Received: 22 Jul 2026 | Accepted: 20 Ago 2026 | Available online: 25 Ago 2026
DOI: http://dx.doi.org/10.18180/tecciencia.2025.38.1

TECCIENCIA

UNIVERSIDAD ECCI

RESEARCH ARTICLE

Engineering , Applied Sciences

Design and Validation of a Calibration Method for Patient Simulators

Miguel A. Castro-Leal [1*]

[1]Department of Biomedical Engineering, Universidad ECCI, Bogotá, Cra 19 no 49 -20, Bogotá - Colombia

Correspondence
Castro-Leal, M.A., Universidad ECCI, Colombia
Email: mcastrol@ecci.edu.co

The publication of this journal is funded by Universidad ECCI, Bogotá-Colombia.

Editor: Christian Alonso Becerra Rivas.

Editorial assistant : Juan Diego Alarcón Forero.

How to cite
Castro-Leal M.A., **Design and Validation of a Calibration Method for Patient Simulators**, TECCIENCIA, Vol. 20, No. 38, 1-11, 2025

ABSTRACT. This study designs and validates a calibration method for patient simulators measuring Heart Rate, Non-Invasive and Invasive Blood Pressure via electrical simulation, enhancing reliability in medical equipment adjustment and verification. Proper calibration prevents adverse events by ensuring device accuracy. The method selects clinically relevant values based on medical significance and international standards. Validation used three-year calibration data analyzed with ANOVA to confirm measurement significance, Tukey's method for confidence intervals, and the Brown-Forsythe test for variance equality. Results demonstrated strong repeatability and reproducibility, proving the method's effectiveness for patient simulator calibration. This approach reduces risks associated with uncalibrated medical devices, improving diagnostic reliability and patient safety.

keywords: Statistical analysis; Methodological design; Calibration; Healthcare equipment; Assessment.

RESUMEN. Este estudio diseña y valida un método de calibración para simuladores de pacientes que miden la frecuencia cardíaca, la presión arterial invasiva y no invasiva mediante simulación eléctrica, lo que mejora la confiabilidad en el ajuste y la verificación de los equipos médicos. Una calibración adecuada previene eventos adversos al garantizar la precisión de los dispositivos. El método selecciona valores clínicamente relevantes basándose en su importancia médica y en las normas internacionales. La validación se llevó a cabo utilizando datos de calibración de tres años, analizados mediante ANOVA para confirmar la significancia de las mediciones, el método de Tukey para los intervalos de confianza y la prueba de Brown-Forsythe para la igualdad de varianzas. Los resultados demostraron una alta repetibilidad y reproducibilidad, lo que comprobó la eficacia del método para la calibración de simuladores de pacientes. Este enfoque reduce los riesgos asociados con los dispositivos médicos no calibrados, lo que mejora la confiabilidad del diagnóstico y la seguridad del paciente.

Palabras clave: Análisis estadístico, Diseño Metodológico, Calibración, Equipo Sanitario, Valoración.

1 | INTRODUCTION

Clinical engineers play a fundamental role in ensuring the quality of medical devices used in hospitals. Patient simulators are vital tools for the design and evaluation of medical equipment in different applications, such as electrocardiographs [1] and invasive pressure monitors [2]. Quality control of biomedical equipment is one of the main functions within the context of the use and application of biomedical technology in hospitals. Metrology, expressed as the science of measurement [3], has been used for several decades in the

*Corresponding Author: mcastrol@ecci.edu.co

quality control of industrial processes and the assurance of commercial transactions worldwide. In the field of healthcare, the metrology of medical equipment, as an essential component and safeguard of quality control for medical products, is increasingly valued by medical institutions at all levels [4]. Metrology, applied to the biomedical field, know as biomedical metrology, provides a valuable tool to determine the real condition of biomedical diagnostic and treatment equipment, preventing erroneous readings of the physical magnitudes measured in patients for diagnosis, and ensuring the accuracy of the physiological quantities measured by biomedical equipment and used by healthcare professionals to support patient diagnosis and treatment. To control biomedical equipment, reference standards [3] with known values of magnitudes such as the heart rate, noninvasive blood pressure, and invasive blood pressure, must be available. These standards, which in the field of biomedical equipment are known as patient simulators, must comply with all the requirements indicated by the science of metrology, such as traceability [3], repeatability [3], and reproducibility [3] to guarantee the quality of the measurements performed on biomedical equipment. To comply with the necessary requirements for the proper use of biomedical simulators, they must be calibrated with methods designed and validated according to their use [5], including the periodic calibration of biomedical equipment, functional testing of new equipment and scientific research. It is for the aforementioned reasons that the design and validation of a calibration method for patient simulators becomes relevant in preventing potential adverse events [6], misdiagnoses, and erroneous treatments that may be caused by these kinds of medical devices. However, while methods for calibrating ECG simulators currently exist [7] [8], these methods are not based on measurements at clinically relevant points, such as when a physician changes a diagnosis from "normal" to "bradycardia" or "tachycardia" [9] [10]. This study aimed to close this technical gap by developing and validating a calibration method that combines clinically relevant calibration points with a metrologically sound validation procedure, thereby ensuring that the simulator performance is evaluated at thresholds that are meaningful for clinical decision making.

2 | DESIGN AND METHODOLOGY

For the design of the calibration method, three common magnitudes in critical patient monitoring were considered: heart rate, noninvasive blood pressure, and invasive blood pressure using electrical simulations [11]. The calibration points for the patient simulator were selected based on the clinical relevance of each measured quantity. The selection considered the physiological conditions and potential pathologies associated with specific values of each quantity. When applicable, the selected values were also based on recommendations issued by internationally recognized organizations.

2.1 | Heart rate

For Heart Rate, the following measurement points were chosen to cover all stages of life, from neonates to adults:
The first measurement point was 60 beats/min, which is the minimum normal heart rate for resting adults. Beyond this point, if the resting heart rate increases, the risk of total and cardiovascular mortality increases [9]. The second measurement point was 120 beats/min, which is the normal maximum heart rate for preschool children [10]. Third measurement point: 180 beats/min, which is the maximum heart rate for an infant and normal for neonates [10].

2.2 | Non-invasive blood pressure

For the non-invasive blood pressure magnitude (NIBP), the calibration points described in OIML Recommendation R149-3 Paragraph 8 [12].The calibration points were used as follows: the first measurement point was 50 mmHg. The second measurement point was set at 150 mmHg. The third measurement was performed at 250 mmHg.

2.3 | Invasive blood pressure

For the invasive blood pressure magnitude (IBP) through electrical simulation, the same points as for noninvasive blood pressure were used, having a gain of 5 μV/V/mmHg for the simulation values in mV and their expression in mmHg as follows: first measurement point: 50 mmHg. The second measurement point was set at 150 mmHg. The third measurement was performed at 250 mmHg.

2.4 | Validation

To validate the patient simulator calibration method in terms of repeatability and reproducibility, three patient simulators of the same reference were used with five measurements under the same conditions in a short period of time to check for repeatability, and the process was repeated over three years, once a year, to check for reproducibility. Heart rate, noninvasive blood pressure, and invasive blood pressure were measured using electrical simulations. The measurements were analyzed using the following statistical tools: standard deviation to determine the repeatability [13]; one-way analysis of variance (ANOVA) [14] to validate the significance of the results of the measurement means; Tukey method to evaluate the confidence intervals between the means of the factor levels in the measurements [15]; and Brown-Forsythe test for the equality of measurement means when there is inequality of variances (heteroscedasticity) [16], using Excel® Microsoft 365.

3 | RESULTS AND DISCUSSION

The hypothesis was that the mean values of the measurements were statistically equal, for which the data were analyzed using the following statistical techniques: ANOVA, Tukey's method, and the Brown-Forsythe test (Tables 1-3).

TABLE 1 Obtained data in the magnitude Heart - Rate

HR Measurement point 180 beats/min	Measurement 1	Measurement 2	Measurement 3
	179.64	179.64	179.64
	179.64	179.64	179.64
	179.64	179.64	179.81
	179.64	179.81	179.64
	179.64	179.64	179.81
Standard deviation	0	0.0077	0.095

TABLE 2 Obtained data in the magnitude NIBP

HR Measurement point 180 beats/min	Measurement 1	Measurement 2	Measurement 3
	50.1	50.0	50.0
	50.0	50.0	50.1
	50.0	50.1	50.1
	50.0	50.0	50.1
	50.0	50.1	50.0
Standard deviation	0.044	0.0548	0.054

TABLE 3 Obtained data in the magnitude IBP

HR Measurement point 180 beats/min	Measurement 1	Measurement 2	Measurement 3
	49.66	49.62	49.66
	49.66	49.62	49.66
	49.66	49.62	49.66
	49.66	49.66	49.66
	49.62	49.66	49.66
Standard deviation	0.017	0.021	0

3.1 | ANOVA analysis

In the ANOVA analysis, for all magnitudes and all measurement points of each magnitude, an $Fcalculated < Fcritical$ was found, confirming the corresponding hypotheses that the means of the measurement values were statistically equal (Tables 4-6).

TABLE 4 ANOVA analysis in the Heart Rate magnitude

Heart Rate Measurement point 180 beats/min

ANOVA Single Factor

Summary	
Average	**Variance**
179.6407186	0.000
179.6754149	0.006
179.7101111	0.009

ANOVA

F Calculated	P-value	F Critical
1.2	0.33	3.88

TABLE 5 ANOVA analysis in the NIBP magnitude

NIPB Measurement point $50mmHg$

ANOVA Single Factor

Summary	
Average	**Variance**
50.02	0.002
50.04	0.003
50.06	0.003

ANOVA

F Calculated	P-value	F Critical
0.75	0.49	3.88

TABLE 6 ANOVA analysis in the IBP magnitude

IPB Measurement point $50mmHg$

ANOVA Single Factor

Summary		
Average	**Variance**	
49.652	0.0003	
49.636	0.0004	
49.660	0.000	
ANOVA		
F Calculated	**P-value**	**F Critical**
2.8	0.1	3.88

3.2 | Tukey analysis.

Applying the Tukey method, the differences between the groups in pairs were lesser than the honestly significant difference (HSD), from which it was concluded that there were no differences between groups (Tables 7-9).

TABLE 7 Tukey method in the Heart Rate magnitude

Heart Rate Measurement (Meas) Point 180 beats/min	
Pairwise Group Differences	
Meas. 1 – Meas. 2	0.034
Meas. 1 – Meas. 3	0.069
Meas. 2 – Meas. 3	0.034
HSD	0.119

.

TABLE 8 Tukey method in the NIBP magnitude

NIBP Measurement (Meas) Point ($50mmHg$)	
Pairwise Group Differences	
Meas. 1 – Meas. 2	0.02
Meas. 1 – Meas. 3	0.04
Meas. 2 – Meas. 3	0.02
HSD	0.087

.

TABLE 9 Tukey method in the IBP magnitude

IBP Measurement (Meas) Point ($50mmHg$)	
Pairwise Group Differences	
Meas. 1 – Meas. 2	0.016
Meas. 1 – Meas. 3	0.008
Meas. 2 – Meas. 3	0.024
HSD	0.0275

.

3.3 | Brown-Forsythe test

The Brown-Forsythe test confirmed that the variances were equal as the probability of applying ANOVA on the deviation from the means was greater than the significance level $\alpha = 0.052$ (Tables 10-12).

TABLE 10 Brown Forsythe test in the Heart Rate magnitude

Heart Rate Measurement Point 180 beats/min		
Medians		
Measurement 1	Measurement 2	Measurement 3
179.64	179.64	179.64
Deviation from medians		
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.17
0.00	0.17	0.00
0.00	0.00	0.17
ANOVA Single Factor Summary		
Average		
0	0.034	0.069
Variance		
0	0.006	0.009
ANOVA		
F Calculated	P-value	F Critical
1.20	0.33	3.88

TABLE 11 Brown Forsythe test in the NIBP magnitude

NIBP Measurement Point ($50mmHg$)		
Medians		
Measurement 1	Measurement 2	Measurement 3
50.0	50.0	50.1
Deviation from medians		
0.1	0.0	0.1
0.0	0.0	0.0
0.0	0.1	0.0
0.0	0.0	0.0
0.0	0.1	0.1
ANOVA Single Factor Summary		
Average		
0.02	0.04	0.04
Variance		
0.002	0.003	0.003
ANOVA		
F Calculated	P-value	F Critical
0.25	0.78	3.88

TABLE 12 Brown Forsythe test in the IBP magnitude

IBP Measurement Point ($50mmHg$)		
Medians		
Measurement 1	Measurement 2	Measurement 3
49.66	49.62	49.66
Deviation from medians		
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.00	0.00
0.00	0.04	0.00
0.04	0.04	0.00
ANOVA Single Factor Summary		
Average		
0.008	0.016	0.00
Variance		
0.0002	0.0004	0.0000
ANOVA		
F Calculated	P-value	F Critical
1.20	0.33	3.88

3.4 | Repeatability and reproducibility $r\&R$

Repeatability
The repeatability is defined by (Eq. 1).

$$r = \sqrt{\sigma^2} \tag{1}$$

Reproducibility
The reproducibility is defined by (Eq. 2).

$$R = \sqrt{(variance)^2} \tag{2}$$

$r\&R$
The $r\&R$ combination is defined by (Eq. 3) as the square root of the sum of repeatability (average internal variability) and reproducibility (variability between conditions) variances (Table 13).

$$r\&R = \sqrt{r^2 + R^2} \tag{3}$$

TABLE 13 Placeholder Caption

	HR Meas 180 beats/min	NIBP Meas ($50mmHg$)	IBP Meas ($50mmHg$)
Repeatability	0.12	0.089	0.0282
Reproducibility	0.01	0.005	0.0005
r&R	0.12	0.089	0.0282

The utility of repeatability and reproducibility ($r\&R$) studies in metrology lies in the processes of evaluation, validation, and analysis of measurements, particularly for the validation of calibration methods [13].
To determine repeatability [3], "the closeness of agreement between the results of successive measurements of the same measurand under the same measurement conditions is considered. These conditions are known as the repeatability conditions. The repeatability conditions included the same measurement procedure, observer, and measuring instrument used under the same conditions, location, and repetition over a short period. Repeatability can be quantitatively expressed in terms of the characteristic dispersion of the results." The standard deviation is a useful measure of repeatability, which is easily corroborated in each of the measurements taken for the defined magnitudes during the same calibration, as observed in Table 1 for heart rate, where the maximum standard deviation is $0.0950197beats/min$; noninvasive blood pressure, where the maximum standard deviation is $0.05477mmHg$; and invasive blood pressure via electrical simulation, where the maximum standard deviation is $0.021909mmHg$. These results directly demonstrate that the method is repeatable for each evaluated magnitude.
To determine reproducibility [3], a one-way analysis of variance (ANOVA) was initially applied to each magnitude of interest, comparing three groups of data collected over three different years for the same equipment during calibration. The hypothesis that all measurements were statistically equal was confirmed because the calculated F-value was less than the critical F-value, as shown in Table 2. For heart rate, the calculated F-value was 1.2, and the critical F-value was 3.88; for noninvasive blood pressure, the calculated F-value was 0.75, and the critical F-value was 3.88; for invasive blood pressure via electrical simulation, the calculated F-value was 2.8, and the critical F-value was 3.88.
Subsequently, Tukey's method was applied to each magnitude of interest to compare the pairwise differences between the groups. The differences were less than the honestly significant difference (HSD), as shown in

Table 3. For the heart rate, the difference between measurements 1 and 2 was 0.03, between measurements 1 and 3 was 0.07, and between measurements 2 and 3 was 0.03, whereas the HSD was 0.12 For noninvasive blood pressure, the difference between measurements 1 and 2 was 0.02, between measurements 1 and 3 was 0.04, and between measurements 2 and 3 was 0.02, whereas the HSD was 0.09. For invasive blood pressure via electrical simulation, the difference between measurements 1 and 2 was 0.016, between measurements 1 and 3 was 0.008, and between measurements 2 and 3 was 0.024, and the HSD was 0.002666. These results indicate no significant differences between the groups, confirming that they were statistically similar. Finally, to ensure that the analysis accounted for potential heteroscedasticity, the Brown-Forsythe test was applied to confirm equal variances in the deviations of the medians for the magnitudes of interest. The obtained p-values were compared with the significance level (α), as presented in Table 4. For heart rate, the p-value was 0.33, and α was 0.052; for noninvasive blood pressure, the p-value was 0.78, and α was 0.052; for invasive blood pressure via electrical simulation, the p-value was 0.33, and α was 0.052. These results confirmed that the variances of the groups were equal.

In the repeatability and reproducibility ($r\&R$) analysis, the individual components of repeatability and reproducibility were calculated for each magnitude studied. The internal variability was estimated using the standard deviation for each measurement condition. Repeatability (r) was derived from the p variances, which estimated the variability for each condition. The following values were obtained for each magnitude (Table 5): repeatability for heart rate: $0.12 beats/min$; for noninvasive blood pressure: $0.09 mmHg$; and for invasive blood pressure via electrical simulation: $0.03 mmHg$. Reproducibility (R), or variability between conditions, was obtained through the variance of the p averages. The following values were obtained for each magnitude (Table 5): the reproducibility of heart rate was $0.01 beats/min$, noninvasive blood pressure was $0.005 mmHg$, and invasive blood pressure via electrical simulation was $0.0006 mmHg$. The reproducibility component was evaluated relative to the repeatability component as an additional indicator of measurement system stability. Although international standards do not establish acceptance limits based on the ratio R/r, the interpretation adopted in this study follows the acceptance philosophy widely used in Measurement System Analysis (MSA) [17], where measurement system variation below 10% is considered acceptable, values between 10% and 30% may be acceptable depending on the application, and values above 30% indicate that improvements to the measurement system are required. Accordingly, the following practical interpretation was adopted for the ratio between reproducibility and repeatability [13]:

$$R < 0.1 * r = \textit{acceptable} \tag{4}$$

$$0.1 * r < R < r = \textit{may be acceptable depending on the application} \tag{5}$$

$$R > 0.3 * r = \textit{improvements to measurement system are required} \tag{6}$$

The following results were obtained:

$R\&R$ Ratio for Heart Rate $R = 0.01$ and $r * 0.1 = 0.01$ Because $R < 0.1 * r$, and according to Eq. (4), the reproducibility of the heart rate was considered acceptable.

$R\&R$ Ratio for Non-Invasive Blood Pressure $R = 0.004$ and $r * 0.1 = 0.009$ Because $R < 0.1 * r$, and according to Eq. (4), the reproducibility of non-invasive blood pressure was considered acceptable.

$R\&R$ Ratio for Invasive Blood Pressure via Electrical Simulation $R = 0.0006$ and $r * 0.1 = 0.003$, because $R < 0.1 * r$, and according to Eq. (4), the reproducibility of invasive blood pressure via electrical simulation was considered acceptable.

4 | CONCLUSIONS

A calibration method for patient simulators was designed and validated for heart rate (HR), non-invasive blood pressure (NIBP), and invasive blood pressure (IBP) through electrical simulation. The calibration points were selected according to the clinical relevance of the measured quantities and, when applicable, recommendations from internationally recognized organizations.

The repeatability results demonstrated low dispersion under the evaluated measurement conditions. The maximum standard deviations were 0.095 beats/min for HR, $0.055 mmHg$ for NIBP, and $0.022 mmHg$ for IBP. These results indicate consistent measurements across repeated measurements for the three evaluated quantities.

The reproducibility analysis showed no statistically significant differences among the measurement groups collected over the evaluated calibration periods. ANOVA resulted in F-values of 1.20, 0.75, and 2.80 for HR, NIBP, and IBP, respectively, all lower than the corresponding critical value of 3.88. The associated p-values were 0.33, 0.49, and 0.10, respectively. Tukey's analysis also showed that the pairwise differences were lower than the corresponding honestly significant differences. In addition, the Brown-Forsythe test indicated no significant differences in variance among the groups, with p-values of 0.335 for HR, 0.783 for NIBP, and 0.335 for IBP.

The combined repeatability and reproducibility analysis yielded values of 0.12 beats/min for HR, $0.089 mmHg$ for NIBP, and $0.028 mmHg$ for IBP. The reproducibility component remained below 10% of the repeatability component for all three quantities, according to the acceptance criterion adopted in this study. This result indicates that the measurement method maintained stability under the evaluated conditions.

Overall, the experimental results support the validity of the proposed calibration method for the evaluated patient simulators and measurement points. The low dispersion, statistical equivalence among measurement groups, homogeneous variances, and low reproducibility relative to repeatability demonstrate that the method provides consistent calibration results for HR, NIBP, and IBP under the conditions evaluated in this study.

5 | ACKNOWLEDGMENTS, FUNDING AND AUTHOR CONTRIBUTIONS

This research was conducted with the support of the Calibration Laboratory at BIOSANCTA company located in Bogota - Colombia, South America and University ECCI project number IN-10-27 2026.

The authors declare that this research received no external funding.

References

[1] R. Muñoz-Quiroz *et al.*, "Sonqusim: Design and performance evaluation of an ecg and arrhythmia simulator32 for medical devices calibration," in *Annu Int Conf IEEE Eng Med Biol Soc.*, pp. 1–4, 2023.

[2] D. Daniele *et al.*, "Design and development of a novel invasive blood pressure simulator for patient's monitor testing.," *Sensors*, vol. 20(1), p. 259, 2020.

[3] T. B. I. des Poids et Mesures (BIPM) JCGM 200, *International vocabulary of metrology – Basic and general concepts and associated terms (VIM)*. The Bureau International des Poids et Mesures (BIPM) JCGM 200, 2012.

[4] F. Weng, J. Tian, X. Fang, *et al.*, "Research on establishment decision of medical equipment measurement standard based on gdm-ahp," *Sci Rep*, vol. 15, p. 9309, 2025.

[5] E. Sastoque, "Standardization and accreditation for the calibration of patient simulators used as working standards in the provision of maintenance and calibration services for electromedical equipment. ministry of science, technology and research of colombia.." https://repositorio.minciencias.gov.co/entities/publication/37bef28b-c3ef-46ae-abdc-d672e7e20cdd/full, 2014. Revised: 2025-20-05.

[6] C. Lalani, E. Kunwar, M. Kinard, S. Dhruva, and R. Redberg, "Reporting of death in us food and drug administration medical device adverse event reports in categories other than death.," *JAMA Intern. Med.*, vol. 181(9), pp. 1217–1223, 2021.

[7] P. Sherwin, "Digitizing the calibration process," in *18th International Congress of Metrology - EDP-Sciences*, 2017.

[8] S. Yang, B. Lam, and M. Chris, "Calibration of electrocardiograph (ecg) simulators," *NCSLI Measure*, vol. 12(1), pp. 46–53, 2018.

[9] E. Melgarejo, "Heart rate and its intervention in the management of ischemic heart disease a new approach and alternative," *Revista Colombiana de Cardiología*, vol. 16(4), pp. 159–169, 2009.

[10] A. Saptra, A. Malik, and P. Bhandari, *Evaluation of vital signs*, ch. Evaluation of vital signs. National Center for Biotechnology Information, 2023.

[11] J. Barreto-Acevedo, *A. Hemodynamic and respiratory monitoring in intensive care*, ch. Fundamentals of critical care medicine. 2023.

[12] I. O. of Legal Metrology OIML R149-3, *Non-invasive automated sphygmomanometers Part 3: Test report format*. International Organization of Legal Metrology OIML R149-3, 2020.

[13] M. . M. Asociados, *Metrological application of RR studies (Repeatability and Reproducibility)*. MetAs Metrólogos Asociados 2003, 2003.

[14] D. C. Montgomery, *Design and Analysis of Experiments, 10th Edition*. John Wiley Sons, 2024.

[15] A. Nanda *et al.*, "Multiple comparison test by tukey's honestly significant difference (hsd): Do the confident level control type i error," *International Journal of Statistics and Applied Mathematics*, vol. 6(1), pp. 59–65, 2021.

[16] A. N. Fazlin and M. Nora, "An overview of homogeneity of variance tests on various conditions based on type 1 error rate and power of a test," *Journal of Quality Measurement and Analysis (JQMA)*, vol. 18(3), pp. 111–130, 2022.

[17] A. I. A. G. (AIAG), *Measurement Systems Analysis (MSA) Reference Manual, 4th ed.* Automotive Industry Action Group (AIAG), 2010.