

Opinion Paper

Jochen K. Lennerz*, Roberto Salgado, Grace E. Kim, Sahussapont Joseph Sirintrapun, Julia C. Thierauf, Ankit Singh, Iciar Indave, Adam Bard, Stephanie E. Weissinger, Yael K. Heher, Monica E. de Baca, Ian A. Cree, Shannon Bennett, Anna Carobene, Tomris Ozben and Lauren L. Ritterhouse

Diagnostic quality model (DQM): an integrated framework for the assessment of diagnostic quality when using AI/ML

<https://doi.org/10.1515/cclm-2022-1151>

Received November 11, 2022; accepted January 13, 2023;

published online January 25, 2023

Abstract

Background: Laboratory medicine has reached the era where promises of artificial intelligence and machine learning (AI/ML) seem palpable. Currently, the primary responsibility for risk-benefit assessment in clinical practice resides with the medical director. Unfortunately, there is no tool or concept that enables diagnostic quality assessment for the various potential AI/ML applications. Specifically, we noted that an operational definition of laboratory diagnostic quality – for the specific purpose of assessing AI/ML improvements – is currently missing.

Methods: A session at the 3rd Strategic Conference of the European Federation of Laboratory Medicine in 2022 on “*AI in the Laboratory of the Future*” prompted an expert roundtable discussion. Here we present a conceptual diagnostic quality framework for the specific purpose of assessing AI/ML implementations.

Results: The presented framework is termed diagnostic quality model (DQM) and distinguishes AI/ML improvements at the test, procedure, laboratory, or healthcare ecosystem level. The operational definition illustrates the nested relationship among these levels. The model can help to define relevant objectives for implementation and how levels come together to form coherent diagnostics. The affected levels are referred to as scope and we provide a rubric to quantify AI/ML improvements while complying with existing, mandated regulatory standards. We present 4 relevant clinical scenarios including multi-modal diagnostics and compare the model to existing quality management systems.

Conclusions: A *diagnostic quality model* is essential to navigate the complexities of clinical AI/ML implementations. The presented diagnostic quality framework can help to specify and communicate the key implications of AI/ML solutions in laboratory diagnostics.

Keywords: artificial intelligence; biomarker; laboratory medicine; machine learning; regulatory science.

***Corresponding author: Jochen K. Lennerz**, MD, PhD, Medical Director, Center for Integrated Diagnostics, Department of Pathology, Massachusetts General Hospital/Harvard Medical School, 55 Fruit Street, Boston, MA 02114, USA, Phone: 617-643-0619, Fax: 617-726-2365, E-mail: JLennerz@partners.org

Roberto Salgado, Department of Pathology, GZA-ZNA Hospitals, Antwerp, Belgium; and Division of Research, Peter Mac Callum Cancer Centre, Melbourne, Australia

Grace E. Kim, Department of Pathology, University of California San Francisco, San Francisco, CA, USA

Sahussapont Joseph Sirintrapun, Pathology Informatics, Memorial Sloan Kettering Cancer Center, NY, USA

Julia C. Thierauf, Department of Pathology, Massachusetts General Hospital/Harvard Medical, Boston, MA, USA; and Department of Otorhinolaryngology, Head and Neck Surgery, German Cancer Research Center (DKFZ), Heidelberg University Hospital and Research Group Molecular Mechanisms of Head and Neck Tumors, Heidelberg, Germany

Ankit Singh, Adam Bard, Yael K. Heher and Lauren L. Ritterhouse, Department of Pathology, Massachusetts General Hospital/Harvard Medical, Boston, MA, USA

Iciar Indave, European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), Lisbon, Portugal

Stephanie E. Weissinger, Alb Fils Clinics GmbH, Institute of Pathology, Göttingen, Germany

Monica E. de Baca, Pacific Pathology Partners, Seattle, WA, USA

Ian A. Cree, International Agency for Research on Cancer (IARC), World Health Organization, Lyon, France

Shannon Bennett, Department of Laboratory Medicine and Pathology (DLMP), Mayo Clinic, Rochester, MN, USA

Anna Carobene, IRCCS San Raffaele Scientific Institute, Milan, Italy

Tomris Ozben, Medical Faculty, Dept. of Clinical Biochemistry, Akdeniz University, Antalya, Türkiye; and Medical Faculty, Clinical and Experimental Medicine, Ph.D. Program, University of Modena and Reggio Emilia, Modena, Italy

Introduction

Human diseases pose challenges that often exceed human capabilities [1]. Personalized medicine is increasingly required for modern laboratory medicine and tissue-based diagnostics, resulting in an increasing amount of health data [2, 3]. When these data are of high quality and presented in the right format, some of the medical challenges can be solved using the computational power of artificial intelligence (AI) and machine learning (ML) solutions [3–7]. Simply put, AI/ML has the potential to revolutionize and optimize laboratory diagnostics [2–7].

In the laboratory, there are two key considerations when implementing AI/ML: clinical utility [5, 8–11] and risk management [12, 13]. While numerous use cases have been explored [14, 15], the euphoria around AI/ML solutions hinges upon quality improvements and unequivocal demonstration of clinical utility [8–11, 16, 17]. It is noteworthy that, at least for now, AI in the clinical laboratory has limited applications when compared to other medical specialties (e.g., radiology) [3, 18–22] and clinical utility definitions vary across settings and implementations [23–29]. Excellent performance metrics alone are no longer sufficient [30]. AI/ML realization now also requires explainability, risk of bias tools, and performance monitoring [2, 18, 31, 32]. Furthermore, algorithms require auditing (and audit management), compliance management, corrective and preventive actions, error tracking, document control, and new approaches to organize competency and proficiency testing [33], and tackle operational challenges, and privacy concerns [18, 23, 34–38]. In addition, we recognized three key hurdles: (1) implementations typically focus only on basic performance metrics (e.g., specificity, accuracy); (2) validations are usually focused on a single context (e.g., local hospital), and (3) meaningful integration relies heavily on the awareness of the complexity and domain knowledge in the laboratory.

The delivery system is characterized by two elements: (1) AI/ML developers who face the complexity of clinical laboratories and their regulatory challenges [28, 39] whereas (2) laboratory directors face the complexity of integrating high complexity solution into the wet-lab and information technology (IT) systems [2, 5]. AI/ML solutions work best when high quality data is continuously available (for improvements); however, laboratory IT-processes are not (yet) interoperable and ready-made guidance documents are not available [5, 28, 32, 36, 40]. From a patient, care-team, and quality governance perspective, these computational solutions need to be scrutinized before implementation, just like any new discovery [37, 39, 41–43]. Even with regulatory approval [39, 44], implementation requires integration of new systems or workflows – and modification of existing IT

systems still carries risks [38, 45]. In other words, laboratories are facing a major implementation problem: How to effectively implement AI/ML into clinical practice?

As outlined in Figure 1, integration of AI/ML solutions into clinical practice is a multidimensional task that relies heavily on the competence of laboratory personnel and in particular the laboratory directors [36, 46–49]. Communicating the specific scope, function, and outcomes relies on many factors (Supplementary Table 1). Specifically, conveying the value proposition is particularly challenging if the key decision-makers and stakeholders do not have a laboratory background. There are several large-scale initiatives underway to tackle these challenges from various aspects. For example, in the USA additional FDA oversight or even CLIA modernization has been proposed [50]. Regulatory agencies have created programs, centers, and initiatives, and are actively engaged in improving approval pathways [51]. Regulatory science initiatives [52], industry representation [53], professional societies, and patient advocacy are also actively working on pathways for AI/ML to enter the clinical laboratory [41, 52, 53]. It is important that AI integration should be defined collaboratively – merging experiences to overcome unintended consequences. It is imperative to include laboratory professionals and their domain expertise when, for example defining algorithms that integrate laboratory parameters. Unfortunately, there is currently no tool or concept that enables conceptualization of diagnostic quality assessments for the various potential AI/ML applications. Until now, an operational definition of laboratory diagnostic quality – for the specific purpose of assessing AI/ML improvements – has been missing.



Figure 1: Implementation of AI in the clinical laboratory. AI integration is not just a new assay. In contrast, when integrating a new AI/ML algorithm into patient care, laboratories face a multidimensional problem. Additional details provided in Supplementary Table 1. AI, artificial intelligence; for simplicity, both AI and ML are used synonymously.

Here, we present a diagnostic quality model that can serve as an operational framework for the integration of AI/ML into clinical practice. Communicating and assessing AI/ML-related diagnostic quality improvements have emerged as a novel responsibility of medical directors. Defining relevant AI/ML objectives and mapping out the scope of a specific solution must be integrated with existing, mandatory laboratory regulatory standards to provide a coherent solution. The presented diagnostic quality model provides a starting point to optimize integration of AI/ML into clinical patient care.

Methods

Design, aims, and starting points

The main aim of this project was to derive an operational definition of diagnostic quality that can be applied to laboratory-based diagnostic testing before and after a given AI/ML solution has been implemented. We convened a roundtable discussion with the following main aims:

- derive an initial version of a checklist that can assist clinical laboratory directors;
- review the proposed conceptual model;
- discuss strengths and weaknesses; and
- outline examples on how the model can be applied.

As a starting point, we used a presentation given on May 25th, 2022, at the 3rd Strategic Conference of the European Federation of Laboratory Medicine (EFLM) in the session entitled: “AI in the Laboratory of the Future” [18]. The presentation by one of the authors (JKL) included an earlier version of the diagnostic quality model. Briefly, the framework was derived from a combination of the relevant CLIA standards [54], definitions from the FDA [55], selected professional societies [56–59], and prior publications [18, 60, 61].

Relevant definitions

Implementation science is the systematic study of methods that support the application of research findings and other evidence-based knowledge into policy and (clinical) practice [62–64]. We focused on the translation of computer science tools (i.e., AI/ML models) and their integration into routine clinical patient care in the laboratory. Despite various definitions, here, the terms *ML/AI* are considered synonymous and refer to models or systems that, when provided with appropriate input, can take actions to achieve a goal and/or show performance improvements over time. An overview of the model is provided in Figure 2.

Diagnostic test/assay: We defined *diagnostic test* as the specific technique (or collection of techniques) to analyze a biological marker used to detect, diagnose, or monitor a disease. A diagnostic test typically contains one (or more) technologies, follows one (or multiple) operating principles, and employs a specific method of analysis (e.g., PCR) [65]. The diagnostic test is the key component(s) of the testing process. It converts the biomarker into data (or vice versa). To avoid confusion with the clinical jargon “diagnostic testing” – the term *assay* and *test* can be used

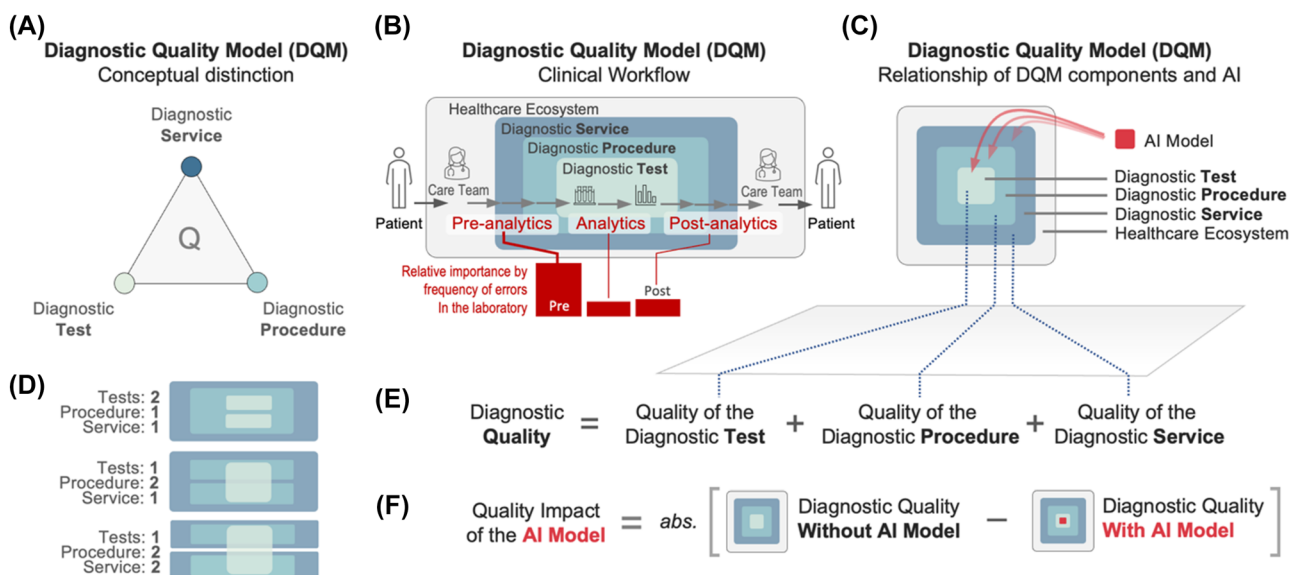


Figure 2: Diagnostic quality model (DQM) overview. (A) The model distinguishes diagnostic tests, diagnostic procedures, and diagnostic services. (B) The diagnostic test forms the innermost layer of the DQM model and is concerned with the analyte and the modality of detection (first layer, diagnostic test layer). The diagnostic test layer is part of a specific set of operations collectively referred to as diagnostic procedures (second layer). Each laboratory typically has numerous diagnostic procedures. The diagnostic procedures interface with the external health care delivery system (third layer, diagnostic service layer). (C) Clinical integration of an AI model requires careful consideration of the relationship of the model to the diagnostic layers. (D) Examples of test/procedure/service combinations (top: two specific PCR tests from the same nucleic acid extraction procedure; middle: same PCR from two different nucleic acids; bottom: specific in- and outpatient services and procedures use the same test). (E) The diagnostic quality can be seen as the combination of the diagnostic quality of the diagnostic test, procedure, and service. (F) Considering the deployment of AI models, the diagnostic quality impact of an AI model can be expressed as the absolute (abs.) difference between the quality with or without the AI model. For simplicity of the conceptualization, the AI model is exemplarily depicted in the diagnostic test layer; however, AI models can be implemented in other and/or multiple layers (e.g., the service layer; see Figure 3). AI, artificial intelligence; ML, machine learning; PCR, polymerase chain reaction.

synonymously for test [66]. For example, a clinician orders a diagnostic test that entails SARS-CoV-2 specific PCR as the key assay (*diagnostic test* component).

Diagnostic procedure: We defined *diagnostic procedure* as the relevant laboratory-specific pre- and post-analytical aspects relevant to a given *diagnostic test*. A diagnostic procedure entails the laboratory specific standard operating procedure(s), relevant validation reports, and management principles applied to this test (e.g., test- and procedure specific competency assessment and proficiency testing). Note that certain components of diagnostic procedures can be shared among multiple diagnostic tests (e.g., nucleic acid extraction for subsequent testing by PCR or Next-Generation Sequencing (NGS)).

Diagnostic service: We defined diagnostic service as the outward facing support services related to one or multiple diagnostic tests/procedures. Diagnostic services entail inclusion of the specific test into the laboratory's menu offerings (on- and offboarding), third party review, contracting, payor, regulatory aspects, and service lines or overarching test-menu offering (e.g., in- and out of network offering). The diagnostic service of a laboratory also entails the user experience (e.g., by the care team, patient, or other laboratory users). The diagnostic service interfaces with the *healthcare ecosystem*, defined here as the larger-scale healthcare delivery system (e.g., hospital network, network of referring institutions, etc.). It is important to emphasize that the final report represents the main output of the laboratory (external facing value). The results (core content) thereby refer to the test and procedural layer; however, the formatting, composition, and data model is core component of the diagnostic service layer.

Diagnostic quality model: We refer to *diagnostic quality model* as the concept of an inter-related, nested relationship between *diagnostic test*, *diagnostic procedure*, and *diagnostic service* where each aspect contributes to the overarching diagnostic quality in the surrounding healthcare ecosystem. The diagnostic quality model is a conceptual model that can be used to specify the scope of an AI/ML intervention and measure diagnostic quality changes.

Scope: We use the term *scope* to specify the extent of an intervention or process improvement; for example, whether a proposed AI/ML tool is primarily affecting the diagnostic test performance (or the diagnostic procedure, diagnostic service, or a combination thereof).

Diagnostic quality: In general terms we follow the established standard that *diagnostic quality* refers to how well the staff and practitioners perform tasks related to the test, procedures, and services including accuracy, completeness, timeliness, and pertinence to the specific task(s) [67, 68]. However, in the specific context of the diagnostic quality model, we defined *diagnostic quality* as the sum of the quality of the diagnostic test(s), the diagnostic procedure(s), and the diagnostic service(s). Diagnostic quality improvement was defined as the difference between the diagnostic quality before and after an intervention (e.g., implementation of an AI/ML tool).

Multimodality workflow: These workflows entail multiple diagnostic modalities. Importantly, AI/ML can help mitigate the increasing complexity of data management and interpretation across test modalities. Multimodal learning involves relating information from multiple sources [69, 70]. We refer to *multimodal prediction tools* as AI/ML models that use data from one modality to predict the results obtained using another downstream modality. The definition of multimodal models is evolving and can entail models that learn as one big task (so-called end-to-end models) or models that require feature extraction followed by learning (so-called decapsulated extra-step multimodal models).

Workflow examples

We chose four clinical applications as examples in which our model could be applied: First, computer tomographic (CT) data (modality 1) predicting the mutational status obtained using molecular diagnostic testing of tissue samples (modality 2) for triggering an insurance test claim via specialty pharmacy (Figure 3) [71–74]. Second, a diagnostic AI-aid for Gleason grading [75, 76] as applied to the multi-biomarker diagnostic workflow of prostate cancer diagnostics (Figure 4) [77, 78]. Third, we constructed a workflow for recent companion diagnostics for abemaciclib (biomarker Ki-67) [79–81] and trastuzumab deruxtecan (biomarker HER2) (Figure 4) [82, 83]. Finally, a clinical decision support system for SARS-CoV-2/COVID prediction from demographic and laboratory data [49, 84–87].

Results

Checklist for diagnostic quality when using AI/ML

The simplest way to define diagnostic quality is through the absence of diagnostic errors or undesirable diagnostic events [88]. The National Academy of Medicine defines diagnostic error as the failure to (a) establish an accurate and timely explanation of the patient's health problem(s) or (b) communicate that explanation to the care team and patient. Simply put, these are diagnoses that are delayed, wrong or missed altogether. Diagnostic errors occur in all settings of care, contribute to ~10% of patient deaths and are the most frequent reason for medical liability claims. It is no surprise that entire branches of the executive side of the government actively pursue risk and safety assessments (FDA), govern clinical practice (CMS), and focus on investments in research to improve diagnostic safety and the reduction in diagnostic error (ARHQ). Key challenges and areas for potential future research have been published [89, 90].

Diagnostic quality is however more than the absence of errors, missed, or incorrect diagnoses [48, 49, 66, 68, 91, 92]. Numerous authors have tackled the concept of diagnostic quality from various angles [67, 93–97]. Notable examples are synoptic reporting [98, 99], harmonization efforts [52, 100], as well as proven approaches for evidence-based disease classification [100]. This compiled preliminary checklist provides an overview of relevant aspects that can influence the diagnostic quality (Table 1). When reviewing the checklist, it becomes clear that integration of AI/ML – or any new technology for that matter – does not solely rely on the specific performance metrics of the algorithm(s). Specifically, the overall integration, function, and objectives are equally relevant. Furthermore, the integration requires careful consideration

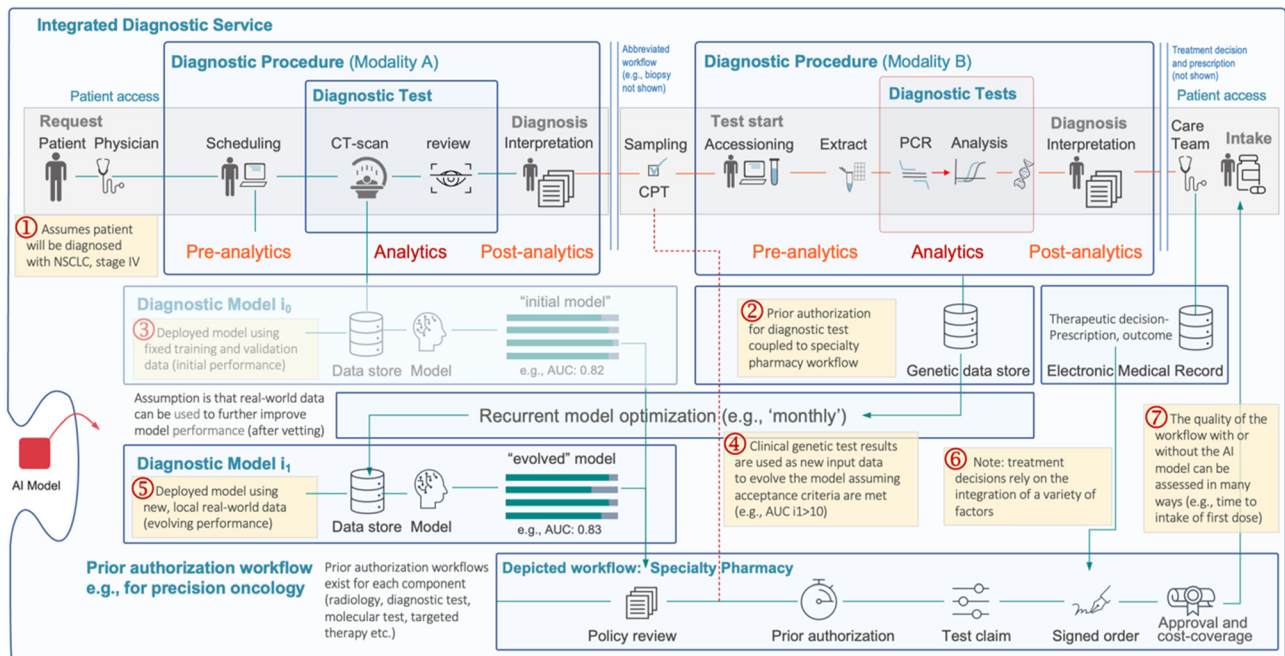


Figure 3: Multimodality workflow. The main aim is to apply the diagnostic quality model to the complexity of a multi-modality learning model. Depicted is an integrated diagnostic service where an AI algorithm predicts molecular testing results based on imaging findings on CT to trigger an administrative process (i.e., test claim). The AI/ML model is implemented in the service layer. Workflow: (1) the target population is exposed to modality A (note the three layers service, procedure, and test). (2) The model result is used to trigger prior authorization for molecular diagnostic testing and prior authorizations via specialty pharmacy workflows. (3) The deployed model with an initial performance is running on the dataset from modality A. (4) The molecular diagnostic testing results represent modality B that serves as continuous input for the AI model running on modality A (note the three layers service, procedure, and test). (5) The deployed model undergoes recurrent model optimization and is used to trigger test claims via specialty pharmacy (administrative process = service layer). (6) The signed orders (e.g., diagnostic or treatment decisions) rely on multiple factors where the AI-triggered test claim and coverage streamlining represent one component. (7) Diagnostic quality assessment across both modalities for a primary outcome measure of choice. Note how each layer of the diagnostic quality model (for each modality) contributes to the overall functionality of the multi-modal workflow. AUC, area under the curve (representative performance measure); CT, computerized tomography; CPT, current procedural terminology; NSCLC, non-small cell lung cancer; PCR, polymerase chain reaction.

of all aspects affecting the specific diagnostic aim and related safety in the laboratory. At first glance, some aspects appear deceptively simple; however, taken together, nearly everyone will recognize the involved complexity (Table 1). Based on these considerations, we observed that an operational definition of diagnostic quality for the specific purpose of assessing AI/ML improvements is currently missing.

Towards an operational definition of diagnostic quality

An operational definition is “ready for use” and specifies concrete, replicable procedures designed to represent a construct that complies with the desired function and routine activities. Furthermore, operational definitions are typically not laboratory specific. We therefore consider an operational definition for diagnostic quality useful because it provides a framework that is flexible enough for local

adoption to be permissible. From the perspective of a laboratory director considering integration of AI/ML, the risk and benefit assessment must align with existing laboratory-specific test, procedural, and service lines. The first element of integration is to distinguish an AI/ML model that primarily affect the test from those that affect the procedure or service-level (Figure 2A). We propose a nested model of diagnostic test quality (Figure 2B) that can be applied when implementing AI/ML models (Figure 2C). The nesting of levels is critically important because each level contributes unique yet valuable combinations of components. From an innovation perspective, the sum of tests, procedures, and services can be regarded as an asset and resource because each existing level can serve as a blueprint for new implementations. In contrast, from a healthcare ecosystem perspective, the service layer is the primary interface zone.

An important consideration is that the model accounts for existing synergies in the laboratory (Figure 2D). Multiple tests use the same (or similar) diagnostic procedures (e.g.,

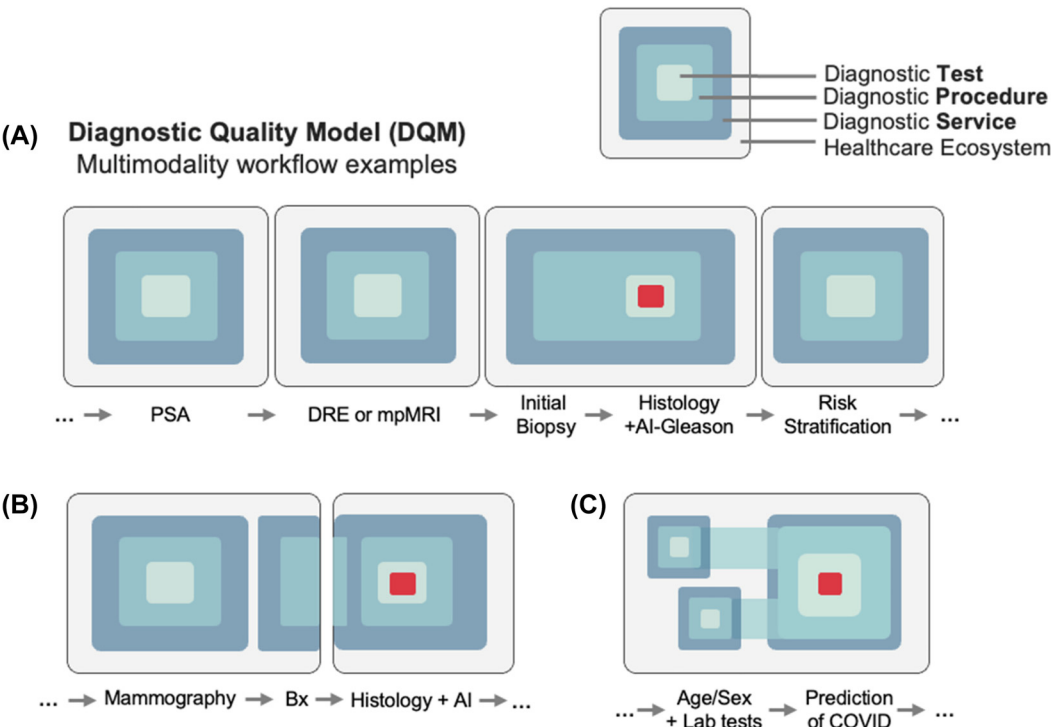


Figure 4: Multimodality workflow examples. (A) Diagnostic care continuum of prostate cancer patients depicted in four different healthcare settings. The depicted AI tool can help with Gleason grading (diagnostic as opposed to predictive biomarker). (B) Diagnostic breast cancer workflow from mammography to biopsy and histologic diagnosis (depicted in two healthcare settings). The depicted AI tool can assist in quantifying the receptor status (diagnostic biomarker). (C) Integration of multiple demographic and laboratory findings into an AI-based prediction tool for SARS-CoV-2/COVID status and prognosis (depicted in one healthcare setting). The AI tool serves as the integrator and decision support tool (predictive biomarker). AI, artificial intelligence tool; Bx, biopsy; COVID, corona virus disease; DRE, digital rectal examination; lab, laboratory; mpMRI, multiparametric magnetic resonance imaging; PSA, prostate specific antigen.

Table 1: Checklist of diagnostic quality aspects for AI/ML implementations.

Diagnostic quality aspects	Key questions for the laboratory director
Care team and patient perspective	What is the benefit?
Baseline	What is our current state (i.e., best practice)?
Responsibility	Who is: affected, responsible, and accountable?
Assurance of patient safety	How can we explain risks and benefits to a patient or provider?
Applicability and bias	Is there a way to assess bias of age, sex/gender, minorities, ...?
Beneficence	Can we assure availability to populations despite socioeconomic variabilities?
Data and measurements	Can we obtain relevant information?
Methods of measurement	Do we have reliable methods?
Data sources and harmonization of data	Are data from two sources comparable? Are the methods harmonized/replicable?
Data characterization	Have the data been characterized at the instrument, method, unit level?
Data standardization	Is data capture following a standard (e.g., LOINC, DICOM, etc.)?
Uncertainty measure	Is the data accompanied by a measure of uncertainty?
Electronic monitoring	Can we continuously monitor the data?
Prioritization of results	How are we prioritizing results?
Ground truth assessment	What do we consider the ground truth and why?
Identification of errors	Can we derive a specific definition of an error?
Diagnosis	How do we distinguish over and underdiagnosis?
Demographic disparities	Can we gain understanding of relative contributions?

Table 1: (continued)

Diagnostic quality aspects	Key questions for the laboratory director
Health information technology (IT)	What aspects of our IT solutions is involved?
Information-gathering	Do we know how to gather the input data?
Information synthesis	Do we have a data synthesis plan?
Detecting safety risks	Are we aware of all data security and safety concerns?
Laboratory information system (LIS)	Can we integrate this into our existing LIS?
EHR integration	If so, how is this presented to the clinician?
Organizational factors	Are there limiting organizational factors?
Teamwork	Who should be on the team?
Leadership	Do we have leadership support at the laboratory and institutional level?
Development methodology	Have we matched the development method to the problem?
Strategies, interventions, timeline	Do we know key strategic and delivery deadlines?
Personnel	Do we have the relevant personnel resources?
Competency	How do we assess competency in use?
Effort and capabilities	Do we have the effort and relevant development capabilities?
Synergy	How does the AI/ML tool work with preceding/following laboratory procedures?
Integration with existing QMS	Are operational quality elements (e.g., training, adverse event reporting, etc.) ready?
Regulatory aspects	Is there a dedicated team for regulatory questions?
Analytical validity (AV)	Is our performance assessment plan complete?
Clinical validity (CV)	Can we support clinical validity claims with data?
Patient safety	Can we document any patient safety concerns?
Intended use	What is the specific purpose of the AI/ML tool?
Indication for use	Can we define the target population?
Proficiency testing	Do we have access to an existing proficiency test or alternative approach?
AI/ML performance monitoring	Can we continuously monitor the AI/ML tool performance?
Performance drift	How do we identify drifts in predictions and ground truth?
Reimbursement	Is this AI/ML tool financially sustainable
Implementation cost	Do we have funding/budget for this implementation?
Ongoing cost	Do we have a micro-costing analysis?
Clinical utility (CU)	Can we make CU claims? What are the outcome benefits?
Applicable diagnoses	Do we have a list of ICD codes that are applicable?
Billing code	Is there a procedural code / billing code?
Payor policies	Do we have existing or future payor policy strategies?

AI, artificial intelligence; AV, analytical validity; CV, clinical validity; CU, clinical utility; DICOM, digital imaging and communications in medicine; EHR, electronic health record; ICD, international classification of disease; IT, information technology; LIS, laboratory information system; LOINC, logical observation identifiers names and codes; ML, machine learning; and QMS, quality management system. The list is preliminary and can help before and during implementation; audits/monitoring/take-down of AI/ML tools require separate checklists.

the procedure for nucleic acid extraction is the same for multiple PCR and NGS tests). Similarly, multiple diagnostic procedures may use the same diagnostic tests (e.g., the nucleic acid extraction from bile vs. bone marrow or other sources may differ but use the same downstream test). Or multiple diagnostic procedures may exist as different service lines (e.g., outpatient vs. inpatient testing services). These considerations are important for troubleshooting – for example, is the root cause at the diagnostic test level, or is it a broader procedural issue? Awareness of the relationship of test, procedure, and service is therefore an integral part of diagnostic quality.

How to concretely measure diagnostic AI/ML quality improvements?

The conceptual starting point is the consideration that the AI model is integrated into the existing healthcare ecosystems. In the simplest case, the AI model (a computer program) becomes part of a specific diagnostic test (or a component of a test). The integration of the model as part of the diagnostic test is considered the first layer (=diagnostic test level). Once accomplished, the model becomes an integral part of this specific laboratory testing process (also known as the laboratory procedures). The diagnostic procedure (that now

entails an AI model) is the second layer (=diagnostic procedure level). With very few exceptions, most laboratories have multiple or numerous diagnostic testing procedures, and these may entail multiple models with their individual use cases, value propositions, and performance metrics (=value add). Each of these testing procedures is facing outward – externally – towards the treating physician, care team, and the patient, which can be considered a diagnostic service (=diagnostic service level). The diagnostic service offered by the laboratory (to a hospital or to outside partners) is considered the third level.

In this conceptualization, the diagnostic quality is a composite of diagnostic test, diagnostic procedure, and diagnostic service (Figure 2E). Consequently, the quality impact of AI models should be considered a function of the improvements related to the diagnostic test(s), the diagnostic procedure(s) and the diagnostic service(s) with their various intended use cases and value propositions. Practically speaking, when choosing a primary endpoint (outcome measure) for a quality improvement initiative (e.g., turnaround time), the contributions of each level can be quantified before and after AI implementation. The absolute difference between the two concatenated measures represents the net change (e.g., in turnaround time; Figure 2F). Secondary endpoint measurement follows a similar logic. Thereby, our model provides a conceptual outline (rubric) to quantify quality improvements related to AI/ML implementations (Figure 2C, E, and F). It is important to recognize that this approach can account for the fact that interventions at one level (e.g., error reduction, efficiency gains) can impact quality through interference at other levels. An outline of ways our model can be applied is provided (Table 2).

Multimodality diagnostic workflows

The diagnostic quality model can successfully be applied to individual improvement initiatives [60, 61] including AI/ML approaches that serve as decision support tools [101]. Applicability to four more challenging scenarios revealed a few noteworthy findings. The first chosen setting, a multimodal radiology/molecular prediction tool (Figure 3), illustrates that various diagnostic quality models can be chained together. While this is nothing unusual in laboratory medicine or pathology, the allocation of data source (feature) and ground truth (label) can be easily visualized. Furthermore, the three-level model – even when in concatenation – can help to map and assign the complexity of quality control procedures and incorporation of performance and

Table 2: Selected applications of the diagnostic quality model.

Domain	Model can be used...
Problem-solving	... to help understand or solve a specific problem or challenge by outlining the different affected layers and structuring the analysis or structure of the AI tool.
Decision-making	... to inform or guide decisions by enabling portrayal of different options to enable selection of best course of action. ... to portray different integration options that enable assessment of benefits and costs during decision-making.
Communication	... to help explain or convey the idea (during the ideation stage) or provide information to new team members more efficiently. ... help to explain and communicate changes of related systems over time.
Concrete tasks	... to visualize which laboratory documents and tasks are primarily affected by the AI tool. ... to distribute code, standard operating procedure, and validation report tasks to team members.
Research	... to serve as a starting point for exploring the topic of quality measurements related to AI implementations ... to investigate characteristics of concatenated AI or ensemble AI approaches.
Education	... as a building block for understanding more complex ideas. ... as an alternative way to explain how complex laboratory operations are organized.

AI, artificial intelligence.

acceptance metrics. Specifically, in this example (Figure 3), the output of the model is not used to replace the genetic test results (diagnostic test) – rather it is used to trigger an insurance test claim via specialty pharmacy (service layer).

As additional multimodal workflow examples we selected prostate cancer diagnostics including recently recommended multiparametric MRI (mpMRI) before biopsy and an AI-based diagnostic aid [26, 75, 77, 78, 102, 103], a breast cancer diagnostic workflow including AI-based quantification of Ki-67 [79–81] and HER2-low [82, 83], and a multianalyte/demographic approach to predict SARS-CoV-2/COVID status and outcomes [49, 84–87, 104]. The simplified diagnostic quality models are provided in Figure 4. The localization of the AI-tool – in these examples at the diagnostic test level – enables the identification of the relevant diagnostic procedures, services, and ecosystems appropriate to the implementation. Key element for each laboratory (and laboratory director) is to establish the baseline performance without AI/ML as a comparator. Establishing performance metrics can then follow recognized regulatory standards and documentation in the quality management system.

Diagnostic quality model vs. quality management system – what is the difference?

A laboratory quality management system (QMS) is a systematic, integrated library of activities that establish and control all work processes. The functions of a QMS extend beyond a single patient or sample and include management of resources, evaluations, audits, and coordination of all activities. The key distinguishing features are provided in Table 3. Briefly, the QMS has a broader scope than the diagnostic quality model and helps to manage and investigate the entire laboratory. We consider a diagnostic quality model that focuses on AI/ML implementations as a specific, operational component of the QMS.

Discussion

We present a conceptual framework termed diagnostic quality model – or DQM for short. The model distinguishes AI/ML improvements at the diagnostic test, diagnostic procedure, diagnostic service, and healthcare ecosystem level. The operational definition of diagnostic quality (as the sum

of the quality at the diagnostic test, procedural, and service level) emphasizes the nested relationship among these three key levels within a laboratory and healthcare ecosystem. By defining scope at the relevant levels, the model can help to define the relevant steps for implementation and how different levels must come together to form coherent, high-quality diagnostics. We provide a simple rubric to quantify AI/ML improvements while complying with existing and mandated regulatory standards. We present several specific clinical scenarios – including multi-modal diagnostics – and provide a comparison of the diagnostic quality model to existing quality management systems.

AI/ML does not exist in a vacuum. The quality framework presented enables us to assess the impact of AI/ML on a process (or multiple processes) to ensure we are maximizing the value over time and are not negatively impacting other layers. For example, AI implementation might increase diagnostic accuracy by 2%; however, integration of results and human exploration of the underlying reasons may increase turnaround time by 200%. Because we can get stuck in siloed thinking, seeing the horizontal and vertical consequences of tasks and work units reveals tremendous inefficiencies in medical practices. Thus, the presented framework enables extrapolating value gains. For example,

Table 3: Distinction between diagnostic quality model (DQM) and quality management system (QMS).

	DQM	QMS
Focus	Individual diagnostic level	Laboratory-level
Key components	DQM model (3 nested components) Diagnostic test Diagnostic procedure Diagnostic service	QMS (12 components) Organization; Personnel; Equipment; Purchasing + inventory; Process control; IT; Documents + records; Org. Mgmt.; Assessment; Quality improvement; Customer service; Facilities & safety
Key emphasis	Balanced emphasis on all layers (test/procedure/service) to maximize diagnostic quality	Covering all aspects of quality management to maximize overall laboratory quality; (includes all diagnostic tests/procedures/services)
Focus	Diagnostic assay or AI-tool	Laboratory operations
Example	Document content and function; step-by-step activities necessary to complete task	Document control: creation, format, review, distribution, versioning, disposal, access, permissions, compliance, audits
Scope	Narrow scope: – Realizing QMS aspects at test, procedure, and service level – Specific diagnostic aspects: Intended use, indications, instructions for use (SOP), performance measures, mitigation strategies	Extended scope: Accreditation organization; case management; clinical resource management; insurance; managed care; patient safety; physician advisor; credentialing; quality assurance; health equity; regulatory environment; risk management; transactions of care; utilization management; compensation; quality improvement
Personnel	Staff-level: specialization, knowledge about process flows and intersections Leadership: synergy of qualifications across test/procedure/service layers	Staff-level: end-to-end view of the diagnostic journey Leadership: extrapolation of AI/ML results for utilization in larger laboratory/institutional context
Education	Less difficult	Difficult
Boundary	Boundaries of diagnostic test, procedure, and service	Laboratory, division, or department

AI, artificial intelligence; IT, information technology; ML, machine learning; Org. Mgmt., organizational management; SOP, standard operating procedure.

efficiency gains in one diagnostic procedure may prompt leveraging the same AI/ML, with appropriate modifications, to other diagnostic procedures under the same diagnostic service. The fact that physicians and laboratorians need to collate ever greater amounts of information to generate a diagnosis is a key hurdle in realizing the promise of personalized medicine. One approach to overcome this hurdle is to clearly define relevant objectives for AI/ML implementations to define coherent diagnostics.

Laboratory safety and process improvements are among the most complicated aspects a medical director needs to manage. Numerous initiatives tackle diagnostic quality at the international, national, regional, local, and individual test levels [39, 51, 105–107]. The terms safety culture and system-level thinking are clearly aimed to ensure appropriate function for patients and clinicians; however, developing and sustaining the systems for reliable diagnostics require time, effort, and dedicated resources [27, 39, 52, 57, 80, 92, 95, 97, 108, 109]. Despite limitations in creativity, ML/AI tools are reliable and exceed human capabilities in terms of computational power and consistency. Thus, there is great interest in integration of AI/ML tools in the laboratory. Our conceptual framework is novel because it attempts the balancing act of recognizing (and valuing) the existing standards while defining the relevant layers for process improvements. The layer of AI/ML integration is consequential. For example, when AI/ML tools trigger administrative processes in the service layer (e.g., test claims in Figure 3) the risk for the patient is substantially lower than when implementing an AI/ML tool at the diagnostic test layer (Figure 4). One classic approach for examining quality improvements are morbidity and mortality (M&M) meetings that have an established history and culture to ensure comprehensive review [110–112]. We consider the presented quality model applicable – especially given recent approaches to modernize M&M conferences [113]. The relationship of test, procedure and service is highly relevant for diagnostic quality because failure in one may deteriorate the overarching (perceived or real) quality. For example, a well-functioning test with a poorly constructed report will be confusing. Similarly, a visually appealing report and great customer service for a test with little clinical relevance is equally misaligned. Importantly, the principle can be applied to assess whether interventions at one layer do negatively affect patient care as a whole [29, 48, 53, 86]. It is the well-balanced, conscious effort to align (and optimize) the quality in all aspects of testing that results in diagnostic excellence. We provide a conceptual framework to map out the scope and help navigate existing, mandatory standards with AI/ML improvements.

The outlined framework can be used for educational purposes. The model might be helpful for smaller laboratories without dedicated QMS personnel. The model is not restricted

to AI/ML improvements. Akin to any machine or computer-assisted device in the laboratory one can distinguish software running on the machine (“test”), software related to the operations (“process”), and interfaces or web-services provided to outside customers, patients, and providers (“service” layer). IT and AI tools have similar layers, and the presented conceptual model can account for these layers. The aim of the checklist (Table 1) is to ensure consistency and completeness when assessing an AI tool as a new diagnostic technology. Note, the list is preliminary and can help before and during implementation; audits/monitoring/take-down of AI/ML tools will require separate checklists.

Numerous limitations apply. First, a conceptual framework cannot be considered a delivery system. Thus, no framework can accomplish high-quality work. In other words, we consider it a key limitation that we only have circumstantial and personal evidence that considering this framework helps conceptualizing process improvement initiatives. Second, the framework cannot replace qualified personnel or competent decision making in an ongoing operation. Third, numerous professional organizations created tools, guidelines, and checklists [26, 29, 31, 57, 64, 83, 106, 114]. For example, the Association for Molecular Pathology has introduced the concept laboratory-developed procedure (as opposed to test) [57]. However, we apply certain terms differently – to emphasize the importance of delivering excellent diagnostic services. Numerous other standards exist (e.g., for reporting diagnostic accuracy) [25, 29, 57, 64]; however, we endorse assigning specific quality metrics at the test, procedure, and service level as key performance indicators. Specifically, standardization and harmonization of laboratory data is critically important [115–117]. Briefly, standardization refers to the process of establishing technical standards and guidelines that aim for and enable interoperability. In contrast, harmonization can involve standardization, but it may also involve other forms of coordination (e.g., local policies or mutual recognition agreements) referring to any process that enables establishing equivalence of reported values among different measurement procedures [115]. Of note, two special CCLM issues were entirely dedicated to the topic of harmonization [117]. Irrespective of the applied approach, integration into the local procedures of a given laboratory requires thoughtful consideration of benefits and risks – especially given the importance of value-based care paradigms and limitation of resources. Fourth, the distinction between QMS and DQM is difficult, and we provide a direct comparison (Table 3). Fifth, the integrated diagnostic framework does not account for the development methodology. For example, some laboratories apply agile development principles [61, 118, 119], while others use more traditional management principles. However, irrespective of the applied development method, the nested relationship

among test, procedure, and service applies. We provide several possible ways to apply the DQM model (Table 2) and at the very least the conceptual framework can serve as an educational tool to illustrate the fascinating and complex world of laboratory testing. Finally, precise measurements of the quality impact of a specific AI implementation will remain challenging. Changes imposed by the implementation of AI in one layer may affect other layers – and consequently the relative contribution of each layer to the specific quality endpoint should be considered. Irrespective of the specific endpoint, we recommend concatenating the quality endpoints of all three layers (test, procedure, and service) to ensure a net benefit. In our experience, cost and value-add contributions can cancel each other out and it is a combination of factors (gains, savings, and strategic benefits) can create utility.

In conclusion, advantages of new technologies require careful alignment with existing practices. Unintentional consequences of well-intended improvements must be avoided, and our *diagnostic quality model* is one approach to navigate the complexities of clinical AI/ML implementations. The presented diagnostic quality framework can help to specify and communicate the key implications of AI/ML solutions in laboratory diagnostics.

Acknowledgments: We thank the organizers of the 3rd Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine EFLM in May 2022.

Research funding: None declared.

Author contributions: All authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Competing interests: LLR has received honoraria from PeerView, Medscape, Clinical Care Options, consulting honoraria from Abbvie, Personal Genome Diagnostics, Bristol Myers Squibb, Loxo Oncology at Lilly, Amgen, Meck, AstraZeneca, Sanofi-Genzyme, and EMD Serono. The other authors declared no conflict of interest.

Informed consent: Not applicable.

Ethical approval: Not applicable.

Disclaimer: The content of this article represents the personal views of the authors and does not represent the views of the authors' employers and associated institutions. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article, and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization. This work was supported by U.S. National Institutes of Health (NIH) grant R37 CA225655 (to J.K.L.). The content of

this article is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health or any other organization.

References

1. Naylor S, Chen JY. Unraveling human complexity and disease with systems biology and personalized medicine. *Per Med* 2010;7:275–89.
2. Sahu M, Gupta R, Ambasta RK, Kumar P. Artificial intelligence and machine learning in precision medicine: a paradigm shift in big data analysis. *Prog Mol Biol Transl Sci* 2022;190:57–100.
3. 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.
4. Khan ZUN, Jafri L, Hall PL, Schultz MJ, Ahmed S, Khan AH, et al. Utilizing augmented artificial intelligence for aminoacidopathies using collaborative laboratory integrated reporting- a cross-sectional study. *Ann Med Surg* 2022;82:104651.
5. Constantinescu G, Schulze M, Peitzsch M, Hofmockel T, Scholl UI, Williams TA, et al. Integration of artificial intelligence and plasma steroidomics with laboratory information management systems: application to primary aldosteronism. *Clin Chem Lab Med* 2022;60:1929–37.
6. Soerensen PD, Christensen H, Gray Worsoe Laursen S, Hardahl C, Brandslund I, Madsen JS. Using artificial intelligence in a primary care setting to identify patients at risk for cancer: a risk prediction model based on routine laboratory tests. *Clin Chem Lab Med* 2021;12:2005–16.
7. Pillay TS. Artificial intelligence in pathology and laboratory medicine. *J Clin Pathol* 2021;74:407–8.
8. Kennedy AG. Evaluating the effectiveness of diagnostic tests. *JAMA* 2022;327:1335–6.
9. Pritchard D, Goodman C, Nadauld LD. Clinical utility of genomic testing in cancer care. *JCO Precis Oncol* 2022;6:e2100349.
10. Svoboda E, Boril T, Rusz J, Tykalova T, Horakova D, Guttman CRG, et al. Assessing clinical utility of machine learning and artificial intelligence approaches to analyze speech recordings in multiple sclerosis: a pilot study. *Comput Biol Med* 2022;148:105853.
11. Tajiri A, Ishihara R, Kato Y, Inoue T, Matsueda K, Miyake M, et al. Utility of an artificial intelligence system for classification of esophageal lesions when simulating its clinical use. *Sci Rep* 2022;12:6677.
12. Banja J. How might artificial intelligence applications impact risk management? *AMA J Ethics* 2020;22:E945–51.
13. Keris MP. Artificial intelligence in medicine creates real risk management and litigation issues. *J Healthc Risk Manag* 2020;40:21–6.
14. Stenzinger A, Alber M, Allgauer M, Jurmeister P, Bockmayr M, Budczies J, et al. Artificial intelligence and pathology: from principles to practice and future applications in histomorphology and molecular profiling. *Semin Cancer Biol* 2022;84:129–43.
15. Car J, Sheikh A, Wicks P, Williams MS. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Med* 2019;17:143.
16. Boon IS, Au YTPT, Boon CS. Assessing the role of artificial intelligence (AI) in clinical oncology: utility of machine learning in radiotherapy target volume delineation. *Basel: Medicines*; 2018:5 p.
17. Fonseka TM, Bhat V, Kennedy SH. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Aust N Z J Psychiatry* 2019;53:954–64.

18. 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.
19. Wen X, Leng P, Wang J, Yang G, Zu R, Jia X, et al. Clinlabomics: leveraging clinical laboratory data by data mining strategies. *BMC Bioinf* 2022;23:387.
20. Lippi G, Da Rin G. Advantages and limitations of total laboratory automation: a personal overview. *Clin Chem Lab Med* 2019;57:802–11.
21. Ialongo C, Bernardini S. Total laboratory automation has the potential to be the field of application of artificial intelligence: the cyber-physical system and “automation 4.0”. *Clin Chem Lab Med* 2019;57:e279–81.
22. Cabitza F, Banfi G. Machine learning in laboratory medicine: waiting for the flood? *Clin Chem Lab Med* 2018;56:516–24.
23. Lidströmer N, Ashradian H. *Artificial intelligence in MEDicine*. New York: Springer; 2022.
24. Thierauf JC, Farahani AA, Indave BI, Bard AZ, White VA, Smith CR, et al. Diagnostic value of MAML2 rearrangements in mucoepidermoid carcinoma. *Int J Mol Sci* 2022;23:4322.
25. Lans A, Pierik RJB, Bales JR, Fourman MS, Shin D, Kanbier LN, et al. Quality assessment of machine learning models for diagnostic imaging in orthopaedics: a systematic review. *Artif Intell Med* 2022;132:102396.
26. Barrett T, de Rooij M, Giganti F, Allen C, Barentsz JO, Padhani AR. Quality checkpoints in the MRI-directed prostate cancer diagnostic pathway. *Nat Rev Urol* 2022;20:9–22.
27. Shetty O, Shet T, Iyer R, Gogte P, Gurav M, Joshi P, et al. Impact of COVID-19 on quality checks of solid tumor molecular diagnostic testing-A surveillance by EQAS provider in India. *PLoS One* 2022;17:e0274089.
28. Thakur V, Akerele OA, Randell E. Lean and Six Sigma as continuous quality improvement frameworks in the clinical diagnostic laboratory. *Crit Rev Clin Lab Sci* 2022;1:1–19.
29. Stahl AC, Tietz AS, Kendziora B, Dewey M. Has the STARD statement improved the quality of reporting of diagnostic accuracy studies published in European radiology? *Eur Radiol* 2023;1:97–105.
30. Reyna MA, Nsoesie EO, Clifford GD. Rethinking algorithm performance metrics for artificial intelligence in diagnostic medicine. *JAMA* 2022;328:329–30.
31. Loh HW, Ooi CP, Seoni S, Barua PD, Molinari F, Acharya UR. Application of explainable artificial intelligence for healthcare: a systematic review of the last decade (2011–2022). *Comput Methods Programs Biomed* 2022;226:107161.
32. Collins GS, Dhiman P, Andaur Navarro CL, Ma J, Hooft L, Reitsma JB, et al. Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence. *BMJ Open* 2021;11:e048008.
33. Naugler C, Church DL. Automation and artificial intelligence in the clinical laboratory. *Crit Rev Clin Lab Sci* 2019;56:98–110.
34. Karekar SR, Vazifdar AK. Current status of clinical research using artificial intelligence techniques: a registry-based audit. *Perspect Clin Res* 2021;12:48–52.
35. Cabitza F, Rasoini R, Gensini GF. Unintended consequences of machine learning in medicine. *JAMA* 2017;318:517–8.
36. 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.
37. Murdoch B. Privacy and artificial intelligence: challenges for protecting health information in a New Era. *BMC Med Ethics* 2021;22:122.
38. Wang D, Li M, Zhang Y. Adversarial data hiding in digital images. *Entropy* 2022;24:749.
39. Beckers R, Kwade Z, Zanca F. The EU medical device regulation: implications for artificial intelligence-based medical device software in medical physics. *Phys Med* 2021;83:1–8.
40. 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.
41. Lennerz JK, Marble HD, Lasiter L, Poste G, Sirintrapun SJ, Salgado R. Do not sell regulatory science short. *Nat Med* 2021;27:573–4.
42. Gallas BD, Chan HP, D’Orsi CJ, Dodd LE, Giger ML, Gur D, et al. Evaluating imaging and computer-aided detection and diagnosis devices at the FDA. *Acad Radiol* 2012;19:463–77.
43. Vizitiu A, Nita CI, Puiu A, Suciu C, Itu LM. Privacy-preserving artificial intelligence: application to precision medicine. *Annu Int Conf IEEE Eng Med Biol Soc* 2019;2019:6498–504.
44. Giudici P. Fintech risk management: a research challenge for artificial intelligence in finance. *Front Artif Intell* 2018;1:1.
45. Alshahrani E, Alghazzawi D, Alotaibi R, Rabie O. Adversarial attacks against supervised machine learning based network intrusion detection systems. *PLoS One* 2022;17:e0275971.
46. Almalawi A, Khan AI, Alsolami F, Abushark YB, Alfakheh AS, Mekuriyaw WD. Analysis of the exploration of security and privacy for healthcare management using artificial intelligence: Saudi hospitals. *Comput Intell Neurosci* 2022;2022:4048197.
47. Gillan C, Milne E, Harnett N, Purdie TG, Jaffray DA, Hodges B. Professional implications of introducing artificial intelligence in healthcare: an evaluation using radiation medicine as a testing ground. *J Radiother Pract* 2018;18:5–9.
48. Plebani M. Exploring the iceberg of errors in laboratory medicine. *Clin Chim Acta* 2009;404:16–23.
49. Campagner A, Famiglini L, Carobene A, Cabitza F. Everything is varied: the surprising impact of individual variation on ML robustness in medicine. 2022. [arXiv:2210.04555](https://arxiv.org/abs/2210.04555).
50. Gottlieb S, McClellan MB. Reforms needed to modernize the US Food and drug administration’s oversight of dietary supplements, cosmetics, and diagnostic tests. *JAMA Health Forum* 2022;3. <https://doi.org/10.1001/jamahealthforum.2022.4449>.
51. Gallas BD, Badano A, Dudgeon S, Elfer K, Garcia V, Lennerz JK, et al. FDA fosters innovative approaches in research, resources and collaboration. *Nat Mach Intell* 2022;4:97–8.
52. Marble HD, Huang R, Dudgeon SN, Lowe A, Herrmann MD, Blakely S, et al. A regulatory science initiative to harmonize and standardize digital pathology and machine learning processes to speed up clinical innovation to patients. *J Pathol Inform* 2020;11:22.
53. Kearney SJ, Lowe A, Lennerz JK, Parwani A, Bui MM, Wack K, et al. Bridging the gap: the critical role of regulatory affairs and clinical affairs in the total product life cycle of pathology imaging devices and software. *Front Med* 2021;8:765385.
54. Clinical Laboratory Improvement Amendments (CLIA). Code of Federal Regulations (CFR), Title 42, Chapter IV, Subchapter G Part CFR 493.1–493.2001 42 USC 263a. US 1988.
55. Federal Drug Administration. Overview of IVD regulation. Administration USFD; 2021.

56. API. Association for Pathology Informatics; 2022. Available from: <https://www.pathologyinformatics.org/> [Accessed 20 Jan 2023].
57. Kaul KL, Sabatini LM, Tsongalis GJ, Caliendo AM, Olsen RJ, Ashwood ER, et al. The case for laboratory developed procedures: quality and positive impact on patient care. *Acad Pathol* 2017;4:2374289517708309.
58. ACLA. Laboratory innovation & operations; 2022. Available from: <https://www.acla.com/laboratory-innovation-operations/> [Accessed 20 Jan 2023].
59. ADASP. Association of directors of anatomic and surgical pathology; 2022. Available from: <https://www.adasp.org/> [Accessed 20 Jan 2023].
60. Marble HD, Bard AZ, Mizrahi MM, Lennerz JK. Temporary regulatory deviations and the coronavirus disease 2019 (COVID-19) PCR labeling update study indicate what laboratory-developed test regulation by the US food and drug administration (FDA) could look like. *J Mol Diagn* 2021;23:1207–17.
61. Lennerz JK, McLaughlin HM, Baron JM, Rasmussen D, Sumbada Shin M, Berners-Lee N, et al. Health care infrastructure for financially sustainable clinical genomics. *J Mol Diagn* 2016;18:697–706.
62. Mazzucca S, Tabak RG, Pilar M, Ramsey AT, Baumann AA, Kryzer E, et al. Variation in research designs used to test the effectiveness of dissemination and implementation strategies: a review. *Front Public Health* 2018;6:32.
63. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Republished research: implementation research: what it is and how to do it: implementation research is a growing but not well understood field of health research that can contribute to more effective public health and clinical policies and programmes. this article provides a broad definition of implementation research and outlines key principles for how to do it. *Br J Sports Med* 2014;48:731–6.
64. Pinnock H, Barwick M, Carpenter CR, Eldridge S, Grandes G, Griffiths CJ, et al. Standards for reporting implementation studies (StaRI) statement. *BMJ* 2017;356:i6795.
65. Huang R, Lasiter L, Bard A, Quinn B, Young C, Salgado R, et al. National maintenance cost for precision diagnostics under the verifying accurate leading-edge in vitro clinical test development (VALID) act of 2020. *JCO Oncol Pract* 2021;17:e1763–73.
66. Bolboaca SD. Medical diagnostic tests: a review of test anatomy, phases, and statistical treatment of data. *Comput Math Methods Med* 2019;2019:1891569.
67. McPherson RA, Pincus MR, Henry JB. Henry's clinical diagnosis and management by laboratory methods, 21st ed. Philadelphia: Saunders Elsevier; 2007.
68. Balogh E, Miller BT, Ball J, Institute of Medicine (U.S.). Committee on Diagnostic Error in Health Care. Improving diagnosis in health care. Washington, DC: The National Academies Press; 2015.
69. Kline A, Wang H, Li Y, Dennis S, Hutch M, Xu Z, et al. Multimodal machine learning in precision health: a scoping review. *NPJ Digit Med* 2022;5:171.
70. Carruthers R, Straw I, Ruffle JK, Herron D, Nelson A, Bzdok D, et al. Representational ethical model calibration. *NPJ Digit Med* 2022;5:170.
71. Ghosh P, Tamboli P, Vikram R, Rao A. Imaging-genomic pipeline for identifying gene mutations using three-dimensional intra-tumor heterogeneity features. *J Med Imaging* 2015;2:041009.
72. Ninatti G, Kirienko M, Neri E, Sollini M, Chiti A. Imaging-based prediction of molecular therapy targets in NSCLC by radiogenomics and AI approaches: a systematic review. *Diagnostics* 2020;10:359.
73. Shen TX, Liu L, Li WH, Fu P, Xu K, Jiang YQ, et al. CT imaging-based histogram features for prediction of EGFR mutation status of bone metastases in patients with primary lung adenocarcinoma. *Cancer Imag* 2019;19:34.
74. Digumarthy SR, Mendoza DP, Lin JJ, Chen T, Rooney MM, Chin E, et al. Computed tomography imaging features and distribution of metastases in ROS1-Rearranged non-small-cell lung cancer. *Clin Lung Cancer* 2020;21:153–9. e3.
75. Rakovic K, Colling R, Browning L, Dolton M, Horton MR, Protheroe A, et al. The use of digital pathology and artificial intelligence in histopathological diagnostic assessment of prostate cancer: a survey of prostate cancer UK supporters. *Diagnostics* 2022;12:1225.
76. Raciti P, Sue J, Retamero JA, Ceballos R, Godrich R, Kunz JD, et al. Clinical validation of artificial intelligence-augmented pathology diagnosis demonstrates significant gains in diagnostic accuracy in prostate cancer detection. *Arch Pathol Lab Med* 2022. <https://doi.org/10.5858/arpa.2022-0066-OA>.
77. Kohaar I, Petrovics G, Srivastava S. A rich array of prostate cancer molecular biomarkers: opportunities and challenges. *Int J Mol Sci* 2019;20:1813.
78. Tikkinen KAO, Dahm P, Lytvyn L, Heen AF, Vernooij RWM, Siemieniuk RAC, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a clinical practice guideline. *BMJ* 2018;362:k3581.
79. Sheffield KM, Peachey JR, Method M, Grimes BR, Brown J, Saverno K, et al. A real-world US study of recurrence risks using combined clinicopathological features in HR-positive, HER2-negative early breast cancer. *Future Oncol* 2022;18:2667–82.
80. Raheem F, Ofori H, Simpson L, Shah V. Abemaciclib: the first FDA-approved CDK4/6 inhibitor for the adjuvant treatment of HR+ HER2- early breast cancer. *Ann Pharmacother* 2022:10600280211073322. <https://doi.org/10.1177/10600280211073322>.
81. Royce M, Osgood C, Mulkey F, Bloomquist E, Pierce WF, Roy A, et al. FDA approval summary: abemaciclib with endocrine therapy for high-risk early breast cancer. *J Clin Oncol* 2022;40:1155–62.
82. Modi S, Jacot W, Yamashita T, Sohn J, Vidal M, Tokunaga E, et al. Trastuzumab deruxtecan in previously treated HER2-low advanced breast cancer. *N Engl J Med* 2022;387:9–20.
83. Baez-Navarro X, Salgado R, Denkert C, Lennerz JK, Penault-Llorca F, Viale G, et al. Selecting patients with HER2-low breast cancer: getting out of the tangle. *Eur J Cancer* 2022;175:187–92.
84. Cabitza F, Campagner A, Ferrari D, Di Resta C, Ceriotti D, Sabetta E, et al. Development, evaluation, and validation of machine learning models for COVID-19 detection based on routine blood tests. *Clin Chem Lab Med* 2021;59:421–31.
85. Cabitza F, Campagner A, Soares F, García de Guadiana-Romualdo L, Challa F, Sulejmani A, et al. The importance of being external. Methodological insights for the external validation of machine learning models in medicine. *Comput Methods Programs Biomed* 2021;208:106288.
86. Campagner A, Carobene A, Cabitza F. External validation of machine learning models for COVID-19 detection based on complete blood count. *Health Inf Sci Syst* 2021;9:37.
87. Famigliani L, Campagner A, Carobene A, Cabitza F. A robust and parsimonious machine learning method to predict ICU admission of COVID-19 patients. *Med Biol Eng Comput* 2022;30:1–13.
88. Olson APJ, Graber ML, Singh H. Tracking progress in improving diagnosis: a framework for defining undesirable diagnostic events. *J Gen Intern Med* 2018;33:1187–91.
89. Henriksen K, Dymek C, Harrison MI, Brady PJ, Arnold SB. Challenges and opportunities from the agency for healthcare research and

- quality (AHRQ) research summit on improving diagnosis: a proceedings review. *Diagnosis* 2017;4:57–66.
90. Horgan D, Plebani M, Orth M, Macintyre E, Jackson S, Lal JA, et al. The gaps between the new EU Legislation on in vitro diagnostics and the on-the-ground reality. *Clin Chem Lab Med* 2023;61: 224–33.
 91. Gale MS. Diagnosis: fundamental principles and methods. *Cureus* 2022;14:e28730.
 92. Morais C, Yung KL, Johnson K, Moura R, Beer M, Patelli E. Identification of human errors and influencing factors: a machine learning approach. *Saf Sci* 2022;146. <https://doi.org/10.1016/j.ssci.2021.105528>.
 93. Lippi G, Plebani M, Simundic AM. Quality in laboratory diagnostics: from theory to practice. *Biochem Med* 2010;20:126–30.
 94. Lavin A, Gilligan-Lee CM, Visnjic A, Ganju S, Newman D, Ganguly S, et al. Technology readiness levels for machine learning systems. *Nat Commun* 2022;13:6039.
 95. Weiss VL, Heher YK, Seegmiller A, VanderLaan PA, Nishino M. All in for patient safety: a team approach to quality improvement in our laboratories. *J Am Soc Cytopathol* 2022;11:87–93.
 96. Harris CK, Chen Y, Jensen KC, Hornick JL, Kilfoyle C, Lamps LW, et al. Towards high reliability in national pathology education: evaluating the United States and Canadian academy of pathology educational product. *Acad Pathol* 2022;9:100048.
 97. Harris CK, Chen Y, Yarsky B, Haspel RL, Heher YK. Pathology trainees rarely report safety incidents: a review of 13,722 safety reports and a call to action. *Acad Pathol* 2022;9:100049.
 98. Renshaw AA, Mena-Allauca M, Gould EW, Sirintrapun SJ. Synoptic reporting: evidence-based review and future directions. *JCO Clin Cancer Inform* 2018;2:1–9.
 99. Sluijter CE, van Lonkhuijzen LR, van Slooten HJ, Nagtegaal ID, Overbeek LI. The effects of implementing synoptic pathology reporting in cancer diagnosis: a systematic review. *Virchows Arch* 2016;468:639–49.
 100. Cree IA, Indave Ruiz BI, Zavadil J, McKay J, Olivier M, Kozlakidis Z, et al. The international collaboration for cancer classification and research. *Int J Cancer* 2021;148:560–71.
 101. Zomnir MG, Lipkin L, Pacula M, Meneses ED, MacLeay A, Duraisamy S, et al. Artificial intelligence approach for variant reporting. *JCO Clin Cancer Inform* 2018;2:CCO.16.00079.
 102. Parker C, Castro E, Fizazi K, Heidenreich A, Ost P, Procopio G, et al. Prostate cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2020;31:1119–34.
 103. Gao J, Zhang Q, Zhang C, Chen M, Li D, Fu Y, et al. Diagnostic performance of multiparametric MRI parameters for gleason score and cellularity metrics of prostate cancer in different zones: a quantitative comparison. *Clin Radiol* 2019;74:895. e17–26.
 104. Ferrari D, Cabitza F, Carobene A, Locatelli M. Routine blood tests as an active surveillance to monitor COVID-19 prevalence. A retrospective study. *Acta Biomed* 2020;91:e2020009.
 105. Horton R. NICE: a step forward in the quality of NHS care. National institute for clinical excellence. national health service. *Lancet* 1999; 353:1028–9.
 106. McGenity C, Bossuyt P, Treanor D. Reporting of artificial intelligence diagnostic accuracy studies in pathology abstracts: compliance with STARD for abstracts guidelines. *J Pathol Inform* 2022;13:100091.
 107. IQN. International quality network for pathology; 2022. Available from: <https://www.iqnpath.org/> [Accessed 20 Jan 2023].
 108. Snead DR, Tsang YW, Meskiri A, Kimani PK, Crossman R, Rajpoot NM, et al. Validation of digital pathology imaging for primary histopathological diagnosis. *Histopathology* 2016;68:1063–72.
 109. Lima-Oliveira G, Lippi G, Salvagno GL, Picheth G, Guidi GC. Laboratory diagnostics and quality of blood collection. *J Med Biochem* 2015;34:288–94.
 110. Misialek M, Heher YK. Culture club: Promoting a culture of safety and quality; 2022. Available from: <https://www.cap.org/member-resources/articles/culture-club-promoting-a-culture-of-safety-and-quality> [Accessed 20 Jan 2023].
 111. Harris CK, Chen Y, Yarsky B, Haspel RL, Heher YK. Pathology trainees rarely report safety incidents: a review of 13,722 safety reports and a call to action. *Acad Pathol* 2022;9:100049.
 112. Pierluissi E. Morbidity and mortality conferences: change you can believe in? *J Grad Med Educ* 2012;4:543–4.
 113. Cifra CL, Miller MR. Envisioning the future morbidity and mortality conference: a vehicle for systems change. *Pediatr Qual Saf* 2016;1: e003.
 114. Pasotti F, Pellegrinelli L, Liga G, Rizzetto M, Azzara G, Da Molin S, et al. First results of an external quality assessment (EQA) scheme for molecular, serological and antigenic diagnostic test for SARS-CoV-2 detection in lombardy region (northern Italy), 2020–2022. *Diagnostics* 2022;12:1483.
 115. Miller WG, Greenberg N. Harmonization and standardization: where are we now? *J Appl Lab Med* 2021;6:510–21.
 116. Vidali M, Carobene A, Apassiti Esposito S, Napolitano G, Caracciolo A, Seghezzi M, et al. Standardization and harmonization in hematology: instrument alignment, quality control materials, and commutability issue. *Int J Lab Hematol* 2021;43:364–71.
 117. Zaninotto M, Graziani MS, Plebani M. The harmonization issue in laboratory medicine: the commitment of CCLM. *Clin Chem Lab Med* 2022 Nov 16. <https://doi.org/10.1515/cclm-2022-1111>. [Epub ahead of print].
 118. Pereira IM, Amorim VJP, Cota MA, Gonçalves CG. Gamification use in agile project management: an experience report. *Agile Methods* 2017; 680:28–38.
 119. Verdugo J, Rodríguez M, Piattini M. Using agile methods to implement a laboratory for software product quality evaluation. In: *Agile processes in software engineering and extreme programming*. Cham, Switzerland: Springer International; 2014:143–56 pp.

Supplementary Material: This article contains supplementary material (<https://doi.org/10.1515/cclm-2022-1151>).