e-ISSN: 0976-822X, p-ISSN:2961-6042

Available online on http://www.ijcpr.com/

International Journal of Current Pharmaceutical Review and Research 2026; 18(6); 2741-2753

Original Research Article

Internal Quality Control in Clinical Biochemistry: From Coefficient of Variation to Levey–Jennings Charts, Westgard Rules, and Corrective Action—A Narrative Review

Vikram Kesar[1], Alpana Goel Kesar[2]

[1]Consultant Biochemistry, Department of Biochemistry, Bhagwan Mahavir Hospital, Pitampura, New Delhi, India

[2]Reader, Department of Anatomy, College of Dental Science and Hospital, Rau, Indore, M.P., India

Received: 01-04-2026 / Revised: 25-05-2026 / Accepted: 12-06-2026

Corresponding Author: Dr. Vikram Kesar

Conflict of interest: Nil

Abstract

Internal quality control (IQC) is the laboratory's principal mechanism for detecting clinically important deterioration in quantitative measurement procedures before unreliable patient results are released. Yet IQC is often reduced to fixed control intervals, universal coefficient-of-variation targets or mechanical application of every Westgard rule. This focused narrative review links the statistical basis of IQC to operational decisions in routine clinical biochemistry. It explains how mean, standard deviation, coefficient of variation, bias, analytical performance specifications and Sigma metrics should be interpreted together; how control materials, concentrations, target values, frequency and rules should be selected; and how Levey-Jennings patterns and multirule signals can be translated into hypotheses about random error, systematic shift, drift and concentration-dependent failure. A structured response to an out-of-control event is presented, covering containment, signal verification, cause-directed investigation, assessment of the affected analytical interval, review of patient results, corrective action and effectiveness monitoring. Conventional control-material IQC remains indispensable, but its limitations require integration with reagent-lot verification, external quality assessment, equipment and quality-system records, and appropriately validated patient-based real-time quality control. The central recommendation is analyte-specific, risk-based IQC: the quality required for clinical use should determine the control strategy, and every rejection should lead to a documented decision about analytical status and patient safety.

Keywords: Internal Quality Control; Clinical Biochemistry; Coefficient of Variation; Analytical Imprecision; Levey-Jennings Chart; Westgard Rules.

DOI: 10.25258/ijcpr.18.6.420

Introduction

Clinical biochemistry results influence diagnosis, treatment thresholds, therapeutic monitoring and prognosis. Their value depends on more than instrument precision: the complete measurement process must remain stable, traceable and fit for the intended clinical use. ISO 15189 therefore requires laboratories to monitor the validity of examination results, manage nonconforming work and prevent release of results when quality cannot be assured [1]. Although automation has reduced many analytical failures, errors still arise from calibration drift, reagent deterioration, pipetting or photometric instability, environmental change, lot variation and inappropriate intervention. Even a modest deviation can be clinically important when patient values lie close to a decision limit [2-4].

Internal quality control (IQC) uses repeated measurements of control material to determine whether a measurement procedure continues to perform as expected. Its immediate purpose is not to prove trueness, but to detect change from an established stable state in time to protect patient results. External quality assessment (EQA) addresses a different question by periodically comparing performance with an assigned value, reference procedure or peer group. IQC and EQA are therefore complementary components of the quality-management system [5,6]. Quality indicators across the pre-examination, examination and post-examination phases remain necessary because acceptable control results cannot detect patient identification, specimen or reporting errors [7,8].

In practice, IQC is often applied as a ritual: two control levels are analysed at fixed times, manufacturer ranges are accepted without local verification, a coefficient of variation (CV) below an arbitrary percentage is labelled satisfactory, and every Westgard rule is activated irrespective of assay capability. These practices can either miss medically important error or generate repeated false rejection. The more defensible approach begins with the quality required for the analyte, estimates stable method performance, and selects control levels, frequency and statistical rules able to detect deterioration before unacceptable patient risk accumulates [5,9]. This review focuses specifically on that decision pathway in quantitative clinical biochemistry, from CV and Levey-Jennings interpretation to investigation, corrective action and result release.

Focused Narrative-Review Approach: A structured narrative search was used to support a practice-oriented synthesis rather than a scoping review or meta-analysis. PubMed/MEDLINE and Google Scholar were searched through 31 July 2026 using combinations of internal quality control, clinical chemistry, coefficient of variation, analytical imprecision, Levey-Jennings, Westgard rules, QC frequency, analytical error, corrective action and patient-based real-time quality control. ISO, CLSI, IFCC and EFLM standards or recommendations and reference lists of key papers were also reviewed. Foundational methodological studies, current guidance and evidence directly applicable to quantitative clinical biochemistry were prioritised. Heterogeneous standards, statistical papers, surveys and reviews were integrated narratively, consistent with accepted guidance for narrative reviews and SANRA principles [10,11].

The Analytical Quality Model for IQC: IQC decisions require clear separation of precision, bias and total uncertainty. Precision describes closeness among repeated results under specified conditions; it does not indicate proximity to the true or reference value. Bias is the systematic difference between the expectation of results and an accepted reference. Measurement uncertainty describes the dispersion of values that could reasonably be attributed to the measurand and is estimated from relevant sources of variation [12-14].

Total analytical error and measurement uncertainty are different models, but both emphasise that imprecision alone cannot establish clinical acceptability [15]. Table 1 summarises the statistics most directly used in routine IQC.

Table 1: Core statistics used in internal quality control [12-20,21]

Measure	Expression or meaning	Primary IQC use	Key limitation
Mean	Average of stable control results	Defines the control centre	Can conceal subgroups or lot-specific shifts
SD	Absolute dispersion around the mean	Sets statistical chart limits	Depends on concentration and data stability
CV (%)	100 x SD / mean	Compares relative imprecision	Unstable near zero; no universal acceptable value
z-score	(result - mean) / SD	Places results on a common SD scale	Valid only if mean and SD represent a stable process
Bias	Difference from an accepted target	Assesses systematic deviation	Target and commutability determine credibility
Sigma metric	(TEa - |bias|) / CV	Supports risk-based rule selection	Highly dependent on TEa, bias and CV estimates

CV, coefficient of variation; SD, standard deviation; TEa, allowable total error.

The mean and standard deviation (SD) define the centre and absolute dispersion of a stable control distribution. A standardised control result can be expressed as $z = (x - \text{mean})/SD$, allowing values at different concentrations to be placed on a common SD scale. The CV, calculated as 100 x SD/mean, expresses imprecision relative to concentration and is useful for comparing precision across control levels, time periods, instruments or laboratories. Its apparent simplicity is also its main hazard: the same numerical CV can be acceptable for one analyte and inadequate for another, and the CV becomes unstable when the mean approaches zero.

Analytical performance specifications (APS) provide the criterion against which observed performance is judged. The Milan consensus describes a hierarchy based on clinical outcomes, biological variation or state of the art [16]. Biological-variation data can support desirable imprecision and bias goals, but the underlying studies require critical appraisal and the derived specification must suit the measurand and clinical application [19,20]. Regulatory or EQA limits may be appropriate when higher-order models are unavailable, but an EQA acceptance limit should not automatically become the laboratory's daily IQC limit. ISO 13528 addresses statistical methods for proficiency testing rather than the design of routine within-laboratory control [22]. Precision estimates must be representative of routine

operation. Short replicate studies primarily estimate repeatability and often understate variation accumulated across days, calibration cycles, reagent lots and operators. CLSI EP05 describes experimental designs for precision evaluation, whereas EP15 supports user verification of precision and estimation of bias [17,18]. For ongoing IQC, a laboratory should review CV over a defined stable interval, exclude only results linked to documented assignable causes, and stratify data when a new reagent or control lot, multiple systems or a changed calibration state creates a distinct population. Multilaboratory data demonstrate that achievable CV varies substantially by analyte and platform [23], while combining results across reagent lots or measuring systems can inflate the estimate and obscure the true source of variation [24].

A practical CV review should distinguish short-term, within-lot and long-term estimates. Short-term CV is useful when verifying immediate precision after installation or repair; within-lot CV reveals whether the current reagent-control combination is stable; long-term CV captures routine variation relevant to uncertainty and service performance. The laboratory should specify the interval, number of observations, treatment of rejected runs and minimum data required before comparison. A falling CV after deleting rejected results is not necessarily improvement: it may reflect selection bias. Conversely, a long-term CV that rises only at a lot transition should prompt separate lot means before attributing the change to random error. Graphs of monthly CV, rejection rate and target mean are more informative than a single annual percentage because they show whether imprecision, centring or both changed.

Sigma metrics may help translate performance into QC design. A commonly used expression is sigma = (TEa - |bias|)/CV, with all terms expressed as percentages [21]. A higher value implies greater tolerance before the allowable total error is exceeded; a lower value indicates that small deterioration may create unacceptable results. However, Sigma is only as credible as the selected allowable error, bias estimate and CV. It should guide, not replace, professional judgement. Evidence-based QC planning therefore combines APS, stable method performance, error-detection probability, false-rejection probability, run size and potential patient harm [21,25].

Designing a Fit-For-Purpose IQC Plan: The control material should challenge the complete analytical procedure and remain sufficiently stable for longitudinal monitoring. Third-party material can provide independence from an instrument-reagent system, but neither third-party nor manufacturer control is automatically commutable with patient serum. Matrix effects may respond differently to reagent-lot changes, calibration or interference. Laboratories should therefore document material type, preparation, storage, open-vial stability and traceability; use independent aliquots where feasible; and distinguish a control-material problem from failure of the patient-sample measurement process [26,27,28].

At least two concentrations are commonly used because a method may perform differently near the low and high parts of its measuring interval. The selected levels should be clinically meaningful: for example, one near a diagnostic or treatment threshold and another in a pathological range. Additional levels may be justified for assays with a wide range, complex calibration, nonlinearity risk or several decision limits. Control concentrations that are both comfortably within the reference interval provide little assurance at medical decision points. Conversely, adding levels without considering detection performance increases cost and can raise the probability of false rejection.

For a new control lot, a provisional target may be taken from the manufacturer only while local data are accumulated. The laboratory should verify or establish its mean and SD using results generated by the routine procedure across representative days and analytical conditions. Grossly erroneous observations may be excluded only with a documented cause; deleting values merely because they widen the SD creates unrealistically narrow limits. A mean inherited from a peer group may conceal local bias, while a broad manufacturer SD may permit deterioration that the laboratory's own stable method could detect. Lot overlap is essential so a new target is characterised before the previous lot expires [5,26,27].

QC frequency should be defined by opportunities for failure and the number of patient results at risk, not only by the clock. Controls are generally appropriate at the beginning of an analytical period and after calibration, major maintenance, reagent replacement or an event capable of changing performance. High-throughput continuous systems may require controls within a run because a failure shortly after the opening QC event could affect hundreds of specimens before the next scheduled control. Low-volume batch testing may appropriately use controls with each batch. Parvin showed that frequency materially affects the number of unreliable results released before detection [29]. Current IFCC recommendations similarly emphasise planned frequency, documented review and response aligned with ISO 15189 [26].

Consider a high-volume electrolyte assay that releases 600 patient results between two daily control events. Even a rule with good detection probability can allow a large number of results to

be reported if failure occurs immediately after the first event. Reducing the interval to every 200 specimens, or adding a validated patient-median alert, shortens the maximum exposure. By contrast, an infrequently requested tumour marker tested in a discrete batch may be safely controlled with each batch and after reagent or calibration changes. These examples show why identical twice-daily schedules can create different risks. Frequency, rule power and run size should be designed together; increasing rule complexity without shortening an excessive patient run may not provide timely protection [21,29,30].

Reagent- and calibrator-lot changes deserve separate verification. Commercial controls may exaggerate or fail to show a lot-related effect because of noncommutability. Comparison of fresh patient specimens spanning the clinically relevant range is therefore the preferred way to assess whether a new lot changes reported results, supplemented by control material and calibration information [28,31].

Long-term CV must also be interpreted with lot history visible; otherwise a stable within-lot method may appear imprecise because two lot-specific means were combined. Global survey data reveal substantial variability in how laboratories establish targets, apply rules and handle materials, reinforcing the need for a written, locally justified plan [32].

Resource limitations change the implementation route but not the underlying quality question. A low-volume laboratory may use single-use or frozen aliquots, analyse controls with each patient batch, share peer-group data and use retained patient specimens for comparability checks. It should not broaden limits simply to avoid material use. When sufficient local observations cannot be accumulated rapidly, provisional targets may be verified in stages with tighter supervisory review until robust estimates are available. Referral or temporary transfer to another validated system may be safer than operating an unstable low-volume assay. Point-of-care devices require the same logic, but control lockout, operator identification, reagent storage, connectivity and central-laboratory oversight become especially important because many nonlaboratory operators share the process [1,5,26].

The plan should state who may accept, reject, troubleshoot and resume testing; which rules are applied within and across runs; how warnings differ from rejection; what event defines the start of the potentially affected interval; and when patient results are held, repeated or amended. The plan must be reviewed after method, workload, software, reagent, clinical use or failure-pattern changes. IQC is therefore a controlled risk intervention, not a static analyser configuration [5,26,33].

Reading the Levey-Jennings Chart: Levey and Jennings adapted statistical process-control charts for laboratory measurements by plotting sequential control results around a target mean with lines at 1, 2 and 3 SD [34]. The chart converts a series of numbers into a time-ordered pattern. A stable process should show irregular dispersion around the mean without sustained direction. One isolated extreme point suggests a large random event, whereas several points displaced to one side suggest a systematic shift. A progressive rise or fall indicates drift; widening scatter without a changed mean suggests increased random error. Figure 1 presents these patterns schematically.

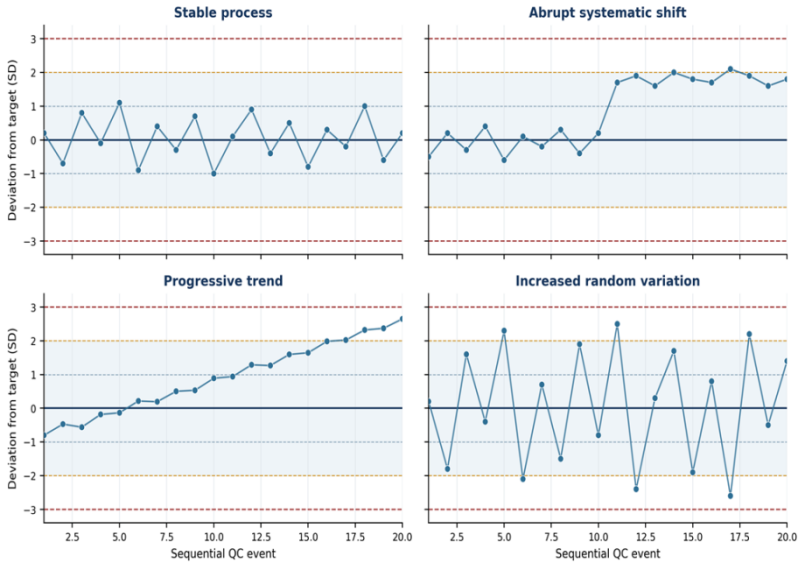

Figure 1: Schematic Levey-Jennings patterns. Values are illustrative standard-deviation units, not patient or assay data

Interpretation must retain the time axis and operational context. The same 2-SD result has different meaning when isolated in months of stable data, when it follows a reagent change, or when three other levels have moved in the same direction.

Reviewing only the current point discards evidence contained in sequence, synchrony across levels and association with calibration, maintenance or lot events. Electronic charts should therefore display annotations and permit adequate history, while manual charts should record relevant changes rather than merely coloured dots.

Multilevel review increases diagnostic value. If low and high controls move by similar z-scores, a common proportional bias is plausible. If their absolute differences are similar but z-scores differ, the effect may be concentration dependent because each level has a different SD. Opposite movement can reflect a curve or calibration-model change, whereas isolated movement of one material may implicate the control lot rather than the assay. For assays run on duplicate instruments, separate charts should be maintained unless comparability and common targets have been demonstrated; a pooled chart can hide a stable offset between systems. Annotations should identify control and reagent lots, calibrations, maintenance and software changes so temporal association can be examined without relying on memory [35-37]. Control limits are statistical, not clinical. If the SD is inflated by unstable historical performance, a clinically important shift may remain inside 2 SD. If the SD is underestimated from a short or selectively edited dataset, acceptable measurements may be repeatedly rejected. The laboratory should compare the chart's scale with the APS and periodically review whether the stored mean and SD still represent stable performance. A control chart should reveal change; repeatedly recentering it after unexplained shifts normalises failure and destroys that function [5,35,36].

Westgard Rules: Selective Use, Not Mechanical Use: A single 3-SD rejection limit detects large errors with relatively low false rejection but may be insensitive to smaller persistent shifts. Westgard and colleagues combined complementary Shewhart rules to improve error detection [7]. The conventional 1_2s rule is a warning, whereas 1_3s, 2_2s, R_4s, 4_1s and consecutive same-side rules are commonly used as rejection or investigation criteria. Table 2 states each rule in plain language. Rule performance depends on the number of control observations and the size and type of error; it is not an intrinsic property of the rule name [35].

Table 2: Common Westgard rules and their practical interpretation [30,38-41]

Rule	Statistical criterion	Usual response	Error mainly suggested
1_2s	One result exceeds +2 SD or -2 SD	Warning; inspect other levels and history	Possible random or systematic error
1_3s	One result exceeds +3 SD or -3 SD	Reject	Large random error or substantial shift
2_2s	Two consecutive results exceed the same 2-SD limit	Reject	Systematic error
R_4s	Within one event, one result exceeds +2 SD and another -2 SD	Reject	Random error; consider curve effects
4_1s	Four consecutive results exceed the same 1-SD limit	Reject	Systematic error
8x/9x/10x	Consecutive results remain on one side of the mean	Reject or investigate as predefined	Small, persistent shift
7T	Seven consecutive results progressively rise or fall	Investigate or reject as predefined	Gradual trend or drift

Rule application depends on the number of control observations, within- versus across-run configuration, method performance and patient risk.

Rules are often grouped by the errors they most readily identify. R_4s is principally sensitive to increased random error because it requires a large difference between two results in the same event. The 2_2s, 4_1s and same-side rules are more sensitive to systematic shift. The 1_3s rule may detect either a large random event or a substantial shift. These labels generate hypotheses rather than diagnoses. For example, opposite low- and high-level changes could violate R_4s but arise from a calibration-curve problem rather than random imprecision. Adding rules generally increases error detection but also increases false rejection. If many independent opportunities are tested, a statistically unusual event eventually occurs even in a stable process. False alarms consume material and staff time, delay reporting and can provoke harmful recalibration. The solution is not to ignore warnings, but to choose the smallest rule set that meets the required detection performance. High-performing assays may need only simple rules; marginal assays may require multirules, more

observations or more frequent QC, but improving or replacing the method is preferable to indefinitely increasing statistical complexity [21,40]. Risk-based rule selection links the probability of detecting a medically important error with the expected number of patient specimens exposed before detection. Graphical and quantitative planning methods can identify combinations of rules, numbers of controls and run size that achieve acceptable risk [30,41]. A 1_2s warning should not automatically stop every run; it should prompt inspection of other levels and recent history. A violated rejection rule should stop release until the signal is resolved. Combining within-run and across-run rules requires clear software configuration, so staff know which observations contributed to a rejection. A worked selection illustrates the principle. Suppose an assay has a small stable CV and negligible bias relative to its APS. A simple 1_3s rule with two control levels may provide sufficient detection of a medically important shift with low false rejection. Activating 2_2s, 4_1s and 10x in addition may add little safety but generate frequent investigations. A second assay operating close to its allowable error has less margin: the laboratory may need two levels, a sensitive multirule procedure and shorter patient runs. If this configuration still cannot achieve acceptable detection, the method is not made safe merely by adding controls; recalibration, maintenance, method improvement or replacement should be considered. The documented plan should state the assumed CV and bias because deterioration in either invalidates the original rule design [21,30,40,41]. CUSUM and exponentially weighted moving average procedures can complement Shewhart rules when small, persistent changes matter. By accumulating or weighting deviations, they can detect shifts that remain individually within conventional limits [42]. Their increased sensitivity also demands validated limits and a defined response. The practical principle is consistent across approaches: use local performance and clinical risk to decide what error must be detected, then verify that the selected procedure can detect it with an acceptable false-alarm burden [25,40].

Translating a QC Signal into an Error Hypothesis: A QC rejection indicates that the observed pattern is unlikely under the assumed stable model; it does not identify the cause. Investigation should first determine whether the event is isolated or coordinated. One extreme low-level control with stable other levels suggests an aspiration event, control preparation problem or level-specific matrix effect.

Movement of all levels in the same direction suggests calibration, reagent or common-system bias. Increased scatter around an unchanged mean suggests deterioration in precision. Table 3 links common patterns to provisional interpretations.

Table 3: Levey-Jennings patterns and provisional error hypotheses [35-37,43]

Observed pattern	Provisional interpretation	First contextual checks
One extreme result; other levels stable	Random event, vial or preparation problem	Aspiration flags, aliquot, mixing, reconstitution
All levels shift in one direction	Calibration, reagent or common systematic error	Calibration, reagent lot, maintenance
Only low level shifts	Low-range sensitivity, nonlinearity or matrix effect	Low calibration point, control lot, LoQ behavior
Only high level shifts	Upper-range calibration, nonlinearity or dilution issue	Curve fit, dilution protocol, high-control stability
Low and high levels move oppositely	Calibration-curve distortion or apparent R_4s pattern	Calibration model, pipetting, replicate precision
One module shows several analyte shifts	Shared fluidic, optical, thermal, water or software failure	Module logs, temperatures, maintenance, water
Mean stable but scatter increases	Increased random error	Probe, pipetting, bubbles, mixing, component wear
Patient results shift; controls stable	Noncommutability, lot effect or patient-specific interference	Patient lot comparison, medians, reaction curves

These are hypotheses, not diagnoses; interpret them with timing, concentration, other analytes and operational records.

The direction and concentration dependence of change are highly informative. A uniform proportional shift may implicate calibration or reagent response; a constant absolute difference can be proportionally larger at low concentration. Divergent movement at low and high levels suggests nonlinearity or calibration-curve distortion. A shift confined to one analyte favours an analyte-specific reagent, calibrator or application problem; concordant shifts across several assays on one module suggest shared pipetting, temperature, optics, water or software.

Recent events should be temporally matched rather than assumed: reagent installation, calibration, maintenance, control reconstitution, environmental excursion and software change all define testable hypotheses [35-37].

Random-error investigation should focus on components that vary from aspiration to aspiration: bubbles, clots or particles in the control; inadequate mixing; intermittent probe obstruction; pipetting imprecision; unstable temperature; electrical disturbance or mechanical wear. Systematic-error investigation should focus on factors that persist across measurements: calibration assignment, reagent preparation or deterioration, lot change, photometric offset, electrode condition, incubation temperature or calculation factor. Trends suggest a progressive mechanism such as reagent ageing, lamp deterioration, electrode drift, evaporation or deposit accumulation. This classification directs the first checks, but it should not become a rigid rule. A severe calibration error can produce an isolated 1_3s signal early, and intermittent hardware failure can mimic a trend if its frequency increases over time.

Control material cannot reveal every clinically relevant failure. It may not mimic patient-sample viscosity, endogenous interference, carryover, antibody effects, biotin exposure or paraprotein behaviour. Irregular analytical errors may affect only a small subset of specimens and evade routine control statistics [43].

If patient medians, delta checks, analyser reaction curves, clinical complaints or interinstrument comparison change while commercial controls remain stable, the laboratory should consider noncommutability or specimen-specific mechanisms rather than concluding that the patient signal is wrong.

Statistical significance must be separated from clinical significance. The risk created by a shift depends on its magnitude and direction, concentration, proximity to decision limits, duration, number of exposed specimens and clinical use. A small sodium bias may reclassify many patients, while a similar percentage change in another analyte may have limited effect. The APS converts the chart signal into a quality judgement; patient-risk assessment determines the urgency and breadth of response. Pattern recognition is therefore only the entry point to an evidence-based investigation [16,21,33].

Managing an Out-of-Control Event: The first action after a valid rejection is containment: stop release of the affected test, hold unreported results, preserve the original QC values and analyser messages, record reagent, calibrator and control lots, and notify the responsible senior person. Testing may continue on a validated alternative system when clinically necessary. Repeating the same control until a value falls inside the limit is not containment; it discards evidence and can convert false acceptance into patient harm. The decision pathway in Figure 2 begins with verification of configuration and material handling, then proceeds to cause-directed investigation and assessment of patient impact.

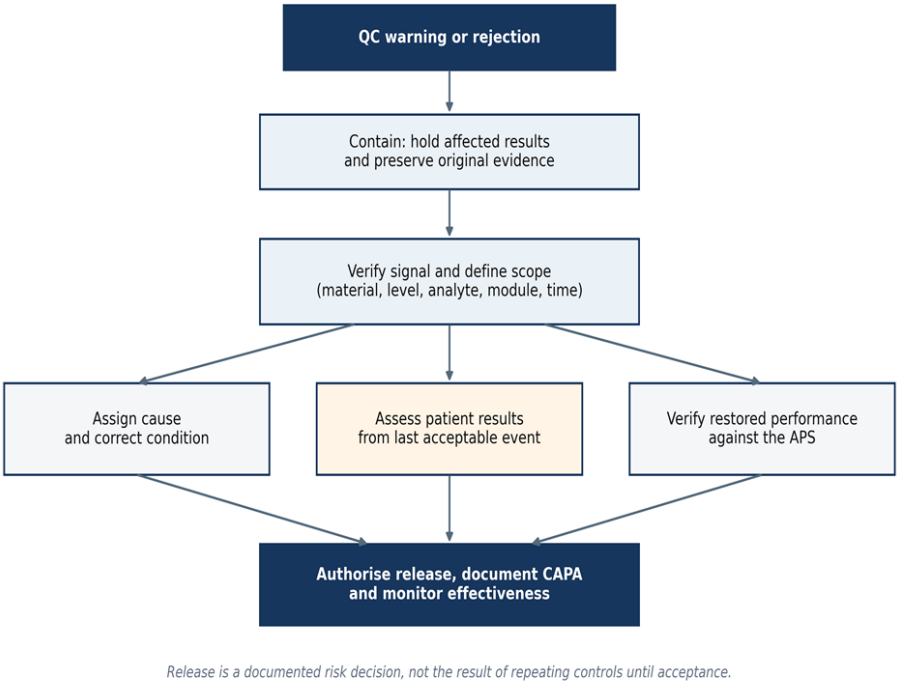

Figure 2: Structured response to an out-of-control IQC event. APS, analytical performance specification; CAPA, corrective and preventive action

Signal verification should be rapid but structured. Confirm the analyte, level, unit, lot, assigned mean and SD, rule configuration, expiry, storage, reconstitution, mixing, instrument flags and recent calibration or maintenance.

A fresh aliquot can distinguish a compromised vial from a persistent analytical failure, but the original result must remain in the record. If a repeat is acceptable, the laboratory still needs a reason to release results: an identified transient sampling event, documented control-vial problem or statistically justified warning response. An unexplained repeat-in-control does not invalidate the first rejection [5,44].

Scope the event before replacing components indiscriminately. Determine whether it affects one vial, one level, one analyte, one reagent lot, one module, the complete platform or multiple sites. Establish the last acceptable control event and inspect the chart, calibration records, lot transitions, service logs, peer comparison and patient-result behaviour to estimate when failure began. Root-cause categories include materials, equipment, method, personnel, environment, information systems and management process. Simple events may be resolved by a Five-Whys analysis; recurrent or multifactorial problems may require an Ishikawa diagram, process map, Pareto analysis or formal risk assessment [33,44,45].

Correction and corrective action are not interchangeable (Table 4). Replacing a deteriorated reagent corrects the immediate condition; changing storage monitoring or inventory control to remove the underlying cause is corrective action.

Preventive or risk-control action addresses a credible failure before it occurs elsewhere. Cause-directed examples are shown in Table 5. CAPA should be proportionate: a single documented aspiration fault does not require system redesign, whereas recurrent control preparation errors may require revised instructions, calibrated dispensers, training and competency reassessment [44-47].

Table 4: Distinguishing correction, corrective action and effectiveness verification [43,44-47]

Activity	Purpose	Example
Correction	Controls the detected condition immediately	Replace an expired reagent or prepare a fresh control vial
Corrective action	Removes the demonstrated underlying cause to prevent recurrence	Revise reconstitution instructions and reassess competency after recurrent preparation errors
Preventive/risk-control action	Reduces the likelihood or consequence of a credible future failure	Introduce continuous refrigerator temperature alarms
Effectiveness verification	Shows that the action produced sustained improvement	Review subsequent QC, CV and event recurrence over a predefined period

Actions should be proportionate to documented risk and supported by evidence.

Table 5: Cause-directed response and tiered development of an IQC programme [26,37,48-57]

Stage or problem	Immediate response	System-level development
Incorrect control preparation	Prepare fresh material correctly	Standardise instructions, dispensers and competency assessment
Reagent deterioration	Replace or quarantine reagent	Improve storage, inventory and temperature alerts
Calibration failure	Resolve cause, then recalibrate	Review calibration preparation, interval and acceptance criteria
Probe or aspiration failure	Clean, repair or replace component	Trend aspiration errors and optimise preventive maintenance
Incorrect stored mean or SD	Correct statistics after verification	Independent lot-entry authorisation and periodic target review
Foundation	Written procedure, valid targets, relevant levels, charts, rules and EQA	Document containment, patient-impact review and authority to resume
Risk-based development	Use APS, throughput and error-detection performance	Add lot verification, uncertainty, peer comparison and optimised frequency
Advanced integration	Use electronic dashboards and linked quality records	Add validated EWMA/CUSUM, PBRTQC or multivariate alerts

The advanced stage complements rather than replaces a sound conventional programme.

Before resuming routine release, restored performance should be demonstrated using fresh control material at clinically relevant levels and, when indicated, replicates across more than one analytical event, calibration verification or comparison of retained patient specimens. The evidence required depends on the failure. A calibration or lot-related bias usually needs patient-sample comparison; a precision problem may require replicates; a repaired module may require controls across all affected assays. The criterion is not merely that one control is within 2 SD, but that the evidence supports removal of the suspected cause and return to performance meeting the APS.

Authorisation to resume should be separated from the technical act of rerunning a control. The authorised person should review the original signal, identified or most probable cause, corrective evidence, restored QC across relevant levels and the patient-impact decision. When the cause remains unknown, release requires stronger evidence and an explicitly documented risk judgement; repeated acceptable controls may demonstrate current stability but do not explain the earlier event. The procedure should define when manufacturer or service support is required, when an alternative analyser must be used and when testing should remain suspended. This separation reduces pressure on bench staff to release a backlog before the investigation is complete [1,44,47].

Patient-result assessment is integral to the investigation. Identify tests produced from the estimated onset of failure until containment, including already reported results. Estimate the magnitude and direction of error and prioritise values near diagnostic or therapeutic limits, critical results and vulnerable populations. Retained specimens should be retested when stable and clinically useful. If a material difference is confirmed, amended reports must preserve the audit trail, identify the correction and trigger timely clinical communication. The rationale for no retesting or no amendment should also be documented. Nonconforming-event records should include containment, scope, cause evidence, actions, authorisation and an effectiveness-review date [1,44,45].

A tiered patient-impact review improves consistency. First, use control and comparison data to estimate the plausible error across concentration. Second, identify results for which that error could cross a reference, diagnostic, critical or treatment threshold. Third, consider specimen stability and the feasibility of repeat measurement. Finally, classify communication urgency from no material effect, through report amendment without immediate harm, to urgent clinician notification. The same numerical bias may require different actions for creatinine used in estimated glomerular filtration rate, bilirubin near a neonatal treatment threshold, or an analyte used only for long-term trend monitoring. Laboratory medical input is especially important when analytical and clinical significance diverge.

Effectiveness monitoring closes the loop. The laboratory should define a period and indicator capable of showing whether recurrence has been prevented: stable CV, absence of the same rejection pattern, reduced calibration failure, compliant storage records, satisfactory lot comparison, competency reassessment or improved EQA performance. Closing an event immediately after control recovery confuses correction with sustained improvement. Recurrent failures should be aggregated for management review because the pattern may reveal procurement, workload, maintenance, design or training weaknesses invisible in a single event [46,47].

Integration and Emerging Approaches: Control-material IQC should be interpreted alongside method verification, measurement uncertainty, EQA, equipment records and quality indicators. Stable IQC with unacceptable EQA suggests persistent bias, target assignment or commutability problems; poor IQC with acceptable EQA may reflect intermittent local instability not present on the survey day. Integration prevents either dataset from being treated as complete evidence [36,37,48]. Measurement-uncertainty estimation can use relevant long-term imprecision and bias information, but the data must represent the measurand and stable operating conditions [49].

Patient-based real-time quality control (PBRTQC) analyses patterns in routine patient results through moving averages, medians, percentiles, outlier counts or related algorithms. It monitors native specimens continuously and may detect failures between scheduled controls or effects not represented by commercial material. Its performance depends on patient distribution, exclusions, block size or weighting, control limits, analytical volume and the magnitude and direction of simulated error [50]. Informatics must preserve chronology, exclude nonpatient records, separate instruments or populations where necessary, provide an audit trail and support prospective alert handling [51].

Algorithm choice should reflect analyte behaviour. A moving mean may work for tightly distributed electrolytes but can be distorted by extreme values unless truncation or winsorisation is validated. A moving median is more robust but may respond more slowly. EWMA assigns greater weight to recent results, while CUSUM accumulates small deviations from a target. Highly skewed analytes, seasonal population changes and mixtures of inpatient and outpatient results may need

stratification or transformation. Performance should be reported as the number of patient results required to detect a simulated bias and the false-alarm rate, not simply whether a plotted curve looks smooth [50-53].

PBRTQC requires local simulation and verification before use. Laboratories should test plausible positive and negative biases, quantify detection delay and false alarms, and assess robustness to changing inpatient, outpatient or emergency case mix [52,53]. Recent work explores next-generation and multivariate models [54], including machine learning for analytical and preanalytical error detection [55]. These methods can improve sensitivity but introduce model drift, explainability, version control and oversight requirements. Slow adoption reflects not only software limitations but also the need for standardised validation, workflow ownership and credible responses to alerts [56]. The most useful future model is hybrid rather than substitutive: conventional controls for known concentrations and rapid large-error detection; patient-based monitoring for continuous surveillance of native specimens; EQA for external comparability; lot verification for transitions; and equipment, middleware and quality indicators for contextual evidence. Integrated concepts such as a QC 'constellation' explicitly combine risk- and patient-based signals [57]. Table 5 provides a tiered implementation framework so laboratories can strengthen foundational practice before adopting advanced algorithms.

Practical Recommendations: A focused IQC programme can be summarised in ten decisions: define an analyte-specific APS; estimate stable local bias and imprecision; select commutable or well-characterised controls at clinically relevant concentrations; establish and periodically verify the mean and SD; choose the minimum rules that deliver adequate error detection; set frequency from throughput, stability and patient risk; verify new reagent and calibrator lots with patient specimens; investigate the original signal rather than repeating to acceptance; assess results generated since the last acceptable event; and document corrective action with effectiveness monitoring. Where resources are limited, these foundations provide greater safety than complex rule panels applied to poorly characterised data. Implementation should be audited at least annually and after significant change. The review should examine target and SD validity, long-term CV and bias, warning and rejection rates, repeat-control frequency, recurrent causes, patient results held or amended, EQA performance, lot comparisons, instrument downtime and staff competency. An unusually low rejection rate may indicate excellent performance, but it may also indicate limits that are too wide or disabled rules; a high rate may indicate poor performance, underestimated SD or an overcomplex panel. Audit findings should result in specific changes to frequency, rules, materials, maintenance or method capability and should be tracked to completion [1,26,32,47].

No universal statement such as 'CV below 5% is acceptable' is scientifically defensible across clinical biochemistry. No fixed Westgard panel is optimal for every method, and acceptable commercial controls do not exclude patient-specific or lot-related failure. Laboratories should make the connection from statistical signal to clinical consequence explicit in their procedure, staff training and authorisation rules. Advanced patient-based monitoring should be added only after the conventional programme is stable, documented and locally understood.

Conclusion

Effective IQC is a clinical risk-control process supported by statistics. Mean, SD and CV describe stable performance; Levey-Jennings charts and selected Westgard rules identify change; APS and patient risk determine whether that change matters. A rejection must lead to containment, investigation, verification of restored performance and assessment of potentially affected results. Conventional IQC remains indispensable, but its blind spots are best addressed through integration with lot verification, EQA, quality-system evidence and validated patient-based monitoring. The strongest programme is not the one with the most rules, but the one that detects medically important failure early and converts every signal into a reproducible, documented decision.

References

1. International Organization for Standardization. ISO 15189:2022: Medical laboratories—Requirements for quality and competence. 4th ed. Geneva: ISO; 2022.
2. Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem. 2002;48(5):691–698. doi:10.1093/clinchem/48.5.691.
3. Plebani M. Errors in clinical laboratories or errors in laboratory medicine? Clin Chem Lab Med. 2006;44(6):750–759. doi:10.1515/CCLM.2006.123.
4. Plebani M. Errors in laboratory medicine and patient safety: the road ahead. Clin Chem Lab Med. 2007;45(6):700–707. doi:10.1515/CCLM.2007.170.
5. Clinical and Laboratory Standards Institute. Statistical quality control for quantitative measurement procedures: principles and definitions. 4th ed. CLSI guideline C24. Wayne (PA): CLSI; 2016.
6. Miller WG, Jones GRD, Horowitz GL, Weykamp C. Proficiency testing/external

quality assessment: current challenges and future directions. Clin Chem. 2011;57(12):1670–1680. doi:10.1373/clinchem.2011.168641.
7. Zemlin AE. Errors in the extra-analytical phases of clinical chemistry laboratory testing. Indian J Clin Biochem. 2018;33(2):154–162. doi:10.1007/s12291-017-0657-2.
8. Sciacovelli L, Lippi G, Sumarac Z, West J, Garcia del Pino Castro I, Furtado Vieira K, et al. Quality indicators in laboratory medicine: the status of the progress of the IFCC Working Group "Laboratory Errors and Patient Safety" project. Clin Chem Lab Med. 2017;55(3):348–357. doi:10.1515/cclm-2016-0929.
9. Loh TP, Lim CY, Sethi SK, Tan RZ, Markus C. Advances in internal quality control. Crit Rev Clin Lab Sci. 2023;60(7):502–517. doi:10.1080/10408363.2023.2209174.
10. Baethge C, Goldbeck-Wood S, Mertens S. SANRA—a scale for the quality assessment of narrative review articles. Res Integr Peer Rev. 2019;4:5. doi:10.1186/s41073-019-0064-8.
11. Green BN, Johnson CD, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. J Chiropr Med. 2006;5(3):101–117. doi:10.1016/S0899-3467(07)60142-6.
12. Joint Committee for Guides in Metrology. International vocabulary of metrology—Basic and general concepts and associated terms (VIM). 3rd ed. JCGM 200:2012. Sèvres: Bureau International des Poids et Mesures; 2012. doi:10.59161/JCGM200-2012.
13. Joint Committee for Guides in Metrology. Evaluation of measurement data—Guide to the expression of uncertainty in measurement. JCGM 100:2008. Sèvres: Bureau International des Poids et Mesures; 2008.
14. Clinical and Laboratory Standards Institute. Expression of measurement uncertainty in laboratory medicine. CLSI guideline EP29-A. Wayne (PA): CLSI; 2012.
15. Westgard JO, Westgard SA. Measuring analytical quality: total analytical error versus measurement uncertainty. Clin Lab Med. 2017;37(1):1–13. doi:10.1016/j.cll.2016.09.001.
16. Sandberg S, Fraser CG, Horvath AR, Jansen R, Jones G, Oosterhuis W, et al. Defining analytical performance specifications: consensus statement from the 1st Strategic Conference of the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med. 2015;53(6):833–835. doi:10.1515/cclm-2015-0067.
17. Clinical and Laboratory Standards Institute. Evaluation of precision of quantitative measurement procedures. 4th ed. CLSI standard EP05. Wayne (PA): CLSI; 2025.
18. Clinical and Laboratory Standards Institute. User verification of precision and estimation of bias. 3rd ed. CLSI guideline EP15-A3. Wayne (PA): CLSI; 2014.
19. Aarsand AK, Røraas T, Fernandez-Calle P, Ricós C, Díaz-Garzón J, Jonker N, et al. The Biological Variation Data Critical Appraisal Checklist: a standard for evaluating studies on biological variation. Clin Chem. 2018;64(3):501–514. doi:10.1373/clinchem.2017.281808.
20. Sandberg S, Coskun A, Carobene A, Fernandez-Calle P, Diaz-Garzon J, Bartlett WA, et al. Analytical performance specifications based on biological variation data—considerations, strengths and limitations. Clin Chem Lab Med. 2024;62(8):1483–1489. doi:10.1515/cclm-2024-0108.
21. Westgard JO, Westgard SA. Six Sigma quality management system and design of risk-based statistical quality control. Clin Lab Med. 2017;37(1):85–96. doi:10.1016/j.cll.2016.09.008.
22. International Organization for Standardization. ISO 13528:2022: Statistical methods for use in proficiency testing by interlaboratory comparison. 3rd ed. Geneva: ISO; 2022. Amended by ISO 13528:2022/Amd 1:2026.
23. Wang W, Zhang Z, Zhang C, Zhao H, Yuan S, Liu J, et al. Evaluation of coefficients of variation for clinical chemistry tests based on internal quality control data across 5,425 laboratories in China from 2013 to 2022. Ann Lab Med. 2024;44(3):245–252. doi:10.3343/alm.2023.0236.
24. Ellis AD, Gross AR, Budd JR, Miller WG. Influence of reagent lots and multiple measuring systems on estimating the coefficient of variation from quality control data: implications for uncertainty estimation and interpretation of QC results. Clin Chem Lab Med. 2020;58(11):1829–1835. doi:10.1515/cclm-2020-0320.
25. Badrick T, Loh TP. Developing an evidence-based approach to quality control. Clin Biochem. 2023;114:39–42. doi:10.1016/j.clinbiochem.2023.01.011.
26. Giannoli JM, Vassault A, Carobene A, et al. Ensuring internal quality control practices in medical laboratories: IFCC recommendations for practical applications based on ISO 15189:2022. Clin Chim Acta. 2025;571:120240. doi:10.1016/j.cca.2025.120240.
27. Giannoli JM, Albarede S, Avellan T, et al. Recommendations for the application and follow-up of quality controls in medical laboratories. Biochem Med (Zagreb).

2021;31(2):020501. doi:10.11613/BM.2021.020501.
28. Miller WG, Erek A, Cunningham TD, Oladipo O, Scott MG, Johnson RE. Commutability limitations influence quality control results with different reagent lots. Clin Chem. 2011;57(1):76–83. doi:10.1373/clinchem.2010.148106.
29. Parvin CA. Assessing the impact of the frequency of quality control testing on the quality of reported patient results. Clin Chem. 2008;54(12):2049–2054. doi:10.1373/clinchem.2008.113639.
30. Bayat H. Selecting multirule quality control procedures based on patient risk. Clin Chem Lab Med. 2017;55(11):1702–1708. doi:10.1515/cclm-2016-1077.
31. Thompson S, Chesher D. Lot-to-lot variation. Clin Biochem Rev. 2018;39(2):51–60.
32. Wheeler SE, Blasutig IM, Dabla PK, et al. Quality standards and internal quality control practices in medical laboratories: an IFCC global survey of member societies. Clin Chem Lab Med. 2023;61(12):2094–2101. doi:10.1515/cclm-2023-0492.
33. International Organization for Standardization. ISO 22367:2026: Medical laboratories—Application of risk management to medical laboratories. 2nd ed. Geneva: ISO; 2026.
34. Levey S, Jennings ER. The use of control charts in the clinical laboratory. Am J Clin Pathol. 1950;20(11):1059–1066. doi:10.1093/ajcp/20.11_ts.1059.
35. Kinns H, Pitkin S, Housley D, Freedman DB. Internal quality control: best practice. J Clin Pathol. 2013;66(12):1027–1032. doi:10.1136/jclinpath-2013-201661.
36. Braga F, Pasqualetti S, Aloisio E, Panteghini M. The internal quality control in the traceability era. Clin Chem Lab Med. 2021;59(2):291–300. doi:10.1515/cclm-2020-0371.
37. van Rossum HH. Technical quality assurance and quality control for medical laboratories: a review and proposal of a new concept to obtain integrated and validated QA/QC plans. Crit Rev Clin Lab Sci. 2022;59(8):586–600. doi:10.1080/10408363.2022.2088685.
38. Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. Clin Chem. 1981;27(3):493–501. doi:10.1093/clinchem/27.3.493.
39. Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH. Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem. 1977;23(10):1857–1867. doi:10.1093/clinchem/23.10.1857.
40. Westgard JO, Westgard SA. Establishing evidence-based statistical quality control practices. Am J Clin Pathol. 2019;151(4):364–370. doi:10.1093/ajcp/aqy158.
41. Bayat H, Westgard SA, Westgard JO. Planning risk-based statistical quality control strategies: graphical tools to support the new CLSI C24-Ed4 guidance. J Appl Lab Med. 2017;2(2):211–221. doi:10.1373/jalm.2017.023192.
42. Westgard JO, Groth T, Aronsson T, de Verdier CH. Combined Shewhart–CUSUM control chart for improved quality control in clinical chemistry. Clin Chem. 1977;23(10):1881–1887. doi:10.1093/clinchem/23.10.1881.
43. Vogeser M, Seger C. Irregular analytical errors in diagnostic testing—a novel concept. Clin Chem Lab Med. 2018;56(3):386–396. doi:10.1515/cclm-2017-0454.
44. Clinical and Laboratory Standards Institute. Nonconforming event management. 2nd ed. CLSI guideline QMS11. Wayne, PA: CLSI; 2015. Reaffirmed 2019.
45. World Health Organization. Laboratory quality management system: handbook. Geneva: WHO; 2011.
46. Motschman TL, Moore SB. Corrective and preventive action. Transfus Sci. 1999;21(2):163–178. doi:10.1016/S0955-3886(99)00088-0.
47. Clinical and Laboratory Standards Institute. A quality management system model for laboratory services. 5th ed. CLSI guideline QMS01. Wayne, PA: CLSI; 2019.
48. Badrick T. Integrating quality control and external quality assurance. Clin Biochem. 2021;95:15–27. doi:10.1016/j.clinbiochem.2021.05.003.
49. International Organization for Standardization. ISO/TS 20914:2019: Medical laboratories—Practical guidance for the estimation of measurement uncertainty. Geneva: ISO; 2019.
50. Badrick T, Bietenbeck A, Cervinski MA, Katayev A, van Rossum HH, Loh TP. Patient-based real-time quality control: review and recommendations. Clin Chem. 2019;65(8):962–971. doi:10.1373/clinchem.2019.305482.
51. Loh TP, Cervinski MA, Katayev A, Bietenbeck A, van Rossum HH, Badrick T. Recommendations for laboratory informatics specifications needed for the application of patient-based real-time quality control. Clin Chim Acta. 2019;495:625–629. doi:10.1016/j.cca.2019.06.009.
52. Loh TP, Bietenbeck A, Cervinski MA, van Rossum HH, Katayev A, Badrick T. Recommendation for performance verification of patient-based real-time quality control. Clin

Chem Lab Med. 2020;58(8):1205–1213. doi:10.1515/cclm-2019-1024.
53. Bietenbeck A, Cervinski MA, Katayev A, Loh TP, van Rossum HH, Badrick T. Understanding patient-based real-time quality control using simulation modeling. Clin Chem. 2020;66(8):1072–1083. doi:10.1093/clinchem/hvaa094.
54. Duan X, Zhang M, Liu Y, et al. Next-generation patient-based real-time quality control models. Ann Lab Med. 2024;44(5):385–391. doi:10.3343/alm.2024.0053.
55. Lorde N, Mahapatra S, Kalaria T. Machine learning for patient-based real-time quality control, analytical and preanalytical error detection in the clinical laboratory. Diagnostics (Basel). 2024;14(16):1808. doi:10.3390/diagnostics14161808.
56. Dean P, Smith J, Badrick T. Are we there yet?—Why is the adoption of patient-based real-time quality control taking so long?. J Lab Precis Med. 2025;10:10. doi:10.21037/jlpm-24-54.
57. Çubukçu HC. QC Constellation: a cutting-edge solution for risk- and patient-based quality control in clinical laboratories. Clin Chem Lab Med. 2024;62(11):2185–2197. doi:10.1515/cclm-2024-0156.