Clinical Biochemistry 73 (2019) 11–25

Contents lists available at ScienceDirect

Clinical Biochemistry

journal homepage: www.elsevier.com/locate/clinbiochem

Review

Autoverification of test results in the core clinical laboratory

Edward W. Randell[a,*], Sedef Yenice[b], Aye Aye Khine Wamono[c], Matthias Orth[d]

[a] Discipline of Laboratory Medicine, Faculty of Medicine, Memorial University, St. John's, NL A1B 3V6, Canada
[b] Gayrettepe Florence Nightingale Hospital, Department of Core Laboratory Medicine, Istanbul 34349, Turkey
[c] NHLS Laboratory, Division of Chemical Pathology, Tygerberg hospital, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa
[d] Arzt für Laboratoriumsmedizin Medizinhygiene, Hämostaseologie F.E.B.M.B., Adlerstr. 7, 70199 Stuttgart, Germany

ARTICLE INFO

Keywords:
Autoverification
Process improvement
Turn-around time
Quality assurance

ABSTRACT

Verification of laboratory test results represents the last opportunity to identify errors before they become part of the electronic medical record. Manual verification of test results places significant reliance on the experience and attentiveness of individual observers to identify errors and is vulnerable to errors through omission and neglect. Peer-reviewed publications have documented gains in process efficiency and quality improvement by use of middleware or laboratory information systems to autoverify test results based on pre-defined acceptability criteria. This review evaluates the acceptability of autoverification (AV) as a safe and reliable alternative to total manual review of laboratory test results. AV schemes developed in accordance with international guidelines and standards are applied throughout the laboratory. Careful design of AV systems involves using multidisciplinary teams to develop test-specific decision algorithms, to assist with programming, to verify programming, and validate programmed algorithms prior to use in evaluation of patient test result profiles. Development of test specific decision algorithms makes use of criteria based on instrument messages and flags, quality control status, result limit checks, delta checks, critical values, consistency checks, and patient-related clinical information. Monitoring of the performance of AV parameters, and regular audits of the AV system integrity is recommended in both the literature and guidelines. The potential for gains to process efficiency, error detection and patient safety, through adoption of AV as part of a laboratories quality assurance tool-case, is well supported in published literature.

1. Introduction

Improving the quality of medical laboratory testing processes is critical to improving healthcare quality. Laboratory error remains a major concern and is usually understood in the context of where in the analytical process (pre-, post-, or during) the error occurred (Table 1). Peer reviewed literature documenting the vulnerability of laboratory test results to error goes back to the early 1950s [46,47]. But over time there has been a transition in focus from analytical phase processes and technology, and quality control processes, to vulnerability of non-analytical phases [43] and predominantly pre-analytical and specimen quality-related [62]. Advances made through automated analyzer systems, have reduced error occurring at the analytical phase. Furthermore the adoption of automated track systems to provide greater standardization and control over centrifugation, pipetting, and aliquoting functions; the use of two unique identifiers and barcoding technologies to address patient and sample identification [82]; and adoption of modern information technologies including Laboratory Information Systems (LIS) and/or middleware to secure data flow [45]; are all directed at reducing the frequency of pre-analytical errors. Nevertheless, pre-analytical errors still impact around 1.5% of clinical chemistry tests [9] and with catastrophic consequences in some cases. For example, patient misidentification errors, which affect between 0.04% and 1% of samples [5,7,28,61,92], are more common in critical care areas [94]; cause medication and diagnostic errors, and account for 11% of all transfusion-related deaths [82]. Hence, downstream checking, still provides potential advantage to limiting the impact of pre-analytical error.

Verification of laboratory test results is the last quality assurance exercise performed before release of test results into the patient's

Abbreviations: AV, Autoverification; HIL, Hemolysis, Icterus, Lipemia; LIS, Laboratory Information System; MLT, Medical Laboratory Technologist; QC, Quality Control; RCV, Reference Change Value; TAT, Turn-Around Time.

* Corresponding author at: Lab. Medicine, Eastern Health Authority & Faculty of Medicine, Memorial University, Health Sciences Centre, Room 1J442, 300 Prince Philip Drive, St. John's, NL A1B 3V6, Canada.
E-mail address: ed.randell@easternhealth.ca (E.W. Randell).

https://doi.org/10.1016/j.clinbiochem.2019.08.002
Received 18 June 2019; Received in revised form 30 July 2019; Accepted 2 August 2019
Available online 03 August 2019

Table 1
Sources of error affecting testing processes.

Pre-analytical	
Incorrect collection vial	Collection tube source (Manufacturer)
Inappropriate specimen quality	Dilution or contamination by IV fluids
Mishandled specimen	Venous stasis/Hemoconcentration (tourniquet use)
Improper fasting status	Improper specimen collection devices
Exercise or activity prior to sampling	Improper mixing
Stress or posture during collection	Inappropriate storage temperature
Patient misidentification and sample mislabeling	Inappropriate transport time or delay in transport
Insufficient sample volume and anticoagulant ratio	Specimen integrity during transport
Blood cell rupture and hemolysis	Centrifugation speed, duration, and force
Anticoagulant contamination	
Analytical	
Sample switching (wrong sample aliquoted/pipetted)	Quality control failure
Procedural error	Delayed analysis
Dilution error	Instrument error
Post-analytical	
Misidentified results	Misinterpretation of results
Result entry error	Delay is reporting results
Failure to notify of critical or problematic results	

Table 2
Data considerations for laboratory test verification.

Focus	Detail
Test result	• Change in result (Delta) • Reference range • Critical values • Compatibility with life • Degree of unusualness
Assay and instrument	• Instrument status • Environment status • QC/Calibration status • Moving averages
Specimen	• Type or source • Collection date and time • Anticoagulant used • Centrifugation status • HIL/interference vulnerabilities
Patient	• Age and sex • Inpatient/outpatient status • Medication history • Clinical care/treatment unit • Diagnosis • Other samples collected

electronic medical record. It's positioning in the post-analytical opens opportunity for detecting error that occurred earlier in the testing process. Result verification makes use of information on the sample, and test results, to determine overall result acceptability for reporting. It provides the mechanism for identifying critical results, but also identifying suspicious and unusual results by comparing against other information concerning the patient or concerning other tests, the sample and previous samples, and the testing system (Table 2). The process can also lead to deeper probing into sample integrity, source, and identification information. It enhances communication and collaboration between the laboratory and healthcare professionals [82], and allows consideration of other patient-specific factors when determining whether a result should be reported or not [6]. The test result verification process, thereby, serves as a vital relationship building opportunity for laboratory professionals and clinical staff as collaborators in preventing harm, and assuring the diagnostic value of laboratory services.

Test verification poses several challenges. In modern core clinical laboratories highly automated instruments perform a large portion of tests, and laboratory results are typically reviewed by medical laboratory technologists (MLT) on a printed report from the instrument, or much more likely, on a computer screen. This review is followed by a decision on whether or not to release the result from the LIS into the electronic medical record. In real life situations, especially for busy laboratories, there are significant limitations to the degree of evaluation that a given sample and test result can receive. Fatigue can easily develop when examining large numbers of results in queue, and the significance of this on error detection and verification accuracy has never been studied but is admittedly a potential risk factor for mistakes [52]. For results under review, the list of possible follow-up actions are extensive but may involve calling a critically abnormal test result; calling a clinical unit or physician to discuss the potential relevance of a result; or various other actions involving the sample, the tube, the test results, including diluting the sample and/or retesting. But in most cases just a brief scan through displayed results precedes the final result release. Despite the many potential results, result combinations, and follow-up actions there is usually little direction provided as to which criteria must be used to inform test result verification decisions. The test verification process is, therefore, subjective and arbitrary with demonstrable differences occurring depending on training level, experience, or "professional judgement" among those verifying test results. Verification of test results is often a mundane, tedious and time-consuming process, and can easily be complicated by shortages of trained and qualified staff. The individual MLT verifying the test result at a particular time and working condition, thereby, becomes the lone determinant of quality at this stage of the testing process. Hence, the overall effectiveness of the manual verification strategy in many clinical laboratories is unknown, but likely variable, costly, and vulnerable to error, and especially if the MLT is tired, interrupted, distracted, or inexperienced. Training and standardized processes are helpful for identifying some result or test profile anomalies, but for those that get through, reliance moves to clinical staff to recognize suspicious results and take appropriate action prior to clinical decision making. Failure at this point raises the level of the "laboratory error" to that of a medical diagnostic or treatment error.

Autoverification (AV) is commonly recognized as a post-analytical workflow improvement tool that limits the number of test results requiring review by an MLT. The analytical phase generates a huge amount of testing data and information that requires review for error [73]. AV provides a well-designed set of rules in algorithmic format, whereby data can be assessed for potential defects, and unacceptable and potentially erroneous results are selected for manual review, and verification, or for other corrective action. AV strategies achieve this by selectively verifying test results that meet criteria and holding those that would be identified by well-trained and experienced laboratorians for closer scrutiny [85].

Possibly the first report describing use of a computer to assist with laboratory test verification was published over 50 years ago and by Lindberg [60] whose algorithms were fixed on identifying "dangerous" results. That report described the evaluation of test results for "compatible with life", expected result patterns and relationships (e.g. urea and creatinine), if results fell outside of defined limits, and by compatibility checks, and delta checks. Other similar "home grown" systems for computer-assisted test verification were subsequently developed and, like the earlier work, involved complex algorithms and notification of laboratory staff about potential problems. More contemporary studies have documented significant gains through implementation of localized or widespread AV of test results. The reported gains are variable across studies and include improved turn-around times [33,92,58,99,84,77,1], reduced labor requirements or improved staff utilization [92,58,43,63,68,85], improved laboratory error detection [92,66], improved operational efficiency [55,58,77,78,92,98]; Sediq et al., [84]; [66]), improved physician satisfaction [33],

improved patient safety [98], and all while providing greater opportunity for MLTs to focus on test result profiles requiring greater scrutiny. The efficiency gains have enabled some laboratories to acquire competitive advantage and larger workloads [66,101].

AV algorithms use data filters or test-specific rules to determine test result suitability. These are based on instrument flags, affecting various instrument and fluidics components, quality control (QC) status, moving average alerts, serum HIL (hemolysis, icterus, lipemia) indices, critical values, delta checks, unusual and extreme test results, and a variety of other checks to evaluate patient results based on expected logical relationships with other test results. The latter set of data filters are typically used to assess unusual test results that can be caused by contamination, incorrect anticoagulant, delayed and un-centrifuged samples, sample mix-ups, or presence of unrealistic results or result relationships [22,55,58]. AV algorithms can also interrogate patient historical, diagnostic, and demographic information, location, age and sex demographics, collection time and date, and collector identification, from the electronic medical record [10]. Because of differences in patient populations served at different health care centers, the data filters or rules are best customized to each laboratory's specific situation, and based on the sophistication of the LIS, middleware or both, whichever is leveraged in the AV process [31]. For example, some laboratories allow high urea test results from dialysis patients to pass AV, while holding other patient groups [58,70,84,92], while other AV systems may apply the same rule across all patients. Because autoverified test results are immediately verified, there is an overall reduction in the number of samples that require manual review, and as a result, TAT decreases, even at peak-times. The effectiveness of AV is also independent of the operator and it avoids the effects (such as delays, second draws) caused by false alarms. As a more systematized and standardized approach to identifying errors, AV minimizes mistakes made because of lack of experience, stress, or a difficult working environment. Studies comparing the level of agreement between expert reviewers and AV software approaches document poor inter-observer agreement [84]. AV based approaches tend to show superior sensitivity and at lease similar specificity to the best performing expert reviewers [22,70]. This is in spite of the potential that experts involved in observation studies spend more time than usual, and possibly achieving greater error detection than in real life situations [70]. By directing attention to fewer samples with results having greater likelihood of error, AV can reduce error rates by 80% or more [92,66,58,103].

2. Verification of laboratory test results

2.1. International guidelines and standards

AV significantly impacts the quality of reported results and has potential impact on patient safety. ISO 15189 [48] addressed AV in section 5.9.2, and indicating that: "*If the laboratory implements a system for automated selection and reporting of results, it shall establish a documented procedure to ensure that: a) the criteria for automated selection and reporting are defined, approved, readily available and understood by the staff.*" It also addressed common components of AV, including the delta checks and result limit checks in:" *Items for consideration when implementing automated selection and reporting include changes from previous patient values that require review and values that require intervention by laboratory personnel, such as absurd, unlikely or critical values.*" The Clinical and Laboratory Standards Institute (CLSI) provides guidance on the safe design and implementation of AV [10,12,13]. For example, the 2006 guideline, Auto 10–1, outlined pathways for design (based on Boolean logic), implementation, validation and customizing AV rules and algorithms. This guideline recommended two phases of validation. One involved use of simulated data covering the analytical measurement range and involving representative and rare patient sample results to confirm intended response from the AV system operating software. This is a robust strategy and may reveal different errors than might be revealed using real patient data and in day to day practice [32]. The second phase of validation involves using real patient results. The call for active supervision by the laboratory director, and for staff involvement during development, implementation, and monitoring of AV was established by the standard. At the time of this writing a new CLSI guideline [13] focusing on AV is under development. Like its predecessor, this guideline leaves laboratories to design their own discipline-specific AV algorithms, criteria and thresholds, but lays out the process for development and implementation, beginning with assembling the team, design and validating algorithms.

Delta checks are widely used components of AV algorithms. The CLSI guideline on delta checks [12], provided the framework for selecting, implementing and evaluating delta checks. In recognition of the subjective nature in design and application of delta checks, a customized purpose-focused approach that considers local populations was recommended. Laboratories were allowed flexibility in selecting calculation modes, thresholds, time intervals, and in determining which patient populations, and tests, on which delta checks should be applied. But to assure safety, the guideline recommended verification involving review of data retrospectively to determine the overall effectiveness in identifying error [12]. The guideline did not characterize the degree effectiveness required.

Laboratory accreditation bodies also address AV related requirements. For example, the College of American Pathologists checklist [15] provided several requirements pertaining to AV and including approved AV policy signed by the laboratory director; documentation of AV initial validation and annual audit; and audit when changes are made to the system. Furthermore, the requirements addressed the need for a defined range of acceptable results subject to AV; acceptable QC result criteria; holding of unrecognized instrument flags; use and approval of delta check criteria; comparisons against absurd and critical values for results that require, repeat testing, telephone notification, or manual dilution; and, like all manually verified results, an audit trail for all autoverified results, including date and time. Other laboratory accrediting bodies, including the Joint Commission, do not have their own standards that specifically address AV, but acknowledge the applicability of quality standards to AV including quality system assessment, the role of laboratory directors in developing verification criteria, and validation of computer functionality; and maintenance of an audit trail for laboratory test result reporting [49].

2.2. Validation of AV

There are no international standards that require validation of the effectiveness of the manual test result review in reducing error. But, considering that observed variation in performance between different staff (discussed above) it is likely that there is significant variability both within and between different organizations and, of course, complicated by differences in local patient mixes. In contrast, the AV algorithm validation process involves pretesting, approval of documentation, followed by aggressive auditing at initial implementation, and then periodic audits [10]. Preliminary "dry testing" using electronically simulated cases is a common approach to validating AV algorithms prior to implementation [21,58,85,98]. This involves designing scenarios with different patient data and result combination to test the AV system, and making use of simulated patient test results that address unusual and hypothetical scenarios. A second approach involves use of previously verified real patient historical data from the LIS [58,84]. This strategy helps to verify consistency of autoverified test results with clinical status and outcomes, determine preliminary pass rates, and assess improvements over traditional processes [58]. A third approach involves using of real samples to challenge specific rules, or groups of rules. This testing is also referred to as "wet testing", and can involve unusual and abnormal patient samples, proficiency testing samples [58,85], or manipulated samples to challenge that AV system from analysis to LIS report.

Depending on the tests and/or laboratory section, other approaches may be necessary to test the effectiveness of the AV process. For example, to validate the performance of an AV system for complete blood count profiles, Zhao et al. [102] prepared three blood smears for each sample to be verified. These were subsequently reviewed microscopically by two experienced MLTs, with a third MLT to resolve discordances. Others have compared a panel of trained and experienced laboratory professionals against performance of AV using simulated cases [84], and or including data from intentionally misidentified patient results to verify effectiveness at identifying sample mix-ups [70]. Pretesting typically uses a test version of the LIS or middleware.

Apart from errors in programming, AV algorithms can yield unexpected results with unforeseen consequences. Risk assessments prior to implementation of an AV system may detect other system vulnerabilities [41,77], but trial runs, and intensive auditing after implementation helps close remaining gaps. Post-implementation auditing is usually intensive during the first couple of days, and may involve all results [75] being scrutinized for AV system failures and defects. The auditing frequency is then slowly tapered off over the coming days [74,92]. Ongoing monitoring audits are, thereafter, carried out at a regular interval, and to confirm intended performance of AV algorithms and the appropriate response of staff to held results [66]. The frequency of audits should also increase with changes to the AV algorithms, or with modifications to the test menu [75].

2.3. AV monitoring and process improvement

Continued process improvement involving use of systematized tools result in significant gains in efficiency of AV [55,66,77,78,4]. The main components of laboratory test quality include its accuracy and precision, timeliness of reports, and authenticity (the right result on the right patient), as well as test error rates, costs, and relevance [54]. Among the metrics used to monitor development, implementation, and post-implementation effectiveness of AV systems are TAT and percentage of results meeting TAT goals [68]; error rates [65]; hold rates for different AV components [77,68]; components of sensitivity and specificity including, true positive rates (intercepted problematic reports), false positive rates (intercepted correct reports) [102], true negative rates (auto-released correct reports) and false negative rates (auto-released problematic reports) [102]; and utilization of MLTs and number of full time equivalent requirements [68]. Also, in connection with the latter is AV productivity gains measured by the total number of hours worked, test volumes, overall payroll impact, and overtime hours [74], and MLT productivity by the number of paid hours per test [4]. These assess the effectiveness of AV in detecting errors, verification of patient information, and assessment of specimen quality, and assuring timeliness of result reporting or overall operational efficiency, all with supporting statistical evidence [50].

Providing rapid laboratory services, especially to critical care areas, is important to improving patient outcomes. TAT can be measured several ways [63], but the most convenient is the time interval between sample arrival at the laboratory to reporting results into the electronic medical record [1]. Essentially all reports on AV indicate gains in TAT and/or increased ability to take on increased workload without compromise to TAT. Increased TAT affects length of stay, especially in critical care areas including emergency departments, where it can complicate and contribute to overcrowding. Increased TAT can also have the unintended effect of increasing repeated and redundant test orders. But a decrease in the percentage of TAT outliers (> 1 h) is a key factor affecting the emergency department workflow and its reduction to $< 3\%$ significantly reduces emergency department length of stay and the number of STAT test requests [4].

The most commonly used metric for AV efficiency is the AV pass rate. This metric is intimately and inversely related to the frequency of false positives. False negative results are difficult to assess as they depend on the identification of errors by other means. Many false negatives that are identified by clinician-mediated repeat testing can go unreported, while others may never be identified. While 100% AV pass rates are not achievable, significant gains to pass rates are possible without compromise to error detection or increase in false negatives [55,65,66,77]. Improvements in AV pass rates in excess of 95% also can be achieved safely [55,65,77] at least in routine clinical chemistry and immunoassay laboratories and through iterative process improvement initiatives. But similar AV approaches can show differences in performance because of variation in patient mix in local settings [77], and over time.

Although AV pass rates may be high, held results because of QC failures, or instrument flags, can cause significant delays as AV is suspended. To address this, Jones [50] recommended use of the “functional autoverification” rate, which is based on the percentage of the time that the system is autoverifying. In situations involving patient misidentification, or where there is specimen quality concern, no results should be released for the sample. AV algorithms address these situations through specimen-specific AV, where all test results are held from release until all test results are completed and evaluated. A weakness of this approach is that a test specific failure in QC affecting one test can result in delay in release of all test results on the same sample. Some laboratories also monitor the effect of specific rule components of the AV algorithm on pass rates, including the effectiveness of limit checks and delta checks, which can be prone to high false positives rates [77,81,85]. The information captured on parameters like serum HIL indices can be directed to other process improvement initiatives, by identification of particular clinical locations or specimen collection staff members for targeted interventions.

2.4. Components

The only true gold standard for error detection requires foreknowledge of the error. While gross errors can be identified by the human “expert”, the more common and more subtle errors pose challenges to even computer assisted approaches. Design and application of AV rules to detect error requires understanding of the causes and effects of errors on test results. These errors fall into three categories [89]: errors that affect a single test result (e.g. instrument failure, or data entry error); errors that affect reliability of more than one test value (e.g. use of incorrect specimen collection vial, inappropriate specimen quality, or due to specimen mishandling); and errors resulting in misidentification of a sample (e.g. sample tube interchange, or collection from the wrong patient). Single test result errors show no relationship to the source of the sample, nor to other test results. Errors affecting more than one test result are often sample processing errors and can have predictable effects on results. The consequence of sample misidentification on test results range from an inappropriate data pattern that are only slightly different from the expected to that which is blatantly obvious when compared to other information. When evaluating results for expected patterns it is important to consider the impacts of biological and analytical variability. Several common rule types comprise most AV algorithms. These include consideration of internal QC and moving averages which mainly identify errors that impact individual tests; instrument flags or warnings affecting single or many tests; checks for sample deterioration or quality which have potential to impact many test results; limit checks, which examines for individual test defects; delta checks, which usually examines individual test results but can be sensitive for sample misidentification or alteration; and evaluations of concordance between test parameters, which is sensitivity to errors affecting usually single test results. To standardize responses to results held by specific rules, automated short and direct “pop-up” comments are used to direct to appropriate corrective action or to troubleshooting algorithms [55,77,78].

2.4.1. Result limit checks

Result limit checks are used to identify test results that exceed

Table 3
Strategies for determining result limit checks.

Strategy
• Reference range[a,b]
• 2nd and 98th percentile of cumulative patient data[c,d]
• Lower reference limit minus total allowable error and upper reference limit plus total allowable error[e]
• Midpoint between the median of reference range and analytical measurement range or critical values[f,g]
• Critical values[b,h,i]
• Expert consensus[c,d]
• Outside of 0.05 and 99.95 percentiles of cumulative patient data[j]
• Outside of 0.0005 and 99.9995 percentiles of cumulative patient data[k]
• Highest and lowest values after exclusion of outliers from cumulative patient data[l]
• Outside of limits of reported patient data[m]
• Absurdity values[n,o]:
o Sodium < 100 mmol/L or > 191 mmol/L
o Potassium < 1.3 or > 9 mmol/L
o Chloride < 65 or > 138 mmol/L
o Ethanol > 97.5 mmol/L
o pH < 6.8 or > 7.8
• Analytical measurement range[c,d,h,i]
• Analytical measurement range plus extended range by auto-dilution[p]

[a] [63]. Clinical biochemistry, 50 [15], 864–869.
[b] *Randell* et al.,*[77]. Clinical Biochemistry, 55, 42–48.*
[c] [85] Laboratory Medicine, 42 [11], 668–673.
[d] Yan et al. [101]. *Journal of clinical laboratory analysis*, e22877.
[e] Jones [50]. Clinics in laboratory medicine, 33 [1], 161–181.
[f] Torke et al. [92]. Clinical chemistry, 51 [12]:2406–8.
[g] Feitosa et al. [31]. Jornal Brasileiro de Patologia e Medicina Laboratorial, 52 [3], 149–156.
[h] Pearlman et al. [74] Clin Leadersh Manag Rev. 2002;16:237–9.
[i] Rimac et al. [81]. Laboratory medicine, 49 [3], 284–291.
[j] Suárez-Vega, & Fuentes-Arderiu [80]. Clinical Chemistry and Laboratory Medicine, 43 [11], 1278–1278.
[k] Solé-Enrech, & Fuentes-Arderiu, [86]. EJIFCC, 21 [3], 74.
[l] Dixon, [23]. *Biometrics*, *9* [1], 74–89.
[m] Contreras et al., [16]. In vitro *veritas*, *7*.
[n] Clot-Silla et al., [14]. EJIFCC, 22 [2], 52.
[o] Vogt & Oesterle, B. [93]. *Wiener klinische Wochenschrift. Supplementum*, *192*, 21–27.
[p] *Krasowski* et al.,*[55]. Journal of pathology informatics, 5* [1]*:13.*

preassigned thresholds based on improbability or unusuality. These are widely used in different AV algorithms, but there is considerable flexibility in how thresholds can be set (Table 3). Thresholds for result limit checks are arbitrarily set or variably set by the use of population-based reference limits, percentile-based approaches, limits associated with pathology, critical results, or a combination of approaches [58] for any particular test, but there are others. For example, Torke et al. [91] used thresholds representing the midpoint between median of the reference range and the upper and lower limits of the analytical measurement range (AMR), or of critical values. A recent novel approach to setting patient-specific limits [101], involved oncology patients, and compared new results, to the limits of values encountered by the same patient over the past year. In many laboratories, different strategies are used for different tests and even for the same test. For example, Zhao et al. [103], used three different levels of result limit checks in an algorithm for coagulation tests, and including the AMR, critical value limits, and limits based on potential for pathology and percentile-based. In these situations, different result limit checks may identify need for different corrective actions for the same test. For example, a serum potassium value > 6.5 mmol/L may prompt a call for critical result, while levels above 10 mmol/L, which are incompatible with life, may prompt investigation for specimen contamination. Thresholds based on reference ranges are a common approach, especially in laboratories starting AV for the first time, but may be rationalized if an abnormal test result requires notification by the laboratory (e.g. cardiac troponin). For most other tests there is little justification for this approach as there is usually no reason to doubt accuracy of a test result simply because it exceeds population-based reference ranges but still remains within the reportable range for an assay [74]. The method used to select result limit thresholds has a significant impact on the AV pass rate [101]. For example, Yan et al. [101] examined the impact of three methods for determining thresholds, in oncology patients. They examined thresholds based on reference ranges, 2.5th and 97.5th percentiles, and based on local consensus. The more extreme limits and higher pass rates were achieved by the third method but varied considerably by test. High performance AV approaches with pass rates beyond 90% require setting of action thresholds well beyond reference intervals [77,38,55]. But in general, AV pass rates, for laboratories servicing mixed outpatient and inpatient groups, show significant decrease as the proportion of inpatients increase, and when moving from weekdays to weekends [8,68,77,102]. Even with thresholds that are based on the AMR [81], there can be remaining opportunity for gain in AV pass rates using auto-dilution functions by the analytical system, especially when results beyond these limits are likely in the patient population served. It is advisable, however, the results from manual dilution be held for review prior to reporting because of greater risk for dilution error.

The challenge in setting alternative thresholds, to reference limits, lies in the adoption of a logical alternative approach. Such alternative approaches should consider how an autoverified test result would affect patient treatment [50], what action it should prompt by the laboratory, and also consider the likelihood that the result represents error. In some cases test results at the lower end of an AMR may indicate short sample or fibrin clot [50], while results exceeding upper limits may indicate need for reflexive intervention, manual dilution of a sample prior to further analysis, or review of sample history. Critical values represent a subset of result limit checks that suggest life-threatening pathology or need for urgent medical intervention. Critical value alerts inform MLTs to generate a telephone notification to clinicians. This is usually done while holding the results for manual review, either with or without further steps to assure result integrity [58,103]. While action on critical values is essential, holding critical results for review prior to AV is not [38,55], provided another system assures that critical results get called, and all other criteria for AV are met. This approach also has the advantage of timely delivery of results to the electronic medical record for discussion with clinicians at the time of the verbal alert. The use of absurd and implausible results is another logical approach to result limit check threshold settings [88]. While there is no consensus on what constitutes an absurd value, a number of publications provide specific test values [14,97], or strategies for their determination [86,90].

2.4.2. Quality control and instrument flags

Quality control (QC) checks are generated by QC software following the application of acceptability criteria like Westgard rules. Failed QC checks or even unavailable QC results halt the release of test results which assures that all QC requirements are met prior to release of patient data. While not all AV software allow results to be held based on QC results [76,74,85], essentially all AV strategies hold results based on QC failure, even if by manual disabling of AV until after acceptable QC results are obtained. Moving averages monitoring flags are also used to hold results [50,58,65], and provide timelier alerts to analytical errors.

Good QC design and robust assays are important for effective AV [18]. QC failures disrupt AV and, especially, false positives reduce operation time and functional AV pass rates [50]. QC tolerance limits based on risk or clinical impact result in fewer flags. Tests that have high sigma metric performance facilitate better AV performance by allowing less demanding QC rules with fewer failures [50].

Instrument flags depend on the specific analytical system capabilities [58]. Instrument flags are generated for various problems and mechanical failures, barcode reading problems, reagent, or sample problems (e.g. clotted sample) and can be used to hold results. For example, various instruments provide flags for insufficient reagent or

sample volume; pressure or liquid-sensing flags when air bubbles, sample clots, reagent crystallization, or when samples of unusual viscosity are encountered; other sensors may generate flags for mechanical blockages in probes, outdated reagent, temperature errors; or for assay specific effects, such as prozone effects, or when manual dilution of a sample is required. Instrument flags are especially important in hematology workstations where a high proportion of analyzer flags are required to alert to specimen anomalies, including suspected cold agglutinins or lipemia interferences; presence of platelet clumps, results outside of linear range, or other findings requiring microscopic WBC differential, or review of the cell count [2,68].

2.4.3. Delta checks

Delta checks compare the current result for a patient to previous results and evaluate the probability of significant change. This concept is widely applied to identify sample mix-ups, but can also be useful in TDM or tumor markers monitoring [88,79]) to identify unexpected absence, presence, or change in measurable levels of an analyte. Many situations cause a sudden and erroneous change in a patient test result over a short period of time. These include consumption of various drugs or herbal remedies; contamination of a sample with EDTA, other incompatible anticoagulant, with skin disinfectants, IV fluid constituents, or by dilution [68]; interference by collection tube constituents like gel separators; clotted samples [69] and infusion of medical imaging contrast media; and various treatments, like intravenous antibody preparations [17]. The most common use of delta checks in AV are for identifying misidentified samples [50]. The best tests to use for this are those with low index of individuality, < 0.6 [44], but this leaves little choice among commonly ordered tests that have sufficient sensitivity, but avoid high false positive rates [79] caused by test result change as a consequence of disease and treatment. For example, delta check rules are prone to frequent failure for common tests like urea and potassium, and especially for patients receiving renal dialysis [77,81]. Although, delta checks rules have been used in AV algorithms to intentionally identify clinically significant change [85], this may lead to a significant number of calls to clinicians, so preliminary buy-in is important. The challenge to more effective use of delta checks in AV rests in identifying and filtering out predictable and expected change.

There are many potential variations of delta checks based on use of different calculation types (Absolute versus percent change, absolute rate versus percent rate change, or complex calculation and multivariate approaches), thresholds, and acceptable interval between the historical result and current test result [79]. This contributes to the considerable variability across laboratories in how delta check rules are applied (Table 4) [42]. Delta checks can be used in two different modes in AV algorithms: 1) In series, as an independent filter with other rules; and/or 2) conditionally applied, based on flagging by other rules (e.g. limit check criteria) (Fig. 1). Conditional mode delta checks are typically applied only to test results exceeding a result limit check – this reduces unnecessary review of a subsequent occurrence of a previously validated high or low test result [58,85,103]. Of note, in the strategy described by Krasowski et al. [55] which achieved > 99% AV test pass rates, delta checks were used minimally, and then only conditionally applied and associated with result limit criteria.

Establishing delta check thresholds can be based on systematic or arbitrary approaches. For example, Dorizzi et al. [26] used a relatively arbitrary approach by applying percentage-based delta check thresholds to 50% to a set of tumor markers. Use of the reference change value (RCV) expressed as a percentage or absolute value is a more common systematic approach to selecting thresholds [42]. RCVs are used to determine the significance of differences between two serial values for a test on an individual [35,36] and typically address one of two levels of significant probability 95% and 99%. Feitosa [31] used RCV determined at both 95% and 99% with results intermediate between the two receiving a result flag for later supervisory review, but with immediate auto-release. Results > 99% were held for immediate review.

Table 4
Examples of various strategies used for delta check thresholds.

Test specific strategies
Local/expert opinion (Complete blood count components)[a]
• White blood cell count: > 20% or > $0.5x10^9$/L
• Mean corpuscular volume: > 5% or > 5 fL
• Platelet count: > 20% or > $15x10^i$/L (within 3 days)
Clinically significant alert
• Sodium[b]: Change > 8 mmol/L (within 24 h; for Osmotic Demyelination Syndrome)
• Creatinine[c]: Change > 27 μmol/L (for Acute Kidney Injury)
Unlikely change[d]
• Troponin: Change from high to normal with ≥50% decrease; change from high to undetectable (within 6 h)
Conditional[e]
• Chloride: Change exceeding ± 20% if current < 80 mmol/L; change exceeding ± 10% if current > 120 mmol/L
• Creatinine: Change exceeding ± 50% if current < 13 mmol/L;
• Glucose: Change exceeding ± 25% if current < 1.1 mmol/L;
• Total calcium: Change exceeding ± 20% if current < 1.88 mmol/L; change exceeding ± 10% if current > 3.75 mmol/L
• Total CO_2: Change exceeding ± 66% if current < 10 mmol/L
• Phosphate: Change exceeding ± 0.19 mmol/L if current < 0.32 mmol/L
• Potassium: Change ± 15% if current < 2.9 mmol/L; or change ± 10% if current > 6.2 mmol/L
• Urea: Change exceeding ± 50% if current < 0.7 mmol/L
General strategies
• Percentile base change based on cumulative patient deltas (within 3 days)[f]
• Reference change value for P < 0.01[g]
• Reference change values for P < 0.05[h,i]

[a] *Zhao* et al. [102]. *Journal of biological regulators and homeostatic agents. 2016;30* [2]*:571–7.*
[b] *Sterns* et al.,*[87]. Overview of the treatment of hyponatremia in adults. UpToDate, edited Post TW, UpToDate, Waltham, MA. (Accessed on March 31, 2018).*
[c] *Garner* et al.,*[39]. Annals of Clinical Biochemistry, 49* [1]*, 59–62.*
[d] *Anectodal/Unpublished.*
[e] *Krasowski* et al. [55]. *Journal of pathology informatics, 5(1)13.*
[f] *Randell* et al.,*[77]. Clinical Biochemistry, 55, 42–48.*
[g] *Fraser* [35]. *Clinical chemistry and laboratory medicine, 50* [5]*, 807–812.*
[h] *Martinez-Nieto* et al. [66]. *Jornal Brasileiro de Patologia e Medicina Laboratorial, 51* [6]*, 369–375.*
[i] *Fernandez-Grande* et al., [33]. *Biochemia medica: Biochemia medica, 27* [2]*, 342–349.*

Use of the 95% RCV was used as the sole AV criteria for hematology, coagulation, and clinical chemistry tests for a hospital emergency department and demonstrated reduced TAT and improved physician satisfaction [33]. In this study, the delta check rule required an unusually long interval of 3 years in order to increase the proportion of patients that could be evaluated, but all new presenters required manual review. This strategy achieved an overall test AV pass rate of 56%. When establishing the interval for application of delta check criteria it is important to consider the patient population involved. While short duration intervals of a few days suit sample mix-up detection on inpatients, much longer intervals are required if delta checking is done on outpatients [69]. Delta checks, especially determined by RCVs, have negative effects on false positive rates [68,77,85]. For example, Fernandez et al. [32] applied delta checks, based on 95% RCV, to over 50 chemistry analytes, and used mismatched data to evaluate performance. While the AV strategy yielded 98.6% AV test pass rates, < 10% of held results represented true errors and the remaining were false alerts, caused by the effects of treatment and disease states.

Multivariate delta check strategies have been used to improve the sensitivity and specificity of delta checks for error detection, and especially for sample mix-ups. It was observed that the more delta check alerts for different tests on a sample, the greater the likelihood of error [27], but univariate approaches still dominate in practice [79,100]. Possibly the most effective demonstration of a multivariate delta check strategy was described by Miller [67] and making use of

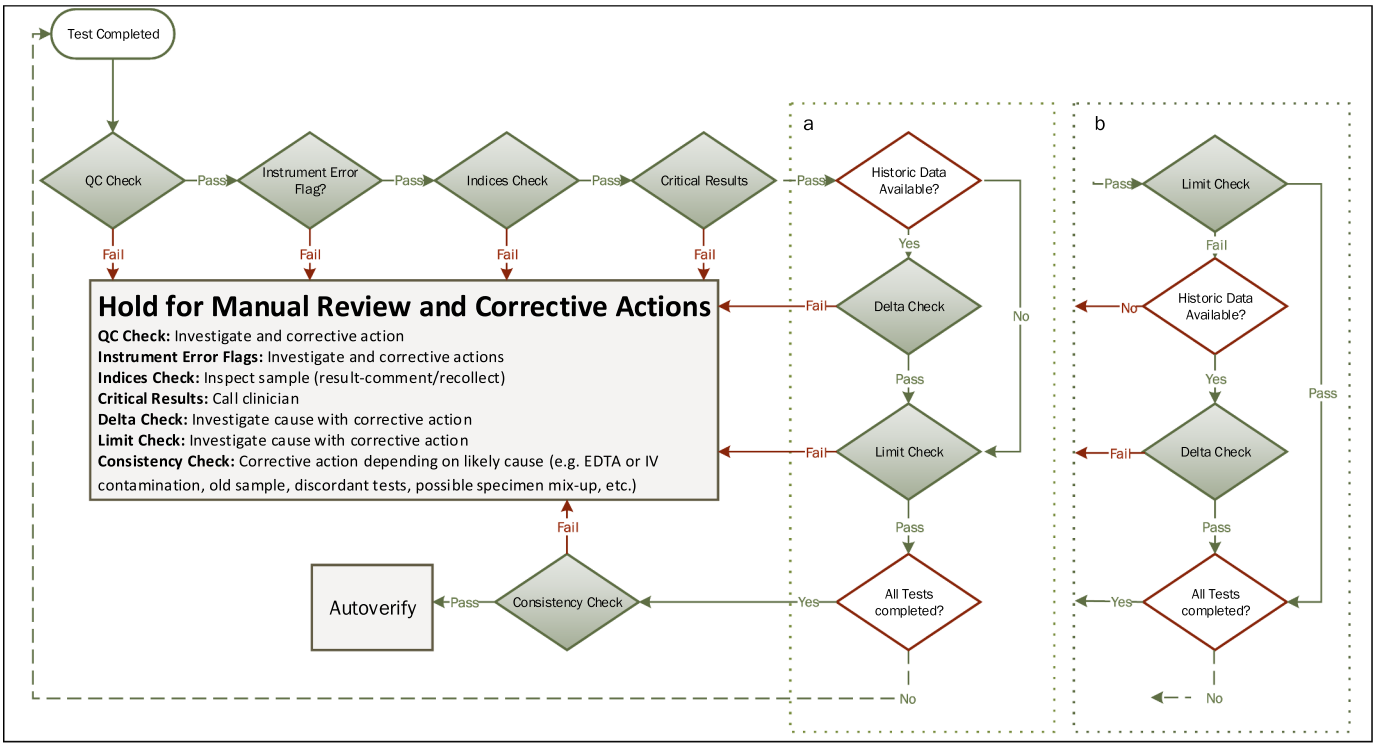

Fig. 1. Typical test AV algorithm. Alternative approaches are presented involving use of delta checks and limit checks. Box "a" presents the decision tree where both criteria are treated independently. Box "b" represents an alternative approach to Box "a" where delta checks are conditional on a limit check failure.

weighed changes in multiple parameters in a calculated parameter called the composite complete blood count delta (CCD). This parameter was defined as the square root of the sum of squares of changes in hemoglobin, MCH, RDW, and platelets and compares the product to assigned thresholds. The report also identified another calculated multivariate parameter based on dropping out the platelet term that when used with the CCD showed discriminatory power that out-performed univariate delta checks based on MCV. It yielded fewer false positive samples and identified twice the number of presumed or confirmed misidentified specimens. Although, not specifically designed for AV it holds potential for integration into AV algorithms in hematology laboratories based on its use of relatively simple calculations and easily applied logic.

2.4.4. Consistency checks

Consistency checks are cross checks involving two or more different correlated analyses performed on the same patient and occasion. These checks are based on predictable relationships shared between tests and/or with patient demographic factors and are used to detect analytical error affecting a single result or pre-analytical error affecting one or more test results (Table 5). For example, levels of certain tumor markers and hormones can be correlated with age and sex demographics [26] and these relationships can be leveraged to assess result validity. For example, the product of a common calculation for anion gap is affected by variation in levels of sodium, potassium, chloride and bicarbonate. Unlikely results for anion gap, such as the occurrence of very low values may be explained by hyponatremia, and hypoalbuminemia, but may also be indicative of errors in one of the four components used in the calculation [40]. Similarly, serum urea and creatinine results are significantly affected by renal function, and changes in one is likely to be accompanied by changes in the other such that the ratio between levels of the two is held within a relatively narrow range [91]. Deviations outside of this narrow range can indicate either pathology or error. Various other ratios between analytes vary within a relatively narrow range include Sodium/Chloride, AST/ALT, LDH/AST, CK/LDH, Calcium/Phosphate, and GGT/ALP [51], and can be used in AV algorithms by applying limit checks around the calculated ratio [77,78]. But like the use of result limit checks and delta checks, numerous caveats impact expected relationships and must be born in mind when deciphering between deviation due to biological variability in disease or the result of treatment versus the impacts of error.

Other consistency checks are directed against specific causes of pre-analytical error (Table 5). For example, the use of potassium salts in EDTA anticoagulant and the binding of divalent cations by EDTA enables the use of results for potassium, magnesium or calcium, or enzymes requiring the metal cofactors, in consistency checks for EDTA contamination of samples, or for use of the wrong tube. Elevated amounts of measured substances or dilution of others form the basis for consistency checks designed to detect intravenous fluid contamination of samples. Pathological factors specific to a patient, such as presence of IgM monoclonal gammopathy, can yield factitious results due to formation of precipitates, binding test reagents, decreasing the aqueous phase due to colloidal effects [19] and impact the accuracy of bilirubin measurements and its relationship with the icterus index, or impact the accuracy of phosphate measurements and its relationship with creatinine and calcium levels. Likewise, a high L-index with low triglycerides can become a consistency check that is highly suggestive of interference by abnormal immunoglobulins [34]. Others are described in Table 5.

Machine learning algorithms and use of sophisticated statistical functions have been applied to discovering new approaches to identifying common pre-analytical problems. For example, Demirci et al. [22] described use of a sensitive artificial neural network approach to detect problematic samples affected by intravenous fluid contamination; containing fibrin clots; EDTA contamination; or delayed analysis. Others have applied learning algorithms toward better detection of contaminated samples [3,64] and have produced strategies with high sensitivity (> 85%) and specificity (99.9%) [64]. Furthermore, machine learning technology has been used to discover unique uses of 30 day mean glucose levels, anion gap criteria, and glucose and electrolyte to identify IV fluid contamination errors [3]. Others have

Table 5
Examples of consistency checks.

Possible error/Defect mode	Basic rule definition
Citrate contamination	• Calcium decreased by 50% and critically low, sodium increased by 5 mmol/L, chloride decreased by 10 mmol/L, and anion gap increase to 30–40 mmol/L in absence of DKA[a]
Clotted sample	• Sodium < 136 mmol/L, potassium < 3.5 mmol/L, calcium < 2.1 mmol/L, glucose < 3.9 mmol/L (and negative deltas)[b]
Clotted sample	• High prothrombin time, high partial thromboplastin time, and low fibrinogen[c]
Delayed sample	• Glucose < 2.21 mmol/L, potassium > 6 mmol/L, and hemolysis index < 50 or negative[b]
Dilution (inappropriate)	• Dilution effect (≥2 analytes ↓↓)[c] • Urine creatinine < 2.2 mmol/L[d]
Discordant results	• Both TSH and fT4 decreased less than lower reference limit[e] • Both TSH and fT4 increased greater than upper reference limit[c] • Normal or increased TSH with increased T3[e] • Normal or decreased T4 and/or T3 with decreased TSH[e] • ALT/AST ratio < 0.25 or > 4[f] • Albumin/total protein ratio < 0.25 or > 1[f] • Direct bilirubin/total bilirubin ratio > 1[c] • HDL-cholesterol/total cholesterol ratio > 0.75[f] • High FSH with high estradiol, progesterone, or testosterone[c] • High FSH with low LH[c] • High LH with low FSH[c] • High LH with high estradiol, progesterone, or testosterone[c] • Low ferritin with low iron saturation[c] • Low ferritin with high iron[c] • High ferritin with low iron[c] • Urea/Creatinine ratio outside of 99.5%ile limits[c] • Mean corpuscular volume > 80 fL with mean corpuscular hemoglobin concentration < 300 g/L[g]
EDTA contamination	• ↑↑ potassium, with or without ↓↓ calcium, magnesium, ALP, CK, iron, zinc[c] • Potassium > 10 mmol/L; calcium < 1 mmol/L with negative delta for calcium and positive delta for potassium[b] • Potassium > 7 mmol/L and calcium < 2 mmol/L, or ALP < 50 U/L, or magnesium < 0.5 mmol/L[f]
Inconsistent macroscopic and microscopic urinalysis	• Comparison of urine dipstick hemoglobin vs. microscopic red blood cell counts: leukocyte esterase vs white blood cell counts; and positive nitrite results vs. white blood cell and bacteria counts[h].
IV dobutamine contamination	• Creatinine < 9 μmol/L[a]
IV glucose contamination	• Negative delta for sodium and chloride and positive delta for glucose and potassium[b] • Sodium < 136 mmol/L, chloride < 98 mmol/L, potassium > 5.5 mmol/L, glucose > 6.0 mmol/L[b] • Sodium < 132 mmol/L, chloride < 103 mmol/L, potassium < 4.1 mmol/L. glucose > 18 mmol/L[f] • Sodium < 131 mmol/L, chloride < 105 mmol/L, potassium < 4.1 mmol/L (reflex glucose > 20 mmol/L)[f] • Any two of sodium, potassium, chloride below lower reference limits[c]. • 30 day mean glucose < 10.3 mmol/L, current glucose > 44 mmol/L but anion gap ≤16.5 mmol/L (inpatients)[i] • 30 day mean glucose between 10.3 and 17.3 mmol/L and current sodium ≥134 mmol/L (inpatients)[i] • Very high glucose result when all previous glucose results are normal or consistent[a]
IV saline contamination	• Any combination of sodium/potassium/chloride↑↑ (or sodium, potassium, chloride all above upper reference limits) [c] • Isolated and critically high chloride without similar sodium abnormality[c]. • Positive delta for sodium and chloride and negative delta for glucose and potassium[b] • Sodium > 160 mmol/L, chloride > 110 mmol/L, potassium < 3.5 mmol/L, glucose < 3.3 mmol/L[c] • Increased chloride, normal sodium, low or critical low potassium[a]
IV parenteral nutrition fluid contamination	• Specimen lipemic and cannot be cleared by airfuge • Increased glucose, and potassium when previous levels were normal • Hematology may be affected with increased mean corpuscular volume and decreased mean corpuscular hemoglobin concentration[a]
Mis-order/mis-labelling	• PSA on a female sex, or CA 125 and CA 15–3 on a male sex[j] • Undetectable levels of therapeutic drugs[c]
Mis-labelling/mis-identification	• Tumor markers showing significant increase or decrease from previous result over 6 month time frame[c]
Monoclonal protein interference	• Elevated phosphate but normal calcium and creatinine (e.g. If phosphate > 2.0 mmol/L, but normal creatinine (and calcium if available)) [c] • Total bilirubin - "I index as concentration" > 65 μmol/L[k,l] • High lipemia index with low/normal triglycerides[m]
Thrombophilia/factor deficiency	• Activated partial thromboplastin time > 33 s and trothrombin time – international ratio < 1.2[n]
Unreliable HIL index	• Hemolysis is 4+ and lipemia flag is ≥1+[o] • Hemolysis is ≥2+ and lipemia flag is ≥1+[o] • All indices (H, I, L) of one plus or greater[o]
Unusual serology	• Isolated HBsAg or isolated HBcAb positivity with other panel constituents negative plus unusual result combinations among the 5 test panel: HBsAg, HBsAb, HBeAg, HBeAb, and HBcAb[p]

(*continued on next page*)

Table 5 (*continued*)

Possible error/Defect mode	Basic rule definition
Unusual urinalysis	• Unusual urine color (for chromogenic interference), specific gravity, pH (impact on urine protein), and bilirubin and urobilinogen (for chromogenic interference) results[q]

[a] Katz et al. [53]. J. Clin. Outcomes Manag, 20 [10], 455–462.
[b] Demirci et al. [22]. American journal of clinical pathology, 146 [2], 227–237.
[c] Anecdotal/Personal experience/Unpublished.
[d] [88]. *Clinics in laboratory medicine, 33* [1], 147–160.
[e] Li et al. [58]. Annals of clinical biochemistry, 55 [2], 254–263.
[f] Randell et al., [77]. Data in brief, 18, 1740–1749.
[g] Zhao et al. [102] Journal of biological regulators and homeostatic agents. 30 [2]:571–7.
[h] Palmieri et al. [71] Clinica Chimica Acta, 485, 275–281.
[i] Baron et al. [3]. American journal of clinical pathology, 138 [3], 406–413.
[j] Dorizzi et al. [26]. Accreditation and quality assurance, 11 [6], 303–307.
[k] Krasowski et al. [55]. Journal of pathology informatics, 5 [1]:13.
[l] Pantanowitz et al., [72]. Archives of pathology & laboratory medicine, 127 [1], 55–59.
[m] Fliser et al. [34]. Biochemia medica: casopis Hrvatskoga drustva medicinskih biokemicara / HDMB. 22 [3]:352–62.
[n] Froom et al. [38]. International journal of laboratory hematology, 37 [5], 680–685.
[o] Spectrophotometric interference by multiple interferences.
[p] Li et al. [59]. Journal of laboratory automation, 21 [5], 642–651.
[q] Davis, [21]. Laboratory Medicine, 30 [1], 56–60.

examined multivariate statistical functions that describe relationships between analytes in order to predict result outliers with high likelihood of error [57]. A probabilistic-based approach making use of relationships between different analyes and using Bayesian Networks were also applied to detecting sample misidentification errors [24,89]. While some of these studies highlight approaches that can be leveraged in development of simplified strategies for AV algorithms, other more complex ones require computing functionality that is not widely available in contemporary AV software.

2.4.5. HIL

Intravascular hemolysis is a life-threatening condition that occurs because of poorly matched blood transfusions, sickle cell crisis, or by mechanical destruction of red blood cells by artificial valves or external circulation. In essentially all other situations, hemolysis is caused artificially by traumatic puncture of blood vessels or inadequate control of pre-analytical factors on blood tubes (e.g. freezing, too fast pneumatic tube transfer, prolonged transit time prior to centrifugation). In the latter causes, a correctly drawn repeat sample will show no hemolysis and the first sample rejected [30]. Hemolysis interferes with many photometric tests by causing absorption at wavelengths used for quantitation of the analyte. Analytes like potassium, and LDH are higher intracellularly and serum levels become significantly contaminated by the release of these contents from hemolyzed red blood cells. While recent reports on AV involve pre-analytical manual inspection for hemolysis, icterus, or lipemia [101], most contemporary automated systems used in clinical chemistry laboratories use multiple spectrophotometric readings to determine HIL indices, and the comparability of the H-index with measured hemoglobin concentrations and across different vendors is good [25]. Quantitative assessment of hemolysis, through measurement of hemoglobin, can be used for decision making on suppressing specific vulnerable test results, or for suppression of all test results, depending on the degree of hemolysis.

Icterus (I-index) is only of importance to a few tests that rely on photometric detection at wavelengths around 400 to 540 nm, but its significance can vary from one analytical platform to the next. Bilirubin's anti-oxidative properties might also interfere with hydrogen peroxide dependent photometric reactions but circumvented by including peroxide acceptors in the test assay. Lipemia (L-index) occurs when very high levels of neutral lipids create an opaque and even milky white sample. This can interfere with spectrophotometry across all wavelengths, and even cause a volume error for indirect electrolyte measurements, used in most clinical laboratories, but not in direct electrolyte measurements. A high L-Index should lead to comment and/or suppression of affected test results. Alternatively, high L-index can be held for inspection and ultracentrifugation prior to reporting potentially affected results. Interference limits for different tests are often provided by the manufacturer and can be programmed into AV algorithms [31] following verification. As the impact of HIL interferences on specific assays are predictable, they do not necessarily require results to be held for review, provided that affected results are flagged for clinicians by automated comments, and reported if the expected impact on results is low, or suppressed if high [55,77,88].

2.4.6. Other compound and customized approaches

Demographic factors like sex and gender, age, clinical ward, clinical presentation, treatment mode or diagnostic investigation, drug history, physiologic status (e.g. pregnancy) can impact test results and, hence, should be considered in electronic verification schemes [10]. For example, Philip [75] described an AV system where there was selective notification of clinical staff to abnormal results for patient in a preoperative unit, or for patients in a labor and delivery unit, like how critical results are addressed. Others have made use of patient care location information to assign special AV rules to CBC results to alert to the need for selective reflexive white blood cell differential counts [68]; or considered drug-related, diagnostic, demographic, and other clinical data for thyroid function tests or sex hormone assays [84,58]. For example, a low estradiol result on an adult female done by immunoassay can be held for manual review and consider adding a cautioning note for follow-up testing; or finding undetectably low levels of testosterone in children and women may prompt a cautioning note for follow-up testing by more sensitive methodologies [88]. In coagulation tests, the ratio of citrate to blood or extremes in hematocrit can affect the reliability of results. It is recommended that fill volumes be checked and that samples for coagulation tests with highly abnormal hematocrit be flagged [11]. Abnormalities in hematocrit (i.e. > 0.55) can affect reliability of common coagulation tests, including aPTT and PT/INR. In recognition of this potential, cross-modality rules have been used whereby cases of elevated hematocrit discovered on a complete blood count can lead to notification of staff working on the coagulation bench of potential for impact on tests [68].

3. Service areas

AV is applied across many laboratory service areas but is customized based on differences in workflow.

3.1. Clinical chemistry and immunoassay

Clinical Chemistry laboratories are the busiest in terms of number of samples and variety of different tests and are also the most complex for AV [65]. AV algorithms components vary from test to test and typically achieve AV test pass rates ranging from 50 to 90% [91,84,26,58,59,101,76], but AV schemes achieving lower test pass rates typically make use of more conservative approaches to result limit checks [33], and/or delta checks, such as using 95% reference change values [81]. In contrast, strategies for clinical chemistry and immunoassay test AV that have achieved 99.5% test pass rates [55] still make use of common AV components (Fig. 1), notably making minimal use of conditional delta checks, autoverifying critical results, and limit checks that extend to the limits of the analytical measurement range for some tests and beyond for others through auto-dilution.

Whenever more than one clinical chemistry or immunoassay test is completed on a sample, and all results are held before AV and release, the test pass rates will exceed sample pass rates. For example, Xia et al. [98] reported a strategy that autoverified > 91% of specific tests, but this translated to only 74% of sample reports. Apart from AV component parameters, patient population and test order mix on individual samples also significantly impact the divergence of test AV pass rates from sample AV pass rates.

In spite of the high test pass rates reported only a small fraction of held clinical chemistry and immunoassay test results may ultimately be found to have error [77,85]. For example, one study reporting AV test pass rates over 92%, indicated that that 93.1% of held tests were manually verified without modification, 6% were suppressed due to instrument error, and < 1% of held results was due to other errors [85]. This suggests that the potential for test AV pass rates in excess of 95% can be safely achieved for clinical chemistry and immunoassay tests.

3.2. Urinalysis

Routine urinalysis is one of the most frequently performed laboratory investigations, but it is complicated by the requirement for two levels of analysis on a sample, and by 2 h or less specimen stability, if not stabilized by chemical preservative or refrigeration. The first level of analysis usually consists of automated macroscopic/chemical analysis, involving 12 or more different chemical or physical analyses. Presence of abnormalities involving hemoglobin, leukocyte esterase, or protein on chemical analysis are often followed by reflexive testing by microscopic examination, which is the second level of analysis. Analysis of the microscopic components are sometimes automated by flow cytometry or by digital image analysis technology.

Differences in instrumentation, workflow, patient populations require customized approaches for AV of urinalysis [21,71,83]. AV of urinalysis operates by variations on one of two modes of operation: 1) the automated release of samples showing no abnormalities by chemical urinalysis, but addressing reflexive microscopy with or without holding samples by AV; or 2) analysis of all samples by total urinalysis technologies [83] followed up with AV of samples showing concordance between chemical and automated microscopic findings [21,71,83,91]. AV pass rates by the former approach is highly dependent on how the approximately 50% of samples that require reflex testing by microscopy are addressed. Davis [21] reported a AV pass rate of 92% for all macroscopic reports, and AV of most samples requiring subsequent reflexive testing, in contrast to about 43% AV pass rate by Torke et al. [91] where samples requiring reflexive testing were also held. Using Davis [21] approach, only samples having cloudiness or unusual/abnormal color (suggestive of erroneous nitrite, glucose, ketone, or blood results), unusually low specific gravity (suggestive of the very dilute sample), pH ≥9.0 with positive protein (potential false positive protein), and positive bilirubin and urobilinogen results were held. This approach, however, loses the potential to correlate microscopic findings with chemical urinalysis findings prior to release of any results. Nevertheless, such review rarely results in modified reports, as careful review of color and pH of the sample, and careful attention to age of the sample, can reduce the occurrence of many misleading results. Improvements in workflow, and reduced labour requirements and stress [21]; decreased error rates from about 0.035% to < 0.006%; but improved TAT [91] were reported benefits of AV of urinalysis.

Reflexive microscopy done manually or by automation, may miss significant pathologic abnormalities [29], however automated chemical urinalysis coupled with automated microscopic technologies makes total urinalysis solutions more feasible. A recent report evaluating this mode of operation reported AV pass rates over 50% by the coupling of automated chemical urinalysis with digital image microscopic system and using complex urinalysis rules and an expert system [71]. The system relied heavily on cross-check rules focusing on discordant findings (e.g. Hemoglobin < 0.3 mg/L with RBC > 5/high power field) and prompted action comments to resolve discordances. The strategy was said to enhance workflow efficiency and sample screening. A factor not addressed by any study evaluating AV and urinalysis so far is how the age of the samples were controlled or whether chemical preservatives were used to provide a more stable sample. As age of the urine sample has a significant impact on result quality, including a check of sample age or controlling its impact is justified.

3.3. Hematology

The complete blood count is the most frequently ordered test panel on which AV algorithms are built [65]. Hematology carries with it unique challenges posed by the specificity of technology used for blood cell measurement; sample quality issues, including coagulation and hemolysis; and the variety of instrument and result flags that require reflexive microscopic examination [102]. But essentially all studies report improvements in TAT, and efficiency by AV, and one study described a decrease by > 90% in error rate [66].

In spite of the availability of a standard rule set [2], there is still considerable variability in the numbers and types of rules used and, as a consequence, significant differences in AV pass rates achievable, typically ranging from about 50% [66] to > 90% [37]. A key factor increasing the number of samples requiring manual review is the number of pathological results requiring a blood smear prior to release of results [2,66,70], hence the ratio of inpatients versus outpatients [70] is important. Other factors include the use of RCV-based delta check thresholds [2], and result limit check thresholds. Achieving higher AV pass rates in hematology requires close attention to the usefulness of criteria in delta checks and result limit checks and evaluating the feasibility of automating components of the white blood cell manual differential. Due to the lower index of individuality of several complete blood count parameters, approaches using univariate and multivariate delta checks tend to be more effective, then clinical chemistry parameters, at identifying mislabeled samples [20,67].

3.4. Coagulation

The results of coagulation tests are particularly important to patients undergoing thrombolytic or coagulation therapy, and in monitoring for disseminated intravascular coagulation disorders. Among the tests most frequently ordered to assess the integrity of the coagulation system, are those for prothrombin time (PT) and calculated international normalized ratio (INR), activated partial thromboplastin time (aPTT), thrombin time (TT) and fibrinogen levels. Several recent reports have described AV systems for coagulation tests [103,69,38,33,68]. The AV pass rates achieved show great variability across studies, but the most commonly ordered tests often achieve pass rates exceeding 90% [38,68,103]. In particular, Froom et al. [38] developed an AV system for a comprehensive coagulation menu including

the most commonly ordered tests and achieved an overall pass rate of about 97%. Achievement of high AV pass rate in this study was attributed to a primarily out-patient population where there would be a lower proportion of abnormal results; use of relatively fresh unfrozen samples; automation of mixing studies; and AV of critical results through use of a separate notification system. Studies using predominantly (83%) hospital inpatient populations [68] have yielded considerably lower AV pass rates (~65%) overall, but high test pass rates are still achievable for common tests like PT/INR (> 97%). Delta checks, based on RCVs and applied over a 15-day interval, accounted for 81.2% of held results, and result limit checks accounted for only 16.7% of held results, similar to studies in hematology and clinical chemistry laboratories. Use of expert systems to perform dilutions of coagulation factors along with AV, have potential to reduce the requirement of human intervention in processing, and hold promise to increased procedural standardization of more specialized coagulation factor tests [80].

3.5. Other service areas

3.5.1. Blood gases

Blood gas analysis is particularly prone to error from inappropriate sampling technique, transport, and storage of samples, or improper mixing prior to analysis [96]. Moreover, the presence of air bubbles, variation in anticoagulant concentration, or contamination of the sample during collections all add to variability in results and to the risk for error. This is further complicated by the variety of different sampling locations (including capillary, arterial, venous, or mixed venous). In the only report to date examining AV of blood gases, Wu et al. [98] described a system based on result limit checks, delta checks, and a variety of consistency checks to achieve sample AV pass rates around 75%. The system reduced TAT and error rates compared with manual verification. Hence, the preselection of samples by the AV system improved the identification of defects by MLTs and led to appropriate action, when assisted by follow-up directions applied to held results.

3.5.2. POCT

While to our knowledge there has been no report specifically examining AV of point-of-care tests (POCT), there is widespread use of AV in many POCT programs. Integration of POCT machines to LIS and hospital system either directly or through the middleware makes other AV benefits possible especially if coupled to a robust system that can link to data from pharmacy, other laboratory results, QC data, and diagnostic imaging and demographic records. With the assistance of middleware, results can be released based on passing QC within acceptable timeframes; with limit checks based on critical values or other actionable limits; and appropriately set delta checks to assist with approval of repeat tests done over a short time interval. For example, accurate glucose measurement by glucose meters can be limited by patient hematocrit, while AV rules can hold samples where hematocrit criteria are not met. Depending on the capabilities of devices, user comments can also be directed back to the operators identifying potential error, need for repeat, to refer the sample to the laboratory, or for flagging a result for review by the laboratory. To the latter point, AV eliminates the need for review of large numbers of test results that require no further action by rule settings that identify instances that require review.

3.6. Implementation of AV

There is no recently published information on the degree of adoption of AV across laboratories, however, results of work done > 5 years ago describe AV as underutilized in its potential. Results of a survey on AV in Spain showed that only 64% of laboratories were using [42]. Around the same time, utilization in the US was comparatively less [65], and in most cases involved only one laboratory section [75]. The reluctance to implement AV arises from lack of supportive IT personnel and/or limitations of available software/LIS, templates or other resources on which to get started; opposition by laboratory personnel and fear of negative impacts on staffing levels; fear of losing control, or distrust of the process; or from uncertainty concerning whether it was worth the costs, software expenses, and the costs for staffing to develop and implement algorithms; too low test workloads; issues related to the complexity of finalizing decision criteria, lack of knowledge and difficulty writing rules, perceived time required; and the burden of regulatory requirements [65,75,95,76]. However, given the benefits to quality and efficiency delivered by AV, continued efforts toward increasing its use and improving AV pass rates should become a priority [65]. Pass rates of 50 to 60% are easily achievable, but rates beyond 90% can be safely implemented across many laboratory areas.

Table 6
Process steps toward implementation or augmentation of AV. (Randell et al., [79, 80, 77,10; CLSI [13]).

1. Choose a process improvement framework
2. Select an AV software (instrument/middleware/LIS)
3. Assemble the team (Include subject matter experts, information technology and frontline staff)
4. Select metrics and gather data on the current state
5. Develop AV algorithms
6. Convert to computer code and complete AV programming
7. Validate programmed algorithms
8. Perform prospective risk assessment
9. Compile and maintain records documenting AV preparation and validation
10. Laboratory director sign-off
11. Train staff, “Go Live” with AV and intensively audit
12. Regularly monitor AV performance and improve

Implementation and improvement in AV have been accomplished across different settings but using similar approaches. Marquardt [65] recommended use of SMART TESTS as an AV design technique. This structured process has similarities to other process improvement techniques, and involved teamwork, structured project management, and selection and monitoring of metrics. By use of continuous improvement frameworks, like Deming's PDSA (Plan, Do, Study, Act) cycle [75], Six Sigma DMAIC [77,78], or others [66], many laboratories have successfully overcome local hesitancy and implemented AV as a quality and process improvement initiative, improving pass rates over time through continued study and improvements [55,66]. Table 6 summarizes the steps involved in developing and implementing AV.

Locally developed AV systems require supporting software and a systematic approach to organization and project management [50]. Several software requirements are important to effective functioning of AV [76,65]. These include the support of AV algorithm operation in real time; flexibility for rapid implementation and testing of algorithm changes; retrieval from multiple different data sources providing patient information, test results and historical laboratory data; and a user-friendly and flexible interface for real time review of AV process activity and information. Crolla and Westgard [18] developed a scorecard to assist with determining whether an analyzer system and LIS were suited for AV. According to the scorecard, low sigma tests were deemed unsuitable for AV, but assay systems that used primary tube sampling; automated test analysis; had verification of sample and reagent volume delivery; used disposable tubes, barcoded reagents, and calibrators; offered direct sample analysis; showed six sigma performance; allowed automated QC tracking; and allowed auto-detection when results were above the reportable range, scored favorably for AV feasibility.

AV development begins with defining goals and realistic expectations, and assembling a team. Clearly defining the problem that needs solving and/or goals for AV upfront helps to focus efforts and activities. Also required is thoughtful change management, involving frontline staff throughout the process to gain buy-in, and using written action plans and testing protocols [95]. Defining, validating, and monitoring process metrics related to goals (e.g. for TAT, staff utilization, error

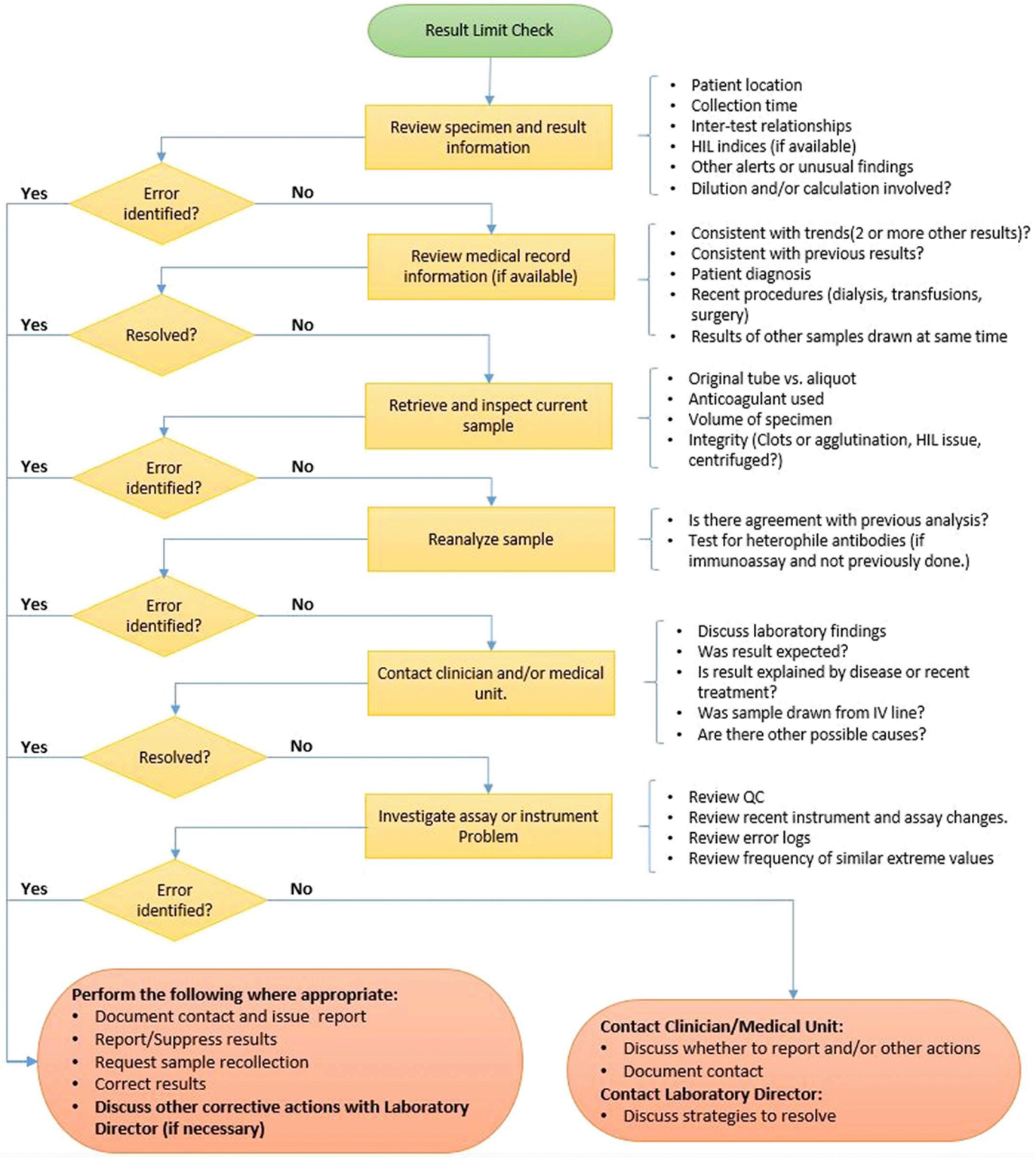

Fig. 2. Example of trouble shooting algorithm of result limit checks.

rates) is important for evaluation [65]. A diversity of skills and input by stakeholders are required, including that of frontline staff [77,78,41]. The accurate identification and assessment of patient information, test requests, date and time of collection, referring clinician's details, and linking patients to their unique account number to track history are among the activities that frontline staff perform during manual test result verification, and must be accurately accommodated by the AV systems. Moreover, input from frontline staff are important to developing a clearer understanding of how the manual verification process works, and understanding why most test results are released, others are held for further action, and what happens next when there is suspicion concerning result integrity [75,77,78].

AV algorithms are built based on learnings from analyzing the existing test verification systems. During early stages, basic flowcharts are constructed to map processes; and fishbone diagrams are used to analyze barriers and root causes to workflow problems; and Pareto charts

are used to determine focus areas and tests for intervention based on impact on the AV process. Preparation of AV rules also requires understanding of the local patient test result distributions, and outpatient and various inpatient populations; how frequently tests are repeated and the typical result changes observed; which tests are ordered most frequently and in which combinations; what instrument flags are produced by analyzers most frequently; what pre- and analytical errors are most frequently encountered and how are they resolved [22,77,78,85]. As it is common to find inconsistencies in practice across staff members, efforts are required to assure consistency in response. Involving frontline staff in developing consensus on how errors need to be addressed helps promote buy-in [75].

Compiling information on each test into spreadsheets assists with presenting the core AV parameter framework in a form that is suitable for AV software programming [50]. Spreadsheets are used to compile information on reference ranges, critical values, delta checks, special notifications based on patient type or location; assay analytical measurement range; clinical reportable range; and special instrument flags according to each test [77,78]. From this information, test specific AV algorithms, with some deviations as required, to accommodate actions from reflexive testing, for example, can be developed and customized from the general algorithm templates (Fig. 1), standardized, and directive action comments developed for held results. Predictive models can be developed once candidate rules have been developed, to determine the adequacy of AV rules and pass rates prior to further actions.

A team member with IT expertise is essential for translating algorithms with test specific parameters into the computer programming code for the AV system. Use of colors and popup comments describing standardized actions to frontline staff improves usability of the system. Offline, decision tree algorithms can address situations where more complex review and troubleshooting is required (Fig. 2). Once computer programming is complete, validation of the system begins. Identified flaws are fixed through redesign and retesting until all challenges pass. This can be followed up with risk analysis to identify and mitigate any remaining risks and uncertainty related to the process [18,41,77,78]. Once significant flaws are addressed, the system is complete and ready for signoff by the laboratory's medical director. Next frontline staff training and “go-live” planning occurs. Going live is another opportunity to test the system through aggressive initial auditing which gets tapered off to less frequent but regular audits. These regular audits become part of the AV quality assurance system over the longer term. The continued capture of metrics related to process effectiveness, or non-conformances, also become part of ongoing process and quality improvement for the AV system.

4. Limitations and conclusions

Despite the many benefits of AV, several cautions deserve noting. Security of information once programmed requires control. Ease of programming can also mean ease for change and inadvertent change errors. AV is usually built on relatively simple Boolean (if-then type) logic routines to support complex decisions [42,75]. These to do not apply abductive reasoning and are incapable of evaluating the probability of a result being in error within the context of other analytical values and conditions [56]. Most AV capable software has relatively limited pattern recognition capability and hence, result flagging or even multiple flags often have little value in predicting the likelihood of error, which if available could help guide a more measured responsive action. Laboratory samples are received from many different settings and a common cause of error affecting samples from one location may not be so for another. For example, some clinical units like neonatal intensive care units or emergency departments tend to experience more hemolyzed samples, while other sample quality challenges face samples coming from an outpatient collection area. Customizing of AV rules to address local patient mix is important to optimizing the effectiveness of the AV strategy. In other instances, rules at different stages of AV algorithms can interact with each other in unpredicted manner. As it is not possible to test all possible combinations of situations during validation of the system, the importance for continued auditing of the system, and updating to address non-conformances and occurrences of error. Finally, there is still little guidance available on which rules and algorithms should be set up first, which specific components should be part of a specific test algorithm, which benchmarks should be set to measure effectiveness of AV algorithms, or to which quality standards or level of error detection the system must attain. While this review has attempted to help fill some of this gap, much work is still required.

But, despite its limitations AV proves superior to the manual review and verification alternative, especially when applied to large core laboratory settings. Most laboratories are faced with the challenge of providing efficient and cost effective services and doing more with less. AV offers and delivers on the opportunity for gains in efficiency and TAT, increase in error detection and reduction in erroneous reports, hence contributing to overall quality of services and patient safety. The software requirements needed to support AV is present in many LIS and/or middleware, hence to potential for benefit by limited or extensive AV exists for most laboratories. These points are important considerations to making the business case for AV, achieving administrative approval, and the making available of IT resources for the project. An appreciation of the benefits of AV and responsive action will lead to continued growth and improvement on AV in Clinical Laboratories as a quality assurance and improvement tool.

References

[1] P. Antal-Szalmás, G. Ivády, A. Molnár, Z. Hevessy, V. Kissné Sziráki, V. Oláh A, ... J. Kappelmayer, “Turnaround time”: a new parameter for the characterization of the overall efficacy of laboratory diagnostic processes, Orv. Hetil. 148 (28) (2007) 1317–1327.

[2] P.W. Barnes, S.L. McFadden, S.J. Machin, E. Simson, The international consensus Group for Hematology Review: suggested criteria for action following automated CBC and WBC differential analysis, Lab. Hematol. 11 (2) (2005) 83–90.

[3] J.M. Baron, C.H. Mermel, K.B. Lewandrowski, A.S. Dighe, Detection of preanalytic laboratory testing errors using a statistically guided protocol, Am. J. Clin. Pathol. 138 (3) (2012) 406–413.

[4] K.E. Blick, Providing critical laboratory results on time, every time to help reduce emergency department length of stay: how our laboratory achieved a six sigma level of performance, Am. J. Clin. Pathol. 140 (2) (2013) 193–202.

[5] P. Bonini, M. Plebani, F. Ceriotti, F. Rubboli, Errors in laboratory medicine, Clin. Chem. 48 (5) (2002) 691–698.

[6] J. Cadamuro, A.M. Simundic, E. Ajzner, S. Sandberg, A pragmatic approach to sample acceptance and rejection, Clin. Biochem. 50 (10 − 11) (2017) 579–581.

[7] P. Carraro, T. Zago, M. Plebani, Exploring the initial steps of the testing process: frequency and nature of pre-preanalytic errors, Clin. Chem. 58 (3) (2012) 638–642.

[8] Valenciano F. Cava, Autovaledación de resultados en el laboratorio clínico, Ed Cont Lab Clín. 13 (2010) 104–135.

[9] R. Chawla, B. Goswami, D. Tayal, V. Mallika, Identification of the types of pre-analytical errors in the clinical chemistry laboratory: 1-year study at GB pant hospital, Lab. Med. 41 (2) (2010) 89–92.

[10] Clinical and Laboratory Standards Institute, Autoverification of Clinical Laboratory Test Results, Approved Guideline. CLSI, Wayne, PA, 2006 (CLSI: AUTO10-A).

[11] Clinical and Laboratory Standards Institute, Collection, Transport and Processing of Blood Specimen for Testing Plasma-based Coagulation Assays ans Molecular Hemostasis Assays, 5th edn, Approved Guideline, 2008 (CLSI: H21-A5).

[12] Clinical and Laboratory Standards Institute, Use of delta checks in the medical laboratory. CLSI document EP33, 1st ed., Clinical and Laboratory Standards Institute; 2016, Wayne, PA, 2016.

[13] Clinical and Laboratory Standards Institute, Autoverification of Medical Laboratory Results for Specific Disciplines (Proposed Draft) CLSI, (2019) (Wayne, PA. CLSI: AUTO15).

[14] E. Clot-Silla, A. Argudo-Ramirez, X. Fuentes-Arderiu, Letter to the editor: measured values incompatible with human life, EJIFCC 22 (2) (2011) 52.

[15] College of American Pathologists, Laboratory General Checklist, CAP Accreditation Program, College of American Pathologists, Northfield, IL, 2012 (Last accessed on July 12, 2018 Available from), http://www.pathologie-online.de/wordpress/wp-content/uploads/2012/11/laboratory_general_checklist.pdf.

[16] M.P. Contreras, A.B. Font, X.F. Arderiu, Resultats inversemblants i la norma ISO 15189, In vitro Veritas 7 (2006).

[17] M.P. Cornes, Exogenous sample contamination. Sources and interference, Clin. Biochem. 49 (18) (2016) 1340–1345.

[18] L.J. Crolla, J.O. Westgard, Evaluation of rule-based autoverification protocols, Clin. Leadership Manage. Rev. 17 (5) (2003) 268–272.

[19] B.I. Dalal, M.L. Brigden, Factitious biochemical measurements resulting from hematologic conditions, Am. J. Clin. Pathol. 131 (2) (2009) 195–204.
[20] G.M. Davis, Autoverification of the peripheral blood count, Lab. Med. 25 (8) (1994) 528–531.
[21] G.M. Davis, Clinical microscopy: autoverification of macroscopic urinalysis, Lab. Med. 30 (1) (1999) 56–60.
[22] F. Demirci, P. Akan, T. Kume, A.R. Sisman, Z. Erbayraktar, S. Sevinc, Artificial neural network approach in laboratory test reporting: learning algorithms, Am. J. Clin. Pathol. 146 (2) (2016) 227–237.
[23] W.J. Dixon, Processing data for outliers, Biometrics 9 (1) (1953) 74–89.
[24] J.N. Doctor, G. Strylewicz, Detecting 'wrong blood in tube' errors: evaluation of a Bayesian network approach, Artif. Intell. Med. 50 (2) (2010) 75–82.
[25] A. Dolci, M. Panteghini, Harmonization of automated hemolysis index assessment and use: Is it possible? Clin. Chim. Acta. 432 (2014) 38–43.
[26] R.M. Dorizzi, B. Caruso, S. Meneghelli, P. Rizzotti, The DNSev™ expert system in the auto-verification of tumour markers and hormones results, Accred. Qual. Assur. 11 (6) (2006) 303–307.
[27] D.R. Dufour, D.L. Cruser, T. Buttolph, The clinical significance of delta checks, Am. J. Clin. Pathol. 110 (1998) 531.
[28] E.J. Dunn, P.J. Moga, Patient misidentification in laboratory medicine: a qualitative analysis of 227 root cause analysis reports in the veterans health administration, Arch. Pathol. Lab. Med. 134 (2) (2010) 244–255.
[29] S. Eisinger, M. Schwartz, M. Rosol, Dam Lisa, S. Riedel, Utility of reflex urine culture based on results of urinalysis and automated microscopy, American Association for Clinical Chemistry, Annual Conference, Los Angeles, CA, 2012.
[30] C.J. Farrell, A.C. Carter, Serum indices: managing assay interference, Ann. Clin. Biochem. 53 (Pt 5) (2016) 527–538.
[31] M.S. Feitosa, D.H. Bücker, S.M.E. Santos, L.S. Vasconcellos, Implementation of criteria for automatic release of clinical chemistry test results in a laboratory at an academic public hospital, J. Brasileiro de Patologia e Med. Lab. 52 (3) (2016) 149–156.
[32] D.C. Fernandez, S.S. Avinash, M. Malathi, A.R. Shivashankara, A. Kumar, P.A. Fernandez, Establishing the reference change values (RCVs) and validating the delta check auto-verification in a clinical biochemistry laboratory, Muller J Med Sci Res 8 (2017) 42–46.
[33] E. Fernández-Grande, C. Varela-Rodriguez, L. Sáenz-Mateos, A. Sastre-Gómez, P. García-Chico, T.J. Palomino-Muñoz, Impact of reference change value (RCV) based autoverification on turnaround time and physician satisfaction, Biochemi. Med. 27 (2) (2017) 342–349.
[34] E. Fliser, K. Jerkovic, T. Vidovic, M. Gorenjak, Investigation of unusual high serum indices for lipemia in clear serum samples on siemens analysers dimension, Biochem. Med. 22 (3) (2012) 352–362.
[35] C.G. Fraser, Reference change values, Clin. Chem. Lab. Med. 50 (5) (2012) 807–812.
[36] C.G. Fraser, H.P. Stevenson, I.M. Kennedy, Biological variation data are necessary prerequisites for objective autoverification of clinical laboratory data, Accred. Qual. Assur. 7 (11) (2002) 455–460.
[37] P. Froom, M. Barak, Auto-validation of complete blood counts in an outpatient's regional laboratory, Clin. Chem. Lab. Med. 53 (2015) 275–279.
[38] P. Froom, E. Saffuri-Elias, M. Barak, Autovalidation rates in an outpatient coagulation laboratory, Int. J. Lab. Hematol. 37 (5) (2015) 680–685.
[39] A.E. Garner, A.J. Lewington, J.H. Barth, Detection of patients with acute kidney injury by the clinical laboratory using rises in serum creatinine: comparison of proposed definitions and a laboratory delta check, Ann. Clin. Biochem. 49 (1) (2012) 59–62.
[40] R.J. Goldstein, N.S. Lichtenstein, D. Souder, The myth of the low anion gap, JAMA. 243 (17) (1980) 1737–1738, https://doi.org/10.1001/jama.1980.03300430039020.
[41] K. Golubjatnikov, Failure Mode Effects Analysis of Auto-verification in the Clinical Laboratory, 341 Capstones, Culminating Projects, 2018 Graduate Master's theses https://scholar.dominican.edu/masters-theses/34.
[42] R. Gómez-Rioja, V. Álvarez, M. Ventura, M.J. Alsina, N. Barba, M. Cortés, ... M. Ibarz, Current status of verification practices in clinical biochemistry in Spain, Clin. Chem. Lab. Med. 51 (9) (2013) 1739–1746.
[43] G.C. Guidi, G. Poli, A. Bassi, L. Giobelli, P.P. Benetollo, G. Lippi, Development and implementation of an automatic system for verification, validation and delivery of laboratory test results, Clin. Chem. Lab. Med. 47 (11) (2009) 1355–1360.
[44] E.K. Harris, Effects of intra- and interindividual variation on the appropriate use of normal ranges, Clin. Chem. 20 (12) (1974) 1535–1542.
[45] J.P. Harrison, G.M. McDowell, The role of laboratory information systems in healthcare quality improvement, Int. J. Health Care Qual. Assur. 21 (7) (2008) 679–691.
[46] R.J. Henry, S. Berkman, M.S. Little, R.J. Winzler, Inaccuracy of four chemical procedures as diagnostic tests for cancer, J. Am. Med. Assoc. 147 (1) (1951) 37–39.
[47] R.J. Henry, S. Berkman, (1953) errors in laboratory procedures, J. Am. Med. Assoc. 152 (12) (1953 Jul 18) 1166–1167.
[48] International Organization for Standardization, Medical Laboratories: Requirements for Quality and Competence, (2012) (ISO).
[49] Joint Commission International, Accreditation Standards for Laboratories ([Last accessed on June 12, 2019] Available from),, 2017. https://www.jointcommissioninternational.org/assets/3/7/JCI_Standards_for_Laboratories_STANDARDS-ONLY.pdf.
[50] J.B. Jones, A strategic informatics approach to autoverification, Clin. Lab. Med. 33 (1) (2013) 161–181.
[51] T. Kanno, I. Takeuchi, K. Sudo, Bivariate ratio monitoring in clinical laboratories, Am. J. Clin. Pathol. 76 (6) (1981) 782–787.
[52] J. Kappelmayer, J. Tóth, Clinical laboratories–production factories or specialized diagnostic centers, EJIFCC 27 (2) (2016) 156.
[53] C. Katz, K. McNicholas, R. Bounds, T. Figurelle, C. Jones, H. Farley, ... S.R. Johnson, Improving patient safety through enhanced communication between emergency department clinicians and medical laboratory staff, J. Clin. Outcomes Manag. 20 (10) (2013) 455–462.
[54] M.F. Kilinckaya, C. Yucel, T. Turhan, Research article an evaluation of efficiency in an emergency laboratory by turnaround time, Int. J. Med. Biochem. 1 (2) (2018) 57–61.
[55] M.D. Krasowski, S.R. Davis, D. Drees, C. Morris, J. Kulhavy, C. Crone, ... D. Voss, Autoverification in a core clinical chemistry laboratory at an academic medical center, J. Pathol. Inform. 5 (1) (2014) 13.
[56] Q.A. Le, G. Strylewicz, J.N. Doctor, Detecting blood laboratory errors using a Bayesian network: an evaluation on liver enzyme tests, Med. Decis. Mak. 31 (2) (2011) 325–337.
[57] T.K. Leen, D. Erdogmus, S. Kazmierczak, Statistical error detection for clinical laboratory tests, 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (pp. 2720-2723), IEEE, 2012, August.
[58] J. Li, B. Cheng, H. Ouyang, T. Xiao, J. Hu, Y. Cai, Designing and evaluating autoverification rules for thyroid function profiles and sex hormone tests, Ann. Clin. Biochem. 55 (2) (2018) 254–263.
[59] J. Li, B. Cheng, L. Yang, Y. Zhao, M. Pan, G. Zheng, ... Y. Cai, Development and implementation of autoverification rules for ELISA results of HBV serological markers, J. Lab. Auto. 21 (5) (2016) 642–651.
[60] D.A. Lindberg, Collection, evaluation, and transmission of hospital laboratory data, Methods Inf. Med. 6 (03) (1967) 97–107.
[61] G. Lippi, N. Blanckaert, P. Bonini, S. Green, S. Kitchen, V. Palicka, ... M. Plebani, Causes, consequences, detection, and prevention of identification errors in laboratory diagnostics, Clin. Chem. Lab. Med. 47 (2) (2009) 143–153.
[62] G. Lippi, C. Mattiuzzi, E.J. Favaloro, Pre-analytical variability and quality of diagnostic testing. looking at the moon and gazing beyond the finger, New Zealand J. Med. Lab. Sci. 69 (1) (2015) 4.
[63] A.H. Lou, M.O. Elnenaei, I. Sadek, S. Thompson, B.D. Crocker, B.A. Nassar, Multiple pre-and post-analytical lean approaches to the improvement of the laboratory turnaround time in a large core laboratory, Clin. Biochem. 50 (15) (2017) 864–869.
[64] J.V. Mancuso, P. Short, E. Reineks, E. Manilich, Learning to Classify Contaminated Blood Tests: A Neural Approach, Retrieved from https://openreview.net/pdf?id=S1vuwKJPz, (2018) ICLR 2018 Workshop submission.
[65] B. Marquardt, A Step-by-Step Process to 95% Autoverification, (2015) (CAP Today December Issue).
[66] O. Martinez-Nieto, A. Lozano-Gaitán, P. Beltrán-Diaz, I.L. Mojica-Figueroa, O.L. Morales-Reyes, M.A. Isaza-Ruget, Autoverification of the automated blood cell counter (CBC) in a reference laboratory in Bogota, Colombia, J. Brasileiro de Patologia e Med. Lab. 51 (6) (2015) 369–375.
[67] I. Miller, Development and evaluation of a logical delta check for identifying erroneous blood count results in a tertiary care hospital, Arch. Pathol. Lab. Med. 139 (8) (2015) 1042–1047.
[68] A. Mlinaric, M. Milos, D.C. Herak, M. Fucek, V. Rimac, R. Zadro, D. Rogic, Autovalidation and automation of the postanalytical phase of routine hematology and coagulation analyses in a university hospital laboratory, Clin. Chem. Lab. Med. (CCLM) 56 (3) (2018) 454–462.
[69] L. Onelöv, E. Gustafsson, E. Grönlund, H. Andersson, G. Hellberg, I. Järnberg, ... J.P. Antovic, Autoverification of routine coagulation assays in a multi-center laboratory, Scand. J. Clin. Lab. Invest. 76 (6) (2016) 500–502.
[70] W.P. Oosterhuis, H.J. Ulenkate, H.M. Goldschmidt, Evaluation of labrespond, a new automated validation system for clinical laboratory test results, Clin. Chem. 46 (11) (2000) 1811–1817.
[71] R. Palmieri, R. Falbo, F. Cappellini, C. Soldi, G. Limonta, P. Brambilla, The development of autoverification rules applied to urinalysis performed on the AutionMAX-SediMAX platform, Clin. Chim. Acta 485 (2018) 275–281.
[72] L. Pantanowitz, G.L. Horowitz, J.N. Upalakalin, B.A. Beckwith, Artifactual hyperbilirubinemia due to paraprotein interference, Arch. Pathol. Lab. Med. 127 (1) (2003) 55–59.
[73] M. Panteghini, The future of laboratory medicine: understanding the new pressures, Clin. Biochem. Rev. 25 (4) (2004) 207.
[74] E.S. Pearlman, L. Bilello, J. Stauffer, A. Kamarinos, R. Miele, M.S. Wolfert, Implications of autoverification for the clinical laboratory, Clin Leadersh Manag Rev 16 (2002) 237–239.
[75] J. Philip, Autoverification: Current Usage in Southern California and an Example Implementation Using Quality Tools and the Deming PDSA Cycle, California State University, Dominguez Hills, 2014.
[76] S. Rajput, S. Jain, Is autoverification of reports the need of the hour in clinical chemistry laboratory? Indian J. Med. Biochem. 22 (1) (2018) 56–60.
[77] E.W. Randell, G. Short, N. Lee, A. Beresford, M. Spencer, M. Kennell, ... D. Parry, Autoverification process improvement by six sigma approach: clinical chemistry & immunoassay, Clin. Biochem. 55 (2018) 42–48.
[78] E.W. Randell, G. Short, N. Lee, A. Beresford, M. Spencer, M. Kennell, ... D. Parry, Strategy for 90% autoverification of clinical chemistry and immunoassay test results using six sigma process improvement, Data Brief 18 (2018) 1740–1749.
[79] E.W. Randell, S. Yenice, Delta checks in the clinical laboratory, Crit. Rev. Clin. Lab. Sci. 56 (2) (2019) 75–97.
[80] P.W. Riley, B. Gallea, A. Valcour, Development and implementation of a coagulation factor testing method utilizing autoverification in a high-volume clinical reference laboratory environment, J. Pathol. Inform. 8 (2017).

[81] V. Rimac, I. Lapic, K. Kules, D. Rogic, M. Miler, Implementation of the auto-validation algorithm for clinical chemistry testing in the laboratory information system, Lab. Med. 49 (3) (2018) 284–291.
[82] P.K. Sandhu, K. Bandyopadhyay, W. Hunt, T.H. Taylor, R.J. Birch, J. Krolak, D.J. Ernst, Effectiveness of laboratory practices to reduce specimen labeling errors at the time of specimen collection in healthcare settings: alaboratory medicine best practices (LMBP™) systematic review, J. Appl. Lab. Med. (2017) jalm.
[83] G. Scott, Improving productivity: total urianlysis vs. reflex to microscopy, Retrieved from: https://pdfs.semanticscholar.org/c533/32e150bc352dbf3f9221b6c402417db76401.pdf.
[84] A.M.E. Sediq, A.G.H. Abdel-Azeez, Designing an autoverification system in Zagazig University hospitals laboratories: preliminary evaluation on thyroid function profile, Anna. Saudi Med. 34 (5) (2014) 427.
[85] M.C. Shih, H.M. Chang, N. Tien, C.T. Hsiao, C.T. Peng, Building and validating an autoverification system in the clinical chemistry laboratory, Lab. Med. 42 (11) (2011) 668–673.
[86] G. Solé-Enrech, X. Fuentes-Arderiu, Letter to the editor: unlikeliness limits estimation, EJIFCC 21 (3) (2010) 74.
[87] R.H. Sterns, M. Emmett, J. Forman, Overview of the treatment of hyponatremia in adults, UpToDate, Edited Post TW, UpToDate, Waltham, MA, 2017(Accessed on March 31, 2018).
[88] J.A. Straseski, F.G. Strathmann, Patient data algorithms, Clin. Lab. Med. 33 (1) (2013) 147–160.
[89] G.B. Strylewicz, Errors in the Clinical Laboratory: A Novel Approach to Autoverification, Doctoral dissertation, University of Washington, 2007.
[90] I. Suárez-Vega, X. Fuentes-Arderiu, Reportable interval of patient examination results and ISO 15189, Clin. Chem. Lab. Med. (CCLM) 43 (11) (2005) (1278–1278).
[91] N. Torke, L. Boral, T. Nguyen, A. Perri, A. Chakrin, Process improvement and operational efficiency through test result autoverification, Clin. Chem. 51 (12) (2005 Dec 1) 2406–2408.
[92] P.N. Valenstein, S.S. Raab, M.K. Walsh, Identification errors involving clinical laboratories: a College of American Pathologists Q-probes study of patient and specimen identification errors at 120 institutions, Arch. Pathol. Lab. Med. 130 (2006) 1106–1113.
[93] W. Vogt, B. Oesterle, Extreme results in electrolyte determination, Wien. Klin. Wochenschr. Suppl. 192 (1992) 21–27.
[94] E.A. Wagar, L. Tamashiro, B. Yasin, L. Hilborne, D.A. Bruckner, Patient safety in the clinical laboratory: a longitudinal analysis of specimen identification errors, Arch. Pathol. Lab. Med. 130 (11) (2006) 1662–1668.
[95] S. Warner, Simple autoverification, Adv. Admin. Lab. 22 (1) (2013) 10.
[96] G. Wennecke, G. Juel, Avoiding Pre-Analytical Errors–in Blood Gas Testing, (2008) (ISBN87–88138–94-1).
[97] P. Whitehurst, T.V. Di Silvio, G. Boyadjian, Evaluation of discrepancies in patients'results–an aspect of computer-assisted quality control, Clin. Chem. 21 (1) (1975) 87–92.
[98] J. Wu, M. Pan, H. Ouyang, Z. Yang, Q. Zhang, Y. Cai, Establishing and Evaluating Autoverification Rules with Intelligent Guidelines for Arterial Blood Gas Analysis in a Clinical Laboratory, SLAS TECHNOLOGY: Translating Life Sciences Innovation, 2018 (2472630318775311).
[99] L.Y. Xia, X.Q. Cheng, Q. Liu, L. Liu, X.Z. Qin, L. Zhang, ... L. Qiu, Developing and application of an autoverification system for clinical chemistry and immunology test results, Zhonghua Yi Xue Za Zhi 97 (8) (2017) 616–621.
[100] T. Yamashita, K. Ichihara, A. Miyamoto, A novel weighted cumulative delta-check method for highly sensitive detection of specimen mix-up in the clinical laboratory, Clin. Chem. Lab. Med. 51 (4) (2013) 781–789.
[101] C. Yan, Y. Zhang, J. Li, J. Gao, C. Cui, C. Zhang, ... X. Han, Establishing and validating of an laboratory information system-based auto-verification system for biochemical test results in cancer patients, J. Clin. Lab. Anal. 33 (5) (2019) e22877, https://doi.org/10.1002/jcla.22877 Epub 2019 Mar 6.
[102] X. Zhao, X.F. Wang, J.B. Wang, X.J. Lu, Y.W. Zhao, C.B. Li, B.H. Wang, J. Wei, P. Guo, J.P. Xiao, J.H. Wang, Multicenter study of autoverification methods of hematology analysis, J. Biol. Regul. Homeost. Agents 30 (2) (2016) 571–577.
[103] Y. Zhao, L. Yang, G. Zheng, Y. Cai, Building and evaluating the autoverification of coagulation items in the laboratory information system, Clin. Lab. 60 (1) (2014) 143–150.