Moving towards total health data integration including quality management: insights from the SIBioC Working Group "Big Data and Artifcial Intelligence" survey

Claudia Bellini[1], Andrea Padoan[2,3], Anna Carobene[4], Roberto Guerranti[5,6]
[1]Clinical Chemistry Laboratory Analysis Unit, Misericordia Hospital Grosseto, South East Tuscany USL, Grosseto, Italy.
[2]Department of Medicine-DIMED, University of Padova, Italy.
[3]Department of Laboratory Medicine, University-Hospital of Padova, Italy.
[4]Laboratory Medicine, IRCCS San Raffaele Scientific Institute, Milano, Italy.
[5]Department of Medical Biotechnologies, University of Siena, Italy.
[6]Clinical Pathology Unit, Innovation, Experimentation and Clinical and Translational Research Department, University Hospital of Siena, Italy.

ABSTRACT

Introduction: recently the Italian Society of Clinical Biochemistry and Clinical Molecular Biology (SIBioC) Big Data and Artificial Intelligence Working Group (BAI-WG) conducted a survey to examine technological and information technology readiness for developing BAI applications and to investigate laboratory professionals' knowledge and skills. This article examines specific survey questions related to supporting data. These data are visible only by laboratory professionals for the purpose of supporting all laboratory processes and, ultimately for the validation of results and they can be generally ascribed under the quality management system (QMS).
Methods: the questionnaire, designed by the BAI-WG, was sent to 1351 SIBioC members. The responses were evaluated using Survey-Monkey software and Google Sheets.
Results: over 90% of respondents work in laboratories with a QMS in place. The participants consider digitisation of QMS as highly advantageous (93%). Nevertheless, computerisation of the QMS is actually often incomplete, and the connection between QMS and Laboratory Information Systems (LIS) is usually lacking or missing. Consequently, alternative systems, separate from the LIS, have been developed to record various QMS data essential for monitoring processes.
Discussion: the integration of medical data sources is crucial for developing BAI applications. The issue of integration is relevant and strongly linked to digitisation. However, algorithms may often consider only reported data, but integration should also be extended to supporting data, which could be correlated with clinical and process outcomes. Current LIS lack the necessary features for BAI applications and QMS digitisation is still too far behind to allow for real-time control. Software vendors should move towards the total integration of health data.

Key words: *artificial intelligence, quality assurance, health care*

INTRODUCTION

Recently, the Big Data and Artificial Intelligence Working Group (BAI-WG) of the Italian Society of Clinical Biochemistry and Clinical Molecular Biology (SIBioC) conducted a survey to study the situation of the Italian clinical laboratories regarding BAI. This survey aimed to examine both technological and information technology (IT) readiness for developing BAI applications and to investigate laboratory professionals' knowledge, skills, and interest in this rapidly growing field (1). The questionnaire received relevant feedback from SIBioC members and the results got the attention of the Clinical Chemistry and Laboratory Medicine, the official journal of the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) (1). Subsequently, a version of the study was translated into Italian and published in Biochimica Clinica, the SIBioC official journal (2). Indeed, the participants' responses uncovered several important barriers to the implementation of artificial intelligence (AI) in Laboratory Medicine (LM). These include deficiencies in the IT infrastructure, both hardware and software, suboptimal accessibility of non-laboratory health data, inadequate integration between the different data sources, and insufficient skills and knowledge of the personnel to carry out projects (1).

Correspondig Author: Claudia Bellini, Clinical Chemistry Laboratory Analysis Unit, Misericordia Hospital, Grosseto, Italy
E-mail: cbellini.md@gmail.com

Received: 31.10.2023 Revised: 11.11.2023 Accepted:04.12.2023

Published Online: 21.12.2023 DOI: 10.19186/BC_2023.093

In addition to the previously published results, the survey inspected some other specific aspects related to supporting data that only laboratory professionals have access to. By supporting data, we intend, as detailed, all auxiliary data that is intended to support all laboratory processes including the clinical validation of results; these data can generally be ascribed to the quality management system (QMS). It is well known that laboratory data do not represent just numbers in themselves, but useful information for clinical evaluation (3). In fact, the reported data is only the tip of the iceberg. This data is visible outside the laboratory, as opposed to all the data we have to deal with as laboratory specialists; i.e. the data reported in middleware, instrumental alarms and serum indexes, quality control data or medical request data, which do not appear in the laboratory report but are a necessary support for clinical validation. To better explain this concept, we can refer to the process approach. According to this approach, laboratory organisations have a primary process known as the production process, which leads to report elaboration. Additionally, there are several supporting processes including activities that interact with primary processes, enabling them to function adequately, generally pertaining to QMS (4). It follows that, transposing this logic to data, we can ideally subdivide the data that LM has to manage into "noble" data, used to answer the clinical question (generally reportable data), and "supporting" data, which are all the accompanying information, aimed at constructing and making reliable the noble data. These fundamental data, apart from being hidden from the user/clinician, are fragmented in various laboratory programmes and documents. On the other hand they are such an important part of the total testing process (TTP) that they must be managed through the QMS (Figure 1). So, beyond the emerging part of the iceberg (noble data) there is a whole submerged world of hidden, often rich, stratified, sometimes labile or transversal and highly informative data, the value of which is only known to the LM professionals (Figure 2).

In the context of BAI applications in LM, we refer to data processing. As it is well known, the integration of medical data sources is crucial for developing BAI

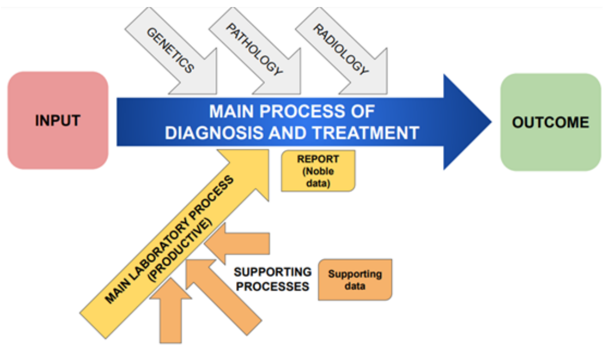

Figure 1
The diagram captures how supporting data from additional laboratory processes enhances the main flow of patient diagnosis and treatment. This supporting information works alongside the core reported (noble) data, collectively informing the outcomes and improving the overall patient care pathway (outcome).

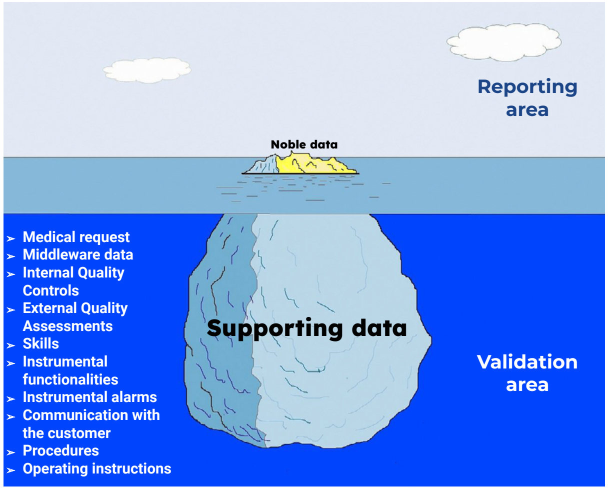

Figure 2
The Data Iceberg of Laboratory Medicine: the diagram metaphorically represents the data ecosystem in Laboratory Medicine as an iceberg. The noble data, visible above the waterline, encompasses the small portion of data included in laboratory reports. Below the surface lies a vast array of supporting data, which, although not directly reported, are crucial for the validation and integrity of the reportable data. These submerged data encompass a variety of operational, quality control, and process information that underpin the reliability of laboratory diagnostics

applications (5,6). Unfortunately, to date BAI algorithms developed on laboratory data may often take into account only the reported data. Therefore, while it is undeniable that in recent years, there has been a significant effort towards multidisciplinary and multiprofessional integration to break down 'silos' with the aim of promoting patient-centred medicine (7), from the perspective of the LM, it is necessary that the laboratory data, which have to be integrated with the other data, are qualitatively reliable (8). Clearly, since before releasing noble data, supporting data are required; along with a greater awareness of the potential of BAI applications in LM, there is an emerging drive to exploit these supporting data and to integrate them with noble data, in order to increase their informational robustness (Figure 3). Hence, the concept of integration has also to be extended to supporting data, which cannot be correlated with clinical and process outcomes if they are not integrated: with BAI analyses, a comprehensive quality assessment should be available, not limited to the review of quality controls. Furthermore, machine learning (ML) algorithms should be enriched by training on datasets including also supporting data (9).

Therefore, one of the aims of the questionnaire on the preparedness of Italian clinical laboratories for BAI applications, was to ascertain the level of integration or integrability of the supporting data with the noble data in everyday practice and their potential for exploitation in BAI applications (10). Because of the close relationship between the collection, integration and thus the analysis of these supporting data and the impact they have on overall quality management in the clinical laboratory, it was deemed appropriate to discuss the answers of the survey dedicated to this specific issue the present paper.

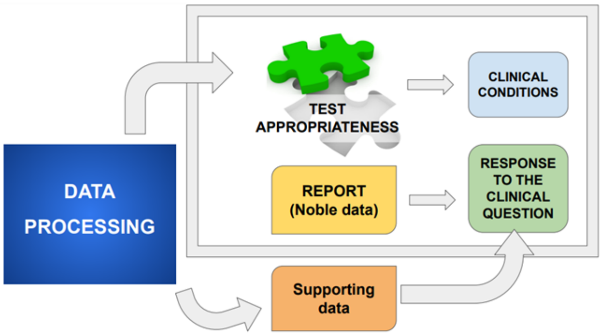

Figure 3
The figure illustrates the current paradigm where Big Data and Artificial Intelligence applications in laboratory medicine predominantly analyze noble (reported) laboratory data. The integration of supporting data into the analytical algorithms is proposed to enhance the clinical decision support system. Such enrichment is poised to refine the decision-making process not only at the individual patient level but also broadly in terms of test appropriateness, thereby optimizing clinical outcomes and resource utilization.

METHODS

The questionnaire was designed by the members of the BAI-WG and sent to 1351 SIBioC members in the modalities described elsewhere (1,2) and consisted of a corpus of 43 questions, divided into five sections. The survey questions analysed in this second paper are part of the section "Laboratory data management and analysis" and are listed in Table 1. The responses were evaluated using Survey-Monkey software and Google Sheets.

RESULTS

The questionnaire obtained relevant feedback as 227 of the 1351 SIBioC members who received this call responded to the survey, representing 47% of the 484 Italian clinical laboratories that were surveyed through SIBioC, mostly from laboratories of public hospitals. Among them, 30% were laboratory directors. The majority of respondents were biologists, followed by laboratory physicians and technologists. A number of barriers to the development of BAI in LM emerged, both infrastructural and software-based, together with a lack of competencies, despite the growing interest in the subject. The dedicated papers can be consulted for in-depth analysis and discussion (1,2).

As for the issues examined in this paper, participants were first asked which data, among those useful for the proper functioning of the clinical laboratory organisation, currently not recorded in the LIS, are most often computer-tracked (185 responses, more than one entry could be selected).

Table 1
Survey questions discussed in this paper.

Question	Options
1. Among the laboratory data not recorded in the Laboratory Information System (LIS), which are collected in computerised mode? (several items can be selected)	1. None, no other registrations are performed 2. None, records are exclusively paper-based 3. Handover at shift change 4. Internal laboratory communications 5. Communications with wards or general practitioners 6. Quality controls 7. EQA results 8. Indicators 9. Non-conformity 10. Maintenance interventions 11. Inventory management 12. Other, specify___________
2. Does the laboratory have a Quality Management System (QMS)?	1. no 2. yes, paper-based 3. yes, computerised 4. yes, paper-based and computerised With answer 1="no" or 2="yes, paper" skip question 3
3. Are the computerised data of the QMS integrated with the LIS? (integration with the LIS means both the connection between the LIS and the possible software/system managing the Quality System and the direct recording of data on the LIS)	1. no 2. yes, totally 3. yes, partially
4. How useful do you believe/would you consider the digitisation of the QMS would be?	Please enter a score from 1 to 4 (1: not useful at all; 4: completely useful)
5. How useful would you consider the management of the QMS on an internet platform accessible from all desktop and mobile devices?	Please enter a score from 1 to 4 (1: not useful at all; 4: completely useful)

In order of frequency the following were given: data relating to internal quality control and external quality assessment (EQA) programmes, followed by those relating to inventory management and the recording of non-conformities. In some cases, maintenance work, internal laboratory communications and quality system indicators are also computer-tracked. A small group also digitally registers communications with wards or general practitioners (GPs) and shift-change handovers. A small part of the respondents report that no further records are made in computerised form, but only in paper form, while in another small group of laboratories there are no records other than the data stored in the LIS (Figure 4).

In over 90% of the cases, the laboratory has a QMS (167 out of 185 answers to this question); this, in the majority of cases is only partially computerised (60%), while it is completely computerised in just over 20% of the cases and, on the contrary, in another 20% it is completely paper-based. Concerning the integration between computerised QMS data, including registrations, and LIS data, only 11% of the respondents benefit from a total integration, another 32% report a partial integration, while in 57% of the cases the QMS data are not integrated at all with the LIS data (76 out of 134 responses). Almost all respondents believe that digitisation (92%) of the QMS (169 out of 183 responses to this question) and the use of a web platform accessible from all desktop and mobile devices (93%) for this purpose (171 out of 183 responses) could be very beneficial.

DISCUSSION

In the previous papers, reporting the results of the survey conducted in the Italian clinical laboratories, a number of barriers to the development of BAI in LM were reported, including inadequate infrastructures, insufficient speed and volume of data extraction from LIS, and inadequate integration between data sources (1,2). Together with the lack of specific education, which was found also by Paranjape et al. in a survey conducted among stakeholders in laboratory medicine (11), these barriers pose challenges to AI implementation in the clinical laboratories. Furthermore, the issue of integration is relevant and strongly linked to digitisation; in particular, it is necessary to take into account the fact that in the microcosm of the LM there are innumerable unreported data that have a valuable information potential.

This part of the survey, exploring the state of the art of QMS in Italian laboratories, offers relevant findings to be considered, especially for planning future implementation of laboratory software. Firstly, the great majority of respondents work in laboratories with a QMS; this fact reflects the increasing need of digital tools for the management of supporting data. Nevertheless, a computerised QMS is lacking in 60% of respondents. Consequently, many Italian laboratories have had to develop alternative systems, separate from the LIS, to collect and record various QMS data essential for monitoring laboratory processes.

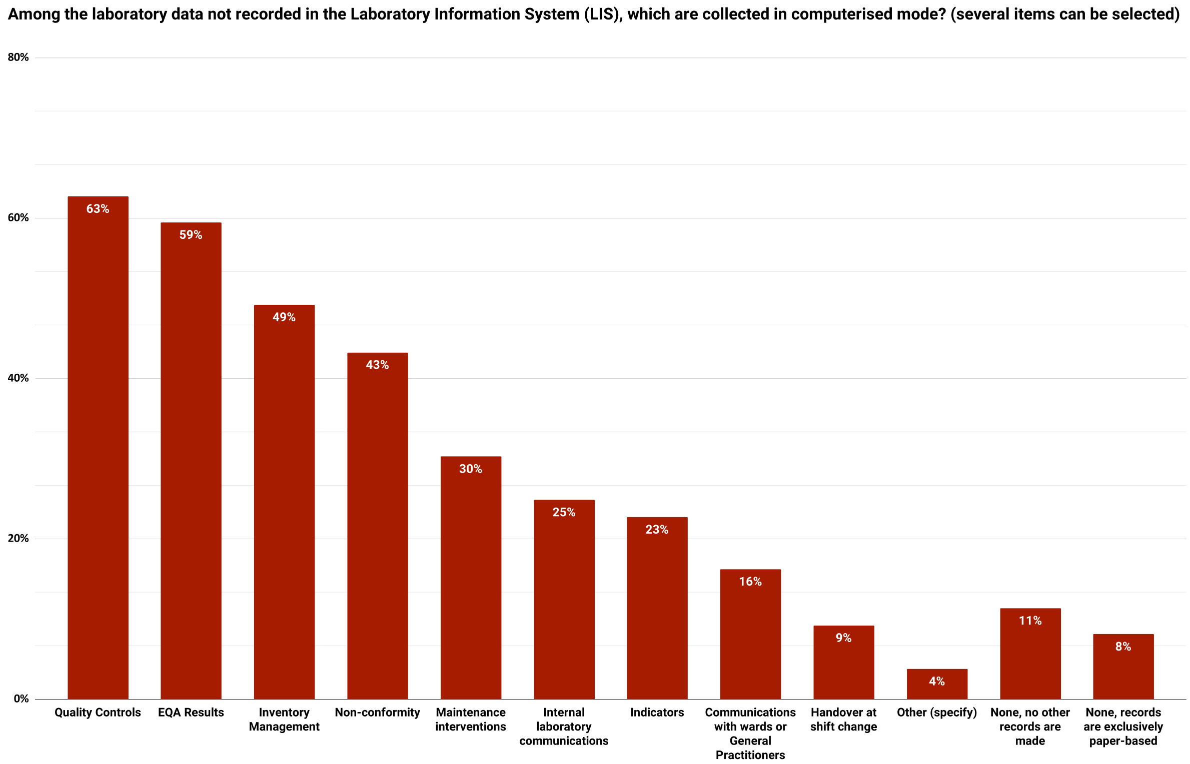

Figure 4
The figure presents the proportion of laboratory data that is computerized but not captured by the Laboratory Information System (LIS), as reported by survey respondents. It highlights the varying degrees of digital record-keeping for different categories of laboratory data that remain outside the competence of the LIS.

Incidentally, it should be noted that, unlike the current LIS, there is a lack of software developed specifically for laboratory QMS. Therefore, even if the QMS is computerised, most applications are designed to meet the requirements of ISO 9001 without taking into account the specific needs of the laboratory.

The participants demonstrate a positive attitude towards the digitisation of QMS, as they perceive the potential for it to be placed on a modern web platform accessible from both fixed and mobile devices as highly advantageous. Nevertheless, the present situation reveals that the computerisation of the QMS is often incomplete, and the linkage between the QMS and the LIS is usually absent. Thus, QMS is separated from LIS, both logically (different functions) and digitally (the various QMS software and databases, when there are any, are different and separate from the LIS). This piece of information deserves special consideration, as it underlines a lack of integration between the main data produced by the laboratory (patients' results) and the supporting data, and it could be of great concern for data generalizability and usability, as well as for reducing costs.

Moving from digitisation to a real-time QMS

While the need to keep the specimen and report flow under control has been driving towards digitisation for a long time now, and LIS have been present in all laboratories for decades, we have seen a delay in the computerisation of QMS. For many years 'computerising' QMS has primarily meant scanning documents and collecting them in folders, possibly shared in a company repository or, more recently, in a cloud system. This certainly makes it easier to keep track of the often-extensive documentation, which constitutes the 'static' part of the quality system. Additionally, digitisation allows documents to be stored in a single, traceable original copy, to be easily retrieved and disseminated to staff. However, it is even more evident that computerisation and digitisation should no longer be regarded as a mere digitisation of documents. Indeed, digitisation of documents is a process mimicking the paper registers, while electronic QMS databases present several advantages, resulting for instance from the collection of structured data, which is vital in the BIA universe. Today, modern database systems offer real-time data recording and analysis, facilitating immediate visualisation of outputs in dashboard form (12). These data processing methodologies should overcome the difficulties of maintaining the system under control on paper-based records or alternatively, made by systems that do not allow for immediate data analysis, with the consequence that monitoring does not actually take place in real-time, unlike what happens on a daily basis with internal quality controls. Thus, monitoring indicators are obtained with a delay with respect to the occurrence of a problem and, consequently, the overall process control cannot take place in real-time. The additional value of modern databases is really understood by laboratory personnel, since 92% of them confirmed their positive feedback in digitised QMS.

As previously hypothesised by some authors (13), in order to have a truly integrated management with the BAI applications, the LIS should be the sole software collector of all noble and supporting data generated in the laboratory.

Many of these supporting data, appropriately used, would also allow benchmarking between different laboratories, e.g. through quality indicators (14-19); in contrast, the participation of laboratories in data sharing was low (20). Possible causes include difficulties with manual, rather than computerised, data collection, since integration between LIS and QMS is usually lacking (17,21).

Why current LIS are inadequate for BAI applications

Theoretically, LIS is not changed more often than 10-20 years because this involves a huge effort (and cost) for the laboratory (22), but with the disruptive evolution of technologies, this concept should be definitively revisited if we want to benefit our laboratories: working with outdated software is a limitation to collecting, analysing and managing the quality data needed to meet the requirements of ISO 9001 and 15189 (23-24). Moreover, in order to develop AI algorithms that can support in improving diagnosis, assessing risk and prognosis, or identifying disease subtypes or personalising therapy, it is first of all necessary to have large integrated multidisciplinary datasets and tools that are able to extract even non-numerical content (25). In this, LM has greatly increased its potential given the enormous amount of data collected on a daily basis, which in the past would have been lost, and it would have even more potential if LIS were up to date with the current need to have datasets in accordance with the five Vs that characterise Big Data (volume, value, variety, velocity, and veracity) (26).

Several years ago, for example, it was described as a feature of the 'ideal LIS' that it could interface with instrument performance data, temperature monitoring systems, water quality parameters, environmental measurements and other data relevant to good laboratory practice and necessary for the periodic documentation required by certification and accreditation bodies (13). The authors further hypothesised that the LIS could manage EQA programmes, from inventory control of materials for verification testing, to documentation of reports and investigation of possible errors, with the possibility of online review and certification of results by relevant management personnel. Ideally, interfacing with external providers of EQA programmes should enable seamless data transmission (13). Nowadays, there has been progress towards computerisation as the data can be found on the providers' websites, but the problem of integration with LIS data and internal quality control data remains. In essence, the LIS should manage accreditation requirements online, including the preparation of appropriate documents, capture and manipulate all required data, with links to quality policies, procedures and other relevant electronic documents as proof of compliance.

In addition the life cycle of the supporting data often does not correspond to that of the noble data, since, being on separate platforms, only the noble data have been safeguarded and archived in the long term over time, while the middleware data archiving sometimes tends to undergo losses, e.g. when changing technology or after the saturation of memory archives. On the other hand, any data currently stored in the QMS software or not transferred from the middleware to the LIS should be gathered by the LIS, serving as a unified repository of laboratory data, as described in the next section, and managed with ML techniques to improve diagnostic algorithms.

Software integration enables real-time gathering and utilisation of data

If our previous papers pointed out a disparity between the current and desirable state of software, hardware, data warehousing, knowledge and skills as a requirement for BAI development (1,2), the integration and use of supporting data can be found in an even more critical situation. To achieve integration in LM, we need to link and combine the information within a single laboratory software, including the LIS, middleware, instrument management software and QMS software. BAI applications will be called upon to manage the laboratory's QMS data and integrate it with the noble data, just as they are required for all healthcare data now. Hence LM must aim towards the development of LIS systems that allow full integration and represent direct quality management tools of the TTP (27). These systems should provide data in a format that can be analysed in real-time by ML technologies to perform real-time process control (28). Furthermore, LIS improvement must serve not only to ensure data quality for algorithms' development, but also to provide those large datasets in which to externally validate these algorithms and prove that they are not biased, thus enabling them to become supportive of medical practice (5,6). In the future, models will need to be trained with large, high-quality datasets; this is a prerequisite for AI applications to provide reliable and relevant information to address clinical utility criteria in improving patient care (29).

CONCLUSIONS

This study highlights an important gap between the requirements for the development of BAI in TTP and the current laboratory situation. Consequently, it represents a first starting point for further investigations and for suggesting to software vendors to move towards the total integration of health data, in order to lay the foundations for a real application of BAI methodologies in the future of LM.

CONFLICT OF INTEREST

None

REFERENCES

1. Bellini C, Padoan A, Carobene A, Guerranti R. A survey on Artificial Intelligence and Big Data utilisation in Italian clinical laboratories. Clin Chem Lab Med 2022;60:2017-26.
2. Bellini C, Padoan A, Carobene A, Guerranti R,. per il Gruppo di Studio SIBioC Big Data e Intelligenza Artificiale. A survey on Artificial Intelligence and Big Data utilisation in Italian Clinical Laboratories. Biochim Clin 2023;47:56-64.
3. Plebani M. Future of laboratory medicine: a Manifesto for laboratory professionals. Biochim Clin 2020;44:263-9.
4. Zima T. Accreditation of Medical Laboratories - system, process, benefits for labs. J Med Biochem 2017;36:231-7.
5. Carobene A, Cabitza F, Bernardini S, Gopalan R, Lennerz JK, Weir C, et al. Where is laboratory medicine headed in the next decade? Partnership model for efficient integration and adoption of artificial intelligence into medical laboratories. Clin Chem Lab Med 2023;61:535-43.
6. Lennerz JK, Salgado R, Kim GE, Sirintrapun SJ, Thierauf JC, Singh A, et al. Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML. Clin Chem Lab Med 2023;61:544-57.
7. Lippi G, Plebani M. Integrated diagnostics: the future of laboratory medicine? Biochem Med (Zagreb) 2020;30:010501.
8. Cabitza F, Campagner A, Balsano C. Bridging the "last mile" gap between AI implementation and operation: "data awareness" that matters. Ann Transl Med 2020;8:501.
9. Carobene A, Milella F, Famiglini L, Cabitza F. How is test laboratory data used and characterised by machine learning models? A systematic review of diagnostic and prognostic models developed for COVID-19 patients using only laboratory data. Clin Chem Lab Med 2022;60:1887-901.
10. Padoan A, Plebani M. Flowing through laboratory clinical data: the role of artificial intelligence and big data. Clin Chem Lab Med 2022;60:1875-80.
11. Paranjape K, Schinkel M, Hammer RD, Schouten B, Nannan Panday RS, Elbers PWG, et al. The Value of Artificial Intelligence in Laboratory Medicine. Am J Clin Pathol 2021;155:823-31.
12. Yadav CP, Sharma A. National Institute of Malaria Research-Malaria Dashboard (NIMR-MDB): A digital platform for analysis and visualization of epidemiological data. Lancet Reg Health Southeast Asia 2022;5:100030.
13. Sepulveda JL, Young DS. The ideal laboratory information system. Arch Pathol Lab Med 2013;137:1129-40.
14. Sciacovelli L, Plebani M. The IFCC working group on laboratory errors and patient safety. Clin Chim Acta 2009;404:79-85.
15. Plebani M. Quality indicators to detect pre-analytical errors in laboratory testing. Clin Biochem Rev 2012;33:85-8.
16. Plebani M, Sciacovelli L, Marinova M, Marcuccitti J, Chiozza ML. Quality Indicators in laboratory medicine: a fundamental tool for quality and patient safety. Clin Biochem 2013;46:1170-4.
17. Plebani M, Astion ML, Barth JH, Chen W, de Oliveira Galoro, CA, Escuer MI, et al. Harmonization of quality indicators in laboratory medicine. A preliminary consensus. Clin Chem Lab Med 2014;52:951-8.
18. Sciacovelli L, Lippi G, Sumarac Z, West J, Garcia Del Pino Castro I, Furtado Vieira K, et al. Working Group "Laboratory Errors and Patient Safety" of International Federation of Clinical Chemistry and Laboratory Medicine (IFCC). Quality indicators in laboratory medicine: the status of the progress of IFCC working group "laboratory errors and patient safety" project. Clin Chem Lab Med 2017;55:348-57.
19. Sciacovelli L, Padoan A, Aita A, Basso D, Plebani M. Quality indicators in laboratory medicine: state-of-the-art, quality specifications and future strategies. Clin Chem Lab Med 2023;61:688-95.

20. Plebani M. The quality indicator paradox. Clin Chem Lab Med 2016;54:1119-22.
21. AlHazme RH, Haque SS, Wiggin H, Rana AM. The impact of health information technologies on quality improvement methodologies' efficiency, throughput and financial outcomes: a retrospective observational study. BMC Med Inform Decis Mak 2016;16:154.
22. Lukić V. Laboratory Information System - Where are we Today? J Med Biochem 2017;36:220-4.
23. ISO 9001:2015 Quality management systems — Requirements. Geneva 2015. 24. ISO 15189:2022 Medical laboratories — Requirements for quality and competence. Geneva 2022.
25. Galozzi P, Basso D, Plebani M, Padoan A.. Artificial intelligence and laboratory data in rheumatic diseases. Clin Chim Acta 2023;546:117388.
26. Gupta S, Kar AK, Baabdullah A, Al-Khowaiter WAA. Big Data with cognitive computing: A review for the future. Int J of Inform Manag 2018;42:78-89.
27. Aita A, Sciacovelli L, Plebani M. Extra-analytical quality indicators - where to now? Clin Chem Lab Med 2018;57:127-33.
28. Zhou R, Wang W, Padoan A, Wang Z, Feng X, Han Z, et al. Traceable machine learning real-time quality control based on patient data. Clin Chem Lab Med 2022;60:1998-2004.
29. Cadamuro J, Cabitza F, Debeljak Z, De Bruyne S, Frans G, Perez SM, et al. Potentials and pitfalls of ChatGPT and natural-language artificial intelligence models for the understanding of laboratory medicine test results. An assessment by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Working Group on Artificial Intelligence (WG-AI). Clin Chem Lab Med 2023;61:1158-66.