MEDINFO 2019: Health and Wellbeing e-Networks for All
L. Ohno-Machado and B. Séroussi (Eds.)

doi:10.3233/SHTI190376

A Knowledge-Based Platform for Assessing Potential Adverse Drug Reactions at the Point of Care: User Requirements and Design

Pantelis Natsiavas[a,b], Marie-Christine Jaulent[b], Vassilis Koutkias[a]

[a] Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
[b] Sorbonne Université, INSERM, Univ Paris 13, Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances pour la e-Santé, LIMICS, F-75006 Paris, France

Abstract

Even though Adverse Drug Reactions (ADRs) constitute a significant public health issue, there is a lack of Information & Communication Technologies (ICT) tools supporting Pharmacovigilance activities at the point of care. In this paper, we present the rationale of a Web-based platform to address this need. The driving user scenario of the proposed platform refers to a clinician who investigates information for a possible ADR as part of a specific patient treatment. The goal is to facilitate this assessment through appropriate tools for searching various relevant data sources, analysing the acquired data, aggregating the obtained evidence, and offering follow-up ADR monitoring over time in a systematic and user-friendly way. In this regard, we describe the adopted user requirements engineering methodology and illustrate the use of Knowledge Engineering (KE) as the platform's main technical paradigm to enable heterogeneous data integration and handle the complexity of the underlying information processing workflow.

Keywords:

Pharmacovigilance; Drug-Related Side Effects and Adverse Reactions; Knowledge Management

Introduction

Pharmacovigilance (PV) is defined as "the science and activities related to the detection, assessment, understanding and prevention of adverse effects or any other possible drug-related problems" [1]. As Adverse Drug Reactions (ADRs) cause a significant social and financial burden [2], PV is widely recognized as an important public health priority. An estimation by the US Office of Disease Prevention and Health Promotion has recently calculated that Adverse Drug Events (ADEs)[1] are responsible for 1 in 3 of all hospital adverse events, related to about 2 million hospital stays each year, and increased hospitalization by 1.7 to 4.6 days[2]. Thus, the detection and prevention of ADRs at the point of care rise as a major clinical issue as the probability of benefit should balance the possibility and cost of potential harm [4].

The assessment of potential new or incompletely documented ADRs (called "signals") is typically performed by national and international drug monitoring/regulatory organizations (e.g., the Food and Drug Administration [FDA] in the US, the Uppsala Monitoring Centre [UMC], World Health Organization collaborating centre for international dug monitoring). These organizations perform statistical analysis of individual case safety reports (ICSRs) gathered in Spontaneous Reporting Systems (SRSs), in order to identify indications of a causal relationship between the drug administration and the adverse effect based on the measures of disproportionality [5], taking also into account other sources of evidence (e.g., scientific literature, clinical trial databases, etc.).

While SRSs are the dominant data source for PV, recent advances in Information and Communication Technologies (ICT) enable the exploitation of new, emerging data sources that can expand the real-world evidence base for PV (e.g., observational healthcare databases, social media, internet search logs). Thus, the need for comprehensive and knowledge-intensive ICT tools supporting the systematic and efficient exploitation of diverse data sources for PV is evident, in order to accommodate the entailed big data challenges [6].

To this end, we develop a Web-based platform aiming to facilitate the early identification and assessment of potential ADRs at the point of care. The main objective is to contribute at "active", post-marketing drug safety surveillance [7], focusing on the timely assessment of potential drug safety risks, supporting clinicians (as well as PV experts and researchers) to explore diverse data sources of interest and obtain actionable insights via knowledge-intensive analytics [8]. The proposed platform is currently in its "user requirements analysis" and "design" phase, which is driven by the real-life scenario according to which a clinician investigates information for a possible ADR as part of a specific patient treatment. The ultimate goal is to facilitate the integration of ADR assessment in routine clinical practice (Figure 1), by introducing tools that facilitate the search of diverse data sources, the analysis of the acquired data, the aggregation of the evidence to conclude with ADR assessment, and follow-up ADR monitoring over time in a systematic and user-friendly way.

In this paper, we present the methodology applied and the main challenges identified during the "user requirements analysis" phase of our development, which were in turn mapped to relevant user goals. We also illustrate the main elements of the platform design through the respective information processing workflow. In this regard, we elaborate on how a Knowledge Engineering (KE) based approach can accommodate the re-

[1] ADEs include side-effects that may or may not have a causal relationship with the respective drug, also referring to cases of adherence failure. ADRs refer only to side-effects caused after legitimate drug use, therefore implying a possible causal relationship between the drug and the adverse effect [3].

[2] https://health.gov/hcq/ade.asp

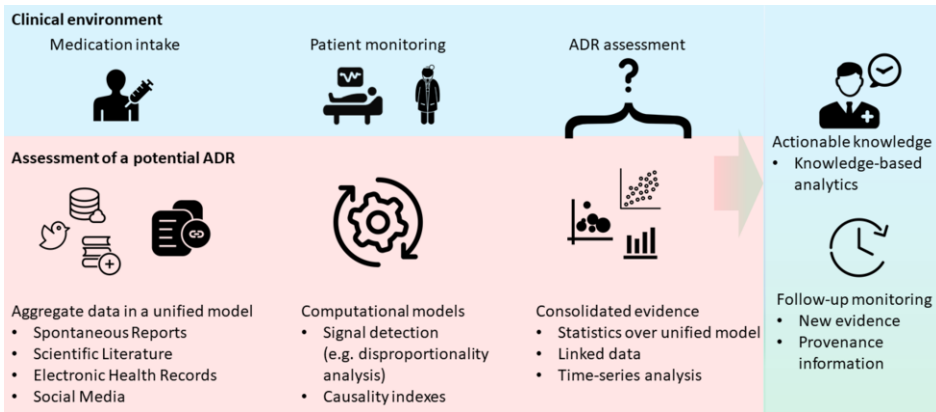

Figure 1 - Integrating the ADR assessment process in the clinical context.

spective design and development challenges. We further discuss practical implications of our work and outline directions for future work, aiming to support a comprehensive learning health system for active, post-marketing drug safety surveillance at the point of care [9].

Compared to relevant works, like the SALUS platform [10], which focused on exploiting Electronic Health Records (EHRs) for ADR detection, or the platform developed in WebRADR [11], which exploited social media for new insights on drug safety, our work relies on multiple, diverse data sources, increasing the search space for real-world evidence with an explicit focus on the clinical environment.

Methods

The employed "user requirements analysis" process is an adjusted version of the methodology described in [12]. It can be summarized as follows:

1. Analysis of the currently applied *Business Processes* (BPs) based on the respective user scenarios.
2. Definition of *User Goals* upon the elaborated BPs based on end-user input.

A *Business Process* (BP) is defined as a collection of relevant and ordered structured activities/tasks aiming to produce a specific outcome [13]. ADR assessment can also be considered as a BP conducted in the context of a hospital, as part of other parallel BPs (e.g., patient treatment, administrative processes). We envisage that the use of the proposed platform could reform the current process of ADR assessment, typically conducted manually and without systematic ICT support, to a well-defined sequence of information processing steps, supporting the overall clinical treatment processes. The ultimate goal is to optimize this BP model by satisfying the so-called *User Goals*. User goals are defined as "abstract user requirements, not directly referring to specific technical solutions or components" [12], associated with specific user actors or roles and facilitating timely identification and resolution of potential conflicts between actors. For the optimization of the ADR assessment process, user goals were elaborated based on feedback provided by clinicians and PV experts in the "user requirements analysis" and the "design" phases.

Regarding the presented platform design, the main BP of interest refers to the assessment of a potential ADR by a clinician. However, other BPs could also interact with it, e.g., concerning the patient's treatment. Given that patient treatment is the topmost priority in the clinical environment and that it is a personalized process, highly dependent on the local context (e.g. the way the specific clinic/hospital is organized), the modelling of these "interacting" BPs is very important and could be rather complex. The proper modelling and early identification of such BP interactions could be critical as they may substantially affect the identified goals and, thus, the platform design.

We model the identified BPs using flowcharts based on a notation similar to Business Process Management Notation (BPMN), in order to identify decision points, possible information processing bottlenecks, interactions with other BPs, etc. These flowcharts are further refined collectively by ICT experts, healthcare professionals (i.e., clinical doctors) and PV professionals (i.e., scientists who investigate potential ADR signals). Furthermore, interviews and workshops among researchers and end users were conducted as part of the overall "user requirements analysis" phase to analyse the established BPs, identify the *User Goals* and refine the platform's information processing workflow accordingly. Meetings were also held in the clinical environment (i.e., in the two hospitals which will host the platform in its pilot phase) to validate these goals and also address deployment issues in practice.

All types of current and emerging data sources considered in PV [6], were found interesting to explore by the end users. These include the local EHR systems, national and international SRSs, reference bibliographic databases as well as social media. From a technical viewpoint, programmatic data access is considered of high priority, as it enables systematic data gathering (e.g., spontaneous reports from the FDA Adverse Event Reporting System via openFDA [14], articles via the PubMed Central Application Programming Interface).

In order to successfully accommodate the imposed challenges regarding the synthesis and analysis of the vast data available, *Knowledge Engineering* (KE) is adopted as the main technical paradigm for our platform development. KE refers to methods, tools and theories for developing knowledge-intensive applications [15], and includes *knowledge extraction*, *knowledge integration*, *knowledge representation*, *knowledge dissemination*, and *knowledge elicitation* as its subdomains.

In the scope of our work, which concerns the systematic exploitation of all the available evidence from multiple PV data sources for the assessment of possible ADRs, we employ two technology artefacts tightly related with KE, namely, "*Linked Data*" [16] and "*Semantic Web*" [17]. Linked Data refer to a group of standards which facilitate the interconnection of data

over the existing Internet infrastructure, while Semantic Web refers to the vision of semantically annotated publicly available data interlinked via Linked Data standards.

The use of Semantic Web and Linked Data standards provides two main technical benefits: (a) *Interoperability:* The use of the Linked Data paradigm provides syntactic and semantic interoperability tools to interlink heterogeneous data sources and unify them in one processing realm. (b) *Reasoning capabilities:* The well-defined semantics upon a robust mathematical infrastructure, i.e., Description Logics [18], enable automatic reasoning through specific software, a.k.a. "reasoners".

Results

Based on this methodology and the choice of KE as the main technical paradigm, several challenges were identified from the end user perspective, leading to concrete user goals and the design of the platform's information processing workflow. In particular, given the characteristics of the clinical environment, the following challenges (enumerated with Cx) regarding the adoption of an ICT-based ADR assessment process were identified:

C1 | *Lack of time*: While the assessment of potential ADRs is identified as an important task, it is often neglected by clinicians due to lack of time.

C2 | *Lack of expertise*: PV entails specialized knowledge, which may not be available in clinical settings. While this argument supports the need for ICT-based support tools, it could also be conceived as a barrier for their adoption as their value might not be evident for the end-users.

C3 | *Adaptation to the clinical workflow*: Workflow diversity among various clinical environments (different hospitals, or even different clinics in the same hospital apply different BPs) could hinder the definition and the adoption of a "one-size-fits-all" workflow of PV information processing.

C4 | *Inadequate evidence*: While spontaneous reports are the dominant source of evidence for PV, other data sources such as EHRs, bibliographic databases, and even social media platforms are interesting for clinicians during ADR assessment. However, systematic access to multiple data sources shall be facilitated through appropriate tools.

C5 | *Coping with "big data"*: Expanding the search space for PV does not only provide a broader evidence space, but it also imposes "big data" challenges.

Overall, challenges C1-C5 have been discussed in the conducted workshops and the following user goals (enumerated with Gx) were identified and mapped to the respective challenges:

G1 | *Flexibility and Unobtrusiveness (mapped to C3):* The ADR assessment process should be flexible and tolerant to interruptions by tasks directly related with patient treatment. Thus, an important feature would be the ability to easily recover from such interruptions. Practically, this can be interpreted as the need to "save" the ADR assessment workflow and continue later. In addition, clinicians stressed that the patient's treatment should not be disrupted. Thus, the designed process should be as unobtrusive as possible, minimizing "alerts"/ "warnings".

G2 | *Balance between assessment depth and speed (mapped to C1 and C5):* As the clinician's time is valuable, the platform should enable both "in depth" assessment capabilities, while also supporting a "quick look" which could provide rigorous information. Although the information provided this way would obviously be more superficial than an "in depth" assessment, it could still provide value for clinicians.

G3 | *Semantic enhancement (mapped to C1, C2 and C5)*: Since the expression of the drug and condition of interest can be ambiguous (e.g., active substances, trade names, synonyms etc., could be used as drug terms), the overall process should be supported by curated standard terminologies and lexicons (e.g., with automatic synonym matching) to accelerate and facilitate information search and synthesis.

G4 | *Heterogeneous data synthesis (mapped to C4 and C5):* Clinicians identified the need to synthesize various and heterogeneous data sources (e.g., scientific literature, drug-information databases, clinical trial information, SRS data, observational healthcare databases). Overloading the end user with incomprehensible data was identified as a major risk and, thus, the need for knowledge-based analytics emerged.

G5 | *Data sharing (mapped to C2):* The need to share data to further elaborate on the collected ADR information and assessment results was also identified. Moreover, the value of data provenance was highlighted, especially for the process of reporting assessment outcomes to regulatory organizations.

G6 | *Follow-up monitoring over time (mapped to C4):* Typically, an ADR assessment produces a report with the conclusion and the supporting evidence. However, the time dimension is critical in PV, especially regarding newly marketed drugs. Thus, a follow-up mechanism for monitoring potential ADRs over time is important and is currently missing.

The main information processing workflow supported by the proposed platform (Figure 2) that was defined based on the abovementioned challenges and user goals are organized in 5 steps. These steps are summarized next, describing the use of the KE methods that are applicable in each case:

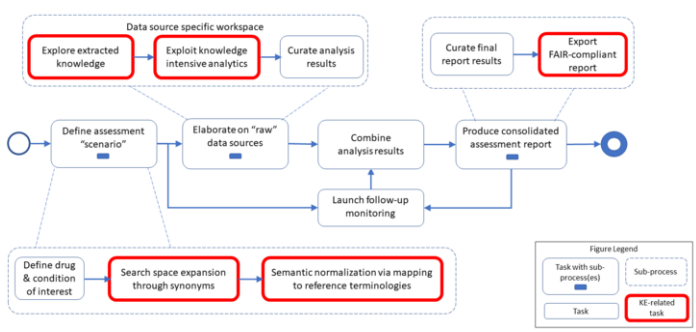

Figure 2 – Outline of the main information processing workflow supported by the proposed platform (tasks related with Knowledge Engineering are highglihted in red).

Step 1 | *Definition of "ADR assessment scenario"*: When the user launches an ADR assessment, the platform shall support the definition of the drug and the condition of interest by automatic suggestions of synonyms and relevant terms obtained from reference terminologies, e.g., the Anatomical Therapeutic Chemical (ATC) classification for drugs, and the Medical Dictionary for Regulatory Activities (MedDRA®) and the International Classification of Diseases (ICD) for the conditions of interest. The use of such well-defined knowledge structures expands the search space, enables the semantic normalization of the overall process, prevents ambiguities, and facilitates automatic information interlinking in the next analysis steps.

Step 2 | *Browse/analyse raw data from each data source*: Each raw data source has its own characteristics. For example, SRS data could be used for disproportionality analysis, while EHR data could be explored through observational healthcare data analytics [14]. KE approaches will be exploited for the analysis of each data source, e.g., text mining techniques can be used to extract and semantically annotate information from unstructured data sources such as the literature or social media. Thus, the end user can browse or analyse the respective data source in a dedicated workspace, providing suitable features and analysis capabilities.

Step 3 | *Combine analysis results from raw data sources*: The results/analysis outcome obtained from each data source workspace shall be integrated in one common processing realm, where all the analysis results could be integrated, compared and evaluated by the end user. Knowledge integration is based on semantic annotations produced in the previous steps and the use of Linked Data standards. Moreover, semantic reasoning can be applied to further elucidate knowledge from the already extracted analysis results.

Step 4 | *Produce a consolidating assessment report*: The overall analysis outcome shall be generated as a consolidated report, facilitating further analysis in collaboration with other clinicians, or even reporting to PV regulatory agencies. The produced report shall be available in both human-readable (e.g., in text form as a PDF document) and machine-readable (e.g., an RDF document) formats. Knowledge dissemination approaches can be used to facilitate the respective information exchange in a way that could promote the automatic reuse of this information. For example, the recently developed OpenPVSignal model [19] could be used in this regard, to enable compliance with the FAIR data principles [20].

Step 5 | *Launch follow-up monitoring*: The end user can launch a monitoring follow-up process, in order to receive potentially new information regarding the assessed ADR from the available data sources. This process would notify the end-user based on his/her notification preferences to avoid over-alerting. Ontology models such as the *Time Ontology* [21] and the *PROV-O Ontology* [22] can be used to enrich the obtained information with semantically enhanced time and provenance annotations, and thus, facilitate further processing/reasoning regarding the time aspects and the origin of the information collected regarding the ADR under assessment.

It should be noted that the presented workflow defines an independent BP, which can be adapted to each organizational context. Furthermore, in each assessment process, the end-user may decide the time spent on each BP step, either selecting to use the automatically retrieved information, drill-down to investigate further or manually curate the produced outcomes, and save or share his/her work with others at any time.

Discussion

Drug safety is an important issue in the clinical environment. Among the common tasks that are routinely performed in PV centres/departments in hospitals, the collection and review of all the available data for a potential ADR of interest is vital. However, there is a lack of comprehensive tools to support PV activities, specifically tailored for use at the point of care. For example, a clinician may ask for a timely evaluation of the respective patient case after a new drug administration, and the PV centre/department shall provide a documented answer with a medical advice about the case, after assessing the eventuality of an ADR [8]. To respond to this challenge, current available sources of information about the drug–event pair have to be searched by PV experts separately and in many cases without using appropriate support tools.

To address this need, we are currently developing a Web-based, knowledge-intensive platform aiming to support the assessment of potential ADRs, experienced during routine patient treatment. In the current paper, we presented the entailed challenges and the goals for such a development from the user perspective. We also presented the platform's main information processing workflow (Figure 2). Its design relies on exploiting various KE-based methods, employed in each step of the workflow.

In particular, the use of Linked Data and Semantic Web technologies provides the following key benefits:

- *Information linking* can be improved and automated by reducing the need for manual data exploration and discovery as data could be automatically retrieved.
- Rich *semantics* enhance information processing capabilities, which are typically limited in the PV domain to statistical measures of disproportionality. The already established statistical methods could be combined with semantically-enhanced knowledge sources to improve outcomes via automatic reasoning capabilities (e.g. regarding causality assessment).
- *Evidence can be strengthened* through the knowledge-intensive, concurrent exploitation of multiple data sources, eliminating false positive findings [6].
- KE-based automatic information linking enables the use of *provenance information* to annotate the generated analysis outcomes. This is important as full supporting evidence shall be explicitly available, e.g. when reporting results to regulatory organizations.

Despite the abovementioned benefits, the implementation of KE-oriented techniques entails complex challenges, both in methodological and technical terms, e.g.:

- Automatic reasoning capabilities based on the Description Logic defined semantics are one of the most prominent features of Semantic Web technologies. However, "reasoners" require significant computational resources and their efficient use in large datasets remains a challenge.
- Various reference knowledge sources (e.g. terminologies/thesauri/vocabularies) are available and can be applicable in the scope of this work. However, since these sources are constantly evolving and refined, their alignment is a complex task as it can lead to semantic inconsistencies.
- Integrating all the collected evidence under one unified knowledge model can be very challenging. This

process engages many heterogeneous data sources, which could be available via standard data exchange interface or not. For example, using proprietary EHRs to retrieve observational healthcare data would typically require specific interface implementations.

To this end, besides the ultimate goal of delivering a robust and evaluated ADR assessment platform, our mid-term goals are: (a) the design of a unifying semantic model enabling the integration of heterogeneous data sources in one information processing realm, and (b) the modelling of ADRs in one ontological model, facilitating advanced reasoning operations upon the collected information.

Conclusions

There is a clear need for comprehensive tools to support PV activities at the point of care. The proposed platform aims to support the assessment of potential ADRs in routine clinical practice, relying on the concurrent exploitation of multiple data sources for appropriate evidence. This entails the analysis of the acquired data, the aggregation of the obtained evidence, and the support of follow-up ADR monitoring over time in a systematic and user-friendly way. In this paper, we presented the main challenges and the goals from the end-user perspective for such a development, identified during the "user requirements analysis" phase of our development. We also presented the main elements of the platform design, i.e. its main information processing workflow, and illustrated the use of KE as the platform's main technical paradigm. Our work contributed to the development of a learning health system for active, post-marketing drug safety surveillance at the point of care.

Acknowledgements

This research has been co-financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneur-ship and Innovation, under the call RESEARCH – CREATE - INNOVATE (project code: T1EDK-03789).

References

[1] W.C.C. for I.D.M. World Health Organization, The importance of pharmacovigilance., World Health Organization, 2002. http://apps.who.int/medicinedocs/en/d/Js4893e/ (accessed Nov. 24, 2018).

[2] D. Formica, et al., The economic burden of preventable adverse drug reactions: a systematic review of observational studies, *Expert Opin Drug Saf.* **17** (2018) 681–95.

[3] M. Lindquist, The need for definitions in pharmacovigilance, *Drug Saf.* **30** (2007) 825–30.

[4] R.E. Ferner, and P. McGettigan, Adverse drug reactions, *BMJ*. **363** (2018) k4051.

[5] J.-L. Montastruc, et al., Benefits and strengths of the disproportionality analysis for identification of adverse drug reactions in a pharmacovigilance database., *Br J Clin. Pharmacol*. **72** (2011) 905–8.

[6] V.G. Koutkias, and M.-C. Jaulent, Computational Approaches for Pharmacovigilance Signal Detection: Toward Integrated and Semantically-Enriched Frameworks, *Drug Saf.* **38** (2015) 219–32.

[7] J.J. Gagne, J.A. et al., Active safety monitoring of new medical products using electronic healthcare data: selecting alerting rules, *Epidemiology* **23** (2012) 238–46.

[8] V.G. Koutkias, et al., Exploiting heterogeneous publicly available data sources for drug safety surveillance: computational framework and case studies, *Expert Opin Drug Saf.* **16** (2017) 113–24.

[9] L.B. Ramsey, et al., Learning Health Systems as Facilitators of Precision Medicine., *Clin Pharmacol Ther.* **101** (2017) 359–67.

[10] M. Yuksel, et al., An Interoperability Platform Enabling Reuse of Electronic Health Records for Signal Verification Studies, *Biomed Res Int.* **2016** (2016) 6741418.

[11] R. Gosh and D. Lewis, Aims and approaches of Web-RADR: a consortium ensuring reliable ADR reporting via mobile devices and new insights from social media, *Expert Opin Drug Saf.* **14** (2015) 1845–53.

[12] P. Natsiavas, et al., Comprehensive user requirements engineering methodology for secure and interoperable health data exchange, *BMC Med Inform Decis Mak.* **18** (2018) 85-101.

[13] M. Weske, *Business process management: concepts, languages, architectures*, *2nd Edition*, Springer, 2012.

[14] T.A. Kass-Hout, et al., OpenFDA: an innovative platform providing access to a wealth of FDA's publicly available data, *J Am Med Inform.* **23** (2016) 596–600.

[15] G. Schreiber, Knowledge Engineering, in: F. Van Harmelen, V. Lifschitz, and B. Porter (Eds.), *Handb Knowl Represent.*, Elsevier, 2008: pp. 929–46.

[16] T. Heath, and C. Bizer, Linked Data: Evolving the Web into a Global Data Space, *Synth Lect Semant Web Theory Technol.* **1** (2011) 1–136.

[17] N. Shadbolt, et al., The Semantic Web Revisited, IEEE Intell Syst. **21** (2006) 96–101.

[18] F. Baader, et al., Description Logics, in: *Handb Ontol.*, Springer, Berlin, Heidelberg, 2004: pp. 3–28.

[19] P. Natsiavas, et al., OpenPVSignal: Advancing Information Search, Sharing and Reuse on Pharmacovigilance Signals via FAIR Principles and Semantic Web Technologies, *Front Pharmacol.* **9** (2018) 609-24.

[20] M.D. Wilkinson, et al., The FAIR Guiding Principles for scientific data management and stewardship, *Sci Data.* **3** (2016) 160018.

[21] S. Cox, et al. Time Ontology in OWL, 2017. https://www.w3.org/TR/owl-time/ (accessed Nov 24, 2018).

[22] Y. Gil, et al., W3C PROV Model Primer, 2013. https://www.w3.org/TR/prov-primer/ (accessed Nov 24, 2018).

Address for correspondence

Pantelis Natsiavas, Institute of Applied Biosciences, Centre for Research & Technology Hellas, 6th Km. Charilaou-Thermi Road, P.O. BOX 60361 GR – 57001, Thermi, Thessaloniki, Greece, email: pnatsiavas@certh.gr