element

WHITE PAPER | 2026

Navigating the Transition: ISO 80601-2-61:2026

What is changing in pulse oximeter clinical validation, why it matters, and how manufacturers should prepare



AUTHOR: **EVAN AMMIDOWN**
ENGINEERING MANAGER AT ELEMENT CONSUMER & INDUSTRIALS

element.com

CONTENTS

Executive summary 03
Introduction 04
- Why pulse oximetry validation is changing 04
- Understanding ISO 80601-2-61 04

The transition from the 2017 to the 2026 edition 05
- Study population requirements 06
- Study dataset requirements 07
- Accuracy and performance requirements 08
- Motion performance claims 09

What this means in practice 09
How manufacturers should prepare 10
Conclusion 11
Why work with Element 12
References 13

EDITION REFERENCED

ISO 80601-2-61:2026 (third edition, published 10 April 2026), which withdraws and replaces ISO 80601-2-61:2017 (corrected version 2018-02). European adoption: EN ISO 80601-2-61:2026

element

EXECUTIVE SUMMARY

ISO 80601-2-61:2026 represents a material increase in the clinical validation expectations applied to pulse oximeter equipment. This change is not limited to a larger study. Instead, it introduces a more structured approach to participant diversity, skin pigmentation assessment, dataset distribution, and statistical confidence.

For manufacturers, the most significant shift is the increase in necessary clinical data. A study designed around the minimum expectations of the 2017 edition is unlikely to be sufficient for the 2026 edition without substantial additional evidence. For example, the minimum study population rises from 10 participants to 24, with the potential need for additional participants to achieve the statistical significance requirement of the accuracy specification[6]. The new edition also introduces minimum representation by biological sex and detailed skin pigmentation categories supported by both subjective and objective assessment.

Additionally, the dataset must now not only be balanced across oxygen-saturation decades, but also across pigmentation categories. Participant-level coverage becomes more important, including requirements for usable low-saturation data. At the same time, the overall accuracy acceptance criterion tightens from an A_{RMS} of no more than 4.0% across 70% to 100% SaO_2 to no more than 3.0% with 95% confidence. A new pigmentation differential bias target is introduced for the 80% to 90% SaO_2 interval, centered at 85%[6].

Manufacturers should treat the transition as a clinical evidence, product development, and program planning issue rather than a documentation update. Early gap assessment, protocol design, recruitment feasibility, statistical planning, and pre-study performance work will be central to reducing the risk of an underpowered, unbalanced, or unsuccessful validation study.

THE PRACTICAL MESSAGE

Start with the evidence, not the test date. Confirm which products, probes, configurations, and claims need to be supported, then design the study population and dataset around the 2026 acceptance criteria. Element Boulder typically sees study success directly correlated to the amount of upfront planning, including leveraging existing data, review of product families, and timeline mapping, performed by manufacturers prior to beginning the clinical study.



INTRODUCTION: WHY PULSE OXIMETRY VALIDATION IS CHANGING

Pulse oximeters estimate arterial oxygen hemoglobin saturation non-invasively by analyzing how light at different wavelengths is absorbed by oxygenated versus deoxygenated blood. Their apparent simplicity can disguise the complexity of demonstrating clinically acceptable performance across people, saturation levels, probe sites, operating conditions, and product configurations.

The 2026 edition responds to a central challenge in pulse oximetry: an overall accuracy result can conceal differences between participant groups or weaknesses in particular parts of the measurement range. A pooled dataset may appear acceptable while representation within that dataset is uneven.

Melanin in the epidermis absorbs light across the visible and near-infrared range, resulting in less light passing through the tissue and reaching the photodetector. To the pulse oximeter, this change in signal appears the same as one caused by changes in saturation, potentially leading it to estimate more oxygen in the blood than there is. This inaccuracy can lead to the disparity reported in clinical practice at saturations guiding clinical decision making.

Pigmentation is not the only contributor. Perfusion, probe fit and placement pressure, ambient light, nail polish, and the measurement site all affect the signal. The revised requirements therefore place greater emphasis on who is included, how pigmentation is characterized, where data is collected across the saturation range, and how statistical significance is reflected in the accuracy requirement.

This matters to engineering and compliance teams because clinical validation is closely coupled to product architecture. Optical component selection, probe geometry, signal processing, calibration, averaging behavior, and artifact rejection can all affect the evidence generated in a controlled desaturation study. Failure to meet the statistical significance requirement is not always solved by recruiting more participants. It may expose a design, calibration, or algorithm issue that requires technical investigation.

UNDERSTANDING ISO 80601-2-61

ISO 80601-2-61 is the particular standard for the basic safety and essential performance of pulse oximeter equipment intended for use on humans. It applies to the equipment needed for normal use, including the pulse oximeter monitor, pulse oximeter probe, and probe cable extender. The scope includes equipment used in professional healthcare institutions, home healthcare, and emergency medical services environments, including remanufactured products.

As a particular standard within the IEC 60601-1 and ISO/IEC 80601 framework, it modifies and supplements the general medical electrical equipment requirements for the specific risks and performance characteristics of pulse oximetry. Its requirements should therefore be considered alongside the applicable general and collateral standards, the manufacturer's risk management process, usability engineering, software lifecycle evidence, and the regulatory expectations of each target market.

The clinical validation provisions are intended to demonstrate that pulse oximeter estimates

of SpO_2 are sufficiently close to reference arterial oxygen saturation values, SaO_2, across the claimed range. Controlled desaturation studies create a structured set of paired observations from the device under test and reference measurements. The design of that study, the selection of participants, the distribution of data, and the statistical analysis determine whether the claimed performance is adequately supported.

IMPORTANT TERMINOLOGY

SpO2 is the pulse oximeter estimate. SaO2 is the arterial oxygen saturation reference value used for comparison in the controlled desaturation study. ARMS is the root-mean-square difference used to express accuracy across the dataset.

THE TRANSITION FROM THE 2017 EDITION TO THE 2026 EDITION

This transition changes both the scale and the structure of clinical validation. The 2017 corrected 2018 edition established a minimum participant count, a minimum pooled dataset, and an overall A_{RMS} criterion. The 2026 edition adds controls intended to make the resulting evidence more representative, statistically robust, and capable of identifying performance differences associated with skin pigmentation.

Area	2017 corrected 2018	2026 edition	Manufacturer impact
Participants	Minimum 10	Minimum 24; statistical significance requirements may increase this	Longer recruitment, greater cost, and more complex scheduling
Biological sex	No stated distribution	At least 33% of each biological sex	Recruitment quotas must be monitored throughout the study
Skin pigmentation	No stated distribution	MST and spectrophotometer-based ITA categories with minimum representation	Objective measurement, trained observers, and controlled recruitment are needed
Data volume	Minimum 200 pairs	Average 20 ±4 pairs per participant; 16 to 30 per participant	Participant-level data yield becomes a protocol and operational control
Accuracy	A_{RMS} ≤4.0% over 70% to 100% SaO_2	A_{RMS} ≤3.0% with 95% confidence over 70% to 100% SaO_2	Less performance margin and greater statistical scrutiny

Differential bias	Not a calculated statistic	Target absolute pigmentation differential bias ≤2.5% for 80% to 90% SaO_2, centered at 85%	Subgroup performance must be quantified, not assumed
Motion	Reporting if claimed; no specified accuracy method	Expanded reporting if claimed; supported by a controlled desaturation data with motion typical of the probe site	Motion claims require a deliberately designed study condition

STUDY POPULATION REQUIREMENTS

A LARGER MINIMUM POPULATION

The minimum number of participants increases from 10 to 24[6]. Note that 24 is not an automatic design target; achieving the required statistical significance may require more participants. The study size should therefore be justified from the analysis plan, anticipated variability, confidence requirements, subgroup balance, and expected rate of usable data.

A study can recruit 24 people and still fail operationally if exclusions, unusable signal periods, or insufficient saturation coverage reduce the analyzable dataset. Recruitment planning should include contingency for replacement participants and for the possibility that pigmentation or biological-sex quotas are no longer met after exclusions.

REPRESENTATION BY BIOLOGICAL SEX

The 2026 edition requires at least 33% representation of each biological sex[6]. This prevents a study population from being heavily weighted towards one sex and makes recruitment balance a formal protocol control. Sites should monitor the analyzable population, not only enrollment, because withdrawals or data exclusions may change the final proportions.

SKIN PIGMENTATION ASSESSMENT

The new edition introduces a combined subjective and objective approach. Three observers assess Monk Skin Tone (MST) at the forehead using printed swatches. Sites should ensure their MST scale swatches are tested with a spectrophotometer against the tolerances listed in the 2026 standard. In parallel, a spectrophotometer is used to obtain L*a*b* values and calculate Individual Typology Angle, or ITA. Sites should ensure their equipment has a valid calibration prior to use in a study[6].

The study population is divided into three pigmentation categories:

- At least 25% in MST A to C with ITA greater than 30.
- At least 25% in MST D to G with ITA from 30 to -30, inclusive.
- At least 25% in MST H to J with ITA less than -30 and at least half of this group with an ITA less than -50.

*In the event the MST and ITA place a participant in different cohorts, the ITA value is used.

Using both methods is significant. The visual scale supports a clinically intuitive categorization, while spectrophotometry provides an objective measurement. Study sites will need procedures for instrument calibration, measurement technique, observer training, lighting, printed swatch control, and data reconciliation.

CALIBRATION-STUDY PARTICIPANTS

Participants who took part in a calibration study for the pulse oximeter under test should be excluded from the validation study. This strengthens independence between algorithm calibration and performance confirmation. Manufacturers should maintain records of datasets used in the calibration of probes and supply appropriate information to clinical sites during study planning. Clinical sites need to maintain records of participants comprising study datasets. Cross referencing these records ensures participants who took part in a calibration for the pulse oximeter under test are not recruited to take part in the validation study.

PROBE-USE STRATEGY

For single-use or single-patient probes, a different probe of the same design is used for each participant. For multiple-patient or multiple-use probes, multiple probes of the same design are used across participants. The intent is to avoid a clinical result being unduly dependent on one individual probe and to capture representative device-to-device variation.

PROTOCOL RISK

A population that satisfies the quotas at enrollment may not satisfy them after exclusions. Build recruitment and statistical contingency around the final analyzable population. In our experience at Element Boulder, tracking data and demographic distributions in real time is critical for study success and ensures the analyzable population reached at the end of enrollment meets all distribution requirements.

STUDY DATASET REQUIREMENTS

PARTICIPANT-LEVEL DATA EXPECTATIONS

The 2017 edition required at least 200 data points[5]. The 2026 edition moves to a participant-level expectation: an average of 20 ±4 data pairs per participant, with a permitted range of 16 to 30 per participant[6]. This changes the operational focus from accumulating a pooled total to ensuring that each participant contributes a suitable and reasonably balanced amount of information.

Over-collection from a small number of participants cannot simply compensate for weak data yield elsewhere. The sampling plan, desaturation plateaus, arterial draws, device stabilization, and data-quality rules should be designed to produce usable pairs for each participant without allowing individuals with unusually large datasets to dominate the analysis.

DISTRIBUTION ACROSS SATURATION DECADES

The 2017 edition placed roughly one-third of pooled data in 70% to <80%, 80% to <90% and 90% to 100%[5]. The 2026 edition adds a firm lower bound by requiring approximately

33%, but not less than 25%, of the total data pairs within each saturation decade. The 2026 edition also links this saturation coverage to pigmentation representation by requiring at least 25% of the data points in each saturation decade to be derived from each skin pigmentation category[6].

The dataset therefore needs to be balanced across two dimensions at once: saturation and pigmentation. A study may meet the overall participant quota yet fail because one pigmentation category has insufficient usable observations in a particular saturation decade.

Through our research at Element Boulder, we have found that the lowest saturation decade of 70% to <80% is most at risk for either the lightest or darkest pigmentation to have insufficient observations. Real time tracking of data distributions in both saturation and pigmentation is paramount for addressing any imbalances early, when they can be corrected with less data collection. Doing so later in the study is much more difficult, as the larger dataset at that time (relative to earlier in the study) means more data collection is required to rebalance proportions.

LOW-SATURATION PARTICIPATION

At least 90% of participants must provide at least one usable pair with an SaO_2 below 85% SaO_2 and at least 75% must also provide at least one usable pair between 70% and 80% SaO_2[6]. These requirements prevent the low-saturation region from being represented by a small subset of the study population and make controlled desaturation execution more demanding.

Sites should establish real-time or near-real-time tracking of usable pairs by participant, saturation decade, and pigmentation category. Waiting until database lock to identify a shortfall in the dataset creates a high risk of repeat recruitment or an additional study phase.

ACCURACY AND PERFORMANCE REQUIREMENTS

TIGHTER A_{RMS} ACCEPTANCE CRITERION

The overall A_{RMS} acceptance criterion tightens from no more than 4.0% across 70% to 100% SaO_2 to no more than 3.0% across the same range, with 95% confidence[6]. The addition of confidence is important. Compliance is not based only on the point estimate but also on the specified level of statistical confidence.

Products that historically reported an A_{RMS} result close to 4.0% should not be assumed to transition through incremental documentation. The tighter limit may require supplemental data, review and adjustment of calibration coefficients, optimization of signal processing, investigation of outliers, and assessment of probe-to-probe variability.

PIGMENTATION DIFFERENTIAL BIAS

The 2026 edition introduces a separate statistic for pigmentation differential bias. The absolute value should be no more than 2.5% for the interval between 80% and 90% SaO_2, centered at 85%[6]. This is designed to identify if performance differs materially between skin pigmentation groups even when the overall A_{RMS} result is acceptable.

While this statistic is a suggested target, not a requirement, it still has direct implications for analysis planning. The study must generate sufficient, well-distributed data in the relevant interval for each pigmentation category. A manufacturer should predefine the statistical model, treatment of repeated observations within participants, confidence interval approach, and handling of missing or invalid pairs before analyzing the final dataset.

ENGINEERING IMPLICATION

Passing overall A_{RMS} does not automatically demonstrate target performance across pigmentation categories. Manufacturers need sufficient margins for both overall accuracy and differential bias.

MOTION PERFORMANCE CLAIMS

The 2017 edition did not prescribe a method for establishing accuracy under motion, although reporting requirements applied when a manufacturer made a motion-performance claim. The 2026 edition expands these reporting expectations and requires data from a controlled desaturation study in which the pulse oximeter probe is exposed to motion typical of its monitoring site to support accuracy claims during motion.

Motion may be participant-driven (active movement) or imposed on the probe monitoring location (passive movement). Motion should begin before the data-sampling period and continue through it so that sampling occurs after the system's averaging response has settled[6]. This is intended to test the device under a sustained, representative condition rather than during a brief transition that the averaging algorithm may obscure.

Manufacturers should define the claimed use condition and translate it into a repeatable motion protocol. Key variables include the probe site, motion type, amplitude, frequency, duration, synchronization with arterial sampling, rejection criteria, and the device's averaging or response-time settings[9]. The protocol and report should make clear that the movement is representative of the intended claim rather than an arbitrary disturbance.

WHAT THIS MEANS IN PRACTICE

Individually, the changes are manageable. Their cumulative effect is a validation study several times the size and complexity of the one it replaces, with acceptance criteria that a proportion of currently marketed devices would not meet.

The independent evaluation of 34 pulse oximeters published in Anesthesia & Analgesia in 2026 is the clearest available indication of scale. Testing in cohorts of 24 participants using a controlled desaturation protocol, the authors found that 28 of the 34 devices met the 2017 A_{RMS} criterion of 4.0% and 21 met the updated 3.0% level[7]. The study concluded that 18 devices could pass or fail depending on which participants were selected for the analysis, demonstrating the influence of cohort selection on performance outcomes[7].

1. Cost and schedule

Controlled desaturation studies are conducted at a small number of specialist research facilities with specific equipment, CO-oximeters, arterial line capability, and medical oversight to run them safely. Capacity is limited, and the new recruitment requirements lengthen the screening phase at every site simultaneously. Manufacturers planning to move several products through the transition should assume that facility availability, rather than internal engineering work, sets the critical path.

2. Portfolio grouping

Not every product in a range needs its own study. Probe variants that share optics, calibration data, and signal processing may be grouped, and monitors that implement the same algorithm with the same probe family may share clinical evidence. What has changed is that the grouping argument now must be made explicitly and evidenced, because the new sampling and distribution requirements make the relationship between the tested configuration and the marketed configuration a matter the study design has to address.

3. Where the risk sits

The important shift is that a study can now pass based on a point estimate, but fail based on the observed confidence. A product with an A_{RMS} of 2.6% and an upper confidence bound of 3.1% fails, and there is no route to compliance through documentation. That failure is expensive if it is discovered after a full 24-participant desaturation study, and comparatively cheap if it is discovered by recomputing the statistics on data the manufacturer already holds.

HOW MANUFACTURERS SHOULD PREPARE

The preparation follows a familiar shape, but the order matters more than usual, because the most informative diagnostic is also the cheapest and can be run before any new data is generated.

RECOMPUTE THE NEW STATISTICS ON THE DATA YOU ALREADY HOLD

Before commissioning anything, take the existing clinical dataset and calculate what the 2026 criteria would produce. Compute A_{RMS} with a 95% confidence interval rather than as a point estimate alone. Where the historical study recorded any pigmentation information, compute the differential bias between groups, even if the categories do not map cleanly onto MST and ITA bands. Establish which participants appear in both the calibration and validation datasets.

This produces three outcomes: a device with comfortable margins on both statistics, a device with a marginal result that needs a well-powered study, or a device with a visible performance shortcoming that needs engineering work before any study is booked. The cost of finding out is only a few days of statistical analysis.

LOOK UPSTREAM AT THE CALIBRATION DATASET

Where differential bias appears, the cause usually lies in the calibration curve rather than in

the validation study, and the fix is upstream. If the calibration cohort was assembled without pigmentation characterization — which was common before this revision — the manufacturer may need additional calibration data and a refitted curve. Identifying this need early will prevent manufacturers from undertaking a futile validation study by addressing device specific changes beforehand, not after.

DESIGN AND COMPONENT CONSIDERATIONS

The requirement to use multiple probes across participants makes production spread part of the result. Before a study, review LED center wavelength distribution and binning, photodiode responsivity spread, optical window and adhesive consistency, and the mechanical design that sets probe pressure and prevents light piping around the tissue. Where the calibration is tuned to a nominal wavelength, quantify how performance varies across the bin limits.

On the signal processing side, the areas that most often affect low-saturation performance in darker pigmentation are front-end dynamic range and noise floor at low signal amplitude, the perfusion index threshold at which readings are suppressed, and the averaging window. Each of these is a design parameter that can be characterized on the bench before a clinical study is committed.

PLAN THE CLINICAL STUDY EARLY

Engage the research facility before the protocol is finalized rather than after. Site availability, ethics approval, screening capacity for the pigmentation distribution, and access to a calibrated spectrophotometer and compliant MST swatches all sit on the critical path. Agree prospectively on operational and statistical plans with the site.

UPDATE DOCUMENTATION, LABELING, AND CLAIMS

The accuracy statement in the instructions for use, the description of the validation population, the declared measurement range, and any claim relating to motion all need review against the evidence that will exist after the study. Where a claim cannot be substantiated, the choice is to generate the evidence or to withdraw the claim, and that decision is better taken during study design than during a Notified Body review.

CONCLUSION

ISO 80601-2-61:2026 changes what a pulse oximeter must demonstrate, and it adjusts the kind of evidence that counts. The updated approach strengthens validation through a larger, predefined study population, two independent methods of pigment characterization, cohort-level analysis, tighter accuracy requirements, and a new measure of differential bias, helping align performance evaluation more closely with real-world clinical findings.

The manufacturers that navigate this transition most efficiently will be those that leverage the data they already hold, evaluate calibration datasets alongside the validation study, design a single clinical study capable of satisfying their most demanding market, and book facility capacity before their competitors do.

WHY WORK WITH ELEMENT

Element supports medical device manufacturers across the product lifecycle, from early regulatory and test strategy through verification, validation, and market access. For the transition to ISO 80601-2-61:2026, the value lies in connecting the clinical evidence requirements to the product, the applicable standards framework, and the manufacturer's wider technical documentation.

1. Advisory and gap assessment

Element can help manufacturers interpret the revised requirements in the context of a specific pulse oximeter, probe portfolio, intended use, and target markets. A structured gap assessment can identify where existing evidence remains useful, where new analysis may be possible, and where additional clinical validation is needed.

2. Test program support

Element's network of medical device laboratories deliver the non-clinical testing that accompanies clinical validation, including IEC 60601-1 basic safety and essential performance, IEC 60601-1-2 electromagnetic compatibility, IEC 60601-1-11 for the home healthcare environment, IEC 60601-1-12 for emergency medical services, usability engineering to IEC 62366-1, and environmental and mechanical testing. For connected pulse oximeters, the same programs can be run alongside radio testing, protocol testing, and product cyber security assessment under the RED, the UK PSTI regime, and the EU Cyber Resilience Act.

3. Integrated compliance support

Element's global testing, inspection, and certification capabilities help manufacturers coordinate product safety, EMC, wireless, cybersecurity, and market-access activities where these apply to the complete device. This integrated approach is particularly valuable for connected pulse oximeters and multi-market programs, where clinical performance evidence is only one part of the overall compliance case.

RELATED SERVICES

- Advisory services for medical device compliance
- Pre-compliance and diagnostic testing
- Medical electrical equipment safety testing (IEC 60601-1 and particular standards)
- EMC and electrical safety testing
- Usability and human factors engineering
- Radio and wireless testing
- Cyber security (RED, PSTI, CRA)
- Global Market Access (GMA)

REFERENCES

[1] Commission Implementing Decision (EU) 2026/1231, amending Implementing Decision (EU) 2021/1182 as regards harmonized standards for medical devices.

[2] EN ISO 80601-2-61:2026, European adoption of ISO 80601-2-61:2026, CEN.

[3] FDA Executive Summary, Performance Evaluation of Pulse Oximeters, Anesthesiology and Respiratory Therapy Devices Panel, 2024.

[4] IEC 60601-1:2005+AMD1:2012+AMD2:2020, Medical electrical equipment — Part 1: General requirements for basic safety and essential performance.

[5] ISO 80601-2-61:2017 (corrected version 2018-02), second edition — withdrawn, retained for comparison.

[6] ISO 80601-2-61:2026, Medical electrical equipment — Part 2-61: Particular requirements for basic safety and essential performance of pulse oximeter equipment. Third edition, published 10 April 2026. ISO/TC 121/SC 3.

[7] Pulse oximeter performance and skin pigment: comparison of 34 oximeters using current and emerging regulatory frameworks. Anesthesia & Analgesia, 2026.

[8] Regulation (EU) 2017/745 on medical devices (MDR), Annex I.

[9] US Food and Drug Administration, Pulse Oximeters for Medical Purposes — Non-Clinical and Clinical Performance Testing, Labeling, and Premarket Submission Recommendations. Draft guidance, January 2025, docket FDA-2023-N-4976.

[10] US Food and Drug Administration, Pulse Oximeters — Premarket Notification Submissions [510(k)s], March 2013.

element

Scan to find out more.

WHY CHOOSE ELEMENT?

Element Connected Technologies & Mobility is a leading provider of testing, inspection, and certification services for connected devices and mobility solutions.

We provide comprehensive testing and certification services that ensure products meet international standards for connectivity, interoperability, and safety.

CONTACT

Our LinkedIn

Telephone: +44 (0) 808 234 1667 Web: element.com Email: contact.us@element.com

Making tomorrow safer than today.