Review Article
Laboratory Informatics

Ann Lab Med 2025;45:1-11
https://doi.org/10.3343/alm.2024.0258
ISSN 2234-3806 • eISSN 2234-3814

ANNALS OF LABORATORY MEDICINE

Toward High-Quality Real-World Laboratory Data in the Era of Healthcare Big Data

Sollip Kim, M.D., Ph.D.[1] and Won-Ki Min, M.D., Ph.D.[1,2]

[1]Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Korea; [2]Future Strategy Division, SD Biosensor, Seoul, Korea

With Industry 4.0, big data and artificial intelligence have become paramount in the field of medicine. Electronic health records, the primary source of medical data, are not collected for research purposes but represent real-world data; therefore, they have various constraints. Although structured, laboratory data often contain unstandardized terminology or missing information. The major challenge lies in the lack of standardization of test results in terms of metrology, which complicates comparisons across laboratories. In this review, we delve into the essential components necessary for integrating real-world laboratory data into high-quality big data, including the standardization of terminology, data formats, equations, and the harmonization and standardization of results. Moreover, we address the transference and adjustment of laboratory results, along with the certification for quality of laboratory data. By discussing these critical aspects, we seek to shed light on the challenges and opportunities inherent to utilizing real-world laboratory data within the framework of healthcare big data and artificial intelligence.

Key Words: Artificial intelligence, Big data, Data quality, Harmonization, Laboratory medicine, Real-world data, Standardization

Received: May 20, 2024
Revision received: July 4, 2024
Accepted: September 4, 2024
Published online: September 30, 2024

Corresponding author:
Won-Ki Min, M.D., Ph.D.
Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea
E-mail: wonkmin@gmail.com

© Korean Society for Laboratory Medicine
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

INTRODUCTION

Industry 4.0 represents a convergence of physical, digital, and biological systems, leading to profound changes in the way we live and work by integrating emerging technologies, such as big data and analytics, robotics, and the internet of things. In the medical field of Industry 4.0, big data and analytics are crucial.

Big data in healthcare are derived from various sources within and external to the hospital setting. These sources include hospital registries and electronic health records (EHRs), as well as data created, reported, or gathered by patients, including from home devices such as wearables and mobile health apps, device-generated data, and omics data including genomics and proteomics. Additional sources include patient portals, social media, search engine data, and insurance payer records. Big data are utilized for various analytical purposes and include four core types of data analytics: descriptive, diagnostic, predictive, and prescriptive. Descriptive and diagnostic analytics consider the past to understand "what happened" and to explain "why it happened," respectively, whereas predictive and prescriptive analytics are future-oriented, forecasting "what will happen" and advising on "what should be done." Artificial intelligence (AI), resembling human cognitive functions, has led to a paradigm shift in healthcare, driven by increased accessibility of healthcare data and advancements in analytical techniques [1]. The integration of big data and AI has markedly enhanced health outcomes across several domains, including diagnostics, preventive medicine, precision medicine, medical research, has led to a reduction in adverse events, and has improved cost efficiency and the management of population health [1] (Fig. 1). Research re-

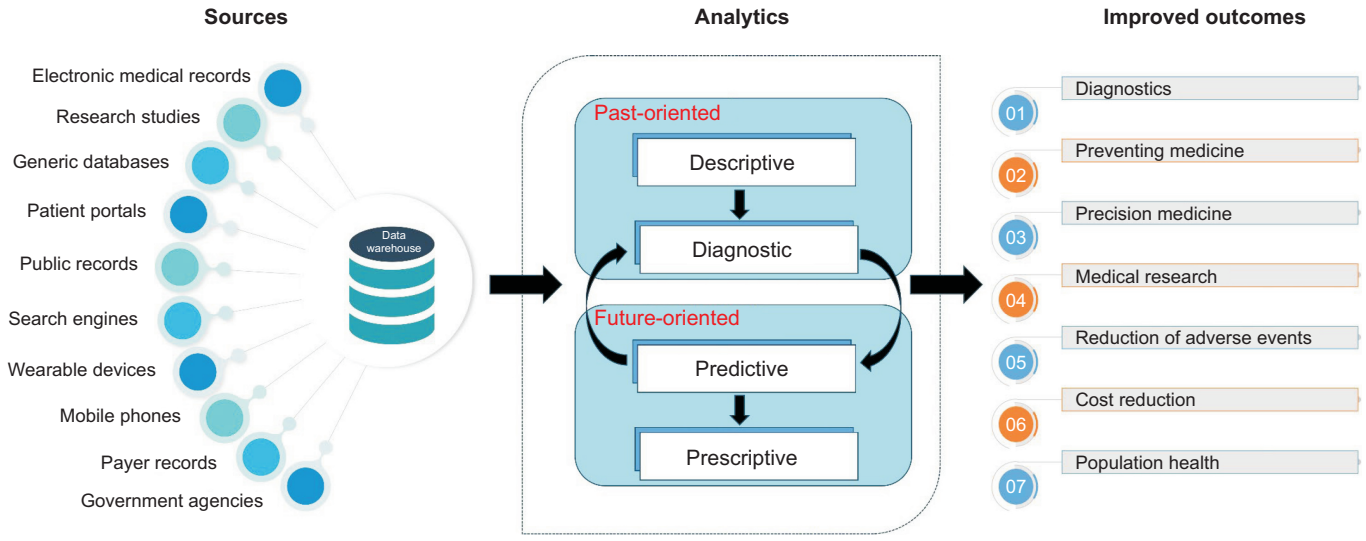

Fig. 1. Improving patient outcomes through analytics performed on big data gathered from various sources.

lated to big data and AI in the medical field has increased explosively. PubMed searches for "big data in healthcare" and "AI in healthcare" in 2023 retrieved 817 and 4,076 papers, respectively.

EHRs are the primary source of medical big data. They typically comprise medical histories of patients, demographic details, medications, immunizations, test results, and progress notes. EHR data comprise seven data types: numerical, categorical, text, image, video, speech, and signal data. The fields of application vary depending on the data type, and the preprocessing techniques for AI differ accordingly. Detailed information is summarized in Table 1. To generate qualified big data from EHRs, stringent quality metrics, clinical guidelines, and data extraction methods are essential. Numerous documents provide guidance on these processes [2-7].

In laboratory medicine, clinical test results are mostly composed of structured data, which can be numerical or categorical. Although structured data are easier to analyze and manipulate because of their standardized nature, real-world numerical data in laboratory medicine still have limitations when utilized in big data applications. The major challenge arises from the lack of standardized test methods and reporting, which complicates the comparison of results across healthcare institutions. Moreover, instances may arise where the quality of reagents used in tests or the quality of the testing facility falls below standard. Current EHR data are of insufficient quality and hardly usable for big-data research. To accumulate clean data suitable for big data research, efforts must focus on constructing a "qualified database" [8, 9].

In the era of Industry 3.0, doctors requested laboratory tests and interpreted results based on their expertise, commonly known as the "brain-to-brain loop." However, in the Industry 4.0 era, i.e., in the healthcare big data and AI era, we are confident that big data will replace the knowledge of doctors, establishing a "big data-to-big data loop" (Fig. 2). In this review, we discuss the essential components that real-world laboratory data, particularly quantitative results, require for integration into big data. These include terminology standards, data format standards, equation standards, standardization and harmonization of laboratory results, transference of laboratory results, adjustment of laboratory results, and certification of laboratory outcomes (Fig. 3).

TERMINOLOGY STANDARDS

The first requirement for establishing laboratory big data is the use of terminology standards, including Logical Observation Identifiers Names and Codes (LOINC) and Systematized Medical Nomenclature for Medicine Clinical Terminology (SNOMED CT) for coding laboratory requests and reporting results.

A LOINC term comprises a LOINC code and a corresponding fully specified name (FSN). The LOINC code serves as a unique, permanent identifier. The FSN consists of five or six primary components: the name of the component or analyte measured (e.g., albumin, vancomycin), the property observed (e.g., substance concentration, mass, volume), the temporal aspect of

https://doi.org/10.3343/alm.2024.0258

Table 1. Key data types, preprocessing techniques, and applications for artificial intelligence in electronic health records

Data type	Examples of data	Preprocessing techniques	Applications
Numerical	Clinically measured parameters (e.g., blood pressure level, blood glucose level)	Outlier removal, imputation for missing value, scaling (standardization, normalization)	Disease/patient status identification, disease occurrence prediction, clinical outcome estimation/prediction, numerical abnormality detection, reasoning of contributing factors to outcomes
Categorical	Coded parameters (e.g., patient type, disease code)	Imputation for missing value, encoding	Same as for numerical data
Text	Nursing note, doctor's note, manual error report	Tokenization, non-words removal, vectorization (e.g., bag of words, term frequency-inverse document frequency, word embedding)	Generating report, disease/patient status identification, disease occurrence prediction, reasoning of contributing factors to outcomes
Image	X-ray, CT, MRI, PET, US (image), tissue image, skin photography	Image conversion (e.g., resampling, bit-depth conversion, domain transformation, normalization, regularization), image processing (e.g., noise reduction, image quality enhancement, image restoration, segmentation), recognition/feature extraction (e.g., region of interest, object/situation, feature)	Improving traditional image processing technology, identification/classification (e.g., cell type), counting/enumeration (e.g., cell, chromosome), numerical estimation (e.g., ventricular volume, lung volume), disease/patient status identification, reasoning of contributing factors to outcomes, data curation (e.g., annotation, labeling, description)
Video	US (video), echo, endo, telemedicine/teleconsultation, surgical video, video for medical education	Same as for image preprocessing and frame processing (e.g., frame extraction and selection, temporal resampling, temporal segmentation)	Same as for image data and captioning
Speech	Psychiatric consultation, diagnostic conversation	Traditional audio processing (e.g., noise reduction, normalization, feature extraction, segmentation. domain transformation)	Generating reports, speech-to-text transformation for medical dialogue, disease identification using voice or contents
Signal	Auscultation (heart and lung sound), ECG, EEG, EMG, EOG, snoring	Signal conversion (e.g., resampling, bit-depth conversion, domain transformation, normalization, regularization), signal conditioning (e.g., noise reduction, signal quality enhancement, signal restoration), feature extraction (e.g., QRS complex)	Improving traditional bio-signal processing technologies, disease/patient status identification, clinical outcome prediction, reasoning of contributing factors to outcomes, data curation (e.g., annotation, labeling)

Abbreviations: CT, computed tomography; MRI, magnetic resonance imaging; PET, positron emission tomography; US, ultrasound; ECG, electrocardiography; EEG, electroencephalography; EMG, electromyography; EOG, electrooculography; Echo, echocardiography; Endo, endoscopic imaging; QRS, Q, R, and S waves.

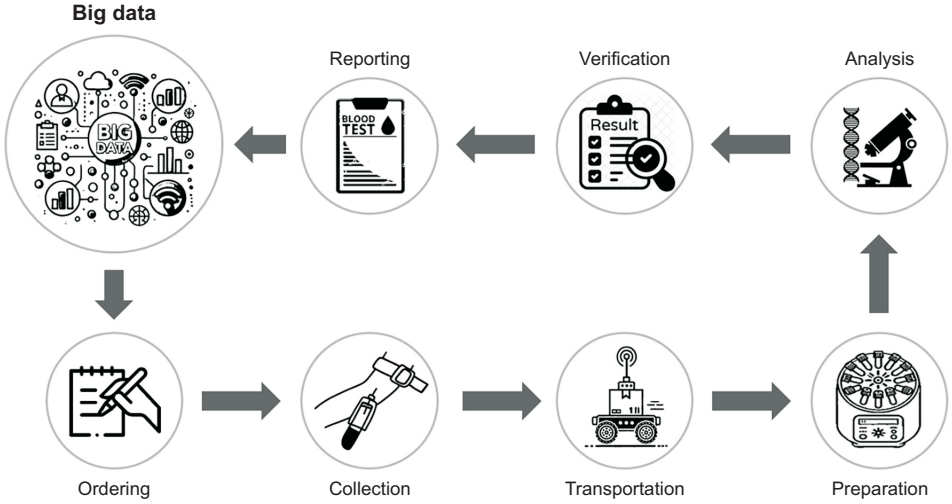

Fig. 2. 'Big data-to-big data loop' of laboratory tests in the Industry 4.0 era.

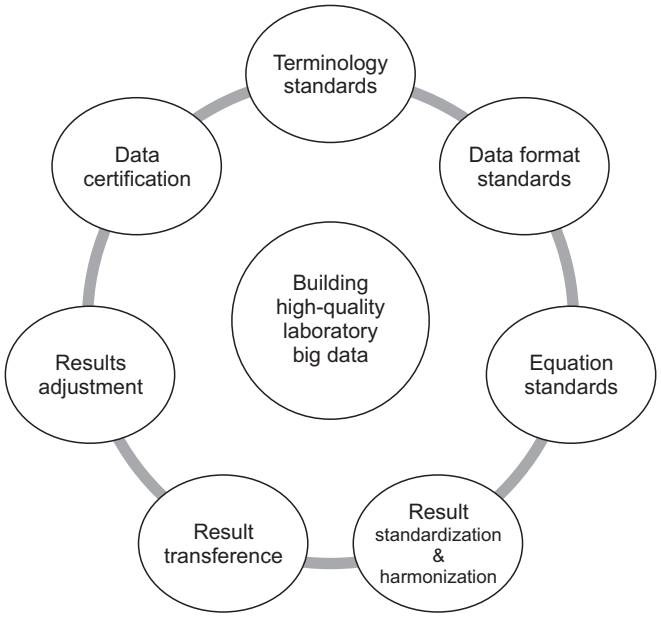

Fig. 3. Essential components for building high-quality laboratory big data.

the measurement (e.g., over a period or at a specific moment), the system or sample type (e.g., plasma, urine), the measurement scale (e.g., qualitative vs. quantitative), and, where applicable, the method of measurement (e.g., chemiluminescence immunoassay, mass spectrometry) [10, 11]. LOINC was initiated by Dr. Clem McDonald in 1994, and updates are released biannually (in February and August). LOINC v2.77, distributed in February 2024, contains 102,465 terms and has been translated into 20 languages, including Korean [11].

Mapping LOINC codes to each institution's codes is a vast and tedious task; thus, developers have created automated mapping tools based on various methods, such as using medical device identifiers or machine learning [12-14]. However, none of these tools have demonstrated an adequate level of accuracy for practical use.

The problem lies in the low accuracy of LOINC mapping within laboratories. In a 2018 College of American Pathologists (CAP) coagulation and cardiac markers survey of 1,916 laboratories, out of 275 reported LOINC codes, 54 (19.6%) were incorrect, and two codes (5934-2 and 12345-1) (0.7%) were not found in the LOINC database [15]. In a LOINC mapping accuracy evaluation conducted by three major institutions in the United States (ARUP, Intermountain, and Regenstrief), out of 884 test codes evaluated, four tests were mapped to completely unrelated LOINC codes, and 36 tests had at least one error in the mapping to one or more of the six LOINC axes [16]. In another study in the United States targeting quantitative tests registered in PCORnet, a large-scale research consortium of more than 60 institutions, the reported error rate for LOINC mapping was 4.6% (0.4% for clinical chemistry tests and 7.5% for hematology tests) [17]. When comparing LOINC mappings prepared for the clinical data model for tests commonly conducted in seven major university hospitals in South Korea, 169 out of 961 LOINC codes (17.6%) did not match [unpublished data].

The primary reasons for LOINC mapping errors are the absence of a LOINC code for the test, omission or neglect of LOINC mapping by the institution, or incorrect mapping to a similar code that is not the gold standard [18]. In Korea, rather than relying on the less accurate bottom-up approach, where each medical institution manually maps LOINC codes, plans are underway to introduce LOINC mapping codes to each institution through a top-down approach via a proficiency testing (PT) provider, the Korean Association of External Quality Assessment Service (KEQAS). In this approach, each laboratory enters their test local codes for the PT tests. Then, the KEQAS links the corresponding LOINC codes to these test codes and provides them to the laboratories. This approach has the advantage of accurate LOINC mapping for common tests, but it cannot be applied to tests not available through the PT provider via external quality assessment (EQA). Similarly, the Royal College of Pathologists of Australasia Quality Assurance Programs has implemented Informatics External Quality Assurance as a pilot program. In this project, using the Health Level Seven (HL7) v2.4 transmission standard, laboratories send EQA test results mapped to LOINC to the PT provider [19].

SNOMED CT is a standard terminology for the laboratory test field that can be utilized for test naming, result reporting, and the coding of clinical concepts, such as clinical findings, body sites, and specimen types [20]. Originating in 1965 as the Systematized Nomenclature of Pathology, SNOMED CT has evolved into an international terminology that includes all medical specialties. It comprises over 352,000 concepts, including clinical findings (32%), procedures (18%), and body structures (11%). Unlike LOINC, concepts are arranged hierarchically, with relationships to similar concepts displaying varying degrees of specificity [20]. SNOMED CT enables the creation of new codes by combining concepts (post-coordinated codes) in addition to using pre-coordinated codes. It remains the preferred terminology for result reporting, result interpretation, and specifying specimen type, source, and condition acceptability in the United States Core Data for Interoperability.

In Australia, the Royal College of Pathologists of Australasia

https://doi.org/10.3343/alm.2024.0258

has been leading the Pathology Terminology and Information Standardisation Projects since 2011 [21]. The standards developed here include test terminology used in Australia, standardized units, secure reporting rendering, information models, harmonized reference intervals, and best practice guidelines for safe pathology requests and reporting. Laboratories are required to use SNOMED CT for requesting tests and standardize reporting using LOINC codes for the test result name [21].

DATA FORMAT STANDARDS

The second requirement for establishing laboratory big data is the standardization of reporting units and unit sizes (i.e., the number of significant figures; integers or decimal places) for reporting test results. An international standard for reporting test results is lacking, and most countries have no national standards [19, 22].

Cho, *et al*. [23] examined the reporting units and significant figures for 99 clinical chemistry tests across 99 Korean laboratories. They discovered that for 93 tests (93.9%), >80% of laboratories utilized the same units, whereas only 46 tests (46.5%) had consistent units across all laboratories. For high-sensitive troponin I, although clinical guidelines [24, 25] recommend reporting results in ng/L, 52% of the Korean institutions used ng/mL. Failure to consider the units used by institutions when comparing troponin results among laboratories may lead to errors in data integration [23]. Similarly, in a troponin survey conducted in Europe, diverse units were utilized [26, 27]. Thus, standardizing units remains a global challenge.

One solution to this issue involves emulating the approach in Australia of nationally standardizing appropriate units for each test item [28]. Alternatively, EQA providers could designate suitable reporting units for each test and receive EQA results from participating laboratories in a uniform manner, with laboratories adhering to the reporting format specified by the EQA provider. KEQAS allows only one unit per test, whereas the American EQA provider, CAP, permits multiple units for a single test. Reaching a consensus among international EQA providers regarding standard reporting units would be beneficial.

When reporting quantitative test results, determining the clinically or statistically appropriate number of significant figures presents challenges. In cases where clinical guidelines specify the reporting unit size, adherence to these guidelines is advisable. For instance, serum creatinine is recommended to be reported as an integer (μmol/L) or with two decimal places (mg/dL), the estimated glomerular filtration rate (eGFR) as an integer (mL/min/1.73 m^2) [29], cardiac troponin as an integer (ng/L) [24, 25], and HbA1c with one decimal place (mmol/L or %) [30]. For tests not specified in clinical guidelines, given the variety of instruments and reagents used in laboratories and the consideration of measurement uncertainty, further research is required to determine the appropriate number of decimal places for reporting results [23]. However, when the performance of testing instruments and reagents do not significantly differ, relevant professional societies or EQA providers may recommend suitable decimal places. In this context, the safety of result interpretation must be considered [31].

EQUATION STANDARDS

Test results such as calculated low-density lipoprotein cholesterol (LDL-C), total globulins, unconjugated bilirubin, corrected total calcium, international normalized ratio, creatinine clearance, plasma osmolality, and anion gap can be obtained using specific tests' results. Additionally, analyzers, especially hematology and blood gas analyzers, calculate and report various test results, such as mean corpuscular volume, mean corpuscular hemoglobin (Hb), and mean corpuscular Hb concentration HCO_3^- [32].

Most calculated tests are based on a single formula, but for some specific tests, multiple formulas are used. For example, in clinical settings, two methods are used to calculate the anion gap; one includes the potassium value, whereas the other does not [33]. Including results from both anion gap formulas in laboratory big data may decrease the accuracy of analyses. Similarly, multiple formulas exist for calculating LDL-C [34, 35].

The GFR is widely accepted as the best overall measure of kidney function with a single numeric expression. Creatinine-based eGFR calculations are widely used in clinical practice globally. However, there are various formulas for calculating the eGFR, and new formulas are continually being developed and evaluated [36-41]. Values can significantly differ depending on the formula used.

Jeong, *et al*. [42] compared the Modification of Diet in Renal Disease (MDRD) and Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equations in 5,822 Koreans. They found a significant difference between the two formulas at stages 1, 2, and 3 of chronic kidney disease (CKD). Levey, *et al*. [43], using data from the National Health and Nutrition Examination Survey (NHANES) from 1999–2006, and Stevens, *et al*. [44], in a study involving 116,321 participants from the Kidney Early Evaluation Program, reported differences in the prevalence of CKD when

using the MDRD and CKD-EPI formulas. Accordingly, using both formulas interchangeably in big data analyses is inappropriate. To include creatinine-based eGFR or anion gap in big data, laboratories must clearly indicate the specific equation used when reporting results. Converting and utilizing the necessary data with a different formula would be straightforward.

LABORATORY RESULT STANDARDIZATION AND HARMONIZATION

The term "standardization" has traditionally been used when equivalent results, within medically meaningful limits, are achieved among different measurement procedures through calibration traceable to a reference measurement procedure or certified reference material [45]. International Organization for Standardization (ISO) standard 17511 outlines five reference system categories. Categories 1, 2, and 3, which have a defined reference method, are classified for standardization, whereas the other categories are aimed at harmonization [46]. The Joint Committee for Traceability in Laboratory Medicine (JCTLM) is the international organization responsible for standardization. As of April 2024, the JCTLM database comprises 291 entries of available higher-order certified reference materials, representing 180 measurands across 12 analyte categories. Additionally, it includes 234 reference measurement methods, representing 110 measurands in 10 analyte categories [47]. In Korea, significant efforts are being made to standardize tests through various activities [48-50]. A reference method, because of its time-consuming, labor-intensive nature, is impractical for routine work. Instead, laboratories indirectly assess the accuracy of their test results through accuracy-based PT [51, 52]. For standardized test items, all qualified results theoretically can be included in laboratory big data. However, although the tests are considered standardized, the degree of standardization in the real world may vary by region, and clinical pathologists should be aware of this [53].

Harmonization is a generalization of the concept of standardization aimed at achieving equivalent results, within medically meaningful limits, among different measurement procedures using a scientifically sound approach [45]. This does not necessarily involve standardizing the entire testing process but aligning outcomes to be clinically comparable [46]. Harmonization does not rely on the availability of primary reference materials or primary reference measurement procedures. It includes standardization and addresses tests that cannot be calibrated through traceability to a primary reference measurement. The International Consortium for Harmonization of Clinical Laboratory Results is responsible for harmonization.

The harmonization status of test items can be assessed using PT. Currently, most harmonization tests are conducted within peer groups rather than the total group to ensure accurate comparisons of performance metrics. Tests that can theoretically be harmonized within a total group include the following categories. The first category includes tests for which most laboratories utilize the same test method, such as protein testing with the biuret method and albumin testing with bromocresol green. The second category includes tests predominantly supplied by a single manufacturer, such as pro-brain natriuretic peptide and troponin I. Lastly, there are tests for which international organizations have established harmonization standards, such as thyroid function tests. However, even for tests within these categories, significant challenges arise owing to substantial variation among instruments or reagents [54]. Therefore, most PT providers evaluate PT results for such tests by dividing the total group into peer groups, limiting harmonization to peer groups.

From a big data perspective, the task of building comprehensive laboratory big data across laboratories that use various test instruments yielding different test values presents significant challenges.

LABORATORY RESULT TRANSFERENCE

The relationship between the volume of big data and the performance of machine learning models is a fundamental aspect of data science. As the amount of big data increases, the performance of machine learning models generally significantly improves [55, 56]. Additionally, larger datasets mitigate the risk of bias that smaller datasets might introduce, leading to more accurate and robust predictions. This is particularly crucial in healthcare, where high accuracy is paramount.

In practical terms, in laboratory medicine, the method of transference through an equation is used when setting reference intervals, as introduced in CLSI guideline C28-A3c [57]. The best example of transference through an equation is the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER), whose main objective is to establish a comprehensive database of reference intervals for blood test results in children and adolescents [58]. The first comprehensive CALIPER direct reference interval study was published in 2012, with age- and sex-specific reference intervals reported for 40 biochemical markers, including common chemistry markers, enzymes, lipids, lipoproteins, and proteins [59]. CALIPER has established a com-

https://doi.org/10.3343/alm.2024.0258

prehensive database of age- and sex-specific reference intervals for more than 100 pediatric disease biomarkers [58]. All direct reference interval studies used a robust statistical method based on CLSI guideline C28-A3. Most CALIPER reference intervals originally established for Abbott assays have since been transferred to assays from other manufacturers, including Beckman Coulter, Ortho, Roche, and Siemens [60-63]. To transfer reference intervals, CALIPER uses an approach in accordance with CLSI guidelines C28-A3c [57] and EP09c [64].

When collecting test results from multiple institutions that employ technologies from various manufacturers to build big data, performing transference ensures that test outcomes resemble those obtained with representative test methods, as demonstrated in the CALIPER study [58]. This approach facilitates the integration of all test results into big data, thereby improving performance of machine learning and big data

However, merely possessing more data does not automatically guarantee improved performance; the quality of data, their relevance to the problem at hand, and the capability of machine learning algorithms to efficiently learn from large volumes of data are equally critical [65, 66].

LABORATORY RESULT ADJUSTMENT

Laboratory big data are accumulated over a long period utilizing various calibrators and reagents. However, manufacturing conditions may not always be identical between different batches in producing, leading to lot-to-lot variations [67].

Clinically significant lot-to-lot variation, when undetected, may affect test outcomes, potentially leading to incorrect patient diagnoses and posing a risk to patient care. Test results can vary depending on the lot [68-72]. Thaler, *et al*. [69] reported a 0.5% difference in two immunoturbidimetric HbA1c reagents due to lot-to-lot variation and found that lot-to-lot variation in these tests can lead to patients being incorrectly diagnosed as having diabetes mellitus and erroneously medicated [69]. Kim, *et al*. [70] analyzed the accuracy-based total cholesterol PT in the KEQAS from 2016 to 2018 and observed that the average percent bias across all participants was +0.14%. However, institutions using Roche products during that period exhibited an average bias of −3.0%. This negative bias for Roche products persisted until 2020. In response, Roche adjusted the total cholesterol calibrator set point by +2.4% in early 2021. Subsequent PT results indicated the correction of the negative bias in accuracy-based total cholesterol PT. When integrating total cholesterol results obtained using Roche products during the period of negative bias into big data, adjusting the values is essential for the accuracy of the big data.

In national surveys conducted over a long period, lot-to-lot reagent variations can lead to inconsistencies in test results. Hence, adjustments are required when such variations occur. There have been cases where the results of long-term national surveys have been adjusted using calibration equations. For example, in the NHANES conducted by the Centers for Disease Control and Prevention (CDC) in the United States since 1971, values for serum creatinine, cystatin C, high-density lipoprotein cholesterol (HDL-C), and 25-hydroxyvitamin D have been corrected using calibration equations [73-75]. In the Korea NHANES, HDL-C values from 2008–2015 showed substantial positive bias when compared with the CDC reference method values. To address this issue, the original HDL-C values were adjusted to CDC reference method values using calibration equations [76].

Lot-to-lot verification in laboratory medicine is a crucial aspect of monitoring the long-term stability of a measurement procedure. In current practice, subsequent reagent lots are generally compared with each other [77]. CLSI EP-26A provides an extensive description of this approach using patient samples to compare subsequent lots [78]. Additionally, EP-31-A-IR provides guidance on how laboratories can verify the comparability of individual patient results within a healthcare system [79]. The practice is challenged by resource requirements and uncertainty regarding the experimental design and statistical analysis that are optimal for individual laboratories [80].

When constructing accurate big data, maintaining long-term test result stability is crucial. Laboratories must identify lots with variations and refrain from using such lots to ensure consistent test outcomes. Loh, *et al*. [80] suggested that collaborations among all stakeholders, including regulatory bodies, manufacturers, and laboratory medicine institutions, are key to developing a balanced system whereby regulatory, manufacturing, and clinical requirements are met to minimize differences among reagent lots and ensure patient safety.

Since 2011, Korean Society of Laboratory Medicine has collaborated with the Korea Disease Control and Prevention Agency to implement the Academia–Government Collaboration for Laboratory Medicine Standardization in Korea. This initiative aims to standardize several tests, including creatinine, total cholesterol, LDL-C, HDL-C, triglyceride, and HbA1c tests. Additionally, since 2017, they have been evaluating and issuing certificates for each combination of calibrator lot, reagent lot, and instrument [48].

When testing is performed using lots for which undetected lot-to-lot variation exists, considering methods to adjust or modify test values becomes essential. For this purpose, data elements required for each test result include test analyzer, test method, calibrator lot number, calibration value for each test, reagent lot number, slope factor, interceptor factor, and software version of the instrument, to facilitate result adjustment or transference.

DATA CERTIFICATION

For building big data based on real-world laboratory results, data reliability is more crucial. [52]. To achieve high-quality outcomes through big data analysis, only high-quality data should be used. The certification of laboratory data, achieved through the periodic reanalysis of accumulated EQA data, could be an option for evaluating the quality of laboratory results [52].

Cho, *et al*. [81] compared the results of participants for seven tests using EQA data. Institutions with low-quality ratings exhibited significantly biased results. They noted that the EQA results of participants could be used as surrogates for the quality of real-world patient data. EQA evaluates results at a single point in time, and many laboratories tend to adhere to relatively lenient criteria in their evaluations. Therefore, passing the EQA criteria does not guarantee the highest data quality. Hence, not all data that pass EQA criteria are suitable for building big data. Therefore, when constructing laboratory big data architecture, evaluating the data quality for big data analysis and excluding low-quality data is imperative.

Kim, *et al*. [52] proposed a model for reanalyzing EQA results to evaluate real-world laboratory results for big data research, emphasizing the reliability of laboratory results for quality management in big data research. Furthermore, they advocated for implementing a certification system for laboratory data quality at the national level.

CONCLUSIONS

With Industry 4.0, integrating real-world laboratory data into high-quality big data is essential but presents challenges like unstandardized terminology and insufficient result harmonization. This review highlights the key elements needed to build high-quality medical big data, emphasizing the vital role of laboratory medicine professionals in preparing for the healthcare big data and AI era.

ACKNOWLEDGEMENTS

We are truly grateful to Prof. Tae-Dong Jeong (Ewha Womans University College of Medicine), Prof. Kyunghoon Lee (Seoul National University Bundang Hospital), Prof. Jae-Woo Chung (Dongguk University Ilsan Hospital), Prof. Eun-Jung Cho (Hallym University Dongtan Sacred Heart Hospital), Prof. Shinae Yu (Haeundae Paik Hospital, Inje University College of Medicine), Prof. Kueny-oul Park (Sanggye Paik Hospital, Inje University College of Medicine), and Prof. Hangsik Shin (Asan Medical Center, University of Ulsan College of Medicine) for reviewing the manuscript drafts and providing valuable feedback.

AUTHOR CONTRIBUTIONS

Conceptualization: Min WK; Investigation: Kim S; Visualization: Kim S and Min WK; Supervision: Min WK; Writing – original draft: Kim S; Writing – review & editing: Kim S and Min WK. All authors have read and approved the final manuscript.

CONFLICTS OF INTEREST

None declared.

RESEARCH FUNDING

None declared.

REFERENCES

1. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2:230-43.
2. Ehsani-Moghaddam B, Martin K, Queenan JA. Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. Health Inf Manag 2021;50:88-92.
3. Daniel C, Serre P, Orlova N, Bréant S, Paris N, Griffon N. Initializing a hospital-wide data quality program. The AP-HP experience. Comput Methods Programs Biomed 2019;181:104804.
4. Aerts H, Kalra D, Sáez C, Ramírez-Anguita JM, Mayer MA, Garcia-Gomez JM, et al. Quality of hospital electronic health record (EHR) data based on the International Consortium for Health Outcomes Measurement (ICHOM) in heart failure: pilot data quality assessment study. JMIR Med Inform 2021;9:e27842.
5. Liaw ST, Guo JGN, Ansari S, Jonnagaddala J, Godinho MA, Borelli AJ, et al. Quality assessment of real-world data repositories across the data life cycle: a literature review. J Am Med Inform Assoc 2021;28:1591-9.
6. Wu J, Wang C, Toh S, Pisa FE, Bauer L. Use of real-world evidence in regulatory decisions for rare diseases in the United States–current status and future directions. Pharmacoepidemiol Drug Saf 2020;29:1213-8.

https://doi.org/10.3343/alm.2024.0258

7. U.S. FDA. Use of real-world evidence to support regulatory decision-making for medical devices. Guidance for industry and food and drug administration staff. FDA-2016-D-2153. Rockville, MD: U.S. Department of Health and Human Services Food and Drug Administration, 2016.
8. Blatter TU, Witte H, Nakas CT, Leichtle AB. Big data in laboratory medicine—FAIR quality for AI? Diagnostics (Basel) 2022;12:1923.
9. Ronzio L, Cabitza F, Barbaro A, Banfi G. Has the flood entered the basement? A systematic literature review about machine learning in laboratory medicine. Diagnostics (Basel) 2021;11:372.
10. Drenkhahn C and Ingenerf J. The LOINC content model and its limitations of usage in the laboratory domain. Stud Health Technol Inform 2020;270:437-42.
11. Regenstrief Institute Inc. Logical Observation Identifiers Names and Codes (LOINC) version 2.77. https://loinc.org/ (Updated on April 2024).
12. Cholan RA, Pappas G, Rehwoldt G, Sills AK, Korte ED, Appleton IK, et al. Encoding laboratory testing data: case studies of the national implementation of HHS requirements and related standards in five laboratories. J Am Med Inform Assoc 2022;29:1372-80.
13. Parr SK, Shotwell MS, Jeffery AD, Lasko TA, Matheny ME. Automated mapping of laboratory tests to LOINC codes using noisy labels in a national electronic health record system database. J Am Med Inform Assoc 2018;25:1292-300.
14. Liu CT, Wang LW, Hsu MH, Wen LL, Lai JS. A unified approach to adoption of laboratory LOINC in Taiwan. Healthcom 2007: Ubiquitous healthcare in aging societies - 2007 9th International Conference on e-Health Networking, Application and Services. 2007:144-9.
15. Stram M, Seheult J, Sinard JH, Campbell WS, Carter AB, de Baca ME, et al. A survey of LOINC code selection practices among participants of the College of American Pathologists Coagulation (CGL) and Cardiac Markers (CRT) proficiency testing programs. Arch Pathol Lab Med 2020;144:586-96.
16. Lin MC, Vreeman DJ, McDonald CJ, Huff SM. Correctness of voluntary LOINC mapping for laboratory Tests in three Large Institutions. AMIA Annu Symp Proc 2010;2010:447-51.
17. McDonald CJ, Baik SH, Zheng Z, Amos L, Luan X, Marsolo K, et al. Mismappings between a producer's quantitative test codes and LOINC codes and an algorithm for correcting them. J Am Med Inform Assoc 2023;30:301-7.
18. Bhargava A, Kim T, Quine DB, Hauser RG. A 20-year evaluation of LOINC in the United States' largest integrated health system. Arch Pathol Lab Med 2020;144:478-84.
19. Hardie RA, Moore D, Holzhauser D, Legg M, Georgiou A, Badrick T. Informatics External Quality Assurance (IEQA) Down Under: evaluation of a pilot implementation. J Lab Med 2018;42:297-304.
20. Rychert J. In support of interoperability: a laboratory perspective. Int J Lab Hematol 2023;45:436-41.
21. The Royal College of Pathologists of Australasia. Pathology terminology and information standardisation projects. https://www.rcpa.edu.au/Library/Practising-Pathology/PTIS (Updated on April 2024).
22. Hauser RG, Gisriel S, El-Khoury J. The surprising absence of a laboratory result standard. Am J Clin Pathol 2022;157:642-3.
23. Cho J, Jeong TD, Moon SY, Chung JW, Nam Y, Lee SG, et al. Current status of reporting units and unit sizes of quantitative test results of clinical chemistry in Korea. Lab Med Online 2022;12:292-303.
24. Thygesen K, Alpert JS, Jaffe AS, Simoons ML, Chaitman BR, White HD, et al. Third universal definition of myocardial infarction. J Am Coll Cardiol 2012;60:1581-98.
25. Barth JH, Panteghini M, Bunk DM, Christenson RH, Katrukha A, Noble JE, et al. Recommendation to harmonize the units for reporting cardiac troponin results. Clin Chim Acta 2014;432:166.
26. McKeeman GC and Auld PW. A national survey of troponin testing and recommendations for improved practice. Ann Clin Biochem 2015;52:527-42.
27. Secchiero S, Sciacovelli L, Plebani M. Harmonization of units and reference intervals of plasma proteins: state of the art from an external quality assessment scheme. Clin Chem Lab Med 2018;57:95-105.
28. National Pathology Accreditation Advisory Council (NPAAC). Requirements for information communication and reporting. 5th ed. TRIM: D22-32436. Australia: Australian Commission on Safety and Quality in Health Care, 2022.
29. Kidney Disease: Improving Global Outcomes CKDWG. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int 2024;105:S117-S314.
30. Jones GRD, Barker G, Goodall I, Schneider HG, Shephard MDS, Twigg SM. Change of HbA1c reporting to the new SI units. Med J Aust 2011;195:45-6.
31. Sinnott M, Eley R, Steinle V, Boyde M, Trenning L, Dimeski G. Decimal numbers and safe interpretation of clinical pathology results. J Clin Pathol 2014;67:179-81.
32. Coskun A. Westgard multirule for calculated laboratory tests. Clin Chem Lab Med 2006;44:1183-7.
33. Kraut JA and Madias NE. Serum anion gap: its uses and limitations in clinical medicine. Clin J Am Soc Nephrol 2007;2:162-74.
34. Hong J, Gu H, Lee J, Lee W, Chun S, Han KH, et al. Intuitive modification of the Friedewald formula for calculation of LDL-cholesterol. Ann Lab Med 2023;43:29-37.
35. Martins J, Rossouw HM, Pillay TS. How should low-density lipoprotein cholesterol be calculated in 2022? Curr Opin Lipidol 2022;33:237-56.
36. Jeong TD, Hong J, Lee W, Chun S, Min WK. Accuracy of the new creatinine-based equations for estimating glomerular filtration rate in Koreans. Ann Lab Med 2023;43:244-52.
37. Meeusen JW, Kasozi RN, Larson TS, Lieske JC. Clinical impact of the refit CKD-EPI 2021 creatinine-based eGFR equation. Clin Chem 2022;68:534-9.
38. Inker LA, Eneanya ND, Coresh J, Tighiouart H, Wang D, Sang Y, et al. New creatinine- and cystatin C-based equations to estimate GFR without race. N Engl J Med 2021;385:1737-49.
39. Miller WG, Kaufman HW, Levey AS, Straseski JA, Wilhelms KW, Yu HE, et al. National Kidney Foundation Laboratory Engagement Working Group recommendations for implementing the CKD-EPI 2021 race-free equations for estimated glomerular filtration rate: practical guidance for clinical laboratories. Clin Chem 2022;68:511-20.
40. Pottel H, Delanaye P, Cavalier E. Exploring renal function assessment: creatinine, cystatin C, and estimated glomerular filtration rate focused on the European Kidney Function Consortium equation. Ann Lab Med 2024;44:135-43.
41. Lee HS, Bae GE, Lee JE, Park HD. Effect of two cystatin C reagents and four equations on glomerular filtration rate estimations after standardization. Ann Lab Med 2023;43:565-73.
42. Jeong TD, Lee W, Chun S, Lee SK, Ryu JS, Min WK, et al. Comparison of the MDRD study and CKD-EPI equations for the estimation of the glomerular filtration rate in the Korean general population: the fifth Korea National Health and Nutrition Examination Survey (KNHANES V-1), 2010. Kidney Blood Press Res 2013;37:443-50.
43. Levey AS, Stevens LA, Schmid CH, Zhang YL, Castro AF 3rd, Feldman HI, et al. A new equation to estimate glomerular filtration rate. Ann Intern Med 2009;150:604-12.
44. Stevens LA, Li S, Kurella Tamura M, Chen SC, Vassalotti JA, Norris KC, et al. Comparison of the CKD Epidemiology Collaboration (CKD-EPI) and Modification of Diet in Renal Disease (MDRD) study equations: risk fac-

tors for and complications of CKD and mortality in the Kidney Early Evaluation Program (KEEP). Am J Kidney Dis 2011;57(S2):S9-16.
45. Miller WG. Harmonization: its time has come. Clin Chem 2017;63:1184-6.
46. International Organization for Standardization. In vitro diagnostic medical devices—requirements for establishing metrological traceability of values assigned to calibrators, trueness control materials and human samples. Geneva: International Organization for Standardization, 2003.
47. Joint Committee for Traceability in Laboratory Medicine. JCTLM Database: higher-order reference materials, methods and services v1.45. https://www.jctlmdb.org/ (Updated on April 2024).
48. Lee S, Yu J, Cho CI, Cho EJ, Jeong TD, Kim S, et al. Impact of academia-government collaboration on laboratory medicine standardization in South Korea: analysis of eight years creatinine proficiency testing experience. Clin Chem Lab Med 2024;62:861-9.
49. Kim S, Lee K, Park HD, Lee YW, Chun S, Min WK. Schemes and performance evaluation criteria of Korean Association of External Quality Assessment (KEQAS) for improving laboratory testing. Ann Lab Med 2021;41:230-9.
50. Jeong TD, Cho EJ, Lee K, Lee W, Yun YM, Chun S, et al. Recent trends in creatinine assays in Korea: long-term accuracy-based proficiency testing survey data by the Korean Association of External Quality Assessment Service (2011–2019). Ann Lab Med 2021;41:372-9.
51. Kim S. Laboratory data quality evaluation in the big data era. Ann Lab Med 2023;43:399-400.
52. Kim S, Cho EJ, Jeong TD, Park HD, Yun YM, Lee K, et al. Proposed model for evaluating real-world laboratory results for big data research. Ann Lab Med 2023;43:104-7.
53. Kim S, Jeong TD, Lee K, Chung JW, Cho EJ, Lee S, et al. Quantitative evaluation of the real-world harmonization status of laboratory test items using external quality assessment data. Ann Lab Med 2024;44:529-36.
54. van Rossum HH, Holdenrieder S, Ballieux B, Badrick TC, Yun YM, Zhang C, et al. Investigating the current harmonization status of tumor markers using global external quality assessment programs: a feasibility study. Clin Chem 2024;70:669-79.
55. Ihde N, Marten P, Eleliemy A, Poerwawinata G, Silva P, Tolovski I, et al. A survey of big data, high performance computing, and machine learning benchmarks. In: Nambiar R and Poess M, eds. Technology Conference on Performance Evaluation and Bechnmarking. TPCTC 2021. Cham: Springer International Publishing, 2022:98-118.
56. Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: opportunities and challenges. Neurocomputing 2017;237:350-61.
57. Horowitz GL, Altaie S, Boyd JC, Ceriotti F, Garg U, Horn P, et al. Defining, establishing, and verifying reference intervals in the clinical laboratory. 3rd ed. CLSI EP28-A3C. Wayne, PA: Clinical and Laboratory Standards Institute, 2010.
58. Adeli K, Higgins V, Trajcevski K, White-Al Habeeb N. The Canadian laboratory initiative on pediatric reference intervals: a CALIPER white paper. Crit Rev Clin Lab Sci 2017;54:358-413.
59. Colantonio DA, Kyriakopoulou L, Chan MK, Daly CH, Brinc D, Venner AA, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. Clin Chem 2012;58:854-68.
60. Estey MP, Cohen AH, Colantonio DA, Chan MK, Marvasti TB, Randell E, et al. CLSI-based transference of the CALIPER database of pediatric reference intervals from Abbott to Beckman, Ortho, Roche and Siemens clinical chemistry assays: direct validation using reference samples from the CALIPER cohort. Clin Biochem 2013;46:1197-219.
61. Abou El Hassan M, Stoianov A, Araújo PAT, Sadeghieh T, Chan MK, Chen Y, et al. CLSI-based transference of CALIPER pediatric reference intervals to Beckman Coulter AU biochemical assays. Clin Biochem 2015;48:1151-9.
62. Araújo PAT, Thomas D, Sadeghieh T, Bevilacqua V, Chan MK, Chen Y, et al. CLSI-based transference of the CALIPER database of pediatric reference intervals to Beckman Coulter DxC biochemical assays. Clin Biochem 2015;48:870-80.
63. Higgins V, Chan MK, Nieuwesteeg M, Hoffman BR, Bromberg IL, Gornall D, et al. Transference of CALIPER pediatric reference intervals to biochemical assays on the Roche cobas 6000 and the Roche Modular P. Clin Biochem 2016;49:139-49.
64. Budd JR, Durham AP, Gwise TE, Hawkins DM, Holland M, Iriarte B, et al. Measurement procedure comparison and bias estimation using patient samples. 3rd ed. CLSI EP09c. Wayne, PA: Clinical and Laboratory Standards Institute, 2018.
65. Gong Y, Liu G, Xue Y, Li R, Meng L. A survey on dataset quality in machine learning. Inf Softw Technol 2023;162:107268.
66. Fenza G, Gallo M, Loia V, Orciuoli F, Herrera-Viedma E. Data set quality in machine learning: consistency measure based on group decision making. Appl Soft Comput 2021;106:107366.
67. Thompson S and Chesher D. Lot-to-lot variation. Clin Biochem Rev 2018;39:51-60.
68. Algeciras-Schimnich A, Bruns DE, Boyd JC, Bryant SC, La Fortune KA, Grebe SKG. Failure of current laboratory protocols to detect lot-to-lot reagent differences: findings and possible solutions. Clin Chem 2013;59:1187-94.
69. Thaler MA, Iakoubov R, Bietenbeck A, Luppa PB. Clinically relevant lot-to-lot reagent difference in a commercial immunoturbidimetric assay for glycated hemoglobin A1c. Clin Biochem 2015;48:1167-70.
70. Kim JH, Cho Y, Lee SG, Yun YM. Report of Korean Association of External Quality Assessment Service on the accuracy-based lipid proficiency testing (2016-2018). J Lab Med Qual Assur 2019;41:121-9.
71. Böttcher S, van der Velden VHJ, Villamor N, Ritgen M, Flores-Montero J, Escobar HM, et al. Lot-to-lot stability of antibody reagents for flow cytometry. J Immunol Methods 2019;475:112294.
72. Kitchen AD and Newham JA. Lot release testing of serological infectious disease assays used for donor and donation screening. Vox Sang 2010;98:508-16.
73. Yetley EA, Pfeiffer CM, Schleicher RL, Phinney KW, Lacher DA, Christakos S, et al. NHANES monitoring of serum 25-hydroxyvitamin D: a roundtable summary. J Nutr 2010;140:2030S-45S.
74. Selvin E, Juraschek SP, Eckfeldt J, Levey AS, Inker LA, Coresh J. Calibration of cystatin C in the National Health and Nutrition Examination Surveys (NHANES). Am J Kidney Dis 2013;61:353-4.
75. Selvin E, Manzi J, Stevens LA, Van Lente F, Lacher DA, Levey AS, et al. Calibration of serum creatinine in the National Health and Nutrition Examination Surveys (NHANES) 1988-1994, 1999-2004. Am J Kidney Dis 2007;50:918-26.
76. Yun YM, Song J, Ji M, Kim JH, Kim Y, Park T, et al. Calibration of high-density lipoprotein cholesterol values from the Korea National Health and Nutrition Examination Survey data, 2008 to 2015. Ann Lab Med 2017;37:1-8.
77. Katzman BM, Ness KM, Algeciras-Schimnich A. Evaluation of the CLSI EP26-A protocol for detection of reagent lot-to-lot differences. Clin Biochem 2017;50:768-71.
78. CLSI. User evaluation of between-reagent lot variation. EP26-A. Wayne, PA: Clinical and Laboratory Standards Institute, 2013.
79. CLSI. Verification of comparability of patient results within one health care system. EP31-A-IR. Wayne, PA: Clinical and Laboratory Standards Institute, 2012.

80. Loh TP, Markus C, Tan CH, Tran MTC, Sethi SK, Lim CY. Lot-to-lot variation and verification. Clin Chem Lab Med 2023;61:769-76.
81. Cho EJ, Jeong TD, Kim S, Park HD, Yun YM, Chun S, et al. A new strategy for evaluating the quality of laboratory results for big data research: using external quality assessment survey data (2010–2020). Ann Lab Med 2023;43:425-33.