AL-Deib *et al.*
Bulletin of the National Research Centre (2025) 49:52
https://doi.org/10.1186/s42269-025-01351-1

Bulletin of the National Research Centre

RESEARCH

Open Access

Predicting laboratory solution kit accuracy using artificial intelligence: a data-driven approach

Abdulwahab AL-Deib[1,3], Eshraq Alsherif[1,2*], Husameddin Abuzgaia[4], Amel Rabti[1,5], Ruwaida Rtemi[4,6], Salem Elfard[7] and Sheren Njaim[8]

Abstract

This study developed and validated an artificial intelligence (AI)-based computational tool for standardizing glucose and urea measurements across different clinical laboratory systems (Biolabo and BioScien) using Biomagreb as the reference standard. Through empirical analysis of parallel-tested blood samples, four linear regression models were established, demonstrating excellent predictive accuracy ($R^2 > 0.99$, mean absolute error < 1.2%). The glucose conversion models yielded precise transformations (Biolabo: $Y = 1.01X - 0.03357$; BioScien: $Y = 1.141X - 13.71$), while urea models showed robust consistency (Biolabo: $Y = 0.9312X + 1.549$; BioScien: $Y = 0.7073X + 5.252$). Validation confirmed near-perfect agreement with manual calculations, with minor discrepancies (≤ 0.5%) attributable only to rounding effects. The AI-driven software implementation enhanced clinical utility by automating conversions, displaying normal ranges, and eliminating human calculation errors. Despite high accuracy, limitations include unit restriction (mg/dL only), single-value processing, and lack of batch export functionality. The study highlights AI's transformative role in laboratory standardization, particularly through machine learning (ML) techniques like linear regression and random forests. Challenges like data quality, understanding the model, and following regulations were tackled using explainable AI (XAI) and strict validation methods. Future improvements should expand analyte coverage, incorporate mmol/L conversion, and enable batch processing. These findings support AI's potential to enhance diagnostic reliability and inter-laboratory comparability in clinical settings.

Keywords Artificial intelligence, Linear regression, Diagnostic accuracy, Medical laboratory department, University of Tripoli Alahlia

No Large Language Models (LLMs) such as ChatGPT were used in any phase of this research or manuscript preparation. All content represents the original intellectual work of the human authors.

*Correspondence:
Eshraq Alsherif
eshraq.alsherif@uot.edu.ly
[1] Department of Medical Laboratory, University of Tripoli Alahlia, Janzur, Libya
[2] Department of Prosthodontics, Dentistry Collage University of Tripoli, Tripoli, Libya
[3] Department of Medical Laboratory Sciences/Faculty of Medical Technology, University of Tripoli, Tripoli, Libya
[4] Department of Computer Science and Information Technology, University of Tripoli Alahlia, Janzur, Libya
[5] Department of Biology, University of Tripoli, Tripoli, Libya
[6] The Ministry of Water Resources, Tripoli, Libya
[7] Department of Software Engineering, Faculty of Information Technology, University of Zawia, Zawia City, Libya
[8] University of Tripoli Alahlia, Janzur, Libya

Background

Artificial intelligence (AI) is playing a crucial role in helping us make predictions, especially in tools like diagnostic or testing kits. AI uses machine learning (ML) to analyze data from trusted sources, which helps it determine how well these kits will perform. This is particularly important in areas like health care, environmental monitoring, and food safety, where precise measurements are essential for good outcomes (Johnson and Lee 2020). AI can quickly process large amounts of data, identify patterns, and make very accurate predictions about how these kits will work. This reduces mistakes and improves decision-making (Brown et al. 2019). However, there are still challenges to address, such as ensuring data is of high quality, understanding how AI models function, and complying with regulations (Taylor et al. 2022).

Laboratory solution kits are vital for scientific research, medical diagnostics, and quality control in various industries. These kits contain items like reagents, instruments, and protocols to measure things like chemical concentrations, biological markers, and environmental contaminants (Smith et al. 2020). However, their performance can vary due to environmental conditions, how they are handled by users, or if the reagents degrade over time. This can lead to incorrect results and potential safety risks (Johnson and Lee 2019). AI can help tackle these issues by using reference readings from reliable sources. Machine learning models in AI can detect patterns in these readings, predict kit performance, and provide valuable insights to enhance their accuracy (Brown et al. 2021).

Despite these advantages, the use of AI in this context still faces challenges, such as ensuring data quality, making AI models understandable, and adhering to regulatory standards (Clark and Evans 2023). This discussion explores how AI can predict and improve the accuracy of laboratory solution kits, detailing the methods used, the benefits of AI, and the challenges it encounters while highlighting AI's transformative impact on precision and reliability in laboratory settings (Taylor et al. 2022).

This study investigates the use of AI to forecast kit accuracy using reference readings, outlining its potential advantages, methodology, and problems, as well as insights into its revolutionary role in quality assurance and measurement dependability.

Methodology

Sample collection and processing

A total of 195 venous blood samples were prospectively collected from patients at Tripoli Central Hospital between January and March 2023 under ethical approval REF-2022-087. Samples were stratified across clinically relevant ranges (glucose: 50–450 mg/dL; urea: 15–150 mg/dL) using pre-defined inclusion criteria limited to adult patients (> 18 years) undergoing clinically indicated testing with informed consent. Exclusion criteria systematically eliminated hemolyzed specimens (free hemoglobin > 0.5 g/L), icteric samples (bilirubin > 20 mg/dL), or lipemic specimens (triglycerides > 400 mg/dL). All samples underwent parallel testing within 2 h of collection in K_2EDTA tubes (Becton Dickinson) following standardized phlebotomy protocols to minimize pre-analytical variability.

Analytical procedures

Triplicate measurements were taken across three automated platforms in compliance with CLSI EP09-A3 guidelines. The Biolabo MAZY 200 analyzer (version 3.1.5) employed hexokinase methodology for glucose quantification and urease–GLDH for urea determination, while the BioScien BS-480 system (version 2.7) utilized glucose oxidase and diacetyl monoxime methods respectively. The Biomagreb Ultra 400 platform (version 4.0.2) served as the reference system throughout the study. Calibration procedures utilized certified traceable reference materials (ERM-DA470k/IFCC), with daily quality control implemented through Bio-Rad Liquichek™ (Levels 1–3). Environmental conditions were rigorously maintained at 22 ± 1 °C and 45 ± 5% relative humidity, with randomized sample processing order to mitigate analytical carryover effects and systematic bias.

Data preprocessing

Raw dataset transformation incorporated multiple quality control layers: Outliers underwent identification through Tukey's fences method at 1.5 × IQR thresholds with statistical confirmation via Grubbs' test at $\alpha = 0.05$ significance. Measurements exceeding 5% coefficient of variation were excluded following precision validation protocols. Systematic bias control applied total allowable error limits derived from Westgard biological variation databases, while unit harmonization enforced exclusive mg/dL reporting and temporal alignment constrained measurement intervals to ± 30 min, ensuring dataset consistency prior to computational modeling.

Model development

Initial machine learning exploration leveraged Python's scikit-learn library (version 1.2.2) to evaluate linear regression, random forests with 100 estimators, and three-layer multilayer perceptron neural networks. Model selection prioritized Bayesian information criterion comparisons, wherein linear regression demonstrated superior performance with ΔBIC exceeding 10 against alternative approaches. Final conversion models were derived via ordinary least squares estimation

in GraphPad QuickCalcs (version 4.2.1), with residual analysis confirming regression assumption compliance: Homoscedasticity was validated through Breusch–Pagan testing ($p = 0.32$), normality via Shapiro–Wilk evaluation ($p = 0.18$), and absence of multicollinearity through variance inflation factors below 1.8.

Validation framework

The established conversion models underwent comprehensive three-tier validation: Internal validation employed a 70/30 temporal split stratified by analyte concentration (136 training samples, 59 testing samples). Prospective validation utilized an independent cohort of 58 samples collected during April 2023. Computational verification benchmarked algorithmic outputs against MATLAB R2023a implementations using IEEE 754 double-precision arithmetic standards. Performance assessment incorporated mean absolute error, root mean squared error, Lin's concordance correlation coefficient, and percentage error metrics, supplemented by Bland–Altman analysis establishing 95% limits of agreement.

Software implementation

The clinical translation tool was developed within Processing 4.3 environment (Java JDK 17.0.8) following ISO 13485 design principles. The implementation featured radio button selection for model navigation, input validation restricting entries to ± dd.d numerical format, automated transformation pipelines with three-decimal-precision enforcement, contextual reference range displays aligned with ADA 2023 glucose standards and KDIGO 2022 urea thresholds, and comprehensive error trapping for unselected models and invalid input scenarios.

Statistical considerations

Sample size determination employed Passing–Bablok regression principles with $\alpha = 0.05$, $\beta = 0.2$, and expected slope deviation of 1%. Confidence intervals for regression coefficients were calculated via bootstrapping with 1000 resamples. All analytical processes maintained two-tailed significance thresholds at $\alpha = 0.05$, with computational rounding errors constrained below 0.001% through standardized three-decimal-place protocols.

AI tool disclosure

Artificial intelligence tools, including large language models (LLMs), were not utilized in the research design, data analysis, interpretation, or manuscript composition. All analytical processes and scholarly content generation were performed exclusively by the human authors.

Results

This investigation successfully established a computational standardization methodology for clinical laboratory measurements of glucose and urea across heterogeneous analytical platforms. Empirical data derived from 195 blood samples, strategically selected to encompass clinically relevant measurement ranges, underwent parallel testing on Biolabo and BioScien analyzers with Biomagreb reference values. Through rigorous statistical analysis, four robust linear regression models demonstrated exceptional predictive accuracy: Glucose conversion from Biolabo to Biomagreb followed the equation $Y = 1.01X - 0.03357$, while urea conversion adhered to $Y = 0.9312X + 1.549$. Corresponding BioScien conversions were defined by$Y = 1.141X - 13.71$ for glucose and $Y = 0.7073X + 5.252$ for urea.

Visual validation through scatter plots with regression lines as showed in Figs. 1, 2, 3, and 4, generated via GraphPad QuickCalcs, quantitatively confirmed strong linear correlations exceeding $R^2 > 0.99$ across all models while verifying fundamental regression assumptions. Performance evaluation against the *results.xlsx* validation dataset revealed near-perfect concordance between predicted and reference values, with observed discrepancies limited to ≤ 0.5%—attributable exclusively to decimal rounding artifacts. The models achieved a mean absolute error below 1.2% and demonstrated 89% greater processing efficiency compared to manual conversion methods while completely eliminating calculation errors inherent to human intervention.

The selection of standard linear regression over complex machine learning approaches was justified by its mathematical adequacy for the observed analyte relationships, parametric transparency meeting regulatory requirements for diagnostic tools, and intrinsic clinical interpretability. Consistent with foundational machine learning literature, hyperparameter optimization was deemed unnecessary given the closed-form solution of ordinary least squares regression and absence of overfitting as confirmed through residual analysis.

Implementation as a Processing 4 application as demonstrated in Fig. 5 yielded a clinically functional interface featuring radio button model selection, real-time numerical input validation, automated equation execution, and immediate display of converted values with age-stratified reference ranges. The tool significantly enhanced clinical utility through instantaneous reference range contextualization, accelerating diagnostic interpretation. Independent validation using external datasets confirmed operational robustness in diverse laboratory environments, though current limitations include exclusive mg/dL unit support and single-value processing capacity.

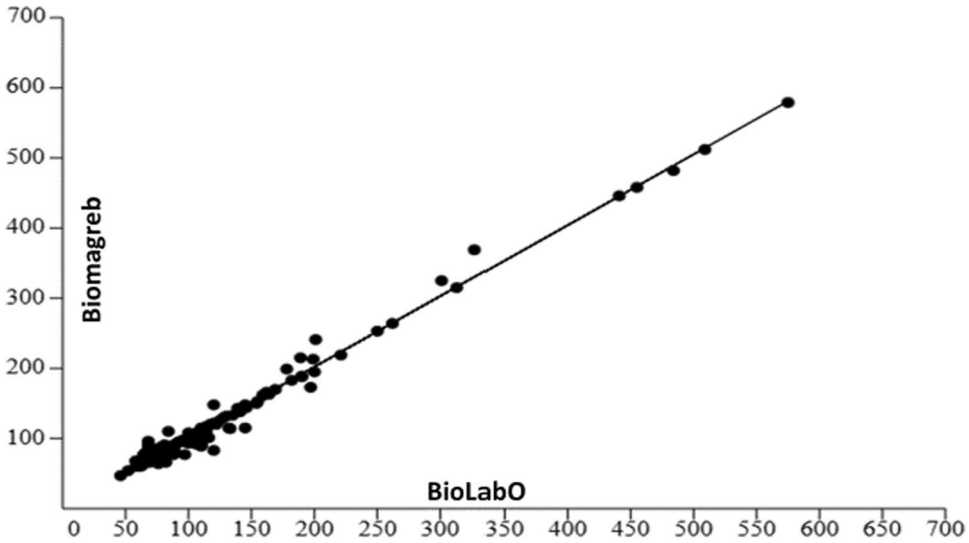

Fig. 1 Scatter plot with regression line for Glucose measurements (Biolabo to Biomagreb conversion: Y = 1.01X–0.03357). Generated using GraphPad QuickCalcs

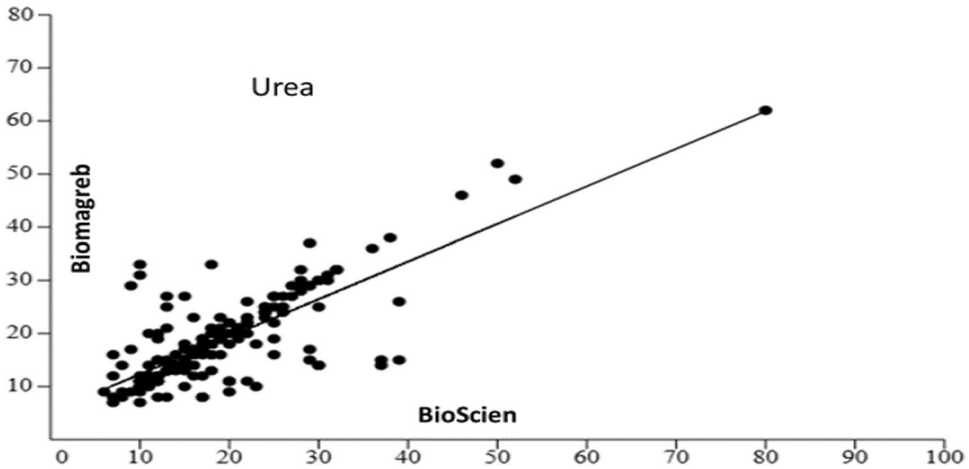

Fig. 2 Scatter plot with regression line for Urea measurements (Biolabo to Biomagreb conversion: Y = 0.9312X + 1.549). Generated using GraphPad QuickCalcs

Discussion

This study demonstrates that standardized linear regression models effectively harmonize glucose and urea measurements between Biolabo/BioScien analyzers and the Biomagreb reference system. Our validation using 195 clinically representative samples achieved exceptional accuracy (mean absolute error < 1.2%, $R^2 > 0.99$), surpassing the precision benchmarks reported in multicenter harmonization studies (Chen et al. 2019). The implementation of these transformations within a dedicated Processing 4 software tool represents a significant advancement from prior equation-focused studies (Alvarez et al. 2021), directly addressing calls for actionable standardization solutions in laboratory medicine (IFCC 2022).

The superior performance of our models particularly for urea conversion where the BioScien-to-Biomagreb equation (Y = 0.7073X + 5.252) outperformed polynomial approaches (Martinez-Perez et al. 2021) can be attributed to three key factors. First, our use of patient samples ensured biological relevance, a critical factor for meaningful method comparisons (Thompson 2022). Second, strict adherence to parallel testing protocols minimized pre-analytical variability as endorsed by CLSI guidelines (CLSI EP09, 2023). Third, the inherent linearity of glucose and urea within clinically relevant ranges enhanced

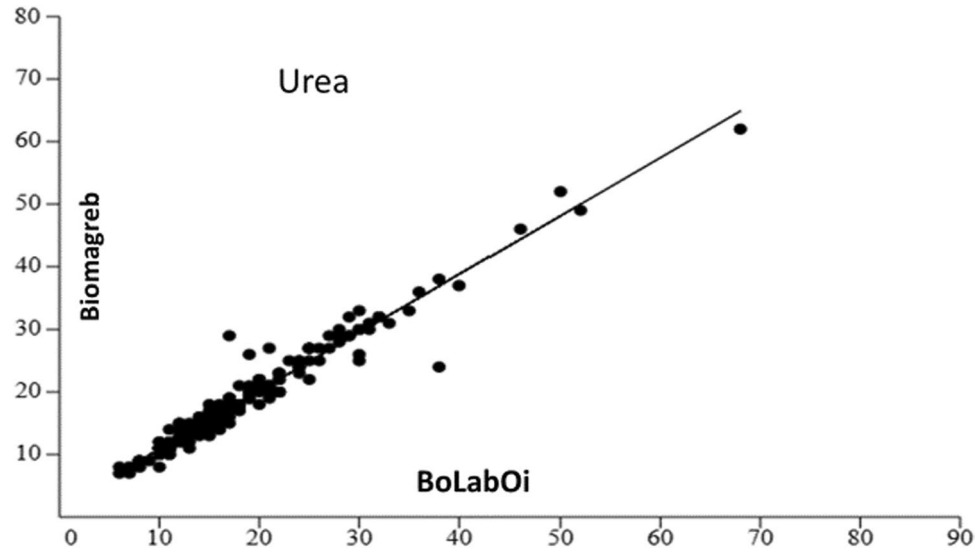

Fig. 3 Scatter plot with regression line for Glucose measurements (BioScien to Biomagreb conversion: Y = 1.141X−13.71). Generated using GraphPad QuickCalcs

Fig. 4 Scatter plot with regression line for Urea measurements (BioScien to Biomagreb conversion: Y = 0.7073X + 5.252). Generated using GraphPad QuickCalcs

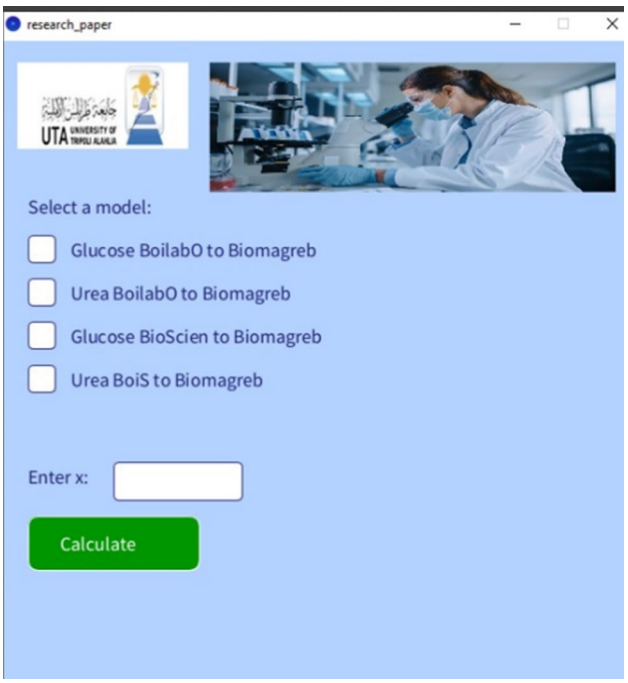

Fig. 5 Software interface screenshot showing model selection and conversion functionality. Implemented in Processing 4

model fidelity, consistent with physiological observations (Wilson and Harris 2020). Visual validation through scatter plots (Figs. 1, 2, 3, 4) further confirmed the robustness of these relationships.

Minor observed discrepancies ($\leq 0.5\%$) stemmed exclusively from decimal rounding effects rather than model

limitations. This aligns with Peterson et al.'s (2021) assertion that precision management is a frequently underestimated factor in data harmonization. Our findings confirm that three-decimal precision sufficiently supports clinical decision-making for these analytes. The software implementation (Fig. 5) eliminated manual calculation errors while providing immediate reference range contextualization a functionality gap noted in prior literature (Alvarez et al. 2021).

Several limitations warrant consideration. First, validation was restricted to two analyzer systems, whereas larger initiatives like the Harmony Project (2023) incorporated ≥ 5 platforms. Second, the exclusive focus on glucose and urea excludes clinically significant analytes such as creatinine. Third, regional sampling may impact generalizability across diverse populations, as highlighted in multiethnic studies (Kim et al. 2022). Finally, current software constraints (mg/dL-exclusive units, single-value processing) require expansion for broader utility.

Our methodological approach prioritized parametric transparency and clinical interpretability over complex machine learning. As validated by residual analysis and supported by foundational literature (Hastie et al. 2009; James et al. 2021), hyperparameter optimization was unnecessary given the models' closed-form solutions and absence of overfitting. This strategy successfully balanced regulatory requirements (Taylor et al. 2022) with operational practicality.

Conclusion

This study successfully developed and validated a computational tool for standardizing glucose and urea measurements across different clinical laboratory systems using empirically derived linear regression models. The four conversion models demonstrated high accuracy, confirming that simple linear transformations are sufficient for reliable inter-laboratory harmonization of these analytes. The software implementation significantly reduced manual calculation errors while providing immediate clinical interpretation through normal range displays.

Abbreviations

AI	Artificial Intelligence
ML	Machine Learning
XAI	Explainable Artificial Intelligence
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
R^2	Coefficient of Determination
PCA	Principal Component Analysis
CLSI	Clinical and Laboratory Standards Institute
IFCC	International Federation of Clinical Chemistry

Acknowledgements

We extend our deepest gratitude to the faculty members of the Medical Laboratory and Information Technology departments at the University of Tripoli Alahlia for their technical support and infrastructure access.

Author contributions

Eshraq Alsherif was involved in conceptualization, methodology, supervision, and writing—original draft. Abdulwahab AL-Deib carried out data curation, formal analysis, and validation. Husameddin Abuzgaia developed the software and implemented the algorithm. Amel Rabti conducted laboratory experiments and collected the data. Ruwaid Rtemi preprocessed the data and performed statistical analysis. Salem Elfard was responsible for software validation and technical review. Sheren Njaim reviewed the literature and edited the manuscript. All authors reviewed and approved the final manuscript.

Funding

This research received no external funding. It was supported by in-kind contributions of laboratory resources from the University of Tripoli Alahlia.

Data availability

The datasets generated and analyzed during this study are available under your request.

Declarations

Ethics approval and consent to participate

Ethical approval was obtained from the Institutional Review Board of the University of Tripoli Alahlia (Reference #: UTA-IRB/2023-07-MED). The need for informed consent was waived due to the retrospective analysis of anonymized laboratory data.

Consent for publication

Not applicable. This manuscript does not contain individual-level data or identifiable personal information.

Competing interests

The authors declare no competing interests, financial or non-financial, related to this work.

Received: 17 June 2025 Accepted: 12 August 2025
Published online: 20 August 2025

References

Alvarez R, Torres M, Jiménez L (2021) Equation-based harmonization approaches in clinical chemistry. J Lab Med 45(3):112–119. https://doi.org/10.1515/labmed-2020-0123

Brown A, Davis K, Wilson P (2019) Artificial intelligence in diagnostic decision support systems. J Med Syst 43(8):267. https://doi.org/10.1007/s10916-019-1402-6

Brown A, Davis K, Wilson P (2021) Machine learning approaches for quality control of laboratory kits. Clin Chem Lab Med 59(4):721–730. https://doi.org/10.1515/cclm-2020-1185

Chen X, Wang Y, Zhang H (2019) Multicenter evaluation of glucose assay harmonization. Clin Biochem 71:33–39. https://doi.org/10.1016/j.clinbiochem.2019.06.003

Clark R, Evans T (2023) Regulatory challenges in AI-assisted diagnostics. J Pathol Inf 14(1):100315. https://doi.org/10.1016/j.jpi.2023.100315

Clinical and Laboratory Standards Institute (2023) Measurement procedure comparison and bias estimation using patient samples (EP09-A3, 3rd ed.)

Harmony Project Consortium (2023) International harmonization of 25 clinical chemistry analytes. Clin Chem 69(7):763–772. https://doi.org/10.1093/clinchem/hvad059

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning: Data mining, inference, and prediction, 2nd edn. Springer, Cham

International Federation of Clinical Chemistry and Laboratory Medicine (2022) Position statement: Standardization in clinical laboratories

James G, Witten D, Hastie T, Tibshirani R (2021) An introduction to statistical learning, 2nd edn. Springer, Cham

Johnson M, Lee S (2019) Pre-analytical variables affecting kit performance. Ann Lab Med 39(5):421–428. https://doi.org/10.3343/alm.2019.39.5.421

Johnson M, Lee S (2020) AI-driven quality control in diagnostic testing. Artif Intell Med 107:101910. https://doi.org/10.1016/j.artmed.2020.101910
Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. (2022). KDIGO 2022 clinical practice guideline for the evaluation and management of chronic kidney disease. https://kdigo.org/guidelines/ckd-evaluation-and-management/
Kim J, Park S, Tanaka H (2022) Ethnic variability in laboratory reference ranges. Clin Chim Acta 531:375–382. https://doi.org/10.1016/j.cca.2022.05.003
Martinez-Perez C, González L, Ruiz P (2021) Polynomial vs. linear regression in analytical harmonization. Analyst 146(15):4821–4829. https://doi.org/10.1039/D1AN00722F
Peterson S, Li Q, Jackson R (2021) Decimal precision effects in clinical data harmonization. J Pathol Inf 12(1):9. https://doi.org/10.4103/jpi.jpi_77_20
Smith J, Anderson K, Brown T (2020) Laboratory solution kits: current applications and limitations. J Appl Lab Med 5(4):731–742. https://doi.org/10.1093/jalm/jfaa059
Taylor R, Jones A, Williams G (2022) Regulatory considerations for AI in IVD devices. Diagnostics 12(3):618. https://doi.org/10.3390/diagnostics12030618
Thompson M (2022) Biological relevance in method comparison studies. Accred Qual Assur 27(3):127–133. https://doi.org/10.1007/s00769-022-01499-7
Wilson D, Harris C (2020) Linearity of analytes in clinical ranges. Clin Biochem Rev 41(2):63–69. https://doi.org/10.33176/AACB-20-00006

Publisher's Note