections of food safety

A similar approach was used to evaluate the efficiency and effectiveness of business inspections in the area of food safety. To measure the outcomes, some statistical data related to the number of cases of infectious and parasitic diseases was used (there is no centralized statistics on food poisoning in Russia). While in Russia a number of government authorities at the federal and regional levels are in charge of food safety issues at least to some extent, for the purpose of this study for measuring intermediate outcomes and outputs the data for the one (major) authority – Rospotrebnadzor (Federal Service for Oversight over Consumer Rights) – was used. The results of the calculations are presented in Table 3.

The table illustrates that the effectiveness of business inspections in the area of food safety is satisfactory: while the threshold-based evaluation exceeds 100 %, dynamics-based evaluation is below 100 %. This demonstrates that the situation in this area of regulatory enforcement and inspections is getting worse. There are negative trends in terms of fatal cases from infectious and parasitic diseases and significant growth in the total number of youth with cases of food poisoning in camps. The dynamics of intermediate outcomes is negative as the ratio of prevented and non-prevented damage has been decreasing for the past few years. Most outputs show positive dynamics with the exception of compliance rate, which tends to decrease, partially due to possibly better selectivity of the inspection process.

Table 3
Effectiveness of business inspections of food safety in Russia

Indicator	Weight	2014	2015	Threshold value	Dynamics-based evaluation	Threshold-based evaluation
Final outcomes	0.5				90.1	100.3
Number of cases with parasitic and some infectious diseases, per 1000 inhabitants	0.4	30.8	28.1	31.8	109.6	113.2
Number of children and youth affected by mass infections/food poisoning in camps	0.2	179	453	345.8	39.5	76.3

Indicator	Weight	2014	2015	Threshold value	Dynamics-based evaluation	Threshold-based evaluation
Number of fatal cases with parasitic and some infectious diseases, per 1000 inhabitants	0.4	21.1	22.0	21.9	95.9	99.4
Intermediate outcomes	0.3				95.3	117.0
Number of cases of health damage	0.4	1241	1090	1507.5	113.9	138.3
Ratio of prevented and non-prevented damage (ratio in the number of legal entities with violations which pose a threat to occupational safety to the number of legal entities with violations which were the cause of accidents)	0.4	18.76	13.96	21.80	74.4	64.1
Percentage of enterprises which claim that product safety has improved due to inspection visit	0.2		35.2	19.5	100.0	180.5
Outputs	0.2				101.7	118.0
Percentage of enterprises which evaluate their awareness about regulatory requirements in occupational safety as high and very high	0.1		60.5	60.3	100.0	100.3
Average administrative costs for business per one inspection (including informal payments)	0.1		40.5	105.3	100.0	260.2
Compliance rate (percentage of enterprises for which no major violations were detected based on inspection)	0.3	74.68	74.10	75.00	99.2	98.8
Percentage of enterprises which claim they have eliminated all violations (out of the total number of enterprises with violations)	0.2		83	75.7	109.6	109.6

Indicator	Weight	2014	2015	Threshold value	Dynamics-based evaluation	Threshold-based evaluation
Percentage of enterprises which agree with the outcome of inspection visits	0.1		90.4	87.3	100.0	103.6
Percentage of inspections which were cancelled (by the management, prosecutor office or court)	0.1	0.01	0.01	0.01	100.0	100.0
Percentage of cancelled inspections which resulted in administrative measures on the inspectors for violations	0.1	100	100	100	100.0	100.0
Evaluation of Effectiveness					94.0	108.9

Sources: calculated by the author based on Rosstat data (www.gks.ru), Rospotrebnadzor data (www.rospotrebnadzor.ru), and RANEPA survey (Dobrolyubova et al. 2017).

This example demonstrates the gap between the positive dynamics of outputs (demonstrating improvements in the way the inspectorate operates) and negative trends on the outcome level. This gap suggests that the system of inspections is not targeting actual risks well, and improvements, including the introduction of a risk-based approach to inspections, are needed.

In terms of business-inspections efficiency, two quite opposite trends are noted, based on the data published by Rospotrebnadzor. On the one hand, the average number of business inspections per one inspector has been decreasing: this indicator dropped from 31.2 inspections per 1 (actual) staff in 2013 to 26.0 inspection per staff in 2015. This trend can be partially explained by the fact that Rospotrebnadzor has been among the control bodies with the highest workload in terms of inspection per staff ratio (Plaksin et al. 2015). At the same time, the average budget expenditures per inspection in 2015 accounted for some 26.0 thousand RUR (compared to 27.3 thousand RUR in 2014), which demonstrates some efficiency gains. Based on these trends, the efficiency may be rated satisfactorily.

While up-to-date cross-country data on food-safety issues is lacking, the data on accidental poisoning published by WHO (the latest available year is 2011) demonstrates that the frequency of poisoning in Russia accounts for 29.34 cases per 100 thousand population, which is significantly higher than in Kazakhstan (19.43), Ukraine (18.5) and the EU countries.⁶ For example, the frequency of accidental poi-

⁶ <https://gateway.euro.who.int/en/hfa-explorer/>

soning in the UK accounted for 8.9 cases per 100 thousand population, in France for 2.42 cases, respectively. This statistics somewhat supports the conclusion made based on the evaluation – the extent of food safety and, therefore, the performance of the regulatory enforcement and inspection authorities is insufficient and needs further improvement.

7. Discussion and conclusions

The proposed evaluation framework aims at addressing the weaknesses of applying traditional approaches to measuring the performance of government-inspection bodies. To eliminate the controversy in corresponding inputs and outputs, an algorithm is proposed, which is based on consecutive evaluation (first, evaluation of effectiveness, including outputs and outcomes, then, if the effectiveness is satisfactory, evaluation of efficiency). To link the operations of inspection bodies with the ultimate public good they produce and to account for other factors influencing the level of risks, intermediate outcome measures are applied. To account for the interests of all key stakeholders, performance indicators reflecting citizen, business, and government (fiscal) perspectives are included in the performance framework. Finally, to minimize the risk of data manipulation on the part of the government inspection bodies, most of the performance indicators used are based on the data from other (independent) sources, such as official statistics or sociological surveys.

The use of several evaluation methods (i.e. target-based evaluation, dynamics-based evaluation and threshold-based evaluation) proposed by the evaluation algorithm allows for developing an interval of acceptable performance levels rather than setting a strict target to be achieved. This approach reduces both incentives and possibilities for data manipulation. Using the interval as a target may also reduce the issues related to possible data manipulation. Interval-based targets could also help to account for natural year-to-year variation of risk factors which are beyond the control of businesses or government bodies (for example, climate and other natural factors).

The proposed framework was tested by evaluating the performance of two Russian government inspection bodies. This pilot evaluation has some limitations. Due to the absence of target values, no target-based evaluation has been conducted. Lack of available data on reimbursing (compensating) the damage affected the list of the outcome indicators used. The weights used for the evaluation were set based on expert assessment rather than in-depth evaluation of historical data on risk events and related damage (due to the lack of relevant data). Nevertheless, the results of the evaluation confirmed that the proposed framework could be useful both in the context of the business inspection reforms in Russia and for similar efforts elsewhere.

The examples presented in this paper demonstrate that the framework for business inspection evaluation should not be limited just to outputs or outcomes.

The practical cases analyzed show that while outcomes may be improving, there may be negative trends in outputs posing a risk for the outcomes in the medium term. The situation also may be the opposite: while performance outputs may be improving, the outcomes presenting the value of inspections to society may deteriorate. Hence, a broader picture with final and intermediate outcomes as well as outputs is needed to provide for a comprehensive performance evaluation.

Applying the performance framework at an agency (inspectorate) level calls for measuring the impact of regulatory enforcement and inspections on the extent of safety and public-interest protection achieved. Given that other factors influencing such risks always exist, an acceptable proxy demonstrating the influence of business inspections on business processes and production safety could be used. It is preferable that such influence is measured based on enterprise data, rather than the data collected by inspection bodies (in the two cases analyzed, the data from a business survey was used to evaluate the impact of government inspections on business processes).

Factoring in the business perspective into the evaluation framework is also important. While the budget also incurs some administrative costs related to conducting inspections, often the administrative costs borne by businesses are significant and incomparable with the risk reduction achieved. Including business administrative costs in the framework used for measuring the performance of government inspection bodies would motivate the latter to better target inspection visits and reduce the unnecessary administrative burden in the economy.

Including performance indicators measuring the compensation of damage into the evaluation framework is another motivating factor. Not only do such measures reflect the interests of citizens and organizations affected by risks; raising the debate on compensation of damage also creates a basis for comparing the effectiveness of government inspections with other regulatory alternatives (i.e. market instruments, self-regulation and the like), which could be an interesting direction for future research.

References

- Alach, Z. 2017. "Towards a Standard Conceptual Typology of Public Sector Performance Measurement." *Kotuitui* 12(1), 56–69.
- Arndt, C., S. Hermanutz, C. Kauffmann and R. Schultz. 2016. "Building Regulatory Policy Systems in OECD Countries." *OECD Regulatory Policy Working Paper* 5. Paris: OECD Publishing. <http://dx.doi.org/10.1787/dbb1a18f-en>
- Barbato, G. and M. Turri. 2017. "Understanding Public Performance Measurement through Theoretical Pluralism." *International Journal of Public Sector Management* 30(1), 15–30.

- Blanc, F. 2009. *Measuring Business Inspections: A Quick Guide to Surveys and Other Methodologies*. IFC. Available at http://www-wds.worldbank.org/external/default/WDSCContentServer/WDSP/IB/2011/02/23/000356161_20110223035404/Rendered/PDF/597910WP02009111public10BOX358294B0.pdf (last accessed October 30, 2017).
- Blanc, F. 2012. *Inspection Reforms: Why, How, and with What Results*. Paris: OECD Publishing. Available at <http://www.oecd.org/regreform/Inspection%20reforms%20-%20web%20-F.%20Blanc.pdf> (last accessed October 30, 2017).
- Bouckaert, G. and J. Halligan. 2008. *Managing Performance: International Comparisons*. London: Routledge.
- Boyne, G., P. Day and R. Walker. 2002. "The Evaluation of Public Service Inspection: A Theoretical Framework." *Urban Studies* 39(7), 1197–1212.
- Charlebois, S. and S. Hielm. 2014. "Empowering the Regulators in the Development of National Performance Measurements in Food Safety." *British Food Journal* 116(2), 317–336.
- Choong, K. K. 2013. "Understanding the Features of Performance Measurement System: A Literature Review." *Measuring Business Excellence* 17(4), 102–121.
- Christensen, T. and P. Lægrend. 2015. "Performance and Accountability: A Theoretical Discussion and an Empirical Assessment." *Public Organization Review* 15(2), 207–225.
- Christie, C. A. and M. C. Alkin. 2008. "Evaluation Theory Tree Re-Examined." *Studies in Educational Evaluation* 34(3), 131–135.
- Coglianesi, C. 2012. *Measuring Regulatory Performance: Evaluating the Impact of Regulation and Regulatory Policy*. Paris: OECD Publishing. Available at https://www.oecd.org/gov/regulatory-policy/1_coglianesi%20web.pdf (last accessed October 30, 2017).
- Cordes, J. J. 2017. "Using Cost-Benefit Analysis and Social Return on Investment to Evaluate the Impact of Social Enterprise: Promises, Implementation, and Limitations." *Evaluation and Program Planning* 64, 98–104.
- Dan, S. and C. Pollitt. 2015. "NPM can Work: An Optimistic Review of the Impact of New Public Management Reforms in Central and Eastern Europe." *Public Management Review* 17(9), 1305–1332.
- Dobrolyubova, E. I. 2016. "International Experience in Evaluating Effectiveness of Public Control in Occupational Safety." *Public Administration Issues* (2), 96–110.
- Dobrolyubova, E. I., N. V. Zybunovskaya, A. N. Pokida and V. N. Yuzhakov. 2017. "Evaluating Impact of Enforcement and Inspections on Business Activity." *Public Administration Issues* 2, 7–25.

- Doroshenko, S. V. 2012. "Assessment of Environment Protection Activity Impact in Russian Regions." *Economy of Region 2*, 116–128.
- FSIS. 2016. *The Food Safety and Inspection Service Strategic Plan for 2017–2021*. Available at <https://www.fsis.usda.gov/wps/portal/informational/aboutfsis/strategic-planning/fy-2017-2021-strategic-plan> (last accessed October 30, 2017).
- Gao, J. 2015. "Performance Measurement and Management in the Public Sector: Some Lessons from Research Evidence." *Public Administration and Development* 35(2), 86–96.
- Greiling, D. 2006. "Performance Measurement: A Remedy for Increasing the Efficiency of Public Services?" *International Journal of Productivity and Performance Management* 55(6), 448–465.
- Grossi, G., M. B. Hansen, J. Johanson, J. Vakkuri and M. J. Moon. 2016. "Introduction: Comparative Performance Management and Accountability in the Age of Austerity." *Public Performance and Management Review* 39(3), 499–505.
- Hammerschmid, G. and L. Löffler. 2015. "The Implementation of Performance Management in European Central Governments: More a North-South than an East-West Divide." *NISPACEE Journal of Public Administration and Policy* 8(2), 49–68.
- Hampton, P. 2005. *Reducing Administrative Burdens: Effective Inspection and Enforcement*. Available at http://news.bbc.co.uk/nol/shared/bsp/hi/pdfs/bud-05hampton_150305_640.pdf (last accessed May 30, 2017).
- HSE. 2016. *Annual Report and Accounts 2015–16*. Available at <http://www.hse.gov.uk/aboutus/reports/ara-2015-16.pdf> (last accessed May 30, 2017).
- Kaplan, R. S. and D. P. Norton. 1992. "The Balanced Scorecard: Measures that Drive Performance." *Harvard Business Review* 70(1), 71–79.
- Kravchuk, R. S. and R. W. Schack. 1996. "Designing Effective Performance-Measurement Systems under the Government Performance and Results Act of 1993." *Public Administration Review* 56(4), 348–358.
- Lewis, J. M. 2015. "The Politics and Consequences of Performance Measurement." *Policy and Society* 34(1), 1–12.
- Ma, S., G. Chen and B.-K. Tan. 2015. "Aspects of the Health Inspection Authority in the People's Republic of China Health Policies, Systems and Management." *BMC Public Health* 15(1), Art. No. 499.
- Magd, H. and A. Curry. 2003. "Benchmarking: Achieving Best Value in Public-Sector Organisations." *Benchmarking* 10(3), 261–286.
- Monk, J. 2012. *Reform of Regulatory Enforcement and Inspections in OECD Countries*. Paris: OECD Publishing.

- Moore, M. H. 2000. "Managing for Value: Organizational Strategy in For-Profit, Nonprofit, and Governmental Organizations." *Nonprofit and Voluntary Sector Quarterly* 29 (SUPPL.), 183–204.
- Moynihan, D. P., S. K. Pandey and B. E. Wright. 2012. "Prosocial Values and Performance Management Theory: Linking Perceived Social Impact and Performance Information Use." *Governance* 25(3), 463–483.
- Northcott, D. and T. Ma'amora Taulapapa. 2012. "Using the Balanced Scorecard to Manage Performance in Public Sector Organizations: Issues and Challenges." *International Journal of Public Sector Management* 25(3), 166–191.
- OECD. 2013. *Governing Effective Prevention and Mitigation of Disruptive Shocks*. Risk Management Police Issues Paper. Paris: OECD Publishing.
- OECD. 2014. *Regulatory Enforcement and Inspections*. Paris: OECD Publishing.
- Plaksin, S. M., S. V. Semenov and E. V. Klimovskaya. 2015. *Onsnovnye tendentsii I itogi kontrol'no-nadzornoj dejatel'nosti v Rossijskoj Federacii v 2011–2014 gg* [Main trends and results of control and oversight activities in the Russian Federation in 2011–2014]. Moscow: HSE.
- Reichborn-Kjennerud, K. and S. I. Vabo. 2017. "Performance Audit as a Contributor to Change and Improvement in Public Administration." *Evaluation* 23(1), 6–23.
- Safe Work Australia. 2013. *The Effectiveness of Work Health and Safety Interventions by Regulators: A Literature Review*. Available at <http://www.safeworkaustralia.gov.au/sites/SWA/about/Publications/Documents/768/Effectiveness-WHS-interventions-by-regulators-literature-review.pdf> (last accessed May 30, 2017).
- Salais, R. 2010. "La donnée n'est pas un donné pour une analyse critique de l'évaluation chiffrée de la performance" [Statistics do not stand alone: Critical analysis of the quantitative assessment of public sector performance]. *Revue Francaise d'Administration Publique* 135(3), 497–515.
- Schwartz, R. and J. Mayne. 2005. "Assuring the Quality of Evaluative Information: Theory and Practice." *Evaluation and Program Planning* 28(1), 1–14.
- Speklé, R. F. and F. H. M. Verbeeten. 2014. "The Use of Performance Measurement Systems in the Public Sector: Effects on Performance." *Management Accounting Research* 25(2), 131–146.
- Ter Bogt, H. J., G. Jan Van Helden and B. Van Der Kolk. 2015. "Challenging the NPM ideas about Performance Management: Selectivity and Differentiation in Outcome-Oriented Performance Budgeting." *Financial Accountability and Management* 31(3), 287–315.

- Virtanen, P. and J. Vakkuri. 2015. "Searching for Organizational Intelligence in the Evolution of Public-Sector Performance Management." *NISPACEE Journal of Public Administration and Policy* 8(2), 89–99.
- WEF. 2016. *The Global Competitiveness Report 2016–2017*. Available at <https://www.weforum.org/reports/the-global-competitiveness-report-2016-2017-1> (last accessed October 30, 2017).
- Williams, D. 2014. "The Evolution of the Performance Model from Black Box to the Logic Model Through Systems Thinking." *International Journal of Public Administration* 37(13), 932–944.
- World Bank. 2006. *Good Practices for Business Inspections: Guidelines for Reformers*. Available at <https://www.wbginvestmentclimate.org/uploads/Bus%20Inspect%20Toolkit.pdf> (last accessed May 30, 2017).
- World Bank. 2011. *How to Reform Business Inspections: Design, Implementation, Challenges*. Washington D.C.: World Bank.
- World Bank. 2016. *Doing Business Report 2017*. Washington D.C.: World Bank.