Human Interaction and Emerging Technologies (IHIET 2023), Vol. 111, 2023, 409–418
https://doi.org/10.54941/ahfe1004047

User Experience of a Web-Based Platform That Enables Ethical Assessment of Artificial Intelligence in the Public Sector

Maria Tsourma[1], Noemi Luna Carmeno[2], Jaime Alessandro Codagnone[2], Sara Manchini[2], Jesper Krognos[3], Anastasios Drosou[1], and Dimitrios Tzovaras[1]

[1]Information Technologies Institute, Centre for Research and Technology Hellas (CERTH), Thermi-Thessaloniki, 57001, Greece
[2]Intellera Consulting, 20145 Milano (MI), Italy
[3]2021.AI, 2200 Copenhagen, Denmark

ABSTRACT

As public sector organizations increasingly adopt Artificial Intelligence (AI) technologies, it is important to ensure that they are used in a responsible and ethical manner. The use of AI systems can have unintended consequences, such as exacerbating existing inequalities or infringing on individuals' privacy rights. Therefore, the use of a web-based platform that enables the ethical assessment of AI helps to identify potential risks and ethical concerns before these technologies are deployed and used by public administrations. This paper presents a web-based platform implemented to support the ethical assessment of AI use in Public Sector, along with its evaluation. The web-based platform implemented for this purpose is designed to address ethical, legal, and social vulnerabilities, allowing Public Sector stakeholders to adopt AI applications in a trustworthy, controlled, and responsible manner. This platform is designed to offer easy transparency of latent risks and the corresponding mitigation measures. The evaluation of this platform was conducted by four public administrations from three different European countries (Italy, Greece and Norway), while the feedback was collected through questionnaires and interviews. The findings of this study can be taken into consideration by developers and research community for the development and adoption of AI applications in public administrations.

Keywords: Ethical assessment of artificial intelligence in public sector, AI in public administration, User evaluation, Web-based platform for ethical AI use assessment, Trustworthy AI

INTRODUCTION

In recent years, Artificial Intelligence (AI) has been increasingly deployed in various public administration domains, ranging from healthcare and public safety to transportation and education, due to its potential to revolutionize the way governments operate. One area in which AI has been widely applied is healthcare, where it has been used for tasks such as medical imaging, drug discovery, and patient triage (Panch et al., 2019). The use of AI in healthcare

can improve the accuracy of medical diagnoses and reduce the workload of healthcare professionals, allowing them to focus on more complex tasks. Another area where AI has been deployed is public safety (Simić, et al., 2020), where it has been used to monitor and analyze crime patterns, predict where crimes are likely to occur, and identify potential security threats. AI can also help law enforcement agencies allocate their resources more effectively and prevent crime before it occurs.

More broadly, AI has been used to improve the delivery of public services and improve the quality of citizens' lives. Examples include the introduction of chatbots in the public administrations, to facilitate communication with the municipality and the citizens (Følstad et al., 2023), or the use of AI in information collection and prioritization process by municipalities. In fact, AI can be used to optimize data collection process to collect all the available data required for decision making, while also evaluating and detecting misinformation to prioritize the counter actions that need to be performed in each case (Tsourma, et al., 2021). AI-powered systems can also be used to foster the green transition, such as to sort waste more effectively, identifying recyclable materials and separating them from non-recyclable waste.

While the potential benefits of using AI in public administration are clear, there are also significant risks associated with its use, including biases, discrimination, and lack of transparency among others (Eitel-Porter, 2021). The public sector is also considered one of the most impacted by the proposal of the EU regulation on AI (AI Act[1]), as most of the prohibited and high-risk applications pertain the PS. To address these challenges and ensure that AI is used in a responsible and ethical manner, it's crucial for public administrations to have tools that evaluate its use. These evaluation tools can help identify potential biases or unintended consequences of AI systems, increase transparency, promote fairness, and ensure continual monitoring and improvement. This paper will explore the importance of such tools in detail and highlight their potential benefits for public administrations, citizens, and society as a whole, through the evaluation of a web-based assessment platform implemented for this scope.

The web-based platform implemented for this purpose, called ETAPAS prototype platform, aims to be used by the public sector during the development or adoption of an AI-based solution. The platform integrates a comprehensive Responsible Disruptive Technologies (RDT) Framework[2] consisting of ethical principles, an analysis of the EU legal framework, an overview of the ethical, social and legal risks and an ensuing RDT Indicators Framework to practically measure those risks and impacts. In particular, the aim of the Indicators Framework is to assess the risks and negative impacts of the adoption of an AI application by a public sector organisation, but also the effectiveness of potential mitigation actions that can be implemented.

Apart from the ETAPAS prototype platform, this paper presents the user evaluation of this platform by four public administrations from three European countries (i.e., Greece, Italy, Norway). The public administrations

[1]https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[2]https://zenodo.org/communities/etapas-project/?page = 1&size = 20

involved are heterogenous in terms of digital maturity, type (both local and central), geographical area (Southern and Northern Europe) and sector (i.e., health, general public services, HR, public relations). This is to widen the applicability of the results and for testing and evaluating the ETAPAS prototype platform beyond the project perimeter by taking into account the different level and type of AI use in the European public sectors. The aim of this study is to understand how each EU country can leverage the results of the ETAPAS prototype platform, in order to reduce the risk of using AI-related applications and the increase in their use in the public sector.

ETAPAS PROTOTYPE PLATFORM

The Public Sector is especially reluctant to AI and other Disruptive Technologies (DTs) at this early stage of adoption, due to its special role and responsibility toward citizens, and considering the evolving regulatory landscape that the European Commission is establishing (e.g., AI Act, Data Governance Act, AI Liability Directive, Digital Services Act, etc). AI-based applications are interactive solutions that are very frequently developed by making extensive use of Learning Functions. Learning Functions typically involve complex user interactions and a degree of learning either from previous user/robotic interactions or historical data. For this reason, it is important that AI-based applications owners identify early the relevant components that should be considered to consistently comply with Ethical, Social, and Legal frameworks. This is an iterative process, starting at the initial inception of the AI-based applications and running all the way through until their production and use.

Aiming to meet the needs that exist for the ethical adoption of such technologies, the ETAPAS Prototype platform was developed. This platform is designed to address ethical, legal, and social vulnerabilities, allowing Public Sector stakeholders (i.e., policymakers and IT managers) to adopt DTs in a trustworthy, controlled, and responsible manner. Furthermore, it is designed to offer easy transparency of latent risks and the corresponding mitigation measures. The ETAPAS Prototype platform works by enabling full transparency of potential risks and impacts, leading to fairness and accountability at the same time, and as a result, enabling trust from both the public and regulators.

The platform is designed to automate tasks, instead of requiring additional work from developers. Since AI based applications typically involve complex interactions and a degree of self-learning or cognition by looking at previous interactions or historical data, it is important that the validation process occurs in an automated way to ensure simplicity, transparency, and traceability. Therefore, the ETAPAS Prototype platform is built upon a horizontal and scalable infrastructure and enables it to accept and validate an almost unlimited number of AI-based applications including pairing them with the right frameworks required for the application in question.

Throughout the implementation of an AI-based application, there will be a need to apply risk mitigation strategies to reduce potential risk concerns

specific to the type of the AI-based application and the type of ethical/social or legal frameworks it must comply with. AI-based applications can be developed in external environments or within the ETAPAS Platform itself, but they use the ETAPAS platform's Connectors to enable tracking of Governance meta-data[3] composing the computational indicators of the ETAPAS prototype platform. These computational indicators are metrics that derive automatically from a set of data an assessment score for a certain risk-aspect of the AI-based application assessed and are tailored to the application. As a matter of fact, where risk aspects can be investigated quantitatively or by ascertaining the occurrence of certain events, the ETAPAS prototype platform can provide more efficiency in managing the assessment autonomously by connecting to each AI-based application and receive computational indicators posted by the application.

While developing AI based applications, it is important to initially identify ethical, social or legal risks for the particular DT based application. This is done by each AI-based application's Owner or Developer logging into the ETAPAS Platform and going through a process of an initial configuration. This process determines the importance of the indicators grouped into risk categories. The data necessary to compute the ethical, social or legal risk indicators should be collected by the AI based application itself. For each risk indicator, it is also possible to identify some mitigation actions that arise from the application's owner's or developer's perspective as a to-do list of tasks that needs to be done to successfully receive the certification of compliance. It is important that the application is re-assessed at certain intervals to monitor how the ethical, social or legal risk level changes over time and how the mitigation actions are effective. This can happen at regular intervals during development (e.g., each six months) or at important project milestones. Depending on the stage in the life cycle only relevant risk and mitigation actions are shown thereby ensuring the continued relevance towards the project.

In the ETAPAS prototype platform, an impact assessment questionnaire is defined based on a pre-defined Responsible Disruptive Technologies (RDT) Framework. The RDT Framework defines Impact Assessments[4] that are supposed to be filled in from the AI-based application's owner's side across 3 stages (Figure 1). Stage 1 of Assessment is tied to pre-development (ex-ante in EU AI act terminology) and is filled in before the development of the AI-based application. Second stage is filled in during development, and third after, during the operations phase. Assessments are targeted toward informing users on what are the main areas of risk, as well as breaking down mitigation measures that AI developers and owners can do to mitigate the risk factors. Upon completion of an assessment user is presented with a report not only showing the answers but also providing the recommendations based on the answer provided for actions that could improve their risk/mitigation scores (Figure 2).

[3]Examples of Governance meta-data: Accuracy score for a model, Bias scores, Feature Impact Scores, Explainability Features

[4]Impact Assessments are documentation tools used to store details and measure risks connected with each AI-based application.

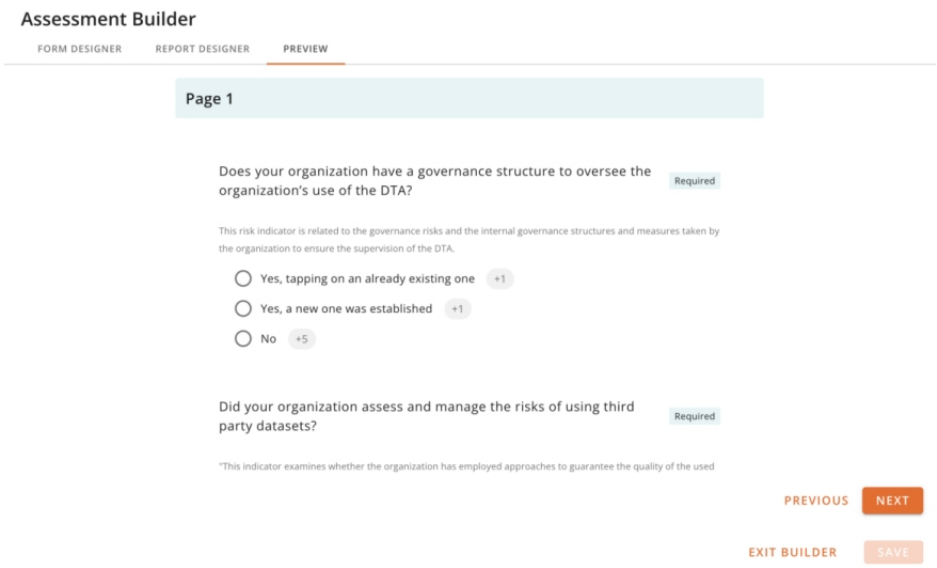

Figure 1: Example of impact assessment as presented in the ETAPAS prototype platform.

Figure 2: Risk assessment results presenting in detail the percentage of risk calculated per risk and mitigation indicator.

Aiming to make the interaction with the platform more user friendly and explainable, the ETAPAS Prototype platform is supported by a "Reports and Dashboards" system that displays the meta-data generated from the development and execution of the AI-based applications. In Figure 3, the outcomes of the system are presented, showing to the user the risk score per risk category, calculated via a predefined formula, and a mitigation score, as they have been resulted from the assessment conducted for the AI-based application. The risk and mitigation scores are reported both in percentage and in absolute terms. As per the RDT methodology, these scores should be considered together as the higher the risk score, the higher should be the mitigation score to ensure that a sufficient number of actions were put in place to reduce such risk. In addition, a view by risk categories will be available with more recommendations for actions and suggestions for mitigation actions that should be implemented based on the results of the assessment.

Figure 3: Presentation of the collected governance meta-data in each risk indicator.

Since the assessment is to be performed more than once during the AI-based application's life cycles, aggregated results will be shown in the overall results dashboard, where the results of each assessment carried out in the various stages of AI-based application's implementation will be shown on a timeline. The report generates similar content as the dashboards but in a more complete and encompassing manner, enabling the application's owners to extract and share these reports with relevant stakeholders and project members. After a request for a report, the platform generates PDF covering sections of results, assessment answers, computational indicators, and historical data.

Moreover, in addition to the risk assessment results presentation depicted in Figure 2, additional graphics have been added to the ETAPAS prototype platform to represent the values of risks and mitigation actions made over time. This historical representation of each assessment's results, allows both the developers of an AI-based application and its owner, to understand if the mitigation actions followed over time led to better results or not, and plan accordingly the next mitigation actions to be performed in each risk.

METHOD

Central and local public administrations from four AI-based applications and three different European countries (i.e., Greece, Italy, Norway) participated in the evaluation process. From each public administration, two employees that have used the ETAPAS prototype platform were engaged, aged from 18–45 years (average 36 years). Two of them were digitally illiterate, while the rest of them had used AI applications in the past. All employees had the chance to test at first the ETAPAS prototype platform and use is to take a risk assessment, more than once.

The feedback collection method that was followed included interviews, accompanied with an evaluation questionnaire. The evaluation questionnaire was used to collect participant demographics, User experience using the User Experience Questionnaire (UEQ) (Brooke, 2016) and knowledge sharing open questions. Initially, the interview's questions were distributed to all participants, while they were given a week to answer it, before the interview

process. The interview process was conducted solely with the public sector employees of each country by two facilitators.

During the interviews, participants were encouraged to discuss about the digital literacy of their country, and also about the readiness, reluctancy or willingness of the public sector to use AI-based applications. After trying each feature, participants gave comments and facilitators asked questions related to the pros and cons as well as further development ideas. Each participant filled in an evaluation questionnaire at the end of the focus group.

PLATFORM EVALUATION RESULTS AND DISCUSSION

This section presents the evaluation evidence collected using the aforementioned methodology along with a discussion of the results, which include the presentation of the UEQ collected. We also present further development ideas emerged from the interviews process.

The analysis of the collected information upon the 11 scales of the UEQ questionnaire is presented in Figure 4. The positive feedback collected from the users has been noted especially in the Trustworthiness of content, concerning the impression of the user about the information provided by the ETAPAS prototype platform, and dependability scales, concerning the subjective impression that the output of the ETAPAS prototype platform responds predictably and consistently to inputs and commands. An above average impression has been noted on the Attractiveness, Efficiency, Visual aesthetics, Quality of content, Trust, Stimulation, and Novelty scales. The importance of these scales, presented in Figure 5, lies in the fact that the user interaction is highly related and connected with the content of the platform, rather than with its aesthetics. In fact, during the interviews the PS representatives underlined the need for a simplified version of the RDT framework and methodology to better grasp its concept and consequently make the best out of the platform too.

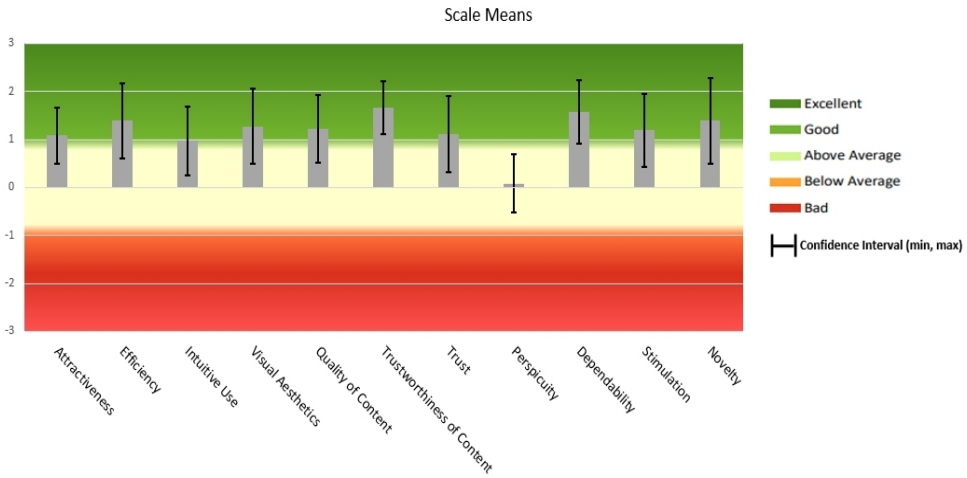

Figure 4: User evaluation results.

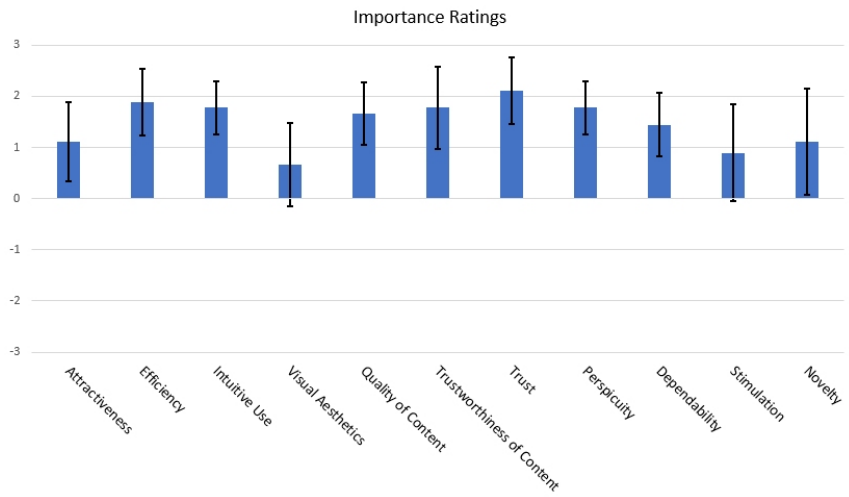

Figure 5: User evaluation results – importance ratings.

According to the results, all public administrations participating in the assessment process had a common vision on focusing more on the content of the assessment and the results, rather than focusing on the aesthetics of the platform. Most PS representatives stated that they need to have a presentation of the risks included in the impact assessment and the mitigation indicators beforehand, in order to have time to better understand the scope of the assessment and become prepared of the answers that they would give. In addition, guidelines on how the impact assessment should be completed must be provided to the corresponding partners, along with a glossary including the terminology of the assessment with their definition.

One of the scales that had a negative impression, but an increased rating in importance, was perspicuity. Through this scale the user has the subjective impression that it is easy to understand and learn how to use the ETAPAS prototype platform. Most of the users found the platform not very easy to use, since not all countries are familiar with using e-governance platforms or similar solutions daily. This results from the insufficient training of the staff who handled the application, identified during the discussions conducted during the interviews process. One basic example could be the public administrations in Greece. As presented by Georgios (2021), because of persistent infections, the Greek public administration has long lagged behind the typical paradigm of operation among the member states of the European Union. On the other hand, it is crucial to emphasize that this discrepancy is decreasing compared to the early years of the Greek state's membership in the European Economic Community and the implementation of numerous restructuring programs of the Greek public sector (Hahamis et al., 2005). Therefore, the use of similar applications in Greece (or other countries with similar digital maturity level) required the performance of additional steps, including the organisation of workshops to present the ETAPAS prototype platform and the preparation of videos and documentation material to help them understand the platform's scope and main functionalities. This material

has been prepared and distributed to all participants, along with the organisation of hands-on lessons in order to train the public officials on how to use the ETAPAS prototype platform.

Another evidence we collected concerns the stakeholders' roles assigned with the responsibility of performing the ETAPAS assessment. Due to the nature of the AI-based applications and the impact assessment part, participants from three out of four public administrations mentioned that it would be of great benefit by the AI-based application's owners to have the option to have different sections answered by different stakeholders, based on the knowledge and involvement of each participant in the AI-based application's development.

Regarding the rest of the countries and their scope on using the ETAPAS Prototype platform, Italy and Norway are more familiar with the use of similar applications and e-government applications in general (Kuziemski et al., 2020, Noordt et al., 2021). According to a study conducted by Datta (2020), the use of applications and AI-based applications in the Italian public administration has been increasing in the last years. Referring to the context of the application, participants from both countries mentioned that it might be useful also to have access to the entire recommendations and best practices when responding to the assessment indicator. Hence, after the user experience evaluations, it should be highlighted that all four public administrations found the ETAPAS prototype application a useful and important tool allowing to avoid long bureaucratic processes for assessing the ethical and legal aspects of the adopted AI-based applications.

CONCLUSION

The ethical adoption of AI or DTs in general can significantly improve not only the quality but also the reliability, speed and accessibility of services provided by public administrations and public service providers. Yet, the real impact of their adoption is largely unknown, so that deploying AI-based applications in public administrations requires a thorough assessment of their potential impact, benefits and risks for the delivery of public goods.

This paper presented the user evaluation of a web-based platform intended to be used by the public sector during the development or adoption of an AI-based solution in order to assess if this solution will be used ethically and identify, monitor and ease mitigation of its ethical risks. This platform has been used by 4 public administrations from three different EU countries, in order to be assessed from the intended final users. The results are common in all four, highlighting the need for training of the civil servants and public administration managers in order to use this application, or similar e-governmental applications. Another result highlighted from this analysis concerns the need for a trustworthy content, presented by the users, in the ETAPAS prototype platform. This aspect is important, as it shows that most public servants do not trust the content of similar applications, which are used to validate other AI based applications content reliability, mostly due to the lack of explanations on each step and lack of AI ethics skills. According to this, clear explanations should be given to users regarding the content of

each question, explaining to them the reason for using this information, but also noting its importance for the use of the application being evaluated.

In the long run, citizens from European Member States will benefit even more from trustworthy AI assessment procedure, since a better trained public sector will be able to reduce digitalization risks. Moreover, not only policy-makers and IT managers need to understand the opportunities and risks of digitalization, since the successful uptake depends on trust in the solutions in the general, which in turn is dependent on knowledge regarding ethical, social, and legal implications. Technology, even when it is not "disruptive", poses many challenges to the public servants and wider civil society that need to develop a digital mindset, while ensuring inclusiveness and addressing technology societal impacts and ethical considerations.

ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101004594 (project ETAPAS).

REFERENCES

Abbas, N., Følstad, A., & Bjørkli, C. A. (2023). Chatbots as part of digital government service provision–a user perspective. In Chatbot Research and Design: 6th International Workshop, CONVERSATIONS 2022, Amsterdam, The Netherlands, November 22–23, 2022, Revised Selected Papers (pp. 66–82). Cham, Switzerland: Springer International Publishing. https://link.springer.com/chapter/10.1007/978-3-031-25581-6_5

Brooke, J. (1996). SUS-A quick and dirty usability scale. In: Usability Evaluation in Industry, pp. 189–194. Taylor & Francis, London.

Datta, P. (2020). Digital transformation of the Italian public administration: A case study. Communications of the Association for Information Systems, 46(1), 11.

Eitel-Porter, R. (2021). Beyond the promise: implementing ethical AI. AI and Ethics, 1, 73–80.

Georgios, K., & Nikolaos, V. (2021). Reinstating greek E-Governance: A framework for E-Government benchmarking, improvement and government policies. Вопросы государственного и муниципального управления, (6), 103–127.

Hahamis, P., Iles, J. and Healy, M. (2005) 'E-government in Greece: Bridging the gap between need and reality' Electronic Journal of e-Government, 3(4), pp. 185–192.

Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications policy, 44(6), 101976.

Panch, T., Pearson-Stuttard, J., Greaves, F., & Atun, R. (2019). Artificial intelligence: opportunities and risks for public health. The Lancet Digital Health, 1(1), e13–e14.

Simić, M., Perić, M., Popadić, I., Perić, D., Pavlović, M., Vučetić, M., & Stanković, M. S. (2020). Big Data and development of Smart City: System architecture and practical public safety example. SJEE, 17(3), 337–355.

Tsourma, M., Kougioumtzidou, A., Drosou, A., & Tzovaras, D. (2021, August 20). On the utilization of Disruptive Technologies for Municipality-wide Policy Making. Zenodo. https://doi.org/10.5281/zenodo.5227650

Van Noordt, C., Misuraca, G., Mortati, M., Rizzo, F. and Timan, T., AI Watch - Artificial Intelligence for the public sector, Publications Office of the European Union, Luxembourg, 2020, ISBN 978-92-76-19438-5, doi: 10.2760/25756, JRC120315.