

Evaluating CardioWatch Performance

Accuracy requirements and validation method for continuous vital sign monitoring

Corsano Health • CardioWatch 287-2 • Version 1.0

1. Introduction

This document sets out the accuracy targets Corsano proposes for clinical accuracy studies of the CardioWatch continuous monitoring system, and the table in section 4 is the starting point for that conversation. It shows what the ISO standards require under laboratory conditions and what we believe is realistically achievable in an ICU and on a General Ward. The reference spot measurements used in such studies carry documented error margins of their own, and any study result reflects both instruments rather than ours alone. Setting targets that recognise this gives both parties a fair basis on which to evaluate the results.

Going forward, we expect accuracy to improve progressively as the system is used across larger patient populations in the EU, US, Africa and Asia. This will generate substantially more real-world data and enable further optimisation of areas such as LED power settings, PPG acquisition and algorithms.

Corsano will continue to improve performance while already delivering the fundamental clinical benefits of continuous trend monitoring and earlier recognition of patient deterioration.

We believe that deploying the technology, monitoring larger patient populations and continuously improving accuracy is the right path toward a new generation of wrist-based continuous patient monitoring. It is also the path to fulfilling our mission of helping save lives through continuous patient monitoring, anytime, anywhere.

2. Objective of the Clinical Validation

A clinical validation of the CardioWatch sets out to establish four things. Stating them explicitly at the outset keeps the analysis focused and prevents the study from being read as something it is not.

- The level of agreement between the CardioWatch and the reference method available in that clinical setting, for each vital sign, expressed as bias, scatter, ARMS and 95 % limits of agreement.
- Whether any disagreement is systematic or random. A systematic offset is generally correctable in software; random scatter is not, and the two call for entirely different responses.
- Where large disagreements concentrate: in motion, in particular patients, measurement range, or in the timing of the reference observation rather than in the measurement itself.
- Whether the system delivers its intended clinical function, which is to make a developing trend visible between observation rounds and to support earlier recognition of deterioration.

It is equally important to state what the study is not. It is a performance and clinical-utility study conducted under real-world conditions, not an ISO validation. It does not produce a regulatory accuracy claim, and its results should not be compared directly with figures from controlled laboratory validation. The reference is not a gold standard: nursing observations are charted rather than recorded simultaneously, and every reference instrument carries its own error, which in several cases is larger than the error of the device under test.

3. Preparation: Nursing Measurement Technique

The quality of the reference measurement sets a ceiling on what any validation can demonstrate. Where the reference is taken inconsistently, the resulting error is attributed to the device even though it did not originate there. Corsano therefore asks that ward staff receive a short briefing before the study begins, covering the points below, and that the agreed technique is recorded in the study protocol.

Timing and clock synchronisation

- Record the time the measurement was actually taken, not the time it was later charted. On a busy ward these can differ by many minutes, and the difference is counted as device error.
- Synchronise the clocks of the reference instruments, the CardioWatch gateway and the electronic record before the study starts, and verify the synchronisation again at the midpoint. A misalignment of sixty to ninety seconds is enough to pair values from different physiological moments.
- Confirm the bracelet is fitted and correctly worn at the moment of each reference observation, and note any observation taken while the device was off the patient.

Technique by vital sign

- Respiratory rate. Careful counting over 15 or 30 seconds is critical. Counting for 15 seconds and multiplying by four quantises the reference into steps of four breaths per minute, and a substantial part of the apparent disagreement is then quantisation of the reference rather than device error.
- Pulse rate. Record the source of the value, whether manual palpation or a bedside monitor, since the two carry different errors. Where the monitor is used, note that it applies its own averaging window.
- Oxygen saturation. Allow the finger probe to stabilise before reading. Note cold extremities, poor perfusion, nail varnish or patient movement, all of which affect the reference as much as the device.
- Blood pressure. Use the correct cuff size, support the arm at heart level, allow the patient to rest before inflation, and record which arm was used. Note the time of cuff inflation rather than the time the result was entered.
- Temperature. Use the same instrument and the same site throughout the study, and allow thermal equilibration before reading. Avoid taking the reference immediately after washing, showering or removal of blankets.

Participant characterisation

- Record Fitzpatrick skin type for every participant. The non-invasive blood pressure algorithm is primarily established on a narrow pigmentation range, Fitzpatrick I to IV, and results should be reported stratified accordingly.
- Record wrist circumference and bracelet fit, arrhythmia where known, and any condition affecting peripheral perfusion. Record also conditions that affect movement, wear compliance or cooperation with the reference measurement, in particular Parkinson's disease, Alzheimer's disease and other dementias, and delirium. Large errors frequently concentrate in a small number of patients, and this is only visible if the relevant characteristics were captured.

4. Accuracy Requirements

ARMS (accuracy root-mean-square) is the combined systematic and random error against a reference: $ARMS = \sqrt{\text{mean of (device - reference)}^2} = \sqrt{\text{bias}^2 + SD^2}$. Because it is measured against a reference, the ARMS that can be achieved in a clinical study is bounded by the accuracy of the reference method available in that setting, not only by device performance.

4.1 ARMS Expectations by Setting

The table below sets out, for each vital sign, (1) the accuracy required under the applicable ISO standard in a controlled laboratory validation, and (2) the ARMS expected for ICU and General Wards clinical studies, with the reasoning for why each is less strict than the laboratory requirement.

Vital sign	Lab / ISO Norm Accuracy requirement*	Lab / ISO Norm Applicable standard	ICU Study ARMS expectation	ICU Study Why a less strict target is expected / justified	General Wards Study ARMS expectation	General Wards Study Why a less strict target is expected / justified
Pulse Rate	ARMS ≤ 5 bpm (manufacturer-declared)	ISO 80601-2-61:2017, cl. 201.12.1.104 (PULSE RATE ACCURACY) The ± 5 bpm or ± 10 % figure commonly quoted is the ANSI/AAMI EC13 criterion.	ARMS ≤ 7 bpm	Reference is the bedside ECG / arterial monitor, which applies its own 4 to 16 beat averaging window and has its own detection error. ICU population includes arrhythmia, ectopy, vasopressors, low peripheral perfusion and delirium, all excluded from the ISO laboratory protocol. Residual time-alignment error between the wearable and the monitor export adds variance that does not exist in a lab study.	ARMS ≤ 9 bpm	Patients move freely; PPG motion artefact is unavoidable. Arrhythmia and low peripheral perfusion are present in the ward population, and tremor in Parkinson's disease, agitation in delirium and reduced wear compliance in Alzheimer's disease and other dementias add artefact and gaps that the ISO protocol excludes by design. Reference is an intermittent spot-check monitor or manual pulse palpation, not synchronised beat-to-beat data. Clinical purpose is trend and deterioration detection.

Vital sign	Lab / ISO Norm Accuracy requirement*	Lab / ISO Norm Applicable standard	ICU Study ARMS expectation	ICU Study Why a less strict target is expected / justified	General Wards Study ARMS expectation	General Wards Study Why a less strict target is expected / justified
SpO ₂	ARMS ≤ 4.0 % (SaO ₂ 70 to 100 %)	ISO 80601-2-61:2017, cl. 201.12.1.101 (SpO ₂ ACCURACY): controlled desaturation study, arterial sampling plus CO-oximetry, full 70 to 100 % range.	ARMS ≤ 4.5 %	Reference is normally the bedside pulse oximeter rather than arterial CO-oximetry; the two error sources combine in quadrature. Against a bedside oximeter of ARMS ≈ 2 % this gives $\sqrt{4^2+2^2} \approx 4.5$ %. Confounders absent from the lab: oedema, hypothermia, vasopressor-induced vasoconstriction, dysaemoglobins, skin pigmentation range. Observed saturations cluster in 88 to 100 %; the low-saturation plateaus that dominate the ISO ARMS calculation cannot be induced in patients.	ARMS ≤ 5.0 %	Reference is a nurse spot-check finger oximeter with no time synchronisation to the continuous wearable trace. Motion, cold extremities and loose wear position are normal ward conditions and are excluded by the ISO protocol. Quadrature of the 4 % laboratory requirement against a ward spot-check reference of ARMS ≈ 3 % gives $\sqrt{4^2+3^2} = 5.0$ %, which is the proposed target. Clinical use is threshold-breach and trend detection (e.g. sustained < 92 %), which tolerates a wider absolute error than a diagnostic spot reading.
Respiratory Rate	No ISO/IEC accuracy criterion exists Our declared accuracy: ± 3 brpm	No dedicated ISO/IEC accuracy criterion for wearable RR. Closest applicable document is IEC 80601-2-49:2018 + AMD 1:2024 (multifunction patient monitoring equipment), under which accuracy is manufacturer-declared and verified against capnography or spirometry.	ARMS ≤ 4 brpm	Reference (capnography, ventilator or thoracic-impedance RR) itself miscounts during ventilator asynchrony, auto-triggering, cardiogenic oscillation and apnoea. Supported and irregular breathing violates the steady-state, paced-breathing assumption of a controlled verification test. No accredited reference method for RR exists that is as tightly specified as Pulse CO-oximetry is for SpO ₂ .	ARMS ≤ 5 brpm vs manual count	The reference is less accurate than the device: only 78 % of manual counts by healthcare professionals fall within 4 brpm of the true rate, inter-observer ICC ≈ 0.64, with systematic over-reporting of 1 to 3 brpm (PLOS ONE 2019). Counts are typically 15 s × 4 or 30 s × 2, quantising the reference to 2 to 4 brpm steps, so part of the measured difference is quantisation of the reference rather than device error.
Systolic BP (SBP)	Mean error ≤ ± 5 mmHg and SD ≤ 8 mmHg (equivalent to ARMS ≈ 9.4 mmHg, our conversion, not an ISO figure)	ISO 81060-2:2018 Continuous and cuffless devices: ISO 81060-3:2022 (intra-arterial reference; bias ≤ 6.0 mmHg, SD ≤ 10.0 mmHg) and IEEE 1708-2025.	ARMS ≤ 15 mmHg	The arterial line reference carries its own error: under- or over-damping, resonance, transducer levelling drift, and site-dependent gradients (radial versus femoral differences of 10 to 20 mmHg in shock or on high-dose vasopressors). ISO 81060-2 explicitly excludes arrhythmia and rapidly changing blood pressure, which is precisely the ICU population. ISO 81060-3 permits bias ≤ 6.0 mmHg with SD ≤ 10.0 mmHg against a fully time-aligned intra-arterial trace, equivalent to ARMS ≈ 11.7 mmHg. The proposed 15 mmHg (clinical and statistical judgement) deliberately carries margin above that figure, because it is applied to a setting where time alignment is imperfect and the confounders above are present. The margin is the price of accepting a real-world reference rather than a controlled one.	ARMS ≤ 20 mmHg	Reference is a single intermittent oscillometric cuff whose own SD against auscultation is permitted to be up to 8 mmHg, so the reference alone consumes most of the ISO error budget. A 30 s cuff inflation cannot be time-aligned with an estimate produced every 30 minutes; genuine physiological BP variability in the interval is counted as device error, and on a general ward this interval term is the largest single contributor. Combining the device, the cuff reference and the interval term in quadrature gives ≈ 14 mmHg, and the proposed 20 mmHg sits deliberately above that budget. The quadrature figure assumes typical conditions, whereas the target has to hold across the whole ward population, including the patients and moments in which the reference itself performs worst. Wrist-versus-upper-arm hydrostatic offset, arm position and white-coat response are uncontrolled on the ward. The non-invasive blood pressure algorithm is primarily established on a narrow pigmentation range, Fitzpatrick I to IV. Intended use is trend and threshold alerting, not diagnosis or titration of antihypertensive therapy.

Vital sign	Lab / ISO Norm Accuracy requirement*	Lab / ISO Norm Applicable standard	ICU Study ARMS expectation	ICU Study Why a less strict target is expected / justified	General Wards Study ARMS expectation	General Wards Study Why a less strict target is expected / justified
Diastolic BP (DBP)	Mean error $\leq \pm 5$ mmHg and SD ≤ 8 mmHg (equivalent to ARMS ≈ 9.4 mmHg, our conversion, not an ISO figure)	ISO 81060-2:2018. Continuous and cuffless: ISO 81060-3:2022, IEEE 1708-2025.	ARMS ≤ 12 mmHg	Invasive DBP is the arterial-line parameter most distorted by catheter damping and tubing compliance, so the reference noise floor is the dominant error term. Same population exclusions as SBP (arrhythmia, unstable haemodynamics) apply. Absolute DBP range in ICU is narrower than SBP, so a smaller absolute ARMS is still a proportionally relaxed requirement.	ARMS ≤ 15 mmHg	Oscillometric cuff DBP is algorithm-derived rather than directly measured and is the least reproducible of all cuff parameters, so the study is reference-limited rather than device-limited. The same time-alignment argument as for SBP applies: a 30 s cuff inflation against an estimate produced every 30 minutes. Ward escalation criteria are typically driven by SBP; DBP is used as a supporting trend.
Temperature	Laboratory error $\leq \pm 0.3$ °C	ISO 80601-2-56:2017 / Amd 1:2018	ARMS ≤ 0.5 °C vs core	A wrist-worn sensor measures peripheral skin temperature, which is physiologically offset from core (bladder, oesophageal or blood) and modulated by vasoconstriction and vasopressors. Active cooling and warming blankets, and a tightly controlled ambient ICU temperature, drive the skin to core gradient independently of device performance. ISO 80601-2-56 assumes measurement at a defined body site in thermal steady state, a condition a continuously worn wrist sensor never satisfies.	ARMS ≤ 0.8 °C	The reference is a tympanic, axillary or temporal spot check. In our measurements with the Genius 3, we observed substantial variability in tympanic temperature values. Ambient exposure, blankets, showering and absence of thermal equilibration before the spot reading add uncontrolled variance.

Notes to the table

- The ISO column states the requirement for a formal laboratory or controlled validation of the device. The ICU and General Wards studies are performance and clinical-utility studies conducted under real-world conditions. They are not ISO validation studies and should not be assessed against ISO acceptance criteria.
- Where the reference method itself has a known error (nurse-counted respiratory rate, spot-check pulse oximetry, intermittent cuff blood pressure, tympanic temperature), the errors add in quadrature: $ARMS_{study} \approx \sqrt{ARMS_{device}^2 + ARMS_{reference}^2}$. For the ward blood pressure and respiratory rate targets a third term is present, the physiological change occurring between an intermittent reference reading and a continuous estimate, and this is stated explicitly in the relevant cells rather than folded silently into the number.
- For blood pressure, respiratory rate and temperature, agreement should additionally be reported as Bland-Altman bias with 95 % limits of agreement, and as trending performance, since these reflect the intended clinical use more directly than ARMS alone. Section 4.2 sets out how each ARMS expectation converts into limits of agreement.
- Values in the ICU and General Wards columns are proposed acceptance targets for discussion and should be confirmed against the final study protocols and statistical analysis plans before being fixed contractually.
- The proposed values are intended for performance and clinical-utility studies of the kind described here. We expect accuracy to improve progressively as the system is used across larger patient populations. This will generate substantially more real-world data and enable further optimisation of the Corsano CardioWatch System.

* ARMS is an ISO term only in pulse oximetry. ISO 80601-2-61 defines and requires it for SpO₂ and pulse rate. ISO 81060-2, ISO 81060-3 and ISO 80601-2-56 do not use ARMS; they specify mean error and standard deviation, or clinical bias and limits of agreement. Where an ARMS-equivalent figure is shown for blood pressure in that column it is our own conversion, $\sqrt{bias^2 + SD^2}$, and is marked as such.

4.2 Bland-Altman Bias and 95 % Limits of Agreement

The expectations above are stated as ARMS. This section expresses the same expectations as Bland-Altman bias and 95 % limits of agreement, which is the form in which agreement is normally reported in clinical publications and the form most statistical analysis plans will ask for. The two are related by a single identity.

$$\text{ARMS}^2 = \text{bias}^2 + \text{SD}^2 \qquad 95\% \text{ LoA} = \text{bias} \pm 1.96 \times \text{SD}$$

An ARMS value therefore does not map to a unique pair of limits: it constrains bias and SD jointly. This matters practically, because the two components are different problems. Bias is a systematic offset and is usually correctable in software, by a constant offset or a calibration curve. SD is the scatter around that offset and no offset improves it; reducing it requires signal quality suppression, motion handling, better sensor contact or algorithm work. A parameter that misses its target on bias and a parameter that misses it on scatter call for entirely different responses, and reporting ARMS alone conceals which case applies.

One caution applies to bias correction. An offset is legitimate where the device and the reference measure genuinely different quantities, for example wrist temperature against core body temperature, because the offset is then a conversion.

Limits of agreement at zero bias

If bias is zero the whole of ARMS is random scatter, so SD equals ARMS and the limits are symmetric at plus and minus $1.96 \times \text{ARMS}$. This is the widest interval consistent with each expectation, and therefore the appropriate figure to quote as an upper bound in a statistical analysis plan.

Vital sign	ICU ARMS expectation	ICU 95 % limits of agreement	General Wards ARMS expectation	General Wards 95 % limits of agreement
Pulse Rate	≤ 7 bpm	−13.7 to +13.7 bpm	≤ 9.0 bpm	−17.6 to +17.6 bpm
SpO ₂	$\leq 4.5\%$	−8.8 to +8.8 %	$\leq 5.0\%$	−9.8 to +9.8 %
Respiratory Rate	≤ 4 brpm	−7.8 to +7.8 brpm	≤ 5.0 brpm	−9.8 to +9.8 brpm
Systolic BP	≤ 15 mmHg	−29.4 to +29.4 mmHg	≤ 20.0 mmHg	−39.2 to +39.2 mmHg
Diastolic BP	≤ 12 mmHg	−23.5 to +23.5 mmHg	≤ 15.0 mmHg	−29.4 to +29.4 mmHg
Temperature	≤ 0.5 °C	−0.98 to +0.98 °C	≤ 0.80 °C	−1.57 to +1.57 °C

Limits of agreement implied by each ARMS expectation in the zero bias case. General wards respiratory rate limits are stated against a manual count reference.

4.3 Clinical Validation Method

The design decisions below change the result materially. Corsano asks that they are agreed in writing, and before the analysis is run, so that the outcome is credible to both parties whatever it shows.

Time matching

Each reference observation is matched to the nearest CardioWatch reading of the same parameter for the same patient. The matching window should reflect how often each parameter is sampled: a few minutes for temperature, pulse rate, respiratory rate and oxygen saturation, and of the order of fifteen minutes for cuffless blood pressure, which is sampled less frequently. The wider window should be declared explicitly, because real physiological change occurs inside it and is counted as device error.

Exclusions and match rate

Two categories of reference observation should be set aside, and both should be reported rather than silently dropped. First, observations charted outside the monitoring period, when the bracelet had not yet been fitted or had already been removed; a device cannot be expected to report when it is not on the patient, and including these measures ward workflow rather than device performance. Second, physiologically impossible values, for example an oxygen saturation of zero.

Report the match rate both ways: against observations taken during monitoring, and against all charted observations without exclusion. The gap between the two figures is itself a finding, because it quantifies how much of the admission was not covered by any wearable at all.

Adjudication of outliers

A comparison that includes every paired point, including those taken during motion, poor peripheral perfusion or device repositioning, measures something different from clinical performance. Where the study is intended to characterise clinical performance, outliers should be adjudicated as physiological or artefactual by a clinician blinded to the device, before the error statistics are computed. Whether adjudication is applied is a legitimate choice, but it must be made in advance and stated.

The structural ceiling of a spot check comparison

A reference spot measurement is a snapshot of a signal that is genuinely moving, so a snapshot compared against a continuously sampled trace will disagree even if both instruments are perfect. This can be quantified directly by comparing the CardioWatch against itself at the same time separation used for the reference matching, across the whole dataset. The resulting agreement is the structural ceiling for that parameter and that band. Observed agreement should be read against that ceiling rather than against perfection, since the ceiling, not one hundred per cent, is what the device is actually competing against.

A related test is worth running. If the charted time of the reference observation is allowed to be approximate, and the best CardioWatch sample within a few minutes is used instead of the nearest one, agreement typically rises appreciably. Charting time is not measurement time on a busy ward, and that difference alone accounts for a material share of what appears at first to be disagreement.

Statistics to report

- Number of patients, number of admissions, monitoring duration and study period.
- Total device records and total reference observations, per parameter.
- Matching window per parameter, and the clock synchronisation method used.
- Exclusions applied, with counts and reasons, and match rate reported both with and without exclusions.
- Per parameter: n, bias, SD, ARMS, 95 % limits of agreement, median absolute difference, and the proportion of readings inside a stated clinical band.
- A Bland-Altman plot per parameter, with bias and both limits marked.
- Agreement split by rest and activity, at minimum for pulse rate, using the on-board accelerometer.
- A per-patient breakdown sufficient to show whether large errors are concentrated in a few individuals, and results stratified by Fitzpatrick skin type.
- The device-against-itself ceiling for each parameter at the matching window used.
- A statement of what the study does not establish.

Two points are worth stating in any statistical analysis plan. ARMS squares the differences and is therefore driven by the tail of the distribution rather than by typical performance, so a median absolute difference several times smaller than ARMS on the same dataset is normal. Both numbers belong in the record because they answer different questions: ARMS describes the worst of the distribution and is the right basis for an acceptance criterion, while the median describes the reading a clinician will actually see most of the time.

5. Working with Corsano on Your Analysis

1. Python scripts are available. Corsano maintains the analysis code used to produce every statistic and plot described in this document: time matching against a reference, exclusion handling, bias, SD and ARMS computation, Bland-Altman plots with 95 % limits of agreement, rest and activity splitting, per-patient breakdown, and the device-against-itself ceiling calculation. The scripts are provided on request so that partners can reproduce our figures on their own data and satisfy themselves that the method is sound.

2. Please contact Corsano Data Science for a discussion. We would rather agree the analysis plan with you at the start than debate the interpretation at the end. Our data science team is available to review a draft statistical analysis plan, to advise on matching windows and exclusion rules for your setting, to run a parallel analysis alongside yours, and to work through any result that is unexpected. There is no result we would prefer not to see. The evaluations that have been most useful to us are the ones that were designed rigorously and reported honestly.

6. References

- ISO 80601-2-61:2017. Pulse oximeter equipment. SpO₂ accuracy cl. 201.12.1.101 (ARMS $\leq 4.0\%$); pulse rate accuracy cl. 201.12.1.104.
- ISO 80601-2-61:2026. Third edition, supersedes the 2017 edition; adds skin-pigmentation-stratified ARMS and a “disparate bias” criterion.
- IEC 60601-2-27:2011. ECG monitoring equipment; requires heart rate meter accuracy to be determined and disclosed. The ± 5 bpm or $\pm 10\%$ criterion commonly quoted for heart rate meters originates in ANSI/AAMI EC13, not in IEC 60601-2-27.
- ISO 81060-2:2018 / Amd 1:2020 / Amd 2:2024. Non-invasive sphygmomanometers, clinical investigation of intermittent automated measurement type (mean error $\leq \pm 5$ mmHg, SD ≤ 8 mmHg).
- ISO 81060-3:2022. Non-invasive sphygmomanometers, clinical investigation of continuous automated measurement type, against an intra-arterial reference (bias ≤ 6.0 mmHg, SD ≤ 10.0 mmHg).
- IEEE 1708-2025. Wearable, cuffless blood pressure measuring devices.
- ISO 80601-2-56:2017 / Amd 1:2018. Clinical thermometers for body temperature measurement (laboratory accuracy cl. 201.101.2, $\leq \pm 0.3$ °C; clinical accuracy validation cl. 201.102).
- IEC 80601-2-49:2018 + AMD 1:2024. Multifunction patient monitoring equipment. Respiratory rate accuracy is manufacturer-declared.
- Latten GHP, Spek M, Muris JWM, Cals JWL, Stassen PM. PLOS ONE 2019;14(10):e0223155. Accuracy and inter-observer agreement of respiratory rate measurements by healthcare professionals.
- Bland JM, Altman DG. Lancet 1986;1(8476):307-310. Statistical methods for assessing agreement between two methods of clinical measurement.