



UNIVERSITAT DE  
BARCELONA

## Estrategias para asegurar la calidad del proceso analítico en el laboratorio clínico

Leonor Guiñón Muñoz

**ADVERTIMENT.** La consulta d'aquesta tesi queda condicionada a l'acceptació de les següents condicions d'ús: La difusió d'aquesta tesi per mitjà del servei TDX ([www.tdx.cat](http://www.tdx.cat)) i a través del Dipòsit Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha estat autoritzada pels titulars dels drets de propietat intel·lectual únicament per a usos privats emmarcats en activitats d'investigació i docència. No s'autoritza la seva reproducció amb finalitats de lucre ni la seva difusió i posada a disposició des d'un lloc aliè al servei TDX ni al Dipòsit Digital de la UB. No s'autoritza la presentació del seu contingut en una finestra o marc aliè a TDX o al Dipòsit Digital de la UB (framing). Aquesta reserva de drets afecta tant al resum de presentació de la tesi com als seus continguts. En la utilització o cita de parts de la tesi és obligat indicar el nom de la persona autora.

**ADVERTENCIA.** La consulta de esta tesis queda condicionada a la aceptación de las siguientes condiciones de uso: La difusión de esta tesis por medio del servicio TDR ([www.tdx.cat](http://www.tdx.cat)) y a través del Repositorio Digital de la UB ([diposit.ub.edu](http://diposit.ub.edu)) ha sido autorizada por los titulares de los derechos de propiedad intelectual únicamente para usos privados enmarcados en actividades de investigación y docencia. No se autoriza su reproducción con finalidades de lucro ni su difusión y puesta a disposición desde un sitio ajeno al servicio TDR o al Repositorio Digital de la UB. No se autoriza la presentación de su contenido en una ventana o marco ajeno a TDR o al Repositorio Digital de la UB (framing). Esta reserva de derechos afecta tanto al resumen de presentación de la tesis como a sus contenidos. En la utilización o cita de partes de la tesis es obligado indicar el nombre de la persona autora.

**WARNING.** On having consulted this thesis you're accepting the following use conditions: Spreading this thesis by the TDX ([www.tdx.cat](http://www.tdx.cat)) service and by the UB Digital Repository ([diposit.ub.edu](http://diposit.ub.edu)) has been authorized by the titular of the intellectual property rights only for private uses placed in investigation and teaching activities. Reproduction with lucrative aims is not authorized nor its spreading and availability from a site foreign to the TDX service or to the UB Digital Repository. Introducing its content in a window or frame foreign to the TDX service or to the UB Digital Repository is not authorized (framing). Those rights affect to the presentation summary of the thesis as well as to its contents. In the using or citation of parts of the thesis it's obliged to indicate the name of the author.

2023

Tesis Doctoral

Leonor Guiñón Muñoz



## Estrategias para asegurar la calidad del proceso analítico en el laboratorio clínico

Leonor Guiñón Muñoz



**FACULTAD DE QUÍMICA**  
**DEPARTAMENTO DE INGENIERÍA QUÍMICA Y QUÍMICA ANALÍTICA**  
**PROGRAMA DE QUÍMICA ANALÍTICA Y MEDIO AMBIENTE**

## **ESTRATEGIAS PARA ASEGURAR LA CALIDAD DEL PROCESO ANALÍTICO EN EL LABORATORIO CLÍNICO**

*Memoria presentada por **Leonor Guiñón Muñoz** para optar al Título de  
Doctora por la Universidad de Barcelona*

**Directoras:**

**Luisa Álvarez Domínguez**

*Consultora Senior*

*Unidad de la Calidad del Centro de Diagnóstico Biomédico*

*Hospital Clínic de Barcelona*

**Ángeles Sahuquillo Estrugo**

*Profesora Agregada*

*Departamento de Ingeniería Química y Química Analítica*

*Universidad de Barcelona*



# ***Agradecimientos***

No hubiese emprendido este proyecto de no haber sido por el estímulo de la Dra. Luisa Álvarez. Fue ella la primera que creyó en el potencial de esta Tesis Doctoral. Además de su excelente labor como directora, debo agradecerle su confianza en mí y todo lo que he aprendido de ella, no sólo a nivel profesional sino también a nivel personal.

A la Dra. Ángeles Sahuquillo, por aceptarme para realizar esta Tesis Doctoral también bajo su dirección y tutoría y, por supuesto, por su implicación y su ayuda en todo el proceso.

A mis excompañeros y excompañeras del Centro de Diagnóstico Biomédico del Hospital Clínic de Barcelona. Nunca pensé que me fuese a dedicar a la gestión de la calidad, pero desarrollarme profesionalmente en un entorno de alto nivel se convirtió en un reto personal. Guardo un gran recuerdo de los momentos que compartimos. Por supuesto, también debo agradecer que me hayan permitido utilizar los datos generados en los laboratorios. En gran parte, eso ha hecho posible la realización de esta Tesis Doctoral. Debo hacer mención especial a la Unidad de la Calidad. Fueron unos años duros, de muchísimo trabajo y exigencia, pero formábamos un gran equipo. Aprendí mucho de cada una de vosotras, pero sobre todo me llevé vuestro cariño.

A mis compañeros y compañeras de los Laboratorios del Hospital de la Santa Creu i Sant Pau, por los ánimos y el apoyo que me han dado en los momentos difíciles, sobre todo en la recta final. Y al margen de esta Tesis Doctoral, por confiar en mí desde el primer día y por hacerme sentir que estoy en mi lugar. Debo hacer un agradecimiento especial al Dr. Francisco Blanco, por haber apoyado en todo momento la realización de este trabajo y por darme la oportunidad de formar parte de un equipo profesionalmente excepcional y con un valor humano inmejorable.

A mis padres, por su lucha para ofrecerme todo lo que ellos no tuvieron, para que yo llegase donde quisiera. Por vuestro esfuerzo, generosidad y dedicación, esta Tesis Doctoral os la dedico a vosotros.

A mi hermano, por su infalible complicidad.

A mis hijas, mis dos tesoros.

A Ángel, por estar siempre, por lo que compartimos y por darle sentido a todo.

## ***Índice***







Listado de figuras y tablas.....	IX
Abreviaturas y Acrónimos.....	XV
Resumen.....	XXI
<b>I. Introducción .....</b>	<b>1</b>
<b>1. Fases del proceso analítico .....</b>	<b>3</b>
<b>2. Errores en medicina de laboratorio.....</b>	<b>5</b>
<b>3. Seguridad en el proceso analítico: gestión del riesgo .....</b>	<b>8</b>
<b>4. Seguridad en el proceso analítico: plan de control de la calidad.....</b>	<b>12</b>
<b>4.1. Control de calidad interno.....</b>	<b>13</b>
<b>4.2. Control de calidad externo .....</b>	<b>22</b>
<b>4.3. Especificaciones de calidad analítica.....</b>	<b>26</b>
<b>4.4. Selección de especificaciones de calidad analítica .....</b>	<b>29</b>
<b>5. Aseguramiento de la calidad según la norma ISO 15189 .....</b>	<b>33</b>
<b>6. Contexto.....</b>	<b>38</b>
<b>II. Objetivos.....</b>	<b>43</b>
<b>III. Resultados y discusión .....</b>	<b>47</b>
<b>1. Primer objetivo: Proporcionar especificaciones de calidad de error total basadas en el estado del arte.....</b>	<b>49</b>
<b>1.1. Primer estudio.....</b>	<b>49</b>
<b>1.1.1. Presentación .....</b>	<b>50</b>
<b>1.1.2. Resultados y discusión.....</b>	<b>52</b>
<b>1.1.3. Conclusión .....</b>	<b>55</b>
<b>1.1.4. Cita al artículo publicado .....</b>	<b>55</b>
<b>1.1.5. Publicación completa .....</b>	<b>55</b>
<b>1.2. Segundo estudio .....</b>	<b>65</b>
<b>1.2.1. Presentación .....</b>	<b>65</b>
<b>1.2.2. Resultados y discusión.....</b>	<b>67</b>
<b>1.2.3. Conclusión .....</b>	<b>70</b>
<b>1.2.4. Cita al artículo publicado .....</b>	<b>71</b>
<b>1.2.5. Publicación completa .....</b>	<b>71</b>
<b>2. Segundo objetivo: Establecer una metodología basada en el análisis de riesgos y la métrica sigma para realizar una evaluación anual del rendimiento de la fase analítica.....</b>	<b>77</b>

2.1. Tercer estudio .....	77
2.1.1. Presentación .....	77
2.1.2. Resultados y discusión.....	79
2.1.3. Conclusión .....	82
2.1.4. Cita al artículo publicado .....	83
2.1.5. Publicación completa .....	83
<b>3. Tercer objetivo: Establecer una metodología para verificar periódicamente la intercambiabilidad de resultados. ....</b>	<b>93</b>
<b>3.1. Cuarto estudio .....</b>	<b>93</b>
3.1.1. Introducción .....	93
3.1.2. Objetivo.....	95
3.1.3. Material y métodos .....	95
3.1.4. Resultados y discusión.....	104
3.1.5. Conclusión .....	109
<b>4. Cuarto objetivo: Establecer una metodología para verificar la fiabilidad de los resultados ante la ausencia de un programa de intercomparación. ....</b>	<b>111</b>
<b>4.1. Quinto estudio.....</b>	<b>111</b>
4.1.1. Presentación .....	111
4.1.2. Resultados y discusión.....	113
4.1.3. Conclusión .....	115
4.1.4. Cita al artículo publicado .....	115
4.1.5. Publicación completa .....	115
<b>5. Quinto objetivo: Establecer una metodología para evaluar la concordancia entre observadores en aquellas actividades que se consideren subjetivas. ....</b>	<b>121</b>
<b>5.1. Sexto estudio.....</b>	<b>121</b>
5.1.1. Presentación .....	121
5.1.2. Resultados y discusión.....	124
5.1.3. Conclusión .....	125
5.1.4. Cita al artículo publicado .....	126
5.1.5. Publicación completa .....	126
<b>IV. Conclusiones .....</b>	<b>135</b>
<b>V. Bibliografía.....</b>	<b>139</b>

## ***Listado de figuras y tablas***





**Figura 1.** Actividades que se desarrollan en las diferentes fases del proceso analítico: pre-pre-analítica, pre-analítica, analítica, post-analítica y post-post-analítica.

**Figura 2.** Gráfica de Levey-Jennings.

**Figura 3.** Descripción gráfica de la fórmula para el cálculo del valor sigma.

**Figura 4.** Diagrama de flujo para la asignación de los modelos de especificaciones de calidad analítica.

**Figura 5.** Variación del error total observada en cada uno de los laboratorios participantes en función del rango de concentración del material de control.

**Figura 6.** Comparación del valor sigma obtenido para las magnitudes incluidas en el AMFE del año 2018 con respecto al año 2017.

**Figura 7.** Diagrama de flujo del protocolo de verificación periódica de la intercambiabilidad de resultados descrito en la guía EP31-A-IR del CLSI.

**Figura 8.** Tarjeta de cribado neonatal con una muestra de calidad aceptable.

**Figura 9.** Electroferograma de hemoglobinas en un recién nacido a término con un patrón normal.

**Tabla 1.** Errores más frecuentes que se observan en las diferentes fases del proceso analítico: pre-pre-analítica, pre-analítica, analítica, post-analítica y post-post-analítica.

**Tabla 2.** Escala de puntuación de los factores de riesgo: ocurrencia, severidad y detectabilidad.

**Tabla 3.** Descripción del nivel de riesgo y las acciones requeridas según el Número de Prioridad del Riesgo (NPR).

**Tabla 4.** Tabla de conversión del número de defectos por millón de oportunidades (DPMO) a la métrica sigma.

**Tabla 5.** Descripción del rendimiento de un procedimiento de medida según su nivel sigma.

**Tabla 6.** Clasificación por categorías de los Programas de Evaluación Externa de la Calidad en función de sus características.

**Tabla 7.** Información sobre los programas de intercomparación: magnitudes incluidas en cada programa, rango de valores de los materiales de control, unidades de medida y número de resultados de los laboratorios participantes.

**Tabla 8.** Procedimientos de medida utilizados por los laboratorios participantes en cada uno de los PEEC.

**Tabla 9.** Percentiles 90 del error de medida obtenido en las concentraciones extremas y especificaciones de error total basadas en la variabilidad biológica para cada magnitud.

**Tabla 10.** Distribución de los procedimientos de medida en los analizadores Alinity c del laboratorio Core.

**Tabla 11.** Estimación del rango de concentración (o actividad) de las diferentes magnitudes para la selección de las muestras del estudio a partir de los datos acumulados de control de calidad.

**Tabla 12.** Criterios de sesgo establecidos, número de series y replicados a realizar, y diferencias máximas permitidas entre analizadores para las diferentes magnitudes a concentraciones cercanas al nivel 2 del material de control.

**Tabla 13.** Criterios de sesgo establecidos, número de series y replicados a realizar, y diferencias máximas permitidas entre analizadores para las diferentes magnitudes a concentraciones cercanas al nivel 3 del material de control.

**Tabla 14.** Diferencias máximas observadas entre analizadores y evaluación de la intercambiabilidad de resultados a concentraciones (o actividad) cercanas al nivel 2 del material de control.

**Tabla 15.** Diferencias máximas observadas entre analizadores y evaluación de la intercambiabilidad de resultados a concentraciones (o actividad) cercanas al nivel 3 del material de control.

**Tabla 16.** Interpretación del coeficiente alfa de *Krippendorff*.





## ***Abreviaturas y Acrónimos***





**AECNE:** Asociación Española de Cribado Neonatal.

**ALT:** Alanina aminotransferasa.

**AMFE:** Análisis modal de fallos y efectos.

**AST:** Aspartato aminotransferasa.

**AT:** Antitrombina.

**BAS:** Basófilos.

**C0:** Carnitina libre.

**C2:** Acetilcarnitina.

**C3:** Propionilcarnitina.

**C4:** Butirilcarnitina.

**C5:** Isovalerilcarnitina.

**C5DC:** Glutarilcarnitina.

**C6:** Hexanoilcarnitina.

**C8:** Octanoilcarnitina.

**C10:** Decanoilcarnitina.

**C14:** Miristoilcarnitina.

**C16:** Palmitoilcarnitina.

**C18:** Estearoilcarnitina.

**CHCM:** Concentración de hemoglobina corpuscular media.

**CK:** Creatina quinasa.

**CLIA:** Clinical Laboratory Improvement Amendments.

**CLSI:** Clinical Laboratory Standards Institute.

**CV:** Coeficiente de variación.

**CV<sub>i</sub>:** Coeficiente de variación basado en la variabilidad biológica intraindividual.

**DC:** Diferencia crítica.

**DE:** Desviación estándar.

**DMP:** Diferencia máxima permitida.

**DPMO:** Defectos por millón de oportunidades.

**EFLM:** European Federation of Clinical Chemistry and Laboratory Medicine.

**EQALM:** European Organization for External Quality Assurance Providers in Laboratory Medicine.

**EM:** Error de medida.

**EOS:** Eosinófilos.

**ES:** Error sistemático.

**ET:** Error total.

**ET<sub>p</sub>:** Error máximo permitido.

**FA:** fosfatasa alcalina.

**FEN:** Fenilalanina.

**FVIII:** Factor VIII.

**FWV:** Factor von Willebrand.

**GGT:** Gamma glutamil transferasa.

**HB:** Hemoglobina.

**HbA<sub>2</sub>:** Hemoglobina A2.

**HCM:** Hemoglobina corpuscular media.

**HDL:** Lipoproteínas de alta densidad.

**HEM:** Hematíes.

**HTO:** Hematocrito.

**IDS:** Índice de desviación estándar.

**IFCC:** International Federation of Clinical Chemistry and Laboratory Medicine.

**ISO:** International Organization for Standardization.

**LEU:** Leucocitos.

**LDH:** Lactato deshidrogenasa.

**LINF:** Linfocitos.

**MON:** Monocitos.

**NEU:** Neutrófilos.

**NPR:** Número de Prioridad del Riesgo.

**PEEC:** Programa de evaluación externa de la calidad.

**PC:** Proteína C.

**PCR:** Proteína C reactiva.

**Pde:** Probabilidad de detección de error.

**Pfr:** Probabilidad de falso rechazo.

**PLT:** Plaquetas.

**RET:** Reticulocitos.

**SEHH:** Sociedad Española de Hematología y Hemoterapia.

**SEQC<sup>ML</sup>**: Sociedad Española de Medicina de Laboratorio.

**SR**: Desviación estándar a corto plazo.

**ST**: Desviación estándar a largo plazo.

**TIR**: Tripsinógeno inmunoreactivo.

**TSH**: Hormona estimulante de la tiroides.

**TP**: Tiempo de protrombina.

**TTPA**: Tiempo de tromboplastina parcial activada.

**TYR**: Tirosina.

**UPLC-MS/MS**: Cromatografía líquida de ultra alto rendimiento-espectrometría de masas en tándem.

**VB**: Variabilidad biológica.

**VCM**: Volumen corpuscular medio.

## ***Resumen***







## **Resumen**

Los estudios realizados en los últimos años han demostrado que la medicina de laboratorio es una pieza clave en la toma de decisiones clínicas y el manejo del paciente. Debido al impacto que se ha reconocido tienen los resultados analíticos en la seguridad del paciente, los profesionales del laboratorio clínico y las sociedades científicas de este ámbito han ido adquiriendo conciencia de la importancia del aseguramiento de la calidad del proceso analítico. Así, se considera necesario que los profesionales del laboratorio clínico se aseguren de que cada uno de los pasos que conforman el proceso analítico se realiza correctamente, de que son capaces de identificar las posibles debilidades que existan en el mismo y de que se diseñan e implementan acciones correctivas o de mejora para abordar tales debilidades.

La necesidad de controlar, monitorizar y mejorar la calidad en los laboratorios clínicos ha crecido de la mano de los avances tecnológicos, la automatización y el alto nivel de especialización. Por otro lado, la aparición de la norma ISO 15189 puso sobre la mesa los requisitos que debía cumplir un laboratorio clínico para asegurar la calidad técnica de sus procedimientos analíticos, así como la competencia del personal para la realización de cualquier actividad perteneciente al proceso analítico. Todo ello ha dado lugar a la necesidad de implementar nuevas herramientas de aseguramiento de la calidad para mejorar la prevención y la detección de errores con la finalidad de garantizar la fiabilidad de la información proporcionada por el laboratorio. Sin embargo, en función del tipo de laboratorio que se considere y de los recursos de los que disponga, su incorporación y compaginación con la práctica asistencial puede resultar compleja o incluso inviable.

En esta Tesis Doctoral se han diseñado e implementado diferentes estrategias para el aseguramiento de calidad de los resultados analíticos en el laboratorio de un hospital de tercer nivel, tras la identificación de algunas debilidades del proceso analítico. Su interés y novedad radican, fundamentalmente, en los aspectos que se consideraron para adaptar las

recomendaciones de las sociedades científicas a las peculiaridades de cada proceso analítico y a la realidad de cada laboratorio.

Los estudios realizados permiten establecer especificaciones de calidad de error total basadas en el estado del arte. Por otro lado, se presenta una metodología basada en la combinación del análisis de riesgos y la métrica sigma aplicable a la evaluación a largo plazo del rendimiento de la fase analítica. Por último, se desarrollan estrategias para verificar periódicamente la intercambiabilidad de resultados entre equipos duplicados, para verificar la fiabilidad de los resultados en ausencia de un programa de intercomparación y para evaluar la concordancia entre observadores en actividades subjetivas.

Más allá de mostrarse su aplicación para asegurar la calidad de los resultados de ciertas magnitudes, la finalidad de esta Tesis Doctoral es ofrecer el conocimiento adquirido durante la implementación de dichas estrategias a la comunidad científica, dada la posibilidad de aplicarlas para otras magnitudes y áreas de especialización del laboratorio clínico.

## **Resum**

Els estudis realitzats en els darrers anys han demostrat que la medicina de laboratori és una peça clau en la presa de decisions clíniques i el maneig del pacient. A causa de l'impacte que s'ha reconegut tenen els resultats analítics en la seguretat del pacient, els professionals del laboratori clínic i les societats científiques d'aquest àmbit han anat adquirint consciència de la importància de l'assegurament de la qualitat del procés analític. Així, es considera necessari que els professionals de laboratori clínic s'assegurin que cadascun dels passos que conformen el procés analític es realitza correctament, que són capaços d'identificar les possibles febleses que existeixin en el mateix i que es dissenyen i implementen accions correctives o de millora per a abordar tals febleses.

La necessitat de controlar, monitorar i millorar la qualitat en els laboratoris clínics ha crescut de la mà dels avanços tecnològics, l'automatització i l'alt nivell d'especialització. D'altra banda, l'aparició de la norma ISO 15189 va posar sobre la taula els requisits que havia de complir un laboratori clínic per a assegurar la qualitat tècnica dels seus procediments analítics, així com la competència del personal per a la realització de qualsevol activitat que formi part del procés analític. Tot això ha donat lloc a la necessitat d'implementar noves eines d'assegurament de la qualitat per a millorar la prevenció i la detecció d'errors amb la finalitat de garantir la fiabilitat de la informació proporcionada pel laboratori. No obstant això, en funció del tipus de laboratori que es consideri i dels recursos que disposi, la seva incorporació i compaginació amb la pràctica assistencial pot resultar complexa o fins i tot inviable.

En aquesta Tesi Doctoral s'han dissenyat i implementat diferents estratègies per a l'assegurament de qualitat dels resultats analítics en el laboratori d'un hospital de tercer nivell, després de la identificació d'algunes febleses del procés analític. La seva novetat i interès radiquen, fonamentalment, en els aspectes que es van considerar per a adaptar les

recomanacions de les societats científiques a les peculiaritats de cada procés analític i a la realitat de cada laboratori.

Els estudis realitzats permeten establir especificacions de qualitat d'error total basades en l'estat de l'art. D'altra banda, es presenta una metodologia basada en la combinació de l'anàlisi de riscos i la mètrica sigma aplicable a l'avaluació a llarg termini del rendiment de la fase analítica. Finalment, es desenvolupen estratègies per a verificar periòdicament la intercanviabilitat de resultats entre equips duplicats, per a verificar la fiabilitat dels resultats en absència d'un programa d'intercomparació i per a avaluar la concordança entre observadors en activitats subjectives.

Més enllà de mostrar-se la seva aplicació per a assegurar la qualitat d'unes certes magnituds, la finalitat d'aquesta Tesi Doctoral és oferir el coneixement adquirit durant la implementació d'aquestes estratègies a la comunitat científica, donada la possibilitat d'aplicar-les per a altres magnituds i àrees d'especialització del laboratori clínic.

## **Summary**

Studies in recent years have shown that laboratory medicine is an essential component for the clinical decision-making and patient management. Due to the recognized impact of analytical results on patient safety, clinical laboratory professionals and scientific societies in this field have become aware of the importance of ensuring the quality of the analytical process. Therefore, it is considered necessary that clinical laboratory professionals ensure the correct performance of each step of the analytical process, that possible weaknesses in the process are identified and that corrective or improvement actions are designed and implemented to address such weaknesses.

The need to control, monitor and improve quality in clinical laboratories has grown hand in hand with technological advances, automation and the high level of specialization. Furthermore, the emergence of ISO 15189 laid out the requirements that a clinical laboratory must fulfil to ensure the technical quality of its analytical procedures, as well as the competence of the staff to perform any of the activities that belong to the analytical process. All this has prompted to implement new quality assurance tools to improve prevention and error detection in order to ensure the reliability of the information provided by the laboratory. However, depending on the type of laboratory that is considered and the resources available to it, its incorporation and compatibility with healthcare practice can be complex or even unviable.

This Doctoral Thesis shows the design and implementation of different strategies for quality assurance of analytical results in the laboratory of a third-level hospital, that were carried out due to the identification of some weaknesses in the analytical process. Fundamentally, its novelty and interest lie in the aspects that were considered to adapt the recommendations of scientific societies to the peculiarities of each analytical process and to the reality of each laboratory.

The studies performed allow us to establish total error quality specifications based on the state of the art. On the other side, it presents a methodology based on the combination of risk analysis and the sigma metric applicable to the long-term evaluation of the performance of the analytical phase. Finally, strategies are developed to verify periodically the comparability of results between duplicate analyzers, to verify the reliability of the results in the absence of proficiency testing, and to assess the agreement between observers in subjective activities.

Beyond showing its application to ensure the quality of certain magnitudes, the purpose of this Doctoral Thesis is to offer the knowledge acquired during the implementation of these strategies to the scientific community, given the possibility of applying them for other magnitudes and areas of clinical laboratory specialization.

## ***I. Introducción***







En el entorno del sistema sanitario, el laboratorio clínico desempeña un papel imprescindible ya que gestiona una información que impacta de forma directa en las decisiones de los profesionales clínicos. Ya por el año 1996, *Forsman* indicó que el 60-70 % de las decisiones clínicas más críticas, como por ejemplo el ingreso hospitalario, el alta de los pacientes o el tratamiento farmacológico, estaban basadas en la información suministrada por el laboratorio (1). En España, en el año 2013, el entonces Ministerio de Sanidad, Servicios Sociales e Igualdad publicó una guía, sobre la calidad y seguridad en el laboratorio clínico, en la que se constataba que los resultados proporcionados por el laboratorio influían en el 70 % de las decisiones clínicas, ya sea en la realización del diagnóstico de una enfermedad, su pronóstico o su tratamiento (2). Estos datos pusieron de manifiesto el impacto que podían tener los errores del proceso analítico en la seguridad del paciente y, por tanto, la necesidad de asegurar la calidad de los resultados de los estudios que se realizan en el laboratorio clínico.

### ***1. Fases del proceso analítico***

El proceso analítico en un laboratorio clínico se ha clasificado tradicionalmente en tres fases: pre-analítica, analítica y post-analítica.

En 1981, *Lundberg* describió el proceso analítico como un conjunto de nueve etapas: la solicitud de la prueba, la obtención/recogida de la muestra, la identificación de la muestra, su transporte al laboratorio, la preparación de la muestra, su análisis, la emisión del informe de resultados, la interpretación de los resultados y la decisión/actuación del profesional clínico (3).

Posteriormente, para referirse a aquellas actividades que no se realizan en el laboratorio y que, por tanto, no están bajo su control directo, se introdujeron los términos referidos a las fases pre-pre-analítica y post-post-analítica (4,5). En estas fases se incluyen las actividades iniciales y finales del proceso analítico que dependen principalmente del profesional clínico. Así, se considera que la elección y solicitud de las pruebas analíticas por parte del profesional clínico

forma parte de la fase pre-pre-analítica. Otras actividades como la obtención /recogida de la muestra, su identificación y su transporte al laboratorio, aun dependiendo de las directrices del laboratorio (como por ejemplo la definición de las características de las muestras y las condiciones de conservación que se precisan para el análisis solicitado), no son realizadas por personal a su cargo. Actividades como la interpretación que realiza el profesional clínico del informe de resultados y las decisiones tomadas al respecto de cara al manejo del paciente forman parte de la fase post-post-analítica. En la Figura 1 se muestran de forma esquemática las fases del proceso analítico en el laboratorio clínico.

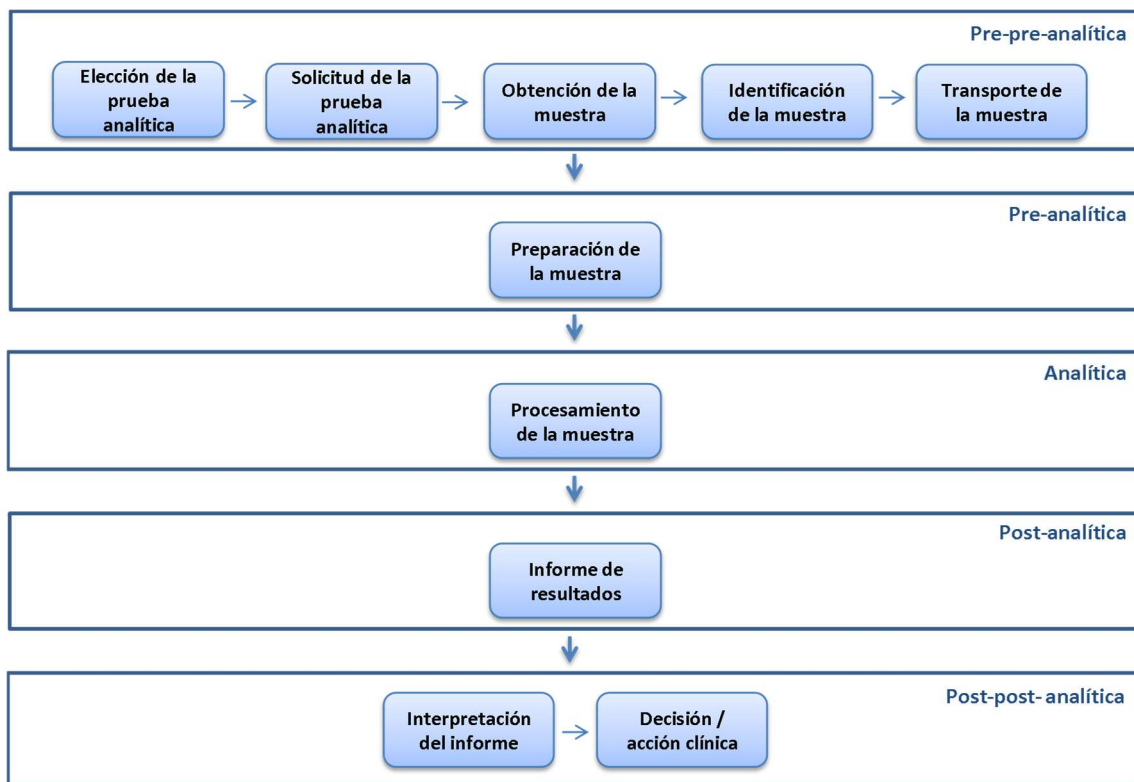


Figura 1. Actividades que se desarrollan en las diferentes fases del proceso analítico: pre-pre-analítica, pre-analítica, analítica, post-analítica y post-post-analítica.

## **2. Errores en medicina de laboratorio**

Hasta la década de los 90, únicamente se consideraban errores del laboratorio clínico aquellos que eran propiamente analíticos y ocurrían dentro de las instalaciones del laboratorio. Este concepto cambió con el paso de los años y actualmente se utiliza el término “error en medicina de laboratorio” para referirse a cualquier error que ocurra desde la solicitud de una prueba analítica, hasta la actuación llevada a cabo en el paciente en base a la interpretación clínica de los resultados; incluyéndose por tanto también los errores del proceso analítico que se originan fuera de las instalaciones del laboratorio (6). En la Tabla 1 se enumeran algunos de los tipos de errores más comunes para cada fase del proceso analítico. Mientras que la mayoría de los errores afectan directamente a la calidad del resultado analítico, algunos tipos de error, como por ejemplo exceder el tiempo de respuesta, tienen un efecto en la calidad del servicio. En este caso en concreto, daría lugar a una demora en la toma de decisión en el diagnóstico del paciente, lo que puede ser de extrema gravedad dependiendo del contexto clínico.

Tabla 1. Errores más frecuentes que se observan en las diferentes fases del proceso analítico: pre-pre-analítica, pre-analítica, analítica, post-analítica y post-post-analítica. Adaptada de (7).

Fase del proceso analítico	Tipo de error
<b>Pre-pre-analítica (fuera del laboratorio)</b>	<ul style="list-style-type: none"><li>♦ Elección de una prueba inadecuada</li><li>♦ Realización incorrecta de la solicitud</li><li>♦ Identificación incorrecta del paciente o de la muestra</li><li>♦ Obtención no adecuada de la muestra: momento de obtención, hemólisis, volumen insuficiente, etc.</li><li>♦ Contenedor inadecuado</li><li>♦ Condiciones de conservación y transporte de la muestra al laboratorio</li></ul>
<b>Pre-analítica (en el laboratorio)</b>	<ul style="list-style-type: none"><li>♦ Clasificación y/o distribución de la muestra incorrecta</li><li>♦ Derrame de la muestra</li><li>♦ Alicuotación incorrecta</li><li>♦ Etiquetado de la muestra (por ejemplo, cruce de muestras)</li><li>♦ Centrifugación incorrecta (por ejemplo, tiempo y/o velocidad)</li></ul>
<b>Analítica (en el laboratorio)</b>	<ul style="list-style-type: none"><li>♦ Funcionamiento incorrecto del analizador</li><li>♦ Contaminación por arrastre</li><li>♦ Interferencias (endógenas y exógenas)</li><li>♦ Fallo del control de calidad no detectado</li></ul>
<b>Post-analítica (en el laboratorio)</b>	<ul style="list-style-type: none"><li>♦ Error en la validación de los resultados</li><li>♦ Error de transcripción de los resultados</li><li>♦ Tiempo de respuesta excedido</li><li>♦ Demora en la comunicación de resultados de alto riesgo</li></ul>
<b>Post-post-analítica (fuera del laboratorio)</b>	<ul style="list-style-type: none"><li>♦ Demora en la actuación desde la emisión del informe de resultados</li><li>♦ Interpretación incorrecta de los resultados</li><li>♦ Seguimiento/tratamiento inadecuado del paciente</li></ul>

Son escasos los estudios en los que se ha cuantificado la tasa de error en el proceso analítico en su totalidad. En el año 2002, *Bonini* publicó una revisión bibliográfica de los estudios en los que para hacer la estimación se habían considerado los errores de todas las fases del proceso analítico (8). Sin embargo, no es posible comparar los resultados entre sí, debido a las diferentes metodologías utilizadas para realizar la estimación, así como a la heterogeneidad de los laboratorios donde se realizaron los estudios.

El estudio citado con más frecuencia es el que realizaron *Carraro y Plebani* en el año 1997 en un laboratorio de urgencias de un hospital de tercer nivel en Italia (9), y el estudio de seguimiento realizado en el mismo hospital en 2007 (10). El porcentaje de error estimado en el estudio más reciente fue del 0,3 % (10). A pesar de ser un porcentaje de error menor que el estimado en otras áreas de especialización de la sanidad (11), dado el elevado número de solicitudes analíticas que se reciben en los laboratorios, el número absoluto de errores que podrían afectar a la seguridad del paciente sigue siendo inaceptable.

A finales de la década de los 90 se constató que del total de errores que ocurrían en el proceso analítico, la proporción de errores asociados a las fases extra-analíticas (pre-pre-analítica, pre-analítica, post-analítica y post-post-analítica) era marcadamente superior a los asociados a la fase analítica (12). Concretamente, se observó que en la fase pre-pre-analítica ocurrían aproximadamente la mitad de los errores del proceso analítico (46-68,2 %) y en la fase post-post-analítica más de una tercera parte de los errores (25-45,5 %). La distribución de errores en las otras fases se estimó que era de un 3-5,3 % en la fase pre-analítica, un 7-13 % en la fase analítica y un 12,5-20 % en la fase post-analítica (7). Esto puso de manifiesto que las etapas más vulnerables del proceso analítico eran las que no se encontraban bajo el control del personal del laboratorio.

La menor proporción de errores en la fase analítica fue debida principalmente a tres factores:

- El elevado grado de automatización implementado por los laboratorios.
- La aparición de mejores procedimientos de medida y su uso mayoritario por parte de los laboratorios, con la consiguiente mejora en el rendimiento analítico (13).
- La aplicación de herramientas de control de la calidad en la fase analítica.

A partir de entonces y hasta la actualidad, se ha insistido en la necesidad de impulsar la cooperación multidisciplinar con los profesionales que participan en las etapas del proceso

analítico que tienen lugar fuera del laboratorio en beneficio del paciente (14). Por lo tanto, a pesar de que sea necesario continuar buscando herramientas para mejorar el rendimiento analítico de los procedimientos de medida, también deben dedicarse esfuerzos a identificar medidas de control dirigidas a aquellos aspectos de las fases extra-analíticas en las que pueda existir un riesgo para el paciente. Y, una vez identificadas las medidas oportunas, éstas deben incorporarse al plan de control de la calidad. De hecho, un plan de control de la calidad se considera sólo adecuado cuando cubre todas las fases del proceso analítico y se utilizan herramientas dirigidas a tres posibles escenarios: a prevenir errores, a detectarlos si ocurren y a corregirlos cuando sea preciso (15). Así, el plan de control de la calidad debe definir herramientas para minimizar el riesgo de que ocurran errores; debe disponer de un método robusto para maximizar la detección de errores; y debe tener establecido cómo actuar cuando el error ya ha ocurrido.

### ***3. Seguridad en el proceso analítico: gestión del riesgo***

Actualmente se reconoce que en los laboratorios clínicos no es suficiente con identificar, analizar y monitorizar los errores, sino que deben gestionarse los riesgos potenciales antes de que ocurra un error que pueda afectar a la fiabilidad de los resultados. Es por ello que hoy en día la gestión del riesgo es uno de los componentes clave de los sistemas de gestión de la calidad de los laboratorios, jugando un papel fundamental en el aseguramiento de la calidad de todas las fases del proceso analítico.

Teniendo en cuenta que tradicionalmente la fase analítica ha sido la más controlada mediante la implementación del plan de control de calidad interno y la participación en programas de intercomparación, actualmente en la gestión del riesgo suelen tener mayor protagonismo los riesgos asociados a las fases extra-analíticas (16) por lo que en los últimos años se ha dado un gran impulso a la monitorización de estas fases para conseguir minimizar los errores. Ejemplo

de ello fue el documento de consenso que publicaron en 2017 los grupos de trabajo “Laboratory Errors and Patient Safety” de la International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) y “Performance specifications for the extra-analytical phases” de la European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) con la finalidad de poner a disposición de los Laboratorios herramientas para la evaluación del rendimiento de las fases extra-analíticas (17). La armonización de indicadores de las fases-extra-analíticas y el establecimiento de especificaciones de calidad han sido de gran ayuda para los laboratorios a la hora de identificar aquellos aspectos que precisan acciones de mejora.

Para implementar la gestión del riesgo en el laboratorio clínico son de gran ayuda algunas normas de la International Organization for Standardization (ISO) (18,19) y guías del Clinical Laboratory Standards Institute (CLSI) (20,21). Si bien en ellas se definen multitud de herramientas para la gestión del riesgo, en los laboratorios clínicos existe bastante consenso en aplicar el Análisis Modal de Fallos y Efectos (AMFE). El AMFE es una herramienta que permite identificar riesgos potenciales y que, además, puede ayudar a solucionar o minimizar riesgos ya conocidos (22). Una vez identificados los fallos reales o potenciales, se realiza una estimación del riesgo mediante diferentes factores. A este respecto, es posible utilizar dos modelos diferentes en función de los factores utilizados para la estimación del riesgo. La guía EP23A “Laboratory Quality Control Based on Risk Management” del CLSI (20) recomienda utilizar únicamente, como factores para la estimación del riesgo, la probabilidad de aparición (ocurrencia) y la gravedad del fallo (severidad), siendo este el modelo de dos factores. El modelo de tres factores, clásicamente utilizado en el ámbito de la industria, considera además incluir la probabilidad de detectar un fallo antes de que impacte en el paciente (detectabilidad) (23).

El uso del modelo de sólo dos factores se basa en que detectar un error cuando ya ha sucedido es demasiado tarde, y los esfuerzos deberían centrarse en reducir la ocurrencia del error y en reducir la severidad de cualquier error que pudiese ocurrir. El modelo de tres factores se basa

en que siempre existe un riesgo residual de que un error ocurra, y en tal caso el laboratorio necesita tener la capacidad de detectarlo. No existe un consenso sobre la aplicación de un modelo u otro en el laboratorio clínico. Sin embargo, el modelo de tres factores parece el más aceptado por ser la detectabilidad la esencia del control de la calidad, ya que gracias a ésta se corrigen los errores antes de que los resultados sean entregados al profesional clínico o al paciente, previniendo el impacto que estos errores podrían suponer (23).

Para realizar la estimación del nivel de riesgo de los fallos identificados, el primer paso es asignar un valor a los factores de riesgo de acuerdo a una escala numérica. Generalmente, cada factor se puntúa de 1 (menor ocurrencia, menor severidad, mayor detectabilidad) a 10 (mayor ocurrencia, mayor severidad, menor detectabilidad). La escala de puntuación puede establecerse de forma individualizada o puede utilizarse alguna de las ya preestablecidas. En la Tabla 2 se presenta una posible forma de puntuar los factores de riesgo.

Tabla 2. Escala de puntuación de los factores de riesgo: ocurrencia, severidad y detectabilidad. Adaptada de (24).

<b>Ocurrencia</b>		<b>Severidad</b>		<b>Detectabilidad</b>	
Frecuente	9 - 10	Catastrófico	9 - 10	Difícil	9 - 10
Ocasional	7 - 8	Mayor	5 - 8	Moderada	7 - 8
Infrecuente	5 - 6	Moderada	3 - 4	Buena	5 - 6
Remota	1 - 4	Menor	1 - 2	Alta	1 - 4

A continuación, se realiza una priorización de los riesgos mediante el cálculo de un índice resultante de la multiplicación de los valores asignados a los factores. Este índice se conoce como Número de Prioridad del Riesgo (NPR).



Para cada modo de fallo se calcula:

$$\text{NPR} = \text{O} \times \text{S} \times \text{D}$$

siendo O: ocurrencia

S: severidad

D: detectabilidad

Si se ha utilizado una escala de puntuación de 1 a 10, se obtendrán valores de NPR entre 1 y 1.000. Deberá darse prioridad a los riesgos que hayan obtenido índices mayores. Cuanto mayor sea el valor de NPR, mayor será el riesgo del modo de fallo y, por lo tanto, más urgente será diseñar una estrategia para mitigarlo. Además, deberá establecerse a partir de qué valor de NPR se considera necesario implementar acciones para reducir el riesgo, así como por debajo de qué valor de NPR el riesgo se considera asumible y, por lo tanto, no es necesaria la implementación de ninguna acción. En la Tabla 3 se muestra un ejemplo de la asociación establecida entre el nivel de riesgo y el valor de NPR habiendo utilizado una escala de puntuación de 1 a 10 para evaluar los factores de riesgo.

Tabla 3. Descripción del nivel de riesgo y las acciones requeridas según el Número de Prioridad del Riesgo (NPR). Adaptada de (25).

NPR	Nivel de riesgo	Acciones requeridas
499 < NPR ≤ 1000	Catastrófico	Requiere acciones inmediatas para mitigar el riesgo
250 < NPR ≤ 499	Alto	Requiere acciones para mitigar el riesgo lo antes posible
100 < NPR ≤ 249	Moderado	Requiere monitorización
NPR ≤ 100	Bajo	No requiere acciones

Por último, después de implementar las acciones correctivas, debe realizarse un nuevo AMFE para comprobar si el NPR ha disminuido, lo que indicaría que las acciones llevadas a cabo para mitigar los riesgos han sido eficaces.

Así bien, a pesar de que el AMFE es la metodología más comúnmente utilizada para la gestión del riesgo, presenta las siguientes limitaciones (26):

- Subjetividad en la asignación de valores a cada uno de los factores de riesgo (ocurrencia, severidad, detectabilidad).
- Falta de estandarización de las escalas de puntuación de los factores de riesgo.
- Falta de estandarización del criterio para evaluar el índice obtenido y por tanto para realizar la priorización de los riesgos.

Por lo tanto, según el criterio de cada laboratorio, un mismo escenario puede tener asignados diferentes niveles de riesgo. Esto dificulta la comparación entre laboratorios, tanto en lo que respecta a la priorización de riesgos; como a la evaluación de la eficacia de las diferentes acciones tomadas para la minimización de riesgos.

#### ***4. Seguridad en el proceso analítico: plan de control de la calidad***

Disponer de un plan de control de la calidad en el laboratorio clínico es esencial para tener la confianza de que los errores se detectan de forma eficaz y, en consecuencia, de que está asegurada la calidad de los resultados emitidos. Para ello, el plan de control de la calidad implementado debe optimizarse al máximo de acuerdo a la información disponible al respecto y a los recursos disponibles en el laboratorio. Éste debe incorporar tanto, herramientas de control dirigidas a la evaluación del rendimiento analítico de los procedimientos de medida, como cualquier actividad complementaria que se considere necesaria para controlar cualquier efecto indeseado en los resultados analíticos (15). A continuación, se detallan algunas de las actividades básicas que forman parte de un plan de control de la calidad en el laboratorio clínico:

- Como mínimo en la fase pre-analítica deben examinarse las muestras para determinar su aceptabilidad para el análisis (por ejemplo, tipo de muestra, volumen, tipo de

contenedor/tubo), y comprobar que la identificación del paciente en la solicitud se corresponde con el de la identificación de la muestra.

- En la fase analítica la medición de materiales de control de calidad interno y la participación en programas de intercomparación se consideran herramientas esenciales para la detección de los posibles fallos en el procedimiento analítico.
- En la fase post-analítica el establecimiento de límites de alerta (intervalo de valores, que cuando se supera, indican un posible resultado erróneo) (27) y de límites de cambio o deltacheck (intervalo de valores de las diferencias obtenidas entre dos resultados consecutivos en el mismo paciente que, cuando se supera, indica un posible resultado erróneo) (27) son de gran ayuda para el control de la plausibilidad de los resultados analíticos. También debe incluirse el seguimiento del tiempo de respuesta como mínimo de las magnitudes que se consideran críticas, así como el tiempo transcurrido hasta la notificación de resultados de alto riesgo para la seguridad del paciente.
- También es fundamental disponer de controles relativos a la ejecución del procedimiento analítico, por ejemplo, que el personal técnico haya sido capacitado para llevarlo a cabo, que se sigan las instrucciones descritas en los procedimientos normalizados de trabajo, que se cumpla la planificación de los programas de mantenimiento del analizador o que se compruebe la concordancia entre observadores si se realizan actividades que requieran una interpretación dado el impacto que puede tener en el resultado o en las decisiones que de éste se derivan.

#### **4.1. Control de calidad interno**

En sus inicios, los planes de control de calidad interno estaban enfocados únicamente a la detección de errores y se basaban en criterios estadísticos. Los límites de tolerancia para el resultado del control se establecían como la media  $\pm 2$  desviaciones estándar (DE), donde debían encontrarse el 95 % de los datos si estos seguían una distribución normal. Estos estadísticos

podían obtenerse a partir de los datos del material de control declarados por la empresa de diagnóstico in vitro en cuestión, o bien podían ser estimados en el propio laboratorio.

La función del modelo estadístico del control de calidad interno es la aceptación o rechazo de las series analíticas. Una serie analítica se define como el intervalo (periodo de tiempo o número de mediciones) dentro del cual se espera que la veracidad y la precisión del procedimiento de medida sean estables (28). Así, la estrategia de control de calidad interno debe estar diseñada para detectar la presencia de errores aleatorios o sistemáticos que comprometan la estabilidad del procedimiento de medida.

A continuación, se describen las definiciones de los términos metrológicos utilizados. Para ello, se han considerado diferentes fuentes bibliográficas con la finalidad de incluir las definiciones que mejor se adaptan al ámbito del laboratorio clínico.

- Veracidad: grado de concordancia existente entre el valor medio de un número infinito de resultados y un valor de referencia de una magnitud. La veracidad está inversamente relacionada con el error sistemático (29).
- Error sistemático: componente del error de medida que, en mediciones repetidas, se mantiene constante o varía de forma previsible (29). Su estimación es el sesgo. Se calcula como la diferencia entre la media de un número definido de medidas de una misma magnitud en una muestra y el valor verdadero (28).
- Precisión: grado de concordancia entre los resultados independientes de mediciones obtenidas en condiciones estipuladas (30). La precisión de medida se expresa mediante medidas de imprecisión tales como la desviación estándar, la varianza o el coeficiente de variación (29).
- Error aleatorio: componente del error de medida que, en mediciones repetidas, varía de manera impredecible. (29). Su estimación es la imprecisión. Se calcula como la

diferencia entre un resultado individual de una magnitud y la media de un número determinado de mediciones de la misma, en condiciones de repetibilidad (28).

- Exactitud: grado de concordancia entre un valor medido y un valor verdadero de la magnitud. Un resultado es más exacto si genera un error de medida menor (29).
- Error de medida: diferencia entre el valor medido y el valor de referencia de una magnitud (29). Su estimación es el error total, siendo este el efecto de la combinación del componente de error aleatorio y sistemático (31).

Las gráficas de Levey-Jennings han sido y siguen siendo las más utilizadas para visualizar de forma rápida la aceptación o rechazo de los resultados del control, así como para realizar el seguimiento de los resultados a lo largo del tiempo (32). Las gráficas se obtienen al representar los resultados de las mediciones del material de control ordenados secuencialmente en el tiempo. En la Figura 2 se muestra una gráfica de Levey-Jennings, donde puede observarse la dispersión de las medidas de los resultados del material de control respecto a la media establecida, siendo  $s$  su desviación estándar.

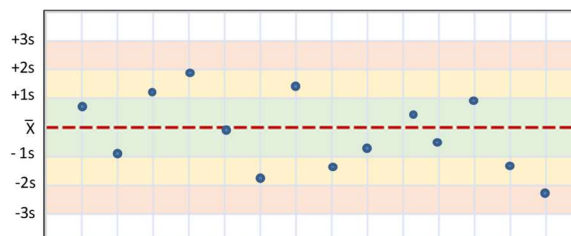


Figura 2. Gráfica de Levey-Jennings.

En los años 80 *James Westgard* introdujo el concepto de regla de control para referirse al criterio de decisión para interpretar los resultados de las mediciones de los materiales de control y para decidir si la serie analítica debía ser aceptada o rechazada. Una regla de control se simboliza como  $A_L$ , donde  $A$  puede indicar la abreviatura de un estadístico particular o bien el número de mediciones del material de control; y  $L$  es el límite de aceptación de los resultados. Así, por ejemplo, la regla de control que implica el rechazo del resultado del control si no está incluido

en el intervalo proporcionado por la media  $\pm 2$  DE es la regla  $1_{2s}$ . Además, reveló que podían calcularse dos características que determinaban el desempeño del control de calidad, que corresponden a la probabilidad de detección de error (Pde) y la probabilidad de falso rechazo (Pfr) (31):

- La Pde indica la probabilidad de detectar un error en una serie analítica cuando está presente, adicional a la imprecisión inherente al procedimiento de medida. Idealmente la Pde debería ser 1, lo que indicaría que se detectan el 100 % de los errores cuando ocurren. En la práctica, generalmente se acepta que la Pde sea superior o igual a 0,9.
- La Pfr indica la probabilidad de rechazar una serie analítica cuando no hay errores adicionales a la imprecisión inherente al procedimiento de medida. Idealmente la Pfr debería ser 0, lo que indicaría que no se rechaza ninguna serie analítica en falso. En la práctica, generalmente se acepta que la Pfr sea inferior o igual a 0,05.

Ambas características se representan en lo que se conoce como gráficas de función de poder. En estas gráficas se muestra la Pde y la Pfr de diferentes reglas de control, lo que permite compararlas y seleccionar la que se considere más apropiada (33).

Fue entonces cuando se puso de manifiesto que, utilizando la regla de control  $1_{2s}$ , la Pfr era muy elevada, concretamente del 4,5 % cuando se analizaba un único nivel de control, del 9 % si se analizaban dos y del 14 % cuando se analizaban tres (34). Esto implicaba rechazar un elevado número de resultados de control que en realidad eran correctos y que además estaban indicando una falsa inestabilidad del procedimiento de medida. Esto evidenció la importancia de la selección de las reglas de control. A partir de ese momento, se aconsejó utilizar la regla de control  $1_{2s}$  únicamente como regla de alerta, y empezaron a utilizarse reglas de control alternativas, o una combinación secuencial de éstas (reglas múltiples), que permitiesen minimizar la Pfr sin limitar la Pde de errores (35). Las reglas de control más utilizadas se describen a continuación:

- $1_{3s}$ : la serie analítica se rechaza cuando un resultado de control excede la media  $\pm 3$  DS. Suele indicar la presencia de un error aleatorio.
- $2_{2s}$ : la serie analítica se rechaza cuando dos resultados de control consecutivos se encuentran por encima de la media + 2 DE o por debajo de la media - 2 DE. Suele indicar la presencia de un error sistemático.
- $R_{4s}$ : la serie analítica se rechaza cuando la diferencia entre dos resultados de control consecutivos supera las 4 DE. Suele indicar la presencia de un error aleatorio.
- $4_{1s}$ : la serie analítica se rechaza cuando cuatro resultados de control consecutivos se encuentran por encima de la media + 1 DE o por debajo de la media - 1 DE. Suele indicar la presencia de un error sistemático.
- $10_x$ : la serie analítica se rechaza cuando diez resultados de control consecutivos se encuentran por encima o por debajo de la media. Suelen indicar la presencia de un error sistemático.

En los últimos años, la aplicación de la métrica sigma, una herramienta que tradicionalmente se utilizaba en el ámbito de la industria, ha sido de gran utilidad en los laboratorios clínicos. La mejora del rendimiento de los procesos mediante el modelo Seis Sigma se basa en reducir la variabilidad hasta conseguir que dentro de la especificación de calidad establecida como error máximo permitido puedan contenerse seis desviaciones estándar. En el laboratorio clínico y concretamente en la fase analítica, el valor de sigma puede calcularse a partir del error total máximo permitido para la magnitud en cuestión, el sesgo y la imprecisión del procedimiento de medida. Para realizar su cálculo, por tanto, es fundamental que los datos de imprecisión y sesgo se hayan obtenido cuando el procedimiento de medida se encuentre en un estado estable, puesto que de ello va a depender el valor de sigma.

Se calcula mediante la fórmula:

$$\sigma = (ETp - ES) / CV$$

siendo:  $\sigma$ : sigma

ETp: error máximo permitido a partir del cual los resultados se consideran defectos o errores

ES: error sistemático que muestra la desviación de la distribución de los resultados respecto al valor verdadero

CV: coeficiente de variación que muestra la amplitud de distribución de los resultados

La fórmula se describe gráficamente de la siguiente manera (36):

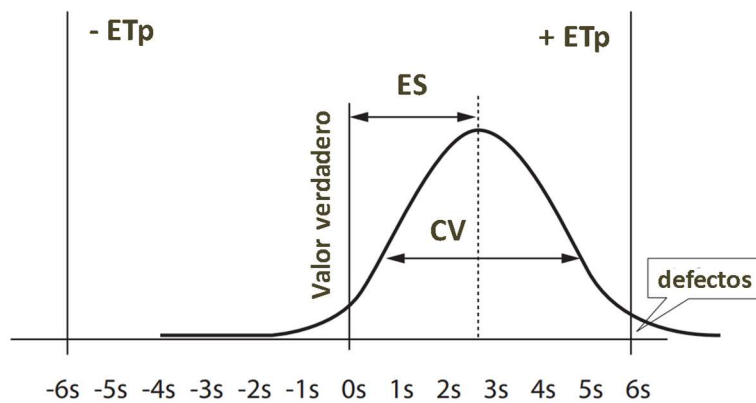


Figura 3. Descripción gráfica de la fórmula para el cálculo del valor sigma. Adaptada de (36).

La combinación del sesgo y la imprecisión permite estimar por tanto la distribución de los resultados y, con ello, cuántos defectos (errores) se producen por millón de resultados; lo que se conoce por el término defectos por millón de oportunidades (DPMO). En caso de conocer el número de DPMO, puede realizarse una estimación del valor sigma a partir de una sencilla transformación (37) o bien seleccionando el número de DPMO más próximo disponible en las tablas de conversión (36). En la Tabla 4 se muestra la conversión del número de DPMO a la métrica sigma.



Tabla 4. Tabla de conversión del número de defectos por millón de oportunidades (DPMO) a la métrica sigma. Adaptada de (36).

DPMO	Sigma
3,4	6
32	5,5
233	5
1.350	4,5
6.210	4
22.750	3,5
66.807	3
158.655	2,5
308.538	2
500.000	1,5
691.462	1
841.345	0,5
933.193	0

Cuanto mayor es el valor de sigma, mejor es el rendimiento analítico del procedimiento de medida, puesto que el número de resultados analíticos cuyo error de medida supera el error máximo permitido (o número de DPMO) es menor. Así, en un procedimiento de medida con un valor sigma de 6 se generan menos de 4 errores por cada millón de resultados.

Según el valor de sigma se han descrito diferentes niveles que corresponden a un determinado rendimiento del procedimiento de medida, tal y como se muestra en la Tabla 5. (38).

Tabla 5. Descripción del rendimiento de un procedimiento de medida según su nivel sigma.

Nivel sigma	Rendimiento del procedimiento de medida
$\sigma < 2$	Inaceptable: el procedimiento de medida no es válido
$\geq 2 \sigma < 3$	Bajo: el procedimiento de medida precisa una mejora de la estrategia de control de calidad
$\geq 3 \sigma < 4$	Moderado: el procedimiento de medida precisa de una estrategia de control de calidad con más de una serie analítica y múltiples medidas de control de calidad por serie.
$\geq 4 \sigma < 5$	Bueno: la utilidad clínica del procedimiento de medida está asegurada si se utilizan reglas múltiples
$\geq 5 \sigma < 6$	Muy bueno: la utilidad clínica del procedimiento de medida está asegurada si se utilizan reglas simples.
$\sigma \geq 6$	Excelente

Una vez obtenido el valor de sigma, es posible planificar la estrategia más adecuada para el control de calidad interno, siendo ésta una combinación de la selección de la regla de control, el número de medidas de control de calidad por serie analítica, y el tamaño de la serie analítica (15,33,39).

Siguiendo esta metodología es posible optimizar el rendimiento del control de calidad. Sin embargo, en la práctica diaria puede resultar de difícil aplicación puesto que supone individualizar el plan de control de la calidad para cada magnitud en cada analizador en la que se mide, lo que implica una inversión de tiempo o de dinero que no siempre es posible (33). Por ello, a la práctica se intenta encontrar el equilibrio entre la capacidad de detección de errores deseada, la complejidad de la regla de control a aplicar y el número de materiales de control a analizar y su coste, buscando que la gestión del plan de control de la calidad sea manejable sobre todo cuando se trata de plataformas analíticas en las que se miden multitud de magnitudes como sucede en los laboratorios clínicos.

Es una práctica habitual que los laboratorios realicen un seguimiento a largo plazo del rendimiento analítico de los procedimientos de medida a partir de los resultados acumulados

del control de calidad interno. Generalmente, éste se basa en la estimación, con una periodicidad mensual, del coeficiente de variación para conocer la imprecisión y del error sistemático para conocer el sesgo. Sin embargo, los materiales de control interno no permiten evaluar la veracidad de los resultados debido a la falta de conmutabilidad entre la matriz de los materiales de control interno y las muestras biológicas; y a que su valor diana no se asigna mediante procedimientos de medida de referencia, sino que se estima a partir de la medición repetida del material de control por un procedimiento de medida seleccionado por el fabricante o por el disponible en el laboratorio. Lo mismo ocurre en el caso de los materiales de control fabricados por una tercera parte (por un fabricante de diagnóstico in vitro distinto al proveedor del instrumento, calibrador o reactivo). En este caso, el proveedor suele aportar información adicional sobre los resultados de otros laboratorios que utilizan el mismo analizador y lote de material de control; lo que permite evaluar indirectamente la desviación de un laboratorio con respecto a los demás, pero tampoco son válidos para estimar el sesgo del procedimiento de medida (40).

Recientemente, se ha demostrado que una buena estrategia puede ser complementar el control de calidad interno con un modelo basado en la aplicación de algoritmos a partir de los resultados de los pacientes. Implementar esta herramienta permitiría superar la limitación de la conmutabilidad de los materiales de control; sin embargo, del mismo modo que en las situaciones anteriores, únicamente resultaría útil para detectar un cambio en el sesgo del procedimiento de medida y no sería posible estimarlo debido a que el valor de una magnitud en la muestra de un paciente es desconocido. En cualquier caso, debido a la complejidad de la aplicación de este tipo de algoritmos, su implantación depende del avance de las tecnologías de la información y de su disponibilidad en los laboratorios (41).

A pesar de la evolución teórica de las herramientas de control de calidad interno a lo largo de los años, la situación real en los laboratorios clínicos demuestra que algunos de los aspectos comentados todavía siguen sin aplicarse en la actualidad (42). Algunos de estos aspectos son:

- No está completamente extendido el uso de materiales de control proporcionados por una tercera parte a pesar de que estén disponibles.
- Todavía se utilizan límites de control para la aceptación o rechazo de los resultados basados en criterios únicamente estadísticos.
- El tamaño de las series analíticas no suele definirse acorde a la probabilidad de entregar resultados incorrectos.
- La regla de control más empleada sigue siendo todavía la  $1_{2s}$ .

#### **4.2. Control de calidad externo**

Desde sus inicios, los Programas de Evaluación Externa de la Calidad (PEEC) han proporcionado al laboratorio clínico una información muy valiosa para la evaluación de su rendimiento analítico. Con el paso del tiempo, los organizadores de PEEC han ido incorporando diferentes metodologías para el análisis estadístico de los resultados de los participantes y en consecuencia la información contenida en el informe de resultados ha ido evolucionando.

En las décadas de los 70 y 80, la evaluación de los resultados de los participantes en PEEC se basaba en la aplicación de criterios estadísticos. Generalmente, el criterio utilizado para la aceptación o rechazo de los resultados era el índice de desviación estándar (IDS) obtenido a partir de la dispersión de los datos y de la comparación de los resultados emitidos por un laboratorio respecto a un valor consenso estimado a partir de los resultados de todos los laboratorios participantes (43). Como norma general, un resultado cuyo IDS era inferior o igual a 2 se consideraba satisfactorio.

Actualmente, suelen incluirse los estadísticos (media/mediana y desviación estándar) teniendo en cuenta el total de los participantes; y, si el número de participantes lo permite, por grupos según el procedimiento de medida, el fabricante y/o el analizador. Esto permite a los laboratorios, además de evaluar su rendimiento analítico, compararlo con el de otros participantes que utilizan el mismo procedimiento de medida y/o analizador.

A partir de los resultados de la participación en los PEECs, los laboratorios suelen realizar un seguimiento del rendimiento analítico de forma periódica acorde a la frecuencia de los envíos de cada programa.

En el ámbito del laboratorio clínico, los PEECs se clasifican en diferentes categorías en función de su capacidad para evaluar el rendimiento analítico (44). La categoría asignada depende de tres características: la conmutabilidad del material de control, la metodología empleada para la asignación del valor diana y la realización de mediciones replicadas de los materiales de control.

La Tabla 6 muestra la clasificación de los PEEC en función de estas características.

Tabla 6. Clasificación por categorías de los Programas de Evaluación Externa de la Calidad en función de sus características.

<b>Categoría</b>	<b>Conmutabilidad</b>	<b>Asignación del valor diana mediante un procedimiento o material de referencia</b>	<b>Solicitud de mediciones replicadas</b>
<b>1</b>	Si	Si	Si
<b>2</b>	Si	Si	No
<b>3</b>	Si	No	Si
<b>4</b>	Si	No	No
<b>5</b>	No	No	Si
<b>6</b>	No	No	No

La mayoría de PEECs solicitan una única medida del material de control, lo que permite realizar únicamente una estimación del error de medida (calculado como la diferencia porcentual entre el resultado emitido y el esperado). En el caso de participar en un PEEC de categoría 1 o 2, debido a que en estos programas se utilizan materiales de control conmutables y los valores diana están

asignados por laboratorios de referencia (45) que utilizan procedimientos de medida de referencia reconocidos (46), el informe de resultados incorpora información adicional sobre el sesgo de los diferentes procedimientos de medida. Los PEECs de categoría 1, 3 y 5 solicitan la realización de mediciones replicadas del material de control, por lo que proporcionan una estimación de la reproducibilidad de cada laboratorio (44).

Ya en la década de los 90, se realizaron algunas propuestas para la armonización de los PEEC. Estas comprendían, entre otras, la asignación del valor diana del material de control a partir de un procedimiento de medida de referencia realizado en un laboratorio de referencia (47), la selección del mejor material de control posible y la evaluación del rendimiento de los participantes mediante especificaciones de calidad (48,49).

Por otro lado, las actividades correspondientes a las fases extra-analíticas no empezaron a considerarse en los PEEC hasta prácticamente el inicio del siglo XXI, donde empezaron a evaluarse aspectos como la idoneidad de las pruebas analíticas solicitadas o la adición de comentarios interpretativos en los informes de resultados como ayuda a la orientación diagnóstica (50, 51).

A pesar de los avances conseguidos con el paso de los años, algunos aspectos siguen siendo un reto en la actualidad (44). Las limitaciones que todavía presentan los PEECs se resumen a continuación:

- La mayoría de PEECs no utilizan materiales de control conmutables por las dificultades inherentes a su preparación, a la preservación de su estabilidad, a la logística de su distribución a los laboratorios y su elevado coste. Si el laboratorio participa en un programa que no utiliza materiales de control conmutables debe tener en cuenta que puede haber diferencias debidas al efecto matriz y que, por tanto, no deben atribuirse a posibles diferencias entre procedimientos de medida.

- Son pocos los PEECs que proporcionan un valor diana obtenido a partir material de referencia certificado y/o de un procedimiento de medida de referencia. Esto es debido fundamentalmente a que conlleva un coste mayor y a que, además, hay magnitudes para las que no están disponibles. Si el laboratorio participa en un programa en el que el valor diana no se ha asignado mediante un procedimiento de medida de referencia, no puede verificar la veracidad de los resultados. Únicamente podrá evaluar si el rendimiento analítico del procedimiento de medida utilizado es acorde a las especificaciones técnicas declaradas por el fabricante y al rendimiento analítico de otros laboratorios que utilicen la misma tecnología.
- Existe una gran diversidad a la hora de indicar especificaciones de calidad en el informe de resultados (52). Los PEECs internacionales que tienen finalidad regulatoria establecen como especificación de calidad el requisito mínimo que está fijado por ley. Esto es debido a que son programas que están dirigidos a identificar laboratorios con un rendimiento analítico inaceptable, de forma que, si el laboratorio no es capaz de cumplir con los requisitos mínimos, el gobierno puede llevar a cabo una inspección del laboratorio y sancionarle si así lo considera. En el caso de los PEECs que no tienen finalidad regulatoria, por lo general, las especificaciones de calidad suelen ser más exigentes. Sin embargo, en este tipo de programas los organizadores utilizan diferentes criterios para la asignación de especificaciones de calidad.
- Es fundamental que los PEECs disponibles en el mercado ofrezcan información actualizada del rendimiento analítico de la tecnología utilizada por los participantes. Esto adquiere especial relevancia para aquellas magnitudes en las que, por ejemplo, debido al tipo de población diana o al tipo de muestra, no sea posible realizar estudios para disponer de especificaciones de calidad basadas en situaciones clínicas concretas o en la variabilidad biológica.

- A pesar de que hay PEECs dirigidos exclusivamente al aseguramiento de la calidad de las fases extra-analíticas (53), todavía no se han empezado a desarrollar PEECs en los que se incluyan actividades de las fases pre-pre-analítica o post-post-analítica y se precise la colaboración de personal externo a los laboratorios (54).
- No hay PEECs disponibles para algunas magnitudes que se realizan sólo en unos pocos laboratorios más especializados. Sin embargo, mediante la colaboración entre los organizadores podría alcanzarse el número de participantes necesario para comercializar programas. Con ello, se conseguiría dar cobertura a un mayor número de magnitudes.

#### **4.3. Especificaciones de calidad analítica**

En los años 90 surge el concepto de especificaciones de calidad analítica, con el que cambia el paradigma del control de calidad, pasando a ser su finalidad el detectar aquellos cambios en el rendimiento analítico que son clínicamente significativos y que, por tanto, pueden impactar en la decisión clínica que se lleve a cabo en el paciente.

En 1999, se realizó una conferencia en Estocolmo en la que se definió una jerarquía de 5 modelos para el establecimiento de especificaciones de calidad analítica en el laboratorio clínico (55). Los 5 modelos, ordenados de mayor a menor impacto sobre las decisiones clínicas, son los siguientes:

- Modelo 1: Especificaciones de calidad basadas en el efecto del rendimiento analítico en situaciones clínicas concretas.
- Modelo 2. Especificaciones de calidad basadas en la evaluación del efecto del rendimiento analítico en las decisiones clínicas generales:
  - a. Datos basados en componentes de la variabilidad biológica.
  - b. Datos basados en la opinión de los profesionales clínicos.



- Modelo 3. Recomendaciones profesionales publicadas:
  - a. Procedentes de guías de grupos nacionales o internacionales.
  - b. Procedentes de guías de expertos individuales o grupos institucionales.
- Modelo 4. Especificaciones basadas en:
  - a. Organismos reguladores.
  - b. Organizadores de programas de intercomparación.
- Modelo 5. Datos basados en el estado del arte:
  - a. Obtenidos a partir de PEECs.
  - b. Obtenidos a partir de metodologías individuales publicadas.

En noviembre de 2014, 15 años después, se celebró en Milán la primera conferencia estratégica de la EFLM sobre “Definición de objetivos de rendimiento analítico 15 años después de la conferencia de Estocolmo sobre especificaciones de calidad en el laboratorio clínico” (56). En esta conferencia se realizó una revisión de la jerarquía propuesta en el año 1999, quedando simplificada a 3 modelos:

- Modelo 1: Especificaciones de calidad basadas en el efecto del rendimiento analítico en los resultados clínicos obtenidos en situaciones clínicas concretas.
- Modelo 2: Especificaciones de calidad basadas en la variabilidad biológica.
- Modelo 3: Especificaciones de calidad basadas en el estado del arte.

El primer modelo emplea diferentes tipos de estudios para establecer las especificaciones de calidad. Se trata de estudios que investigan el impacto del rendimiento analítico bien sobre el diagnóstico clínico (estudios directos) o bien sobre las pruebas de clasificación o decisión clínica afectando a la probabilidad diagnóstica (estudios indirectos). La principal limitación de este modelo es que hasta el momento se han establecido especificaciones de calidad para muy pocas patologías y magnitudes biológicas (57). Sin embargo, se considera el “gold standard” en la definición de especificaciones de calidad por su aplicación en base a las necesidades clínicas.

El segundo modelo está basado en la obtención de especificaciones de calidad a partir de datos de variabilidad biológica. Estas se establecen a partir de los componentes intraindividual (variabilidad de una magnitud biológica en cada individuo) e interindividual (variabilidad de una magnitud biológica entre diferentes individuos). Estas especificaciones se basan en el siguiente concepto: mantener la variabilidad analítica por debajo de la fluctuación biológica de la magnitud en cuestión. Basándose en este modelo, el laboratorio se asegura de que cuando se produce un cambio significativo en un resultado, éste se debe a un cambio en la situación clínica del paciente y no a un error de medida. Estas especificaciones están definidas para más de 300 magnitudes analizadas habitualmente en los laboratorios clínicos; y, para cada magnitud, existen tres niveles (mínimo, deseable y óptimo), pudiendo escogerse un nivel u otro en función del rendimiento que se necesite alcanzar según el uso previsto. Las principales limitaciones de este modelo son dos:

- La falta de estandarización de los estudios de variabilidad biológica, lo que puede dar lugar a la obtención de estimados inadecuados.
- La tecnología más avanzada disponible en los laboratorios clínicos en ocasiones no es capaz de alcanzar ni siquiera el nivel mínimo de las especificaciones de calidad.

El tercer modelo para establecer especificaciones de calidad es el que se basa en el estado del arte. Se obtienen a partir de los datos de la participación en PEECs por lo que dependen del rendimiento de la tecnología disponible en el momento de su obtención. Se calculan a partir de la distribución del error de medida (o error total) de los participantes. Los laboratorios son los encargados de establecer un punto de corte que debería ser más o menos exigente dependiendo del uso previsto de los resultados analíticos. Entre las principales ventajas de este modelo se encuentra la facilidad de obtención de datos para el establecimiento de especificaciones de calidad y el hecho de que se puede inferir en qué punto se encuentra la tecnología actual y

establecer unas especificaciones adaptadas al rendimiento analítico. La mayor desventaja de este modelo es su falta de conexión con criterios clínicos o biológicos.

En diversos países se utiliza el modelo 3 para establecer el nivel mínimo de calidad que deben cumplir las pruebas analíticas. En España, en el año 2006 tres sociedades científicas nacionales crearon un Comité de Expertos Interdisciplinar de Especificaciones de Calidad y emitieron las primeras especificaciones mínimas de consenso para algunas magnitudes de bioquímica (58). Dos años más tarde, se incorporó al Comité una cuarta sociedad, y se proporcionaron especificaciones mínimas de consenso para magnitudes de hematimetría y hemostasia, así como para nuevas magnitudes bioquímicas (59). El comité las estableció utilizando un algoritmo de decisión en el que se comparaba la especificación obtenida con la fijada en dos países en las que estas especificaciones mínimas son un requisito de obligado cumplimiento, concretamente los regulados por Richtlinie en Alemania y por Clinical Laboratory Improvement Amendments (CLIA) en Estados Unidos (60). Por el hecho de depender de forma directa de la tecnología disponible en el momento de su estimación, se han publicado actualizaciones de estas especificaciones a lo largo de los años (61,62). Aunque en España no son de obligado cumplimiento, su finalidad es servir como herramienta de evaluación para entidades de acreditación o certificación, así como para que aquellos laboratorios que no puedan darles cumplimiento realicen acciones correctivas al respecto. Por contra, no deberían ser utilizadas para fijar objetivos de calidad analítica.

#### ***4.4. Selección de especificaciones de calidad analítica***

La selección de especificaciones de calidad analítica para las diferentes magnitudes que se analizan en el laboratorio clínico es un paso fundamental en el establecimiento del plan de control de calidad, puesto que va a condicionar muchas de las decisiones que se tomen con respecto al rendimiento analítico de los procedimientos de medida. Dada la importancia que

tiene la selección de las especificaciones de calidad, la EFLM formó un grupo de trabajo en el año 2017 que se encargó de establecer un algoritmo para la asignación de los diferentes modelos de especificaciones de calidad (63). El diagrama de flujo que pauta la selección de las especificaciones de calidad se muestra en la Figura 4.

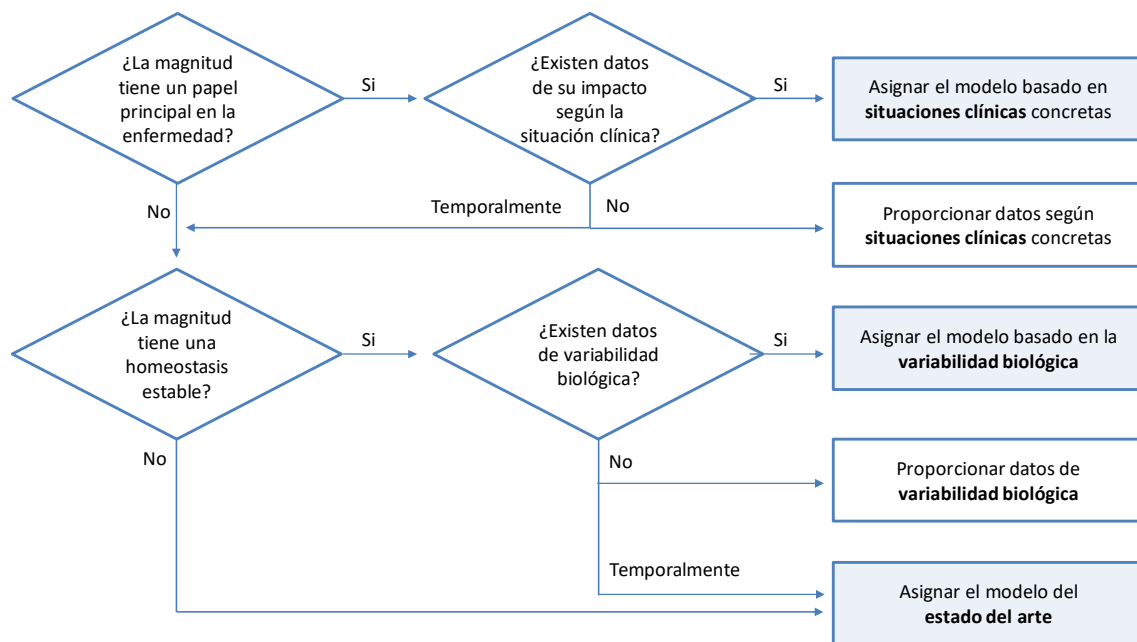


Figura 4. Diagrama de flujo para la asignación de los modelos de especificaciones de calidad analítica. Adaptada de (63).

Tal y como puede observarse, para las magnitudes que son clave para el diagnóstico de una patología, se recomienda utilizar las especificaciones de calidad basadas en situaciones clínicas concretas (modelo 1).

Si el organismo humano mantiene una estricta homeostasis de la magnitud en cuestión (muy baja variabilidad biológica), se recomienda utilizar las especificaciones de calidad basadas en la variabilidad biológica (modelo 2). Si la homeostasis no es estricta, se recomienda utilizar las especificaciones de calidad basadas en el estado del arte (modelo 3). Es el caso por ejemplo de las magnitudes presentes en orina.

Por último, si no existen especificaciones de calidad basadas en la variabilidad biológica (modelo 2), se recomienda utilizar las basadas en el estado del arte (modelo 3).

Algunos de los aspectos del plan de control de la calidad en los que la selección de especificaciones de calidad impacta de forma directa son los siguientes:

- La selección de la especificación de calidad de error total para establecer la estrategia de control interno más adecuada ya que, tal y como se ha mostrado anteriormente, en la ecuación para el cálculo de sigma, el valor sigma es directamente proporcional al error total máximo permitido. Por lo tanto, la optimización del rendimiento del control de calidad depende siempre de la idoneidad de la especificación de calidad establecida.
- La selección de las especificaciones de calidad de imprecisión y sesgo va a condicionar el resultado del seguimiento a largo plazo del control de calidad interno. Esto es debido a que, cuando se exceden las especificaciones de calidad seleccionadas, debe realizarse un Análisis de Causa Raíz (Root Cause Analysis) en búsqueda de las posibles causas de la desviación, lo que conlleva tiempo y dedicación. Por ejemplo, el análisis de las causas debería incluir la revisión de los datos diarios de control de calidad, la verificación de la correcta aplicación de las reglas de control, la comprobación del cumplimiento de los mantenimientos preventivos que estaban programados, entre otros.
- Las conclusiones sobre el rendimiento analítico derivadas de la evaluación de los informes de los PEECs también pueden variar en función de las especificaciones de calidad que se hayan establecido en cada caso. En algunas ocasiones el organizador del programa establece o recomienda, bajo su criterio, las especificaciones de calidad que no deberían exceder los participantes. En otros programas, dada la diversidad de opciones, los organizadores indican en el informe de resultados las diferentes especificaciones de calidad existentes para cada magnitud, algo que incluso puede llegar a confundir a los participantes. En tal caso, el profesional del laboratorio clínico debe

decidir cuál considera más apropiada teniendo en cuenta, por un lado, el impacto que puede tener en el paciente y, por otro lado, si la tecnología disponible permite alcanzarla. Igual que para el control interno, cuando se exceden las especificaciones de calidad seleccionadas debe iniciarse una investigación de las causas para llevar a cabo las acciones correctivas pertinentes. Esta investigación requiere un esfuerzo considerable dada la larga lista de problemas potenciales que deben descartarse: errores de dilución, deterioro del material de control durante el almacenamiento, un efecto arrastre de la muestra previa, un error de cálculo, o un simple error de transcripción del resultado al suministrar los datos al organizador (64).

En definitiva, la selección de especificaciones de calidad más estrictas de lo que la tecnología permite alcanzar puede dar lugar, por un lado, a la implementación de una estrategia de control compleja, cara y de difícil gestión; y por otro lado a la realización de investigaciones innecesarias con los recursos que esto implica. Conseguir que el rendimiento analítico de los procedimientos de medida cumpla con las especificaciones de calidad basadas en el impacto clínico o en la variabilidad biológica debe ser el objetivo de las empresas de diagnóstico in vitro. Este no debe ser el objetivo de los laboratorios, que sí deben esforzarse para que su rendimiento analítico sea el mejor posible de acuerdo a la tecnología disponible y al uso previsto de la magnitud.

Y a la inversa, el establecimiento de especificaciones de calidad demasiado permisivas puede dar lugar una falsa seguridad en la emisión de resultados y va en detrimento de la detección de errores y, por tanto, de la seguridad del paciente.

Por último, hay que destacar que, además de las implicaciones comentadas con respecto a la estrategia de control de calidad y a la evaluación o seguimiento del plan control interno y externo, las especificaciones de calidad se utilizan también como criterio de evaluación de muchos de los estudios que se realizan de forma más o menos eventual en los laboratorios.

Ejemplo de algunos de ellos son los estudios de validación o verificación de los procedimientos de medida o los estudios de intercambiabilidad de resultados entre equipos duplicados (65).

## **5. Aseguramiento de la calidad según la norma ISO 15189**

Uno de los objetivos principales de la implantación de sistemas de gestión de la calidad en los laboratorios clínicos es el de asegurar la calidad de los resultados dado su impacto en la seguridad del paciente. A este respecto la aparición de la norma UNE-EN ISO 15189 en el año 2003 (66) fue de gran ayuda, ya que en ella se establecieron de forma clara los requisitos para la calidad y la competencia técnica en los laboratorios clínicos.

Un laboratorio clínico que da cumplimiento a la norma ISO 15189 (67) ofrece la garantía de que emite resultados fiables porque ha demostrado que las actividades técnicas las realiza personal competente, que utiliza procedimientos analíticos adecuados para el uso previsto y que dispone de un plan de control de calidad que abarca el proceso analítico en su totalidad.

Concretamente, en el capítulo 7.3.7 de la norma se indican los requisitos que el laboratorio debe tener en cuenta al establecer el plan de control de calidad para poder asegurar la validez de los resultados. Este capítulo está estructurado en tres bloques:

- A) Control de calidad interno (requisito 7.3.7.2. *Control interno de la calidad*)
- B) Control de calidad externo (requisito 7.3.7.3. *Evaluación externa de la calidad*)
- C) Intercambiabilidad de los resultados de los análisis (requisito 7.3.7.4. *Comparabilidad de los resultados del análisis*)

A continuación, se comentan los aspectos más relevantes de cada uno de estos requisitos.

### **A) Control de calidad interno**

A la hora de planificar una estrategia de control de calidad interno, la norma indica que el laboratorio debe tener en cuenta las siguientes consideraciones; entre otras:

- La selección de los materiales de control interno debe realizarse de forma que sean lo más parecidos posible a las muestras de los pacientes. Además, siempre que sea posible sus concentraciones deben ser cercanas a los niveles de decisión clínica. En el caso que se considere que los materiales de control disponibles en el mercado no son apropiados, deberán considerarse otras alternativas de control de calidad utilizando muestras de pacientes como puede ser el análisis de tendencias de sus resultados.
- El laboratorio debe establecer las reglas de control para la aceptación o rechazo de resultados, así como la frecuencia de medida de los materiales de control. De la definición del tamaño de la serie analítica va a depender, en gran parte, la probabilidad de emisión de resultados incorrectos. Para establecer la frecuencia, deberá considerarse:
  - La estabilidad del procedimiento de medida: cuanto menor sea la estabilidad de un procedimiento de medida, con mayor frecuencia será necesaria la medición de los materiales de control para poder asegurar la fiabilidad de los resultados de los pacientes.
  - Riesgo de daño al paciente si se entrega un resultado erróneo: riesgo de que un resultado erróneo impacte en la seguridad del paciente antes de que el resultado pueda identificarse como incorrecto por la siguiente medida de control de calidad y éste pueda corregirse. A este respecto, por ejemplo, para aquellas magnitudes cuyos resultados supongan una actuación más o menos inmediata por parte del profesional clínico, podría aumentarse la frecuencia de medida de los materiales de control con la finalidad de prevenir la entrega de resultados incorrectos, por el



hecho de no disponer de un margen de tiempo que permita corregir los errores. Esto podría ser menos exigente para determinadas magnitudes que entrañen un riesgo bajo para el paciente (por ejemplo, el colesterol) o en determinadas situaciones (por ejemplo, algunas pruebas solicitadas de rutina cuyos resultados analíticos no van a ser consultados por los profesionales clínicos hasta el día de la próxima visita médica del paciente).

- El laboratorio debe disponer de un procedimiento de control interno de calidad para realizar el seguimiento de la validez de los resultados de forma que se asegure que éstos disponen de la calidad necesaria para el uso clínico previsto. Esto implica que el laboratorio:
  - Debe definir las especificaciones de calidad a aplicar en cada caso como criterio de aceptación.
  - Debe ser capaz de detectar variaciones en el rendimiento analítico debidas a un cambio de lote de reactivo y/o calibrador.
  - Debe considerar la opción de utilizar materiales de control fabricados por una tercera parte independiente, ya sea en uso exclusivo o adicional a los materiales de control suministrados por el proveedor del reactivo o del instrumento. Esto permite detectar algunos errores sistemáticos que podrían pasar desapercibidos si los materiales de control han sido fabricados con una formulación similar al material de calibración del procedimiento de medida.
  - Debe demostrar el seguimiento de las interpretaciones realizadas en actividades del proceso analítico que sean subjetivas mediante la realización periódica de estudios entre los diferentes observadores. La realización de este tipo de revisiones permite, además, demostrar la competencia técnica de personal.

- El laboratorio debe guardar registro de los resultados de control de calidad interno, así como de las revisiones realizadas periódicamente para examinar los datos acumulados de control de calidad con la finalidad de detectar posibles tendencias que puedan sugerir problemas del rendimiento analítico.
- El laboratorio debe tener establecida la actuación con respecto a las muestras de los pacientes cuando se detecta un fallo del control de calidad. En aquellos casos en que se confirme que la vulneración de la regla de control indica un problema real del procedimiento de medida (y no, por ejemplo, un problema de estabilidad del material de control), el laboratorio debe disponer de una estrategia que impida la comunicación de aquellos resultados que se hayan obtenido después del fallo del control de calidad. Por otro lado, una vez se hayan realizado las acciones correctivas pertinentes y se haya verificado el cumplimiento de las especificaciones requeridas, deben analizarse de nuevo las muestras de los pacientes que se midieron después del último resultado de control de calidad correcto.

## **B) Control de calidad externo**

Uno de los requisitos fundamentales de la norma ISO 15189 es la participación en PEEC para las magnitudes o estudios que se encuentren dentro del alcance de la acreditación del laboratorio. A partir del resultado obtenido en la participación, es posible demostrar el desempeño en la realización de los procedimientos analíticos. Algunas de las consideraciones indicadas en la norma que deben tenerse en cuenta son las siguientes:

- Siempre que exista la posibilidad, el laboratorio debe optar por participar en programas de intercomparación acreditados según la norma ISO/IEC 17043 (68), en los que el material de control sea lo más parecido posible a las muestras de los pacientes y, si es posible, que contemplen el proceso analítico en su totalidad. Además, al realizar la selección deberá considerarse la metodología empleada para la asignación del valor

diana. A este respecto, se especifica que en aquellos PEECs en los que no se proporcione el valor verdadero del material de control, es posible realizar la interpretación de los resultados a partir del valor de consenso obtenido a partir de los resultados de todos los participantes o, si es posible, del grupo que utiliza el mismo procedimiento de medida.

- Debe tenerse en cuenta que el procedimiento para analizar las muestras de los estudios de intercomparación debe ser, siempre que sea posible, el mismo que el que se sigue para analizar las muestras de los pacientes.
- Cuando para una magnitud / estudio no exista ningún programa de intercomparación disponible, se requiere emplear alguna estrategia alternativa que permita asegurar la calidad de los resultados. Entre estas, se acepta la realización de intercambios de muestras con otros laboratorios, así como el análisis de materiales de referencia considerados conmutables con las muestras de pacientes. Independientemente de la estrategia elegida, es fundamental que el laboratorio justifique su uso y que demuestre su utilidad para realizar el seguimiento del desempeño del procedimiento analítico.
- El laboratorio debe definir las especificaciones de calidad a aplicar en cada caso como criterio de aceptación. Además, debe disponer de evidencias que demuestren que el seguimiento de los resultados de la participación se realiza de forma regular. Cuando los resultados no sean aceptables deben analizarse las posibles causas, y deben llevarse a cabo las acciones correctivas correspondientes. A este respecto, si la desviación observada se considera clínicamente significativa, deberán revisarse los resultados de los pacientes que pudiesen resultar afectados, tendrá que decidirse si es preciso o no modificarlos y por último las modificaciones realizadas deberán comunicarse a los profesionales clínicos.

### **C) Intercambiabilidad de los resultados de los análisis**

El laboratorio debe tener establecido un procedimiento para evaluar la intercambiabilidad de los resultados cuando éstos puedan ser obtenidos en analizadores o ubicaciones diferentes o bien mediante procedimientos de medida distintos. La norma detalla los siguientes aspectos:

- Siempre que sea posible es preferible realizar los estudios de intercambiabilidad de resultados con muestras de pacientes, puesto que así se evitan las limitaciones derivadas de la falta de conmutabilidad de los materiales de control interno.
- La intercambiabilidad de los resultados debe realizarse periódicamente. Por lo tanto, no es suficiente con realizar una única verificación inicial de la intercambiabilidad cuando se incorpora un nuevo analizador o procedimiento de medida para la medición de una magnitud.
- Debe conservarse el registro de los estudios realizados y la evaluación de los resultados de los mismos. Si los resultados del estudio indican que no hay intercambiabilidad, deben evaluarse las implicaciones que pueden suponer las diferencias observadas en la práctica clínica. En caso de considerarse que las diferencias son relevantes deben realizarse las acciones correctivas que se consideren oportunas; así como informar a los profesionales clínicos al respecto.

## **6. Contexto**

Si bien la norma ISO 15189 (67) define los requisitos para el aseguramiento de la calidad de los resultados de las pruebas o estudios analíticos, no es objeto de la misma el describir la estrategia a seguir para darles cumplimiento. Ésta va a depender de las actividades que forman parte del proceso analítico en cuestión e incluso de las características del laboratorio clínico en el que se vaya a implementar. Por lo general, en laboratorios de menor envergadura es un proceso más

sencillo, siendo la limitación mayoritaria la disponibilidad de recursos económicos para su implementación. En los laboratorios clínicos que dan servicio a hospitales de tercer nivel, en cambio, la búsqueda de la estrategia más adecuada es más complicada.

Los laboratorios clínicos que dan servicio a hospitales de tercer nivel, como lo es el Hospital Clínic de Barcelona, se caracterizan, entre otros, por los siguientes aspectos:

- Suelen disponer de áreas muy especializadas donde se realizan procedimientos analíticos poco frecuentes en el mundo o bien que precisan de tecnologías emergentes y un alto grado de especialización del personal. Así, además de utilizar las metodologías más tradicionales (por ejemplo, microscopía óptica para el estudio de una muestra de orina, citometría de flujo para la realización de un hemograma, espectrometría de absorción molecular para la cuantificación de la concentración de bilirrubina, albúmina o colesterol, potenciometría indirecta para determinar la concentración de los iones sodio, potasio o cloruro, la electroforesis capilar para la realización de un proteinograma, inmunoanálisis electroquimioluminiscente para la cuantificación de anticuerpos IgG o IgM contra antígenos del virus de la hepatitis B, VIH o toxoplasma, entre muchos otros), disponen también de la tecnología más avanzada (por ejemplo, espectrometría de masas con plasma acoplado inductivamente para la cuantificación de la concentración de metales, cromatografía líquida acoplada a espectrometría de masas en tándem para la cuantificación de la concentración de fármacos inmunosupresores, secuenciación Sanger y análisis de fragmentos del genoma para el estudio de enfermedades metabólicas u oncológicas, o la secuenciación masiva del ADN fetal libre circulante en sangre materna para la detección prenatal de aneuploidías fetales, entre muchas otras).
- Suelen estar estructurados en servicios o unidades independientes atendiendo a las diferentes especialidades del laboratorio clínico (bioquímica, hematología, inmunología, microbiología) y disponer de áreas operativas transversales. Así, generalmente cuentan

con un área denominada comúnmente “core”, que se caracteriza por disponer de un alto nivel de automatización y por tener la capacidad de dar respuesta a una gran demanda de actividad. Esta área suele centralizar acerca del 90 % de la actividad asistencial de las diferentes áreas de especialización, pudiendo llegar a procesar más de 5.000 muestras diarias y a realizar más de 20.000 pruebas analíticas al día. Es frecuente que además integre la actividad de las pruebas analíticas solicitadas con urgencia. También suelen contar con un área transversal de biología molecular organizada en diferentes plataformas automatizadas, con la finalidad de optimizar los procesos y los recursos para el análisis de ácidos nucleicos. En este tipo de áreas se centraliza la extracción de ácidos nucleicos, la secuenciación Sanger, el análisis de fragmentos, la secuenciación masiva y el análisis bioinformático.

- Suelen ser centro de referencia para programas de cribado poblacional (por ejemplo, el programa de cribado prenatal, el programa de cribado neonatal o el programa de cribado de cáncer colorrectal) debido a la mayor disponibilidad de recursos tanto tecnológicos como humanos.

Estas características hacen que, a la hora de implementar una estrategia de aseguramiento de la calidad que se adecúe a los requisitos de la norma ISO 15189, se tengan que abordar situaciones peculiares.

Por un lado, en los laboratorios tipo “core” pueden realizarse más de 300 pruebas analíticas diferentes y suelen disponer de analizadores replicados para poder dar respuesta a la gran demanda de actividad. A pesar de la existencia de herramientas informáticas para la gestión de los datos de control de calidad interno, individualizar la estrategia de control para cada magnitud en cada uno de los analizadores donde se miden, así como realizar un seguimiento exhaustivo de los resultados, requiere de un tiempo de dedicación que no siempre se dispone.

Por otro lado, en función del procedimiento analítico en cuestión y de la tecnología empleada pueden ser necesarias actividades de aseguramiento de la calidad menos convencionales. A este respecto, en los laboratorios en los que se realizan procedimientos analíticos complejos para dar respuesta a la solicitud de estudios analíticos altamente especializados, la búsqueda de un programa de intercomparación puede resultar un reto. Esto puede conducir a la realización de actividades alternativas para el aseguramiento de la calidad que no suelen realizarse de forma planificada ni suelen evaluarse con objetividad. Además, la realización de procedimientos analíticos mediante el uso de técnicas más tradicionales puede no estar exenta de la implementación de actividades de control de calidad adicionales. Así, en los estudios analíticos que utilizan metodologías en las que es preciso que los profesionales del laboratorio realicen una interpretación, deben planificarse actividades para asegurar la concordancia entre observadores que, igual que en el caso anterior, debe evaluarse de forma objetiva. Este escenario puede abarcar desde un estudio citológico de sangre periférica por microscopía óptica para el diagnóstico de enfermedades hematológicas hasta un análisis cromosómico mediante citogenética convencional.

A estas particularidades hay que añadir que los procedimientos recomendados por las sociedades científicas para la realización de estudios con fines de aseguramiento de calidad, suelen resultar de difícil aplicación en la práctica diaria de los laboratorios clínicos. Generalmente esto es debido a que son extremadamente complejos, a que precisan de unos recursos económicos y/o humanos de los que no se dispone, o bien porque el nivel de exigencia teóricamente requerido no es alcanzable por los laboratorios a pesar de disponer de la última tecnología de diagnóstico in vitro.





## ***II. Objetivos***





En el contexto de un laboratorio clínico de un hospital de tercer nivel, y partiendo de la premisa que cualquier actividad no controlada adecuadamente comporta un riesgo añadido en los resultados analíticos y por tanto en la seguridad del paciente, en esta Tesis Doctoral se ha llevado a cabo una revisión del proceso analítico para determinados estudios o pruebas con la finalidad de identificar aspectos relevantes para el aseguramiento de la calidad que, hasta el momento, no se hubiesen contemplado en el plan de control de la calidad o bien que pudiesen optimizarse. Una vez detectadas las carencias, se plantea la introducción de nuevas estrategias para el aseguramiento de la calidad de los resultados analíticos, adaptadas a las necesidades de las distintas áreas de especialización del laboratorio clínico, pero sin perder de vista su factibilidad.

Por tanto, el objetivo de esta Tesis Doctoral es establecer y poner a disposición de la comunidad científica de laboratorio clínico las estrategias dirigidas al aseguramiento de la calidad de los resultados analíticos implementadas en distintas áreas de especialización de un laboratorio que da servicio a un hospital de tercer nivel.

Objetivos específicos:

- 1) Proporcionar especificaciones de calidad de error total basadas en el estado del arte:**
  - para magnitudes de hematología y coagulación, considerando diferentes rangos de concentración.
  - para las magnitudes incluidas en el cribado neonatal, en primer lugar, sin tener en cuenta la concentración y, en segundo lugar, considerando diferentes rangos de concentración.
- 2) Establecer una metodología basada en el análisis de riesgos y la métrica sigma para realizar una evaluación anual del rendimiento de la fase analítica** de algunas de las magnitudes bioquímicas que se miden en el laboratorio Core.

- 3) **Establecer una metodología para verificar periódicamente la intercambiabilidad de resultados** de las magnitudes bioquímicas que se miden indistintamente en cualquiera de los analizadores disponibles en el laboratorio Core.
  
- 4) **Establecer una metodología para verificar la fiabilidad de los resultados ante la ausencia de un programa de intercomparación.** Valorado en la magnitud heparán sulfato.
  
- 5) **Establecer una metodología para evaluar la concordancia entre observadores en aquellas actividades subjetivas** del proceso analítico del cribado neonatal, concretamente: la clasificación de la calidad de las muestras neonatales recibidas de los diferentes centros maternos y la interpretación de los electroferogramas realizados para el cribado de hemoglobinopatías.

### ***III. Resultados y discusión***





## **1. Primer objetivo: Proporcionar especificaciones de calidad de error total basadas en el estado del arte.**

Este objetivo se ha llevado a cabo para magnitudes de diferentes áreas de especialización con diferentes puntos de partida.

El primer estudio se realizó para comprobar si las dificultades que se presentan en los laboratorios clínicos para alcanzar especificaciones de calidad mínimas de error total basadas en el estado del arte, para determinadas magnitudes de hematología y coagulación, podían ser debidas a que el rendimiento analítico varía en función de la concentración. En caso que fuese así, lo mismo podría estar ocurriendo para magnitudes de otras áreas de especialización.

El segundo estudio realizado para las magnitudes incluidas en el cribado neonatal surgió a raíz de la inexistencia de especificaciones de calidad de error total basadas en alguno de los modelos propuestos en la reunión de consenso de Milán. A pesar de la existencia de programas de intercomparación a nivel nacional e internacional, los organizadores no proporcionaban esta información a los participantes. Por lo tanto, la finalidad principal fue estimar el error de medida cometido por los laboratorios inscritos en un programa de intercomparación nacional para obtener y proporcionar a los participantes unas especificaciones de error total mínimas basadas en el estado del arte. En función de los resultados del primer estudio, se valoraría posteriormente si podría ser interesante conocer la variación del rendimiento analítico en función de la concentración de las magnitudes del programa de cribado.

### **1.1. Primer estudio**

**Estudio del rendimiento analítico a diferentes concentraciones de las magnitudes hematológicas a partir de los datos de los programas de evaluación externa de la calidad en España.**

### **1.1.1. Presentación**

Un aspecto esencial para asegurar la fiabilidad de los resultados que se emiten desde el laboratorio clínico es establecer el error máximo permitido en la medida de las magnitudes que se analizan. Actualmente, la opción más viable para establecer estos límites de aceptación es utilizar las especificaciones de calidad basadas en la variabilidad biológica o en el estado del arte.

La participación en programas de intercomparación es una de las herramientas más útiles para controlar el rendimiento analítico de los procedimientos de medida. Los participantes pueden comparar el error de medida que han cometido respecto a las especificaciones de calidad de error total. Un hecho a destacar es que la mayoría de las publicaciones sobre especificaciones de calidad basadas en el estado del arte recomiendan un único valor de error total para cada magnitud independientemente del nivel de concentración ensayado (69,70), a pesar de que se conoce que la precisión de los procedimientos de medida puede variar en función de la concentración, y que la precisión puede ser un factor que afecte al error de medida cometido.

El objetivo de estudio fue evaluar si el error de medida variaba en función de la concentración de las magnitudes de hematología y coagulación y, en caso afirmativo, proporcionar especificaciones de calidad de error total para diferentes rangos de concentración a la comunidad científica. Dado el interés de la propuesta de este objetivo, la Sociedad Española de Hematología y Hemoterapia (SEHH) cedió a la doctoranda una base de datos en la que se recogían más de 380.000 resultados de la participación de más de 500 laboratorios españoles durante los años 2015 y 2016 en distintos PEEC de hematología y coagulación.

En la Tabla 7 se detallan la información relativa a cada uno de los PEEC. La periodicidad de todos los programas fue trimestral, excepto en los programas de hematimetría y coagulación cuya periodicidad fue mensual. En todos los programas se incluían dos materiales de control por envío. El tipo de material de control variaba en función del programa. En los programas de hematimetría y recuento diferencial leucocitario se utilizó sangre humana con estabilizantes,



mientras que para el resto de programas se utilizaron muestras de plasma humano liofilizadas.

En la Tabla 8 se especifican los procedimientos de medida utilizados por los laboratorios participantes.

Tabla 7. Información sobre los programas de intercomparación: magnitudes incluidas en cada programa, rango de valores de los materiales de control, unidades de medida y número de resultados de los laboratorios participantes.

Programa	Magnitudes	Valores	Unidades	Nº Resultados
<b>Hematimetría</b>	LEU	1,36 - 25,84	10 <sup>9</sup> /L	35.056
	HEM	2,58 - 5,03	10 <sup>12</sup> /L	35.316
	HB	71 - 151	g/L	35.080
	HTO	0,216 - 0,482	L/L	35.081
	VCM	82,9 - 104,6	fL	35.408
	HCM	28,3 - 32,1	pg	35.494
	CHCM	297 - 366	g/L	35.541
	PLT	19 - 259	10 <sup>9</sup> /L	35.330
<b>Coagulación</b>	TP	1,02 - 5,49	Ratio	27.401
	TTPA	1,04 - 2,60	Ratio	27.167
<b>Recuento diferencial leucocitario</b>	NEU	0,70 - 15,05	10 <sup>9</sup> /L	4.042
	LINF	1,16 - 5,24	10 <sup>9</sup> /L	4.042
	MON	0,28 - 2,51	10 <sup>9</sup> /L	4.050
	EOS	0,21 - 1,28	10 <sup>9</sup> /L	4.007
	BAS	0,04 - 3,26	10 <sup>9</sup> /L	3.950
<b>Reticulocitos</b>	RET	33,51 - 206,17	10 <sup>9</sup> /L	3.107
<b>Hemoglobina A<sub>2</sub></b>	HbA <sub>2</sub>	3,12 - 5,93	%	966
<b>Factor VIII</b>	FVIII	29,32 - 90,07	%	1.233
<b>Antitrombina</b>	AT	32,02 - 96,59	%	1.564
<b>Factor von Willebrand</b>	FVW	43,74 - 95,71	%	578
<b>Proteína C</b>	PC	32,91 - 99,67	%	660

LEU: leucocitos, HEM: hematíes; HB: hemoglobina; HTO: hematocrito; VCM: volumen corpuscular medio; HCM: hemoglobina corpuscular media; CHCM: concentración de hemoglobina corpuscular media; PLT: plaquetas; TP: tiempo de protrombina, TTPA: tiempo de tromboplastina parcial activada; NEU: neutrófilos; LINF: linfocitos; MON: monocitos; EOS: eosinófilos; BAS: basófilos; RET: reticulocitos; HbA<sub>2</sub>: hemoglobina A<sub>2</sub>; FVIII: factor VIII; AT: antitrombina; FVW: factor von Willebrand; PC: proteína C.

Tabla 8. Procedimientos de medida utilizados por los laboratorios participantes en cada uno de los PEEC.

<b>Programa</b>	<b>Procedimiento de medida</b>
<b>Hematimetría</b>	Citometría de flujo
<b>Coagulación</b>	Coagulometría
<b>Recuento diferencial leucocitario</b>	Citometría de flujo
<b>Reticulocitos</b>	Citometría de flujo
<b>Hemoglobina A<sub>2</sub></b>	Cromatografía líquida de alta resolución Electroforesis capilar
<b>Factor VIII</b>	Coagulometría
<b>Antitrombina</b>	Colorimetría
<b>Factor von Willebrand</b>	Inmunoturbidimetría
<b>Proteína C</b>	Coagulometría

### **1.1.2. Resultados y discusión**

Tras calcular el error de medida de cada uno de los resultados, teniendo en cuenta los diferentes procedimientos de medida y sistemas analíticos, se agruparon los errores según los diferentes rangos de concentración de los materiales de control. Para cada rango de concentración se calculó la concentración media y el percentil 90 del error de medida, correspondiendo éste al error de medida que era capaz de alcanzar el 90 % de los laboratorios.

En primer lugar, se representaron gráficamente los errores de medida en función de la concentración de los materiales de control. Se observó que la mayoría de las magnitudes mostraban variaciones del rendimiento analítico. Al comparar el percentil 90 del error de medida en las concentraciones extremas se observó que para las siguientes magnitudes las diferencias eran destacables: leucocitos, plaquetas, neutrófilos, linfocitos, monocitos, eosinófilos, basófilos, tiempo de protrombina, reticulocitos, hemoglobina A<sub>2</sub>, antitrombina y proteína C. De todas ellas sólo el tiempo de protrombina mostró un mejor rendimiento a concentraciones bajas, probablemente debido a que los procedimientos de medida comerciales están mejor estandarizados para el rango normal de concentración (71). En el caso de los

monocitos y los basófilos los datos mostraron justo lo contrario, el rendimiento analítico fue peor a concentraciones bajas. Las diferencias en el error de medida fueron tan llamativas (un 38 % y un 94 % mayores en la concentración más baja, respectivamente) que utilizar diferentes especificaciones de calidad parece la mejor opción para poder realizar una correcta evaluación del rendimiento analítico. Esto mismo propuso la European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM) para los leucocitos, las plaquetas y los reticulocitos en un estudio llevado a cabo por el grupo de trabajo de hematología en 2017 (72). En estos casos, establecer una única especificación de calidad intermedia podría resultar demasiado estricto o permisivo en función de la concentración del material de control. Este hallazgo puede ser especialmente relevante para aquellos laboratorios de otros países como Alemania y Estados Unidos en los que por regulación estatal deban cumplirse unos requisitos de calidad mínima.

En segundo lugar, los percentiles 90 de los errores de medida obtenidos para las concentraciones extremas se compararon con las especificaciones de error total basadas en la variabilidad biológica con la finalidad de comprobar si era posible alcanzarlas. En la Tabla 9 se muestra, para cada magnitud, el error de medida cometido a la menor y mayor concentración estudiada y los valores de error total según los diferentes niveles de la variabilidad biológica (mínimo, deseable y óptimo). Se observó que, para algunas magnitudes como el volumen corpuscular medio, la hemoglobina corpuscular media, la concentración de hemoglobina corpuscular media, el tiempo de tromboplastina parcial activada, la hemoglobina A<sub>2</sub> y el factor VIII, no era posible alcanzar ni siquiera el nivel mínimo de error total basado en la variabilidad biológica independientemente de la concentración. Para plaquetas, monocitos, basófilos, reticulocitos, antitrombina y factor von Willebrand se observó que en las concentraciones más bajas no era posible alcanzar este nivel mínimo pero que en las concentraciones más altas si era alcanzable. Justo lo contrario se observó para el tiempo de protrombina. Por último, para el resto de las magnitudes del estudio se observó que las especificaciones de error total basadas

en la variabilidad biológica eran alcanzables, pero a diferentes niveles (mínimo, deseable u óptimo) en función de la concentración. Concretamente, en los rangos de concentración más altos el rendimiento analítico demostró ser mejor, por lo que se lograba el cumplimiento de especificaciones de calidad más exigentes; mientras que, al evaluar las concentraciones más bajas se observó que el nivel de variabilidad biológica alcanzable se situaba en el mínimo, estando en ocasiones dos criterios por debajo del nivel alcanzable a altas concentraciones.

Tabla 9. Percentiles 90 del error de medida obtenido en las concentraciones extremas y especificaciones de error total basadas en la variabilidad biológica para cada magnitud.

Magnitud	EM (%)		ET (%)		
	Menor concentración	Mayor concentración	VB mínima	VB deseable	VB óptima
LEU	10,4	3,8	16,7	11,2	5,6
HEM	3,2	2,5	5,8	3,8	1,9
HB	3,0	2,6	5,8	3,9	1,9
HTO	4,3	3,6	5,8	3,9	1,9
VCM	3,1	2,9	2,4	1,6	0,8
HCM	3,3	3,8	2,1	1,4	0,7
CHCM	3,5	4,9	1,9	1,3	0,6
PLT	26,1	7,5	12,8	8,5	4,3
TP	7,2	14,5	7,9	5,3	2,6
TTPA	8,2	11,5	6,7	4,5	2,2
NEU	14,4	2,2	22,0	14,6	7,3
LINF	13,5	5,8	20,6	13,7	6,9
MON	53,7	15,7	21,3	14,2	7,1
EOS	30,1	5,5	39,2	26,1	13,1
BAS	100,0	5,7	23,4	15,6	7,8
RET	31,7	14,8	21,6	14,4	7,2
HbA <sub>2</sub>	19,2	12,7	3,8	2,5	1,3
FVIII	21,9	17,5	13,3	8,9	4,4
AT	22,3	11,1	12,5	8,3	4,2
FVW	15,0	11,0	13,4	8,9	4,5
PC	22,5	11,8	28,0	18,7	9,3

EM: error de medida; ET: error total; VB: variabilidad biológica.

LEU: leucocitos, HEM: hematíes; HB: hemoglobina; HTO: hematocrito; VCM: volumen corpuscular medio; HCM: hemoglobina corpuscular media; CHCM: concentración de hemoglobina corpuscular media; PLT: plaquetas; TP: tiempo de protrombina, TTPA: tiempo de tromboplastina parcial activada; NEU: neutrófilos; LINF: linfocitos; MON: monocitos; EOS: eosinófilos; BAS: basófilos; RET: reticulocitos; HbA<sub>2</sub>: hemoglobina A<sub>2</sub>; FVIII: factor VIII; AT: antitrombina; FVW: factor von Willebrand; PC: proteína C.

Estas diferencias de rendimiento analítico encontradas en función de la concentración pueden crear confusión a la hora de seleccionar especificaciones de calidad. Además, una vez

seleccionada, al tratarse de una única especificación de error total, puede conducir a una evaluación inadecuada de los resultados del programa de control de calidad externo, si no se tiene en cuenta que para algunas magnitudes a bajas concentraciones el rendimiento analítico no permite alcanzar la especificación de calidad establecida. Ello, a su vez, puede derivar en la búsqueda de las posibles causas que puedan estar ocasionando el bajo rendimiento y con ello la realización innecesaria de intervenciones en el analizador por parte del laboratorio o del proveedor.

### **1.1.3. Conclusión**

Este estudio demuestra que el error de medida de algunas magnitudes de hematología y coagulación varía en función de la concentración. Por lo tanto, las especificaciones de calidad de error total propuestas, basadas en el estado del arte y para diferentes rangos de concentración, suponen una mejora añadida para la correcta evaluación del rendimiento analítico.

Es importante puntualizar que a pesar de que el estudio ha sido realizado para magnitudes de hematología y coagulación, sería posible encontrar resultados similares para magnitudes de otras áreas de especialización.

### **1.1.4. Cita al artículo publicado**

Molina A, Alcaraz J, Guiñón L, Pérez A, Segurana A, Reverter JC, Merino A. **Study of the analytical performance at different concentrations of hematological parameters using Spanish EQAS data.** Clinical Chemistry and Laboratory Medicine. 2019;57:1980-7.

### **1.1.5. Publicación completa**



Angel Molina\*, José Alcaraz, Leonor Guiñón, Aránzazu Pérez, Anna Segurana, Joan Carles Reverter, Josep Lluís Bedini and Anna Merino

# Study of the analytical performance at different concentrations of hematological parameters using Spanish EQAS data

<https://doi.org/10.1515/cclm-2019-0108>

Received January 29, 2019; accepted June 17, 2019; previously published online July 24, 2019

## Abstract

**Background:** External quality assessment programs are one of the currently available tools to evaluate the analytical performance of clinical laboratories, where the measurement error (ME) obtained can be compared with quality specifications to evaluate possible deviations. The objective of this work was to analyze the ME behavior over the analytical range to assess the need to establish concentration-dependent specifications.

**Methods:** A total of 389,000 results from 585 laboratories and 2628 analyzers were collected from the Spanish external quality assessment schemes (EQAS) in hematology during the years 2015–2016. The parameters evaluated included white blood cells, red blood cells, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, platelets, prothrombin time, activated partial thromboplastin time, neutrophils, lymphocytes, monocytes, eosinophils, basophils, reticulocytes, hemoglobin A<sub>2</sub>, antithrombin, factor VIII, protein C and

von Willebrand factor. The 90th percentile of ME was calculated for every concentration evaluated of each parameter.

**Results:** We found a significant variation in the analytical performance of leukocytes, platelets, neutrophils, lymphocytes, monocytes, eosinophils, basophils, prothrombin time, reticulocytes, hemoglobin A<sub>2</sub>, antithrombin and protein C. Furthermore, this ME variation may not allow complying with the same biological variability requirements within the whole analytical range studied.

**Conclusions:** Our work shows the importance of implementing concentration-dependent specifications which can help laboratories to use proper criteria for quality specifications selection and for a better external quality control results evaluation.

**Keywords:** biological variability; external quality assessment; hematology; quality specifications; state of the art; total error.

## Introduction

Clinical laboratories play an essential role in the health system ambit due to the fact that 70% of clinical decisions are based on results provided by the laboratory [1]. Therefore, because of the impact they can have on patient's health, it is necessary for laboratories to ensure that their results reach the quality level that guarantees reliable information for clinical diagnosis. For this reason, it is important for clinical laboratories to keep the possible sources of error in their analytical determinations under control. Analytical errors can be estimated through different indicators such as random error, systematic error and total error [2].

External quality assessment schemes (EQAS) are one of the currently available tools to evaluate the analytical performance of clinical laboratories analyzers, since it allows the identification and quantification of significant deviations for the different analytical assays [3].

---

\*Corresponding author: Angel Molina, Hematology External Quality Assessment Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, C/ Villarroel 170, Barcelona 08036, Spain; and CORE Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain, Phone: +34 932272175, E-mail: molinaborras@gmail.com. <https://orcid.org/0000-0002-9584-3646>

**José Alcaraz:** Hematology External Quality Assessment Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain; and CORE Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain

**Leonor Guiñón:** Quality Department, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain

**Aránzazu Pérez, Anna Segurana and Joan Carles Reverter:** Hematology External Quality Assessment Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain

**Josep Lluís Bedini and Anna Merino:** CORE Laboratory, Biomedical Diagnostic Center, Hospital Clinic of Barcelona, Barcelona, Spain

These deviations are considered as total error, since they simulate the routine patient samples processing, in which random and systematic factors can affect together [4].

An objective way to establish a minimum level of quality that ensures reliability in the results consists in defining acceptance limits as maximum allowable errors for the different parameters. These tolerance limits are known as quality specifications and they are an essential requirement for laboratories whose activity is accredited according to the international standard UNE-EN ISO 15189 [5]. International recommendations about the selection of quality specifications have been hierarchized into three models [6]. The first model considers the effect of analytical performance on each specific clinical situation; the second model recommends analytical specifications based on biological variability of the measurand; the third model does so in reference to the so-called state of the art. Due to the easy accessibility and quantity of parameters studied, models 2 and 3 have become the basis for the implementation of analytical quality specifications.

The measurement error (ME) obtained from EQAS results (calculated as the difference between the

participant result and the target value in percentage) can be compared with quality specifications for total error to determine if the ME committed exceeds the maximum accepted limits. Most of the published works on quality specifications recommend a single specification value for each measurand [7–11], and therefore, only one criterion for each parameter is used by EQAS organizers in the evaluation of the participants' performance [12]. However, it is proved that the precision of the analyzers may differ depending on the concentration of the analyte studied [13–15]; thus, it is reasonable to assume that the ME accomplished by laboratories may also be concentration-dependent.

EQAS are considered a valuable tool of information, as they manage a global data on the state of the art of current technology. The study of the ME distribution can provide new data that can help laboratories to use proper criteria for quality specifications selection and for a better evaluation of EQAS results. The objective of this work is to evaluate the relationship between the concentration of the measurand and the ME for hematology test assays using Spanish EQAS data.

**Table 1:** Data summary.

Scheme	QC manufacturer	n	Analyzers	Labs	Measurand	Data
CBC	Spanish blood and tissue bank	48	816	570	WBC, 10 <sup>9</sup> /L	35,056
					RBC, 10 <sup>12</sup> /L	35,316
					HB, g/L	35,080
					HCT, L/L	35,081
					MCV, fL	35,408
					MCH, pg	35,494
					MCHC, g/L	35,541
					PLT, 10 <sup>9</sup> /L	35,330
BCT	Siemens Healthineers	48	688	542	PT Ratio	27,401
					APTT Ratio	27,167
DLC	R&D Systems	16	311	193	NEU, 10 <sup>9</sup> /L	4042
					LYM, 10 <sup>9</sup> /L	4042
					MON, 10 <sup>9</sup> /L	4050
					EOS, 10 <sup>9</sup> /L	4007
					BAS, 10 <sup>9</sup> /L	3950
Reticulocytes	Grifols Movaco	16	248	197	RET, 10 <sup>9</sup> /L	3107
Hemoglobin A <sub>2</sub>	Bio-Rad	16	81	77	HBA <sub>2</sub> , %	966
Antithrombin	Siemens Healthineers	16	127	117	AT, %	1564
Factor VIII	Siemens Healthineers	16	91	86	FVIII, %	1233
Protein C	Siemens Healthineers	16	51	48	PC, %	660
von Willebrand factor	Grifols Movaco	16	46	43	VWF, %	578

Résumé of the data extracted from the Spanish external quality assessment program in hematology. CBC, complete blood count; WBC, white blood cells; RBC, red blood cells; HB, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PLT, platelets; BCT, blood coagulation tests; PT, prothrombin time; APTT, activated partial thromboplastin time; DLC, differential leukocyte count; NEU, neutrophils; LYM, lymphocytes; MON, monocytes; EOS, eosinophils; BAS, basophils; RET, reticulocytes; HBA<sub>2</sub>, hemoglobin A<sub>2</sub>; AT, antithrombin; FVIII, factor VIII; PC, protein C; VWF, von Willebrand factor; QC, quality control; n, number of control materials evaluated (each survey includes two different samples, so the number of surveys is half the number of control materials); Data: number of individual results included.



## Materials and methods

A total of 389,000 results from 585 laboratories and 2628 analyzers were collected from the Spanish EQAS in hematology [16] during the years 2015–2016. Data used in this study were part of the following quantitative schemes: complete blood count (CBC), blood coagulation tests (BCTs), differential leukocyte count (DLC), reticulocytes, hemoglobin A<sub>2</sub>, antithrombin, factor VIII, protein C and von Willebrand factor. All the schemes have a quarterly periodicity, with the exception of the CBC and BCT, which are scheduled monthly.

Two different samples to be tested are sent in each survey. The nature of the control materials used depends on each program. In CBC, DLC and reticulocytes modules, stabilized human blood samples are distributed, while in the rest of the programs, lyophilized human plasma samples are used. Normal plasma samples are citrated normal human pooled plasmas from selected healthy blood donors. Pathological plasma samples are made of normal samples where concentrations of all factors have been decreased by dilution to simulate a pathological sample.

To ensure the proper pre-analytical management of the samples, handling and shipping conditions of external quality control materials were always carried out following the recommendations of each manufacturer. These samples were sent along with the corresponding instructions about their appropriate conservation to ensure a correct management of the samples prior to their analysis.

Before the analysis, every laboratory result identified as an outlier was excluded according to the statistical methods recommended by the international standard UNE-EN ISO 13528 [17].

Data obtained from the different schemes are summarized in Table 1.

The following processing of the data was carried out:

1. For all the results reported by the participants, the ME (in percentage) was calculated by comparing each result with its target value (robust mean) of the corresponding peer group, according to the following formula:

$$\text{ME (\%)} = (\text{result} - \text{target value}) * 100 / \text{target value}.$$

Target values have been calculated by means of consensus values from participants grouped by the same methodology. Algorithm A was the statistical methodology selected to calculate the robust mean, as recommended by the international standard UNE-EN ISO 13528 [17]. It is an iterative calculation that updates the value of the mean and standard deviation by decreasing the weight of the most extreme values. This procedure is repeated until the mean and standard deviation values converge.

2. ME results for each measurand were clustered into different intervals based on the concentration of its target value.
3. For each interval obtained, the concentration average and the 90th percentile of ME were calculated.
4. The concentration average against its 90th percentile of ME was plotted in different graphs for each measurand.

Once the results of the analytical performance have been obtained, the ME calculated were compared with the quality specifications for total error based on biological variability. Biological variability specifications for total error have been calculated based on the three-level model proposed by Fraser [18] and using the interindividual and intraindividual

**Table 2:** Analytical performance specifications for total error based on biological variability for hematological parameters.

	Optimum	Desirable	Minimum
WBC <sup>a</sup>	5.6	11.2	16.7
RBC <sup>a</sup>	1.9	3.8	5.8
Hemoglobin <sup>a</sup>	1.9	3.9	5.8
Hematocrit <sup>a</sup>	1.9	3.9	5.8
MCV <sup>a</sup>	0.8	1.6	2.4
MCH <sup>a</sup>	0.7	1.4	2.1
MCHC <sup>a</sup>	0.6	1.3	1.9
Platelets <sup>a</sup>	4.3	8.5	12.8
PT <sup>b</sup>	2.6	5.3	7.9
APTT <sup>b</sup>	2.2	4.5	6.7
Neutrophils <sup>a</sup>	7.3	14.6	22.0
Lymphocytes <sup>a</sup>	6.9	13.7	20.6
Monocytes <sup>a</sup>	7.1	14.2	21.3
Eosinophils <sup>a</sup>	13.1	26.1	39.2
Basophils <sup>a</sup>	7.8	15.6	23.4
Reticulocytes <sup>a</sup>	7.2	14.4	21.6
Hemoglobin A <sub>2</sub> <sup>b</sup>	1.3	2.5	3.8
Factor VIII <sup>b</sup>	4.4	8.9	13.3
Antithrombin <sup>b</sup>	4.2	8.3	12.5
VWF <sup>b</sup>	4.5	8.9	13.4
Protein C <sup>b</sup>	9.3	18.7	28.0

Different severity levels of quality specifications for total error (TEAPS) based on biological variability. WBC, white blood cells; RBC, red blood cells; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PT, prothrombin time; APTT, activated partial thromboplastin time; VWF, von Willebrand factor. <sup>a</sup>TEAPS calculated from the EFLM data. <sup>b</sup>TEAPS calculated from the Ricós group data.

variabilities published by the European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) and the Ricós group [19, 20]. These referenced specifications are summarized in the Table 2.

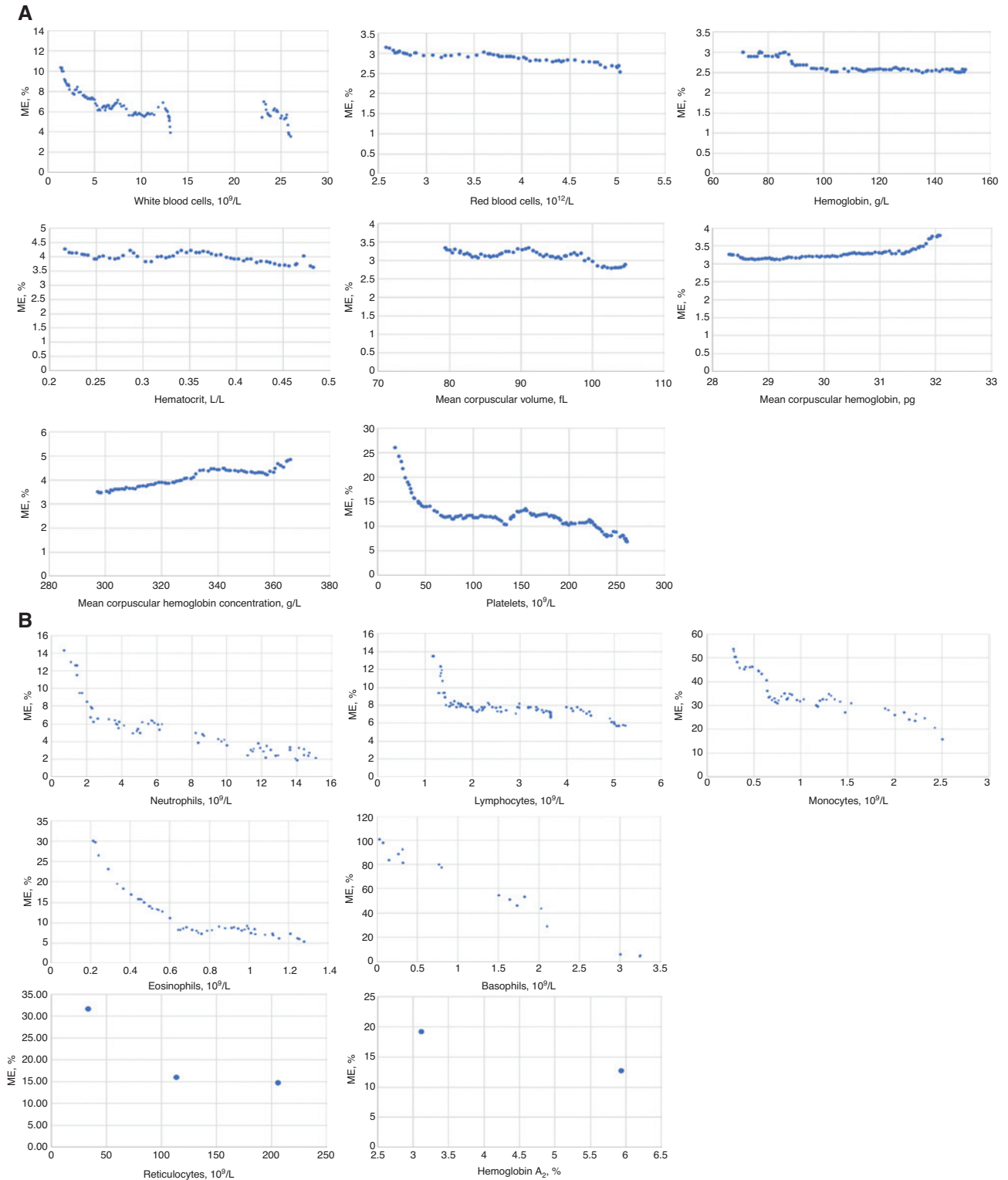
No ethical approval was required since this study did not involve human subjects.

## Results

The distribution of the analytical performance throughout the concentration range studied is represented in Figure 1

for each parameter, which shows the different concentrations (x-axis) against the 90th percentile of ME values (y-axis). Graphs are divided into three sections:

1. Figure 1A shows the values of the 90th percentile of ME corresponding to the following hematological parameters: white blood cells, red blood cells, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration and platelets.



**Figure 1:** Measurement error (ME) distribution graphs: analytical performance.

For each measurand, the analytical concentration (x-axis) is represented against the 90th percentile of the ME (y-axis). (A) General hematology parameters (white blood cells, red blood cells, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration and platelets); (B) differential leukocyte count (neutrophils, lymphocytes, monocytes, eosinophils and basophils), reticulocytes and hemoglobin  $A_2$ ; (C) hemostasis parameters (prothrombin time, activated partial thromboplastin time, antithrombin, factor VIII, protein C and von Willebrand factor).

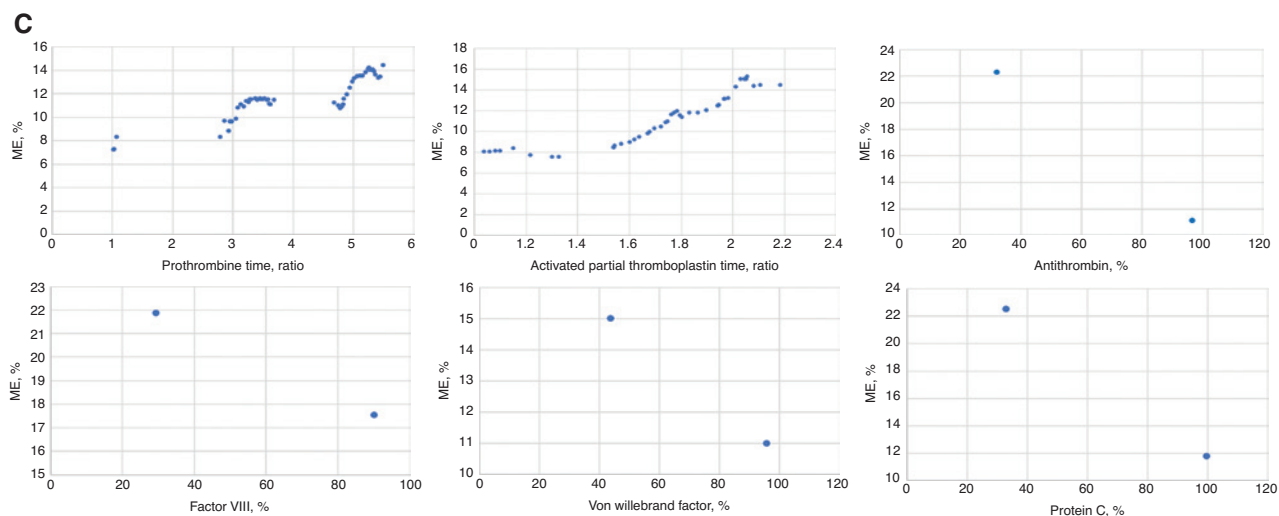


Figure 1: (Continued)

- Figure 1B shows the values of the 90th percentile of ME of DLC (neutrophils, lymphocytes, monocytes, eosinophils and basophils), reticulocytes and hemoglobin A<sub>2</sub>.
- Figure 1C shows the 90th percentile of ME distribution corresponding to the following hemostasis parameters: prothrombin time, activated partial thromboplastin time, fibrinogen, antithrombin, factor VIII, protein C and von Willebrand factor.

Parameters included in the complete blood count, BCTs and DLC programs show a greater number of data since control materials were available in a more varied range of analytical concentrations. However, in reticulocytes, hemoglobin A<sub>2</sub>, antithrombin, factor VIII, von Willebrand factor and protein C schemes, the control materials are only available in two different concentrations (normal and pathological), which usually have similar concentrations in the different surveys. In these cases, only two intervals are considered (with the exception of reticulocytes, in which three different levels of control materials have been distributed).

Most of the parameters considered show an analytical performance that varies depending on the concentration (see Figure 1). These variations are summarized in Table 3, which shows how the 90th percentile of ME varies between the highest and lowest concentrations evaluated. According to this variation, measurands can be divided in two different groups:

- Parameters with a relatively constant analytical performance for the different concentrations analyzed: red blood cells, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin

and mean corpuscular hemoglobin concentration. In these parameters, the difference in the ME values between the lowest and highest concentrations studied is around 1% point. Other parameters such as APTT, factor VIII and von Willebrand factor have an absolute difference lower than 5% points.

- Parameters with a substantial change in the ME values when comparing the lowest and highest concentrations: leukocytes, platelets, neutrophils, lymphocytes, monocytes, eosinophils, basophils, PT, reticulocytes, hemoglobin A<sub>2</sub>, antithrombin and protein C. In these parameters we find ME differences higher than 5% points.

If we compare the data from Table 3 with the quality specifications for total error based on BV, it is possible to see how the variation in the analytical performance affects the ability to accomplish each criteria of analytical performance specification for total error based on BV. A scheme of this impact can be seen in Figure 2. If we look at the white blood cells as an example, at high concentrations, the 90th percentile of ME can reach the optimum specification of biological variability. However, at low concentrations, the 90th percentile of ME can only achieve the desirable criteria. The measurands where the ME variation may involve different biological variability specifications targets depending on the concentration are the following: white blood cells, hematocrit, platelets, prothrombin time, neutrophils, lymphocytes, monocytes, eosinophils, basophils, reticulocytes, antithrombin, von Willebrand factor and protein C. In some of these parameters, as occurs with platelets, monocytes, basophils or reticulocytes, the concentration of the control material can make the difference

**Table 3:** Summary of the effect of the analytical concentration on the measurement error (ME).

Measurand	Lowest concentration		Highest concentration	
	Mean value	ME, %	Mean value	ME, %
WBC	1.36	10.4	25.84	3.8
RBC	2.58	3.2	5.03	2.5
Hemoglobin	71.0	3.0	151.0	2.6
Hematocrit	0.216	4.3	0.482	3.6
MCV	82.9	3.1	104.6	2.9
MCH	28.3	3.3	32.1	3.8
MCHC	297	3.5	366	4.9
Platelets	19	26.1	259	7.5
PT	1.02	7.2	5.49	14.5
APTT	1.04	8.2	2.60	11.5
Neutrophils	0.70	14.4	15.05	2.2
Lymphocytes	1.16	13.5	5.24	5.8
Monocytes	0.28	53.7	2.51	15.7
Eosinophils	0.21	30.1	1.28	5.5
Basophils	0.04	100.0	3.26	5.7
Reticulocytes	33.51	31.7	206.17	14.8
Hemoglobin A <sub>2</sub>	3.12	19.2	5.93	12.7
Factor VIII	29.32	21.9	90.07	17.5
Antithrombin	32.02	22.3	96.59	11.1
VWF	43.74	15.0	95.71	11.0
Protein C	32.91	22.5	99.67	11.8

The data show how the 90th percentile of the ME varies between the lowest and the highest concentrations evaluated. Mean value corresponds to the average value, whereas ME represents the 90th percentile of all ME results for that mean value. WBC, white blood cells; RBC, red blood cells; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PT, prothrombin time; APTT, activated partial thromboplastin time; VWF, von Willebrand factor.

of complying with a quality specification based on BV or not. Furthermore, platelets, eosinophils and basophils are the only cases where the ME variation involves more than one criteria of analytical performance specifications for total error based on BV. Undoubtedly, this disparity in analytical performance is likely to cause confusion in the evaluation and establishment of quality specifications. Finally, the ME variation does not suppose different biological variability specifications targets in the following parameters: red blood cells, hemoglobin, mean corpuscular volume, mean corpuscular hemoglobin, mean corpuscular hemoglobin concentration, activated partial thromboplastin time, hemoglobin A<sub>2</sub> and factor VIII.

## Discussion

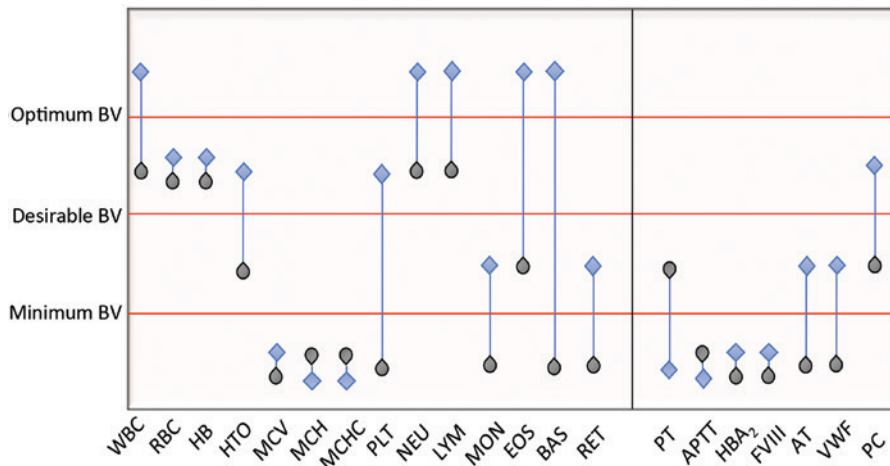
Our data reflect the state-of-the-art analytical performance for hematology and hemostasis parameters.

Comparing our results with the published Spanish minimum consensus specifications (MCS) [11], which are also based on the state-of-the-art performance, we find that our ME data are slightly lower than MCS. This can be explained by the fact that the data used to calculate MCS are based in EQAS results from the period 2005–2010, whereas our data are from 2015–2016. In this more recent period, analyzers and reagents may have been subjected to a certain degree of improvement, leading to better performance and, therefore, to lower ME results. Apart from this remark, both ME results are absolutely comparable, so this can be considered as a validation for the results obtained in our study.

This study shows that for many parameters there are differences in the analytical performance that depend on the concentration of the analyte. This difference is strikingly notable for monocytes and basophils, where huge differences in the ME, up to 38% and 94% respectively, can be observed. In these cases, a single value for quality specifications may be inadequate to properly assess the analytical performance of these tests. The hematology working group of the European Organization for External Quality Assurance Providers in Laboratory Medicine (EQALM) also explored the need to establish maximum acceptance limits based on the concentration [21]. They also did not find any differences in the analytical performance for red blood cells parameters. Nevertheless, they showed the need to set up two different acceptance limits for white blood cells, platelets and reticulocytes. This leads to similar conclusions than the ones our results provide, which reaffirms the need of studying the effect of concentration over the analytical performance. This can be especially important in those countries where quality requirements are laid down by regulation, as in the United States, Germany or Switzerland.

Our data also show that in almost all cases, a higher value on the concentration of the parameter implies a better analytical performance, which means a decrease in the ME. However, this is not the case for prothrombin time and activated partial thromboplastin time, in which the ME increases as the average concentration value become higher. This could be due to the fact that in these coagulation tests the different commercial methods that are available in the market are better standardized in the normal range, not being so as the sample becomes more pathological [22, 23], where the diverse nature of the reagents show different sensitivity to factor deficiency or anticoagulant treatment.

Quality specifications based on biological variability have acquired considerable relevance due to their clinical basis and their daily laboratory applicability [24]. Figure 2 shows that ME changes do not always



**Figure 2:** Effect of analytical performance at different concentrations on quality specifications for total error based on biological variability. Points displayed show the capability of complying with any of the analytical performance specifications for the total error based on biological variability. A point below any biological variability criteria (horizontal lines) indicates that the 90th percentile of the measurement error (ME) results cannot reach that criteria. Circles: 90th percentile of ME for the lowest concentration studied. Squares: 90th percentile of ME for the highest concentration studied. BV, biological variability specification for total error; WBC, white blood cells; RBC, red blood cells; HB, hemoglobin; HCT, hematocrit; MCV, mean corpuscular volume; MCH, mean corpuscular hemoglobin; MCHC, mean corpuscular hemoglobin concentration; PLT, platelets; PT, prothrombin time; APTT, activated partial thromboplastin time; NEU, neutrophils; LYM, lymphocytes; MON, monocytes; EOS, eosinophils; BAS, basophils; RET, reticulocytes; HBA<sub>2</sub>, hemoglobin A<sub>2</sub>; AT, antithrombin; FVIII, factor VIII; PC, protein C; VWF, von Willebrand factor.

allow reaching the same specifications criteria based on biological variability. This variation could have important practical applications, since a laboratory could exceed their quality specifications at low concentrations without implying that its analyzer or methodology is becoming out of control. All this would entail the realization of numerous corrective actions that are actually unnecessary because the deviations are not a consequence of factors that are worsening the instrument performance. Attention should be paid to those parameters in which the analytical performance varies enough to change the maximum allowable ME criteria, especially in those in which there is a change of at least two or three criteria of biological variability requirement, as occurs with platelets, eosinophils and basophils. It is important to highlight that biological variability specifications have been obtained from healthy patients, and consequently, they may not be completely suitable for pathological ranges [25]. It is also important to mention that there are some parameters where it is considerably hard to achieve any biological variability specification based on current technology in any analytical range, as seen in other works [26, 27].

One of the limitations of this work is that, due to the short stability of the cellular components of blood, it is necessary to add stabilizers in order to improve the conservation of the sample. This may present issues of

commutability, so that it requires grouping of instruments with similar methodology. Furthermore, there are no international reference standards available for most of the parameters studied; hence, the target values have had to be designated through the results of the participants. Another limitation of this study is that this work has been focused on purely analytical aspects, but it should be pointed out that analytical performance serves clinical goals. In the evaluated surveys, control materials representing normal and pathological samples have been included. Therefore, with the results provided, it is also possible to assess the analytical performance in most clinical decision intervals. However, for some parameters, the concentrations studied do not cover all possible clinical decision targets.

The goal of this study was to use the findings as aid to clinical laboratories that wish to establish quality specifications to maintain the performance of their analyzers under control. Analyzing the effect of concentration on analytical performance could be remarkably useful to make a correct assessment of the external quality control results. Our work shows the importance of implementing concentration-dependent specifications, because the acceptance limits that are not concentration-stratified can lead to an intermediate specification that would be undemanding at high concentrations or too strict at low concentrations.



**Acknowledgments:** This research was partially supported by FPCQLC (Fundació pel Control de Qualitat dels Laboratoris Clínics).

**Author contributions:** All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.

**Research funding:** This work is part of a research project funded by the Directory of Science, Technology and Innovation of the Ministry of Economy and Competitiveness of Spain, with reference DPI2015-64493-R.

**Employment or leadership:** None declared.

**Honorarium:** None declared.

**Competing interests:** The funding organization(s) played no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the report for publication.

## References

- Laboratorio Clínico Central. Estándares y recomendaciones de calidad y seguridad. Ministerio de Sanidad, Servicios Sociales e Igualdad. Gobierno de España. 2013. [https://www.msbs.gob.es/organizacion/sns/planCalidadSNS/docs/Laboratorio\\_Clinico\\_EyR.pdf](https://www.msbs.gob.es/organizacion/sns/planCalidadSNS/docs/Laboratorio_Clinico_EyR.pdf). Accessed: January 2019.
- Krouwel JS. Setting performance goals and evaluating total analytical error for diagnostic assays. *Clin Chem* 2002;48:919–27.
- Libeer JC. Role of external quality assurance schemes in assessing and improving quality in medical laboratories. *Clin Chim Acta* 2001;309:173–7.
- Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem* 2011;57:1670–80.
- EN ISO 15189:2012. Medical laboratories – Particular requirements for quality and competence. Geneva: ISO, 2012.
- 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;53:833–5.
- Clinical Laboratory Improvement Act 2004. Part 493 laboratory requirements. Subpart I – proficiency testing programs for nonwaived testing. [https://www.ecfr.gov/cgi-bin/text-idx?SID=bed215d568a567e7f0c2f8710d44331c&mc=true&node=sg42.5.493\\_1905.sg1&rgn=div7](https://www.ecfr.gov/cgi-bin/text-idx?SID=bed215d568a567e7f0c2f8710d44331c&mc=true&node=sg42.5.493_1905.sg1&rgn=div7) Accessed: January 2019.
- German Medical Association. Revision of the “Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations–RiliBAEK”. *J Lab Med* 2015;39:26–69.
- Jones GR, Sikaris K, Gill J. ‘Allowable limits of performance’ for external quality assurance programs – an approach to application of the Stockholm Criteria by the RCPA Quality Assurance Programs. *Clin Biochem Rev* 2012;33:133–9.
- Ricós C, Álvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, et al. Current Database on biologic variation: prons, cons and progress. *Scand J Clin Lab Invest* 1999;59:491–500.
- Salas A, Ricós C, Prada E, Ramón F, Morancho J, Jou JM, et al. State of the art approach to goal setting. *Clin Lab Med* 2017;37:73–84.
- Ricós C, Baadenhuijsen H, Libeer JC, Hyltoft Petersen P, Stockl D, Thienpont L, et al. External quality assessment: currently used criteria for evaluating performance in European countries, and criteria for future harmonization. *Eur J Clin Chem Clin Biochem* 1996;34:159–65.
- Hübl W, Tlustos L, Bayer PM. Use of precision profiles to evaluate precision of the automated leukocyte differential. *Clin Chem* 1996;42:1068–73.
- Buttarelli M, Bulian P, Farina G, Temporin V, Toffolo L, Trabuio E, et al. Flow cytometric reticulocyte counting: parallel evaluation of five fully automated analyzers: an NCCLS-ICSH approach. *Am J Clin Pathol* 2001;115:100–11.
- Buttarelli M. Quality specification in haematology: the automated blood cell count. *Clin Chim Acta* 2004;346:45–54.
- Vives-Corrons JL, Gutiérrez G, Jou JM, Reverter JC, Martínez-Brotos F, Domingo A, et al. Characteristics and performance of the external quality assessment scheme (EQAS) for hematology in Spain. Ten years of experience. *Ann Ist Super Sanita* 1995;31:95–101.
- EN ISO 13528:2015. Statistical methods for use in proficiency testing by interlaboratory comparisons. ISO 13528. Geneva: ISO, 2015.
- Fraser CG, Petersen PH, Libeer JC, Ricós C. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8–12.
- Ricós C, Álvarez V, Cava F, García-Lario JV, Hernandez A, Jimenez CV, et al. Desirable specifications for total error, imprecision, and bias, derived from intra- and inter-individual biologic variation. The 2014 update. Original publication available from (Spanish only). [http://www.seqc.es/es/comisiones/comision-de-calidad-analitica/\\_id:4/](http://www.seqc.es/es/comisiones/comision-de-calidad-analitica/_id:4/). Accessed: January 2019. Also available on (English): <http://www.westgard.com/biodatabase1.htm>. Accessed: January 2019.
- Coşkun A, Carobene A, Kilercik M, Serteser M, Sandberg S, Aarsand AK, et al. Within-subject and between-subject biological variation estimates of 21 hematological parameters in 30 healthy subjects. *Clin Chem Lab Med* 2018;56:1309–18.
- Soumali MR, Van Blerk M, Akharif A, Albarède S, Kessler D, Gutierrez G, et al. A new approach to define acceptance limits for hematology in external quality assessment schemes. *Clin Chem Lab Med* 2017;55:1936–42.
- Horsti J, Uppa H, Vilpo JA. Poor agreement among prothrombin time international normalized ratio methods: comparison of seven commercial reagents. *Clin Chem* 2005;51:553–60.
- Bowyer A, Kitchen S, Makris M. The responsiveness of different APTT reagents to mild factor VIII, IX and XI deficiencies. *Int J Lab Hematol* 2011;33:154–8.
- Ricós C, Álvarez V, Minchinela J, Fernández-Calle P, Perich C, Boned B, et al. Biologic variation approach to daily laboratory. *Clin Lab Med* 2017;37:47–56.
- Ricós C, Iglesias N, García-Lario JV, Simón M, Cava F, Hernández A, et al. Within-subject biological variation in disease: collated data and clinical consequences. *Ann Clin Biochem* 2007;44:343–52.
- Molina A, Guiñón L, Pérez A, Segurana A, Bedini JL, Reverter JC, et al. (2018). State of the art vs. biological variability: Comparison on hematology parameters using Spanish EQAS data. *Int J Lab Hematol* 2011;40:284–91.
- Westgard J. 2016 State of the Art Hematology Performance Specifications. <https://www.westgard.com/rcpa.htm>. Accessed: January 2019.

## **1.2. Segundo estudio**

**Especificaciones de calidad analíticas basadas en el estado del arte para el cribado neonatal.**

### **1.2.1. Presentación**

El objetivo del programa de cribado neonatal es la detección de trastornos congénitos antes de la aparición de los síntomas, de forma que pueda iniciarse el tratamiento en los recién nacidos lo antes posible con la finalidad de prevenir o paliar las consecuencias a largo plazo (73). El Sistema Nacional de Salud tiene establecido que cómo mínimo se realice el cribado neonatal de 7 enfermedades. Es competencia de las comunidades autónomas el ampliar esta cartera de servicios mínima. Concretamente en Cataluña, el programa de cribado neonatal consta de 24 enfermedades (74). A continuación, se citan las enfermedades incluidas actualmente en del programa de cribado de Cataluña y se indica cuáles de ellas son obligatorias a nivel estatal (\*):

#### Trastornos del metabolismo de los aminoácidos:

Hiperfenilalaninemias – Fenilcetonuria (\*)

Enfermedad del Jarabe de arce

Tirosinemia tipo I

Citrulinemia tipo I

Homocistinuria

#### Trastornos del metabolismo de los ácidos orgánicos:

Acidemia propiónica

Acidemia metilmalónica (Mut, Cbl A, CblB)

Acidemia metilmalónica (Cbl C, CblD)

Acidemia isovalérica

Deficiencia de 3-OH-3-metilglutaril-CoA liasa

Deficiencia de  $\beta$ -cetotilasa

Acidúria glutárica tipo I (\*)

Trastornos del metabolismo de los ácidos grasos

Deficiencia de acil-CoA deshidrogenasa de cadena media (\*)

Deficiencia de acil-CoA deshidrogenasa de cadena muy larga (\*)

Deficiencia de proteína trifuncional mitocondrial

Deficiencia múltiple de acil-CoA deshidrogenasas

Deficiencia en la captación celular de la carnitina

Deficiencia de carnitina palmitoiltransferasa 1

Deficiencia de carnitina palmitoiltransferasa 2

Deficiencia de carnitina- acilcarnitina translocasa

Otras enfermedades

Hipotiroidismo congénito (\*)

Fibrosis quística (\*)

Enfermedad de células falciformes (\*)

Inmunodeficiencia combinada grave

El proceso analítico del cribado neonatal incluye actividades en las que interactúan, entre otros, los centros maternos en los que se obtienen las muestras, los laboratorios donde se analizan y las unidades de referencia donde se estudian los casos de cribado positivos. Para asegurar la calidad del proceso analítico debe disponerse, por tanto, de un plan de control de calidad que considere todas las actividades que lo componen.

En lo que respecta a la actividad que se realiza en los laboratorios, un aspecto clave para el buen funcionamiento del programa es la adecuación del rendimiento analítico de los procedimientos de medida al uso clínico previsto. Para ello, y tal como requiere la norma ISO 15189 (67), el primer paso es la selección de especificaciones de calidad. Sin embargo, para las magnitudes incluidas en el cribado neonatal no existen especificaciones de calidad documentadas acorde a ninguno de los modelos definidos en la conferencia estratégica de la EFLM realizada en 2014



(56). En este caso en particular, no es factible disponer de datos de variabilidad biológica debido a que el cribado va dirigido a la población de recién nacidos. Por tanto, las posibilidades se limitan a obtener especificaciones de calidad basadas en el estado del arte (modelo 3) utilizando los datos proporcionados por los programas de intercomparación.

El objetivo de este estudio fue proporcionar especificaciones de calidad de error total basadas en el estado del arte para las magnitudes incluidas en el programa de cribado neonatal. Para la realización de este estudio, la Asociación Española de Cribado Neonatal (AECNE) permitió a la doctoranda el uso de los datos de la participación de 19 laboratorios españoles en el programa de intercomparación de cribado neonatal en el periodo comprendido entre los años 2015 y 2018. Para analizar los resultados se generó de forma manual una base de datos con más de 20.000 resultados.

El programa de cribado neonatal nacional incluía 16 magnitudes : hormona estimulante de la tiroides (TSH), tripsinógeno inmunoreactivo (TIR), fenilalanina (FEN), tirosina (TYR), carnitina libre (C0), acetilcarnitina (C2), propionilcarnitina (C3), butirilcarnitina (C4), isovalerilcarnitina (C5), glutarilcarnitina (C5DC), hexanoilcarnitina (C6), octanoilcarnitina (C8), decanoilcarnitina (C10), miristoilcarnitina (C14), palmitoilcarnitina (C16), estearoilcarnitina (C18). Cada envío constaba de tres materiales de control. Los materiales de control presentaban la peculiaridad de ser muestras de sangre humana a las que se habían añadido los analitos en diferentes concentraciones y que, posteriormente, habían sido impregnadas en papel de filtro validado para este uso.

### **1.2.2. Resultados y discusión**

Se obtuvieron unas especificaciones de calidad analítica de error total a partir del cálculo del percentil 90 del error de medida teniendo en cuenta el 75 % de los mejores resultados de cada laboratorio. Para ello, para cada magnitud primero se ordenaron los errores de medida de cada

laboratorio y se eliminaron el 25 % de los valores más elevados. A continuación, se calculó el percentil 90 de los errores de medida teniendo en cuenta todos los laboratorios.

En primer lugar, se realizó la estimación del error total teniendo en cuenta todos los procedimientos de medida. En segundo lugar, se obtuvieron considerando únicamente los procedimientos de medida más utilizados por los participantes y, además, se calcularon para dos rangos de concentración diferentes acordes a los valores de decisión clínica establecidos en el laboratorio para cada magnitud.

Al analizar las especificaciones de calidad obtenidas se observó que, al calcularlas teniendo en cuenta todos los procedimientos de medida de los participantes, el error total obtenido era mayor que si se consideraba únicamente el procedimiento de medida empleado por la mayoría. Este hallazgo no se observó para la TSH ni la TIR por el hecho que sólo uno de los laboratorios de los que emitieron resultados para estas magnitudes utilizaba un procedimiento de medida alternativo. La disminución del error total observada para el resto de las magnitudes sugería que los procedimientos de medida utilizados por una minoría tenían un peor rendimiento analítico. Sin embargo, por no disponer del número de participantes necesario para agrupar los resultados en función del procedimiento de medida utilizado (mínimo 10 participantes), no se pudieron calcular especificaciones de calidad de error total para todos los procedimientos de medida. Ante esta situación, la opción más adecuada para evaluar el rendimiento analítico de los métodos minoritarios sería que los laboratorios que los utilicen comparasen su error de medida respecto a la especificación de error total calculada incluyendo todos los procedimientos de medida. De todos modos, puesto que las especificaciones de calidad propuestas respondían a unos requisitos mínimos de calidad analítica, si los procedimientos de medida minoritarios no lograsen alcanzar la calidad necesaria para el uso previsto, y ante la existencia de otras metodologías con un mejor rendimiento analítico, debería considerarse su reemplazo.

Al obtener las especificaciones de calidad para diferentes rangos de concentración a partir de los datos de los participantes que utilizaron el método mayoritario, se puso de manifiesto que, en general, el rendimiento analítico era mejor en el rango alto de concentraciones. Este hecho supone una mayor seguridad en lo que respecta a la emisión de resultados falsos negativos, ya que la mayoría de los valores patológicos corresponden a concentraciones elevadas. Para algunas magnitudes como TSH, C5DC, C6, C8 y C14 se observó que las diferencias en el error total obtenido en función del rango de concentración eran destacables. En la Figura 5 se muestra, para las magnitudes citadas, la variación del error total observada para cada uno de los laboratorios participantes, teniendo en cuenta diferentes rangos de concentración acorde a los niveles de decisión clínica.

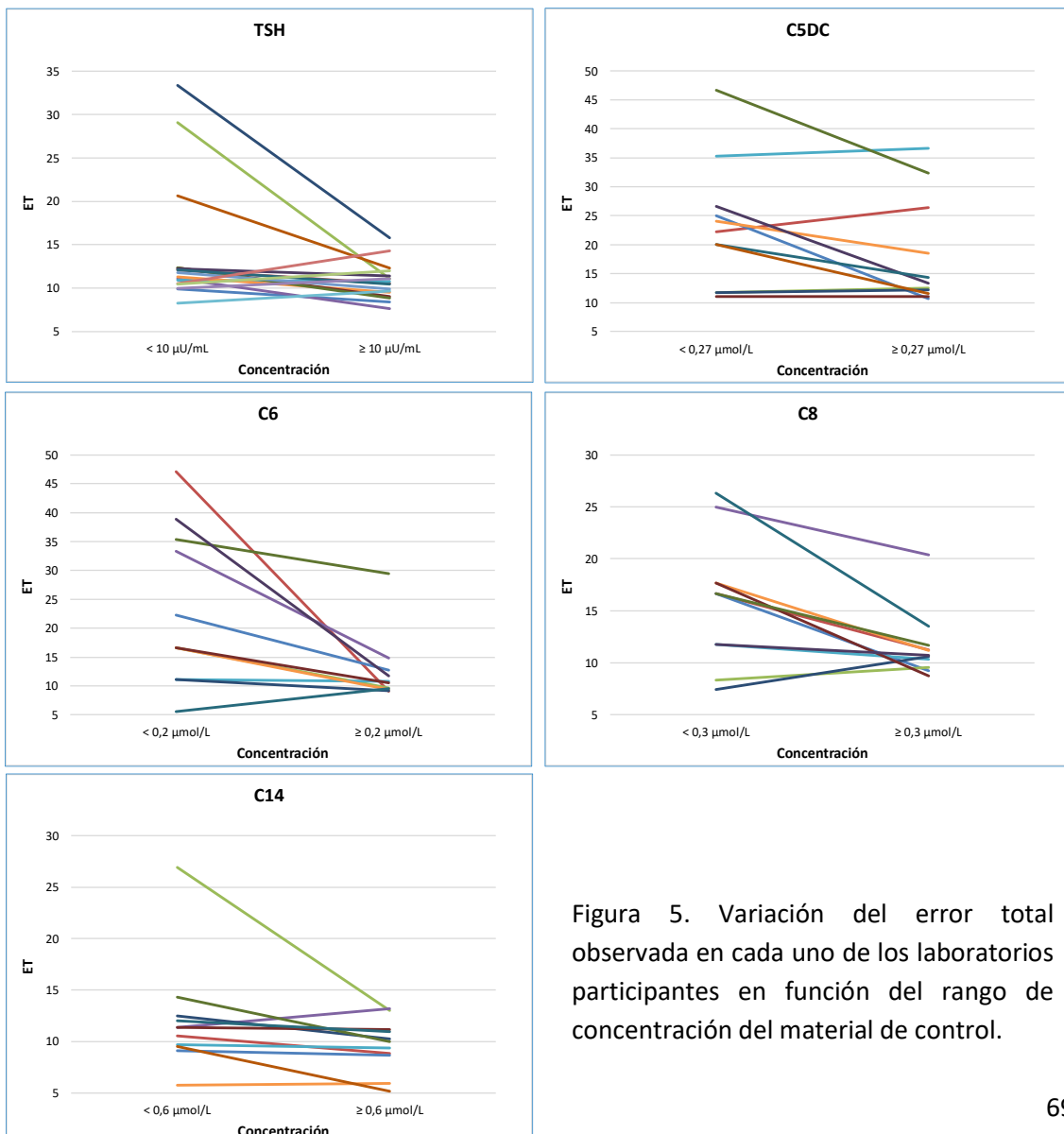


Figura 5. Variación del error total observada en cada uno de los laboratorios participantes en función del rango de concentración del material de control.

Este hallazgo implica que, en caso que un laboratorio establezca especificaciones de calidad de error total sin considerar que éstas pueden ser dependientes de la concentración, cuando no logre alcanzarlas interprete que tiene un problema analítico cuando, en realidad, lo que ocurre es que a bajas concentraciones el rendimiento analítico de todos los laboratorios que utilizan el mismo procedimiento de medida es peor. Por este motivo la información aportada por este estudio es de gran valor para realizar una evaluación adecuada de los resultados de los programas de intercomparación de cribado neonatal y, en último término, para tomar decisiones basadas en el rendimiento analítico del procedimiento de medida que sean apropiadas.

Las especificaciones de calidad propuestas presentan algunas limitaciones. La primera se debe a que no pueden relacionarse con la situación clínica del paciente ni con la variabilidad biológica. Aun así, en ocasiones se ha observado que éstas pueden resultar más estrictas que las basadas en la variabilidad biológica (75, 79). También se ha observado que, para algunas magnitudes de otras áreas de especialización, la estimación de especificaciones de calidad usando un percentil más exigente puede ser más acorde a las necesidades clínicas que las especificaciones basadas en la variabilidad biológica (48). La segunda limitación es que se ha de tener presente que no se han proporcionado diferentes especificaciones de calidad para cada posible uso clínico, pero esta información no suele ser proporcionada por los PEECs, a pesar de que sería de un gran interés. Por lo tanto, si el uso previsto difiere de lo habitual, es el profesional del laboratorio clínico quien debería adecuar las especificaciones a la calidad requerida (77).

### **1.2.3. Conclusión**

Gracias a la realización de este estudio ha sido posible establecer especificaciones de calidad de error total basadas en el estado del arte para las magnitudes incluidas en el programa de cribado neonatal.

Las especificaciones de calidad propuestas han sido calculadas como un mínimo de calidad, por lo que pueden ser de referencia para otros laboratorios de cribado neonatal. Además, conocer cómo varía el error de medida en función de la concentración, puede evitar acciones innecesarias derivadas de una evaluación incorrecta del rendimiento analítico.

#### **1.2.4. Cita al artículo publicado**

Guiñón L, Soler A, Molina A, Marín JL, López RM, García-Villoria J, Sahuquillo A, Alvarez L.

**Analytical performance specifications based on the state-of-the-art for the newborn screening.** Clinica Chimica Acta. 2020;510:117-20.

#### **1.2.5. Publicación completa**





## Analytical performance specifications based on the state-of-the-art for the newborn screening



Leonor Guinón<sup>a,b,c,\*</sup>, Anna Soler<sup>a,1</sup>, Angel Molina<sup>d,1</sup>, José Luis Marín<sup>e,1</sup>, Rosa María López<sup>e,1</sup>, Judit García-Villoria<sup>e,1</sup>, Ángeles Sahuquillo<sup>f,2</sup>, Luisa Alvarez<sup>a,1</sup>

<sup>a</sup> Quality Department. Biomedical Diagnostic Center. Hospital Clínic of Barcelona

<sup>b</sup> Quality Department. Laboratories. Hospital de la Santa Creu i Sant Pau of Barcelona

<sup>c</sup> Biochemistry Department. Hospital de la Santa Creu i Sant Pau of Barcelona

<sup>d</sup> Core Laboratory. Biochemistry and Molecular Genetics Department. Biomedical Diagnostic Center. Hospital Clínic of Barcelona

<sup>e</sup> Division of Inborn Errors of Metabolism. Biochemistry and Molecular Genetics Department. Biomedical Diagnostic Center. Hospital Clínic of Barcelona

<sup>f</sup> Section of Analytical Chemistry. Faculty of Chemistry. University of Barcelona

### ARTICLE INFO

#### Keywords:

Analytical performance specifications  
Total error  
State-of-the-art  
Measurement error  
Quality specifications

### ABSTRACT

**Introduction:** For the measurands of the newborn screening (NBS), there are no analytical performance specifications (APS) available based on the Milan consensus Models. The objective is to provide total error (TE) APS based on the state-of-the-art (SOTA) for the NBS.

**Material and Methods:** 23,662 results were collected from the Spanish NBS EQA scheme between May 2015 and September 2018. Measurands included: thyroid-stimulating hormone (TSH), immunoreactive trypsinogen (IRT), phenylalanine (Phe), tyrosine (Tyr), free carnitine (C0), acetylcarnitine (C2), propionylcarnitine (C3), butyrylcarnitine (C4), isovalerylcarnitine (C5), glutarylcarnitine (C5DC), hexanoylcarnitine (C6), octanoylcarnitine (C8), decanoylcarnitine (C10), myristoylcarnitine (C14), palmitoylcarnitine (C16), stearoylcarnitine (C18).

TE APS were calculated as the 90th percentile of the measurement errors, considering 75% of the best results from each laboratory only. It was also studied whether the analytical performance was concentration-dependent. **Results:** When TE APS were calculated including all methods, TSH, IRT, C16 and C18 showed the best analytical performance and Phe, C5DC and C10 showed the worst. Generally, TE APS decreased when considering only majority methods and higher TE APS were obtained for lower concentrations.

**Discussion:** Due to the lack of APS based on superior models, the proposed TE APS based on the SOTA can help NBS laboratories to set quality specifications.

### 1. Introduction

Newborn screening (NBS) is a public health activity intended to identify congenital disorders prior to the onset of symptoms. The aim of the NBS is, as in other screening programs, to provide early intervention for affected individuals to assure the best possible health outcome [1]. The NBS involves health professionals from different areas and different institutions that must work together to make the program a success. The information given to the population, the regular training of the health personnel, the logistics and the laboratory methodology are the main components of NBS [2,3], so measures have to be taken to control the quality of these procedures.

The NBS has some peculiarities that must be considered. Among them, samples must be collected at a specific time since some analyte levels vary during the neonatal period. A few drops of blood from a heel prick are collected onto an absorbent paper card. These peculiarities increase the variability of the analytical procedure but, as in any screening program, every positive result requires confirmation [4].

Ensuring the quality of the analytical procedures in NBS is one of the essential steps for the program to function properly [2]. In this regard, the international standard for accreditation of clinical laboratories ISO 15189:2012 requires the establishment of quality specifications [5] to ensure that the performance of the analytical procedures in clinical laboratories allows an adequate patient care [6]. In order to

\* Corresponding author at: Sant Quintí, 89 - 08041 Barcelona Spain.

E-mail addresses: [leonorguinon@yahoo.es](mailto:leonorguinon@yahoo.es) (L. Guinón), [soler@clinic.cat](mailto:soler@clinic.cat) (A. Soler), [amolinas@clinic.cat](mailto:amolinas@clinic.cat) (A. Molina), [jlmarin@clinic.cat](mailto:jlmarin@clinic.cat) (J.L. Marín), [rmlopez@clinic.cat](mailto:rmlopez@clinic.cat) (R.M. López), [jugarcia@clinic.cat](mailto:jugarcia@clinic.cat) (J. García-Villoria), [angels.sahuquillo@ub.edu](mailto:angels.sahuquillo@ub.edu) (Á. Sahuquillo), [alvarez@clinic.cat](mailto:alvarez@clinic.cat) (L. Alvarez).

<sup>1</sup> postal address: Villarroel 170 - 08036 Barcelona, Spain.

<sup>2</sup> postal address: Martí i Franquès 1-11, 08028 Barcelona, Spain.

<https://doi.org/10.1016/j.cca.2020.07.006>

Received 11 April 2020; Received in revised form 17 June 2020; Accepted 3 July 2020

Available online 09 July 2020

0009-8981/ © 2020 Elsevier B.V. All rights reserved.

keep the analytical error under control, the laboratory must compare their performance characteristics (precision, trueness and accuracy) against well-documented analytical performance specifications (APS). Different types of error can affect analytical measurements. The measurement error (ME) is a combination of random and systematic errors, and it is defined as the closeness of the agreement between a measured quantity value and a reference quantity value [7]. It can be estimated by means of the External Quality Assessment (EQA) schemes. A ME that exceeds the total error (TE) APS should alert the laboratory that their analytical procedure could be producing erroneous results with the consequent impact on clinical decisions and therefore on patient safety [8].

Three different models to assign APS were defined in the 1st European Federation of Clinical Chemistry and Laboratory Medicine (EFLM) Strategic Conference held in 2014 in Milan [9]. Model 1 is based on the effect of the analytical performance on the clinical outcome and Model 2 on biological variation of the measurand. Either of these models is preferred when setting APS, since they ensure that analytical performance meets clinical needs. Model 3, which is based on the state-of-the-art (SOTA) of the methodology, is supposed to be used only when there is no data available in models 1 or 2, or when it is not possible to apply either of the two models. This model is related to the analytical performance that can be achieved with the current technology [9,10]. One of the ways to establish APS based on the SOTA is to use the information provided by the EQA schemes [10]. However, as EQAs cannot provide APS for every possible intended use of a measurand, laboratory specialists should establish their own required response if the use differs from the generally expected one [11].

Despite the importance of ensuring the quality of the NBS programs, there are no published quality specifications available for the measurands included [12]. The aim of this work is to provide TE APS based on the SOTA for measurands included in the NBS in order to control their analytical performance.

## 2. Material and methods

A total of 23,662 results from 19 laboratories were collected from the NBS EQA scheme which was organized by the Spanish Association of Newborn Screening (AECNE) during the period between May 2015 and September 2018. The control material samples provided by the EQA were prepared from human whole blood of 50% hematocrit enriched with predetermined quantities of selected analytes and dispensed in 100 µL aliquots on a Whatman™ 903 card (GE Healthcare Life Sciences Corporation (Marlborough, MA)). Each survey included three samples that were analyzed separately; therefore, a total of 123 quality control materials were provided for each measurand. Due to the particularities of the measurands studied, it was not possible to send commutable materials or certified reference materials by the EQA provider.

Measurands included in the study were thyroid-stimulating hormone (TSH), immunoreactive trypsinogen (IRT), phenylalanine (Phe), tyrosine (Tyr), and acylcarnitines profile: free carnitine (C0), acetylcarnitine (C2), propionylcarnitine (C3), butyrylcarnitine (C4), isovalerylcarnitine (C5), glutaryl carnitine (C5DC), hexanoylcarnitine (C6), octanoylcarnitine (C8), decanoylcarnitine (C10), myristoylcarnitine (C14), palmitoylcarnitine (C16) and stearoylcarnitine (C18).

The number of participating laboratories for each measurand, as well as the number of results sent, the concentration range of the control materials analyzed and the units in which the results are expressed, are shown in Table 1. The control materials included normal and pathological samples. Therefore, clinical decision intervals of each measurand were represented in the concentration ranges.

The methodologies used by the different laboratories are shown in Table 2. Most of the participants perform the measurement of TSH and IRT by Time-resolved fluorescence and the measurement of Phe, Tyr and Acylcarnitines (non-derivatized method) by tandem mass

**Table 1**

Number of participating laboratories, number of results obtained, concentration range of the control materials analyzed and units in which the results are expressed.

Measurand	Number of laboratories	Number of results	Concentration range	Units
TSH	18	1,847	2.71–114.53	µU/mL
IRT	17	1,789	24.48–414.74	ng/mL
Phe	19	1,903	41.19–389.94	µmol/L
Tyr	17	1,581	88.45–713.33	µmol/L
C0	16	1,461	12.13–95.97	µmol/L
C2	15	1,172	0.86–73.87	µmol/L
C3	15	1,443	0.32–19.57	µmol/L
C4	15	1,195	0.23–4.70	µmol/L
C5	16	1,436	0.20–4.14	µmol/L
C5DC	16	1,523	0.08–2.11	µmol/L
C6	16	1,136	0.16–2.39	µmol/L
C8	16	1,420	0.17–3.45	µmol/L
C10	16	1,438	0.11–2.87	µmol/L
C14	16	1,302	0.21–2.71	µmol/L
C16	16	1,512	2.66–22.69	µmol/L
C18	16	1,504	0.86–7.44	µmol/L

(TSH: thyroid-stimulating hormone; IRT: immunoreactive trypsinogen; Phe: phenylalanine; Tyr: tyrosine; C0: free carnitine; C2: acetylcarnitine; C3: propionylcarnitine; C4: butyrylcarnitine; C5: isovalerylcarnitine; C5DC: glutaryl carnitine; C6: hexanoylcarnitine; C8: octanoylcarnitine; C10: decanoylcarnitine; C14: myristoylcarnitine; C16: palmitoylcarnitine; C18: stearoylcarnitine.)

**Table 2**

Methods used by the participants for the measurands included in the Spanish newborn screening External Quality Assessment scheme.

Method	Measurand				
	TSH	IRT	Phe	Tyr	Acylcarnitines
Time-resolved fluorescence	17	16	–	–	–
Tandem mass spectrometry	–	–	13*	15**	16***
Fluorimetry	–	–	4	–	–
Thin layer chromatography	–	–	1	1	–
High performance liquid chromatography	–	–	1	1	–
Non specified	1	1	–	–	–

(TSH: thyroid-stimulating hormone; IRT: immunoreactive trypsinogen; Phe: phenylalanine; Tyr: tyrosine.)

\* non-derivatized method = 12, derivatized method = 1.

\*\* non-derivatized method = 13, derivatized method = 2.

\*\*\* non-derivatized method = 12, derivatized method = 4.

spectrometry. Most participants used certified commercial kits with labeled internal standards; only two participants (that used derivatized methods) did not have commercial procedures.

For each result of the EQA surveys, the ME was calculated as the distance in percentage from the value reported by the participant to the target value:

$$ME (\%) = |(\text{Result} - \text{Target value})| * 100 / \text{Target value}$$

The recommendations described in ISO 13528:2015 were followed to calculate the target values used to estimate the ME [13]. Thus, the target values were calculated as the mean values derived from participant results, since no international standards were available and the number of participating laboratories in the EQA scheme was small.

The 90th percentile of the ME for each measurand was calculated as an indicator of analytical performance of current technology, considering 75% of the best results from each laboratory only. The obtained values were proposed as TE APS. These TE APS were first calculated by taking all methods into account and then, by considering only majority methods. Moreover, in order to assess whether the analytical performance was concentration-dependent, TE APS were also calculated (considering only the methods most used) for two different concentration ranges according to the clinical decision values used in



**Table 3**

Total error analytical performance specifications obtained for the measurands included in the Spanish newborn screening External Quality Assessment scheme.

Measurand	TE APS (%)	
	All methods	Majority methods
TSH	17.02	17.60
IRT	15.19	15.24
Phe	42.03	18.33
Tyr	24.43	20.56
C0	30.15	18.12
C2	22.59	19.19
C3	27.88	15.02
C4	32.13	15.49
C5	23.23	18.85
C5DC	36.61	35.51
C6	25.00	15.08
C8	23.96	15.78
C10	38.31	20.12
C14	29.16	16.01
C16	18.50	12.42
C18	18.46	11.27

(TE: total error; APS: analytical performance specifications; TSH: thyroid-stimulating hormone; IRT: immunoreactive trypsinogen; Phe: phenylalanine; Tyr: tyrosine; C0: free carnitina; C2: acetylcarnitine; C3: propionylcarnitine; C4: butyrylcarnitine; C5: isovalerylcarnitine; C5DC: glutaryl carnitine; C6: hexanoylcarnitine; C8: octanoylcarnitine; C10: decanoylcarnitine; C14: myristoylcarnitine; C16: palmitoylcarnitine; C18: stearyl carnitine.)

our laboratory.

### 3. Results

TE APS obtained for each measurand considering the entire analytical range are shown in Table 3. When analyzing the data that took all the participating laboratories into account, it was observed that the measurands that had the best analytical performance were TSH, IRT, C16 and C18, with a TE APS lower than 20%. For a greater number of measurands such as Tyr, C0, C2, C3, C4, C5, C6, C8 and C14, the TE APS obtained were between 20 and 35%. Phe, C5DC and C10 showed the worst analytical performance, with a TE APS higher than 35%. When estimating the TE APS that took only the methods most used into account, the TE APS decreased for all the measurands except for TSH and IRT in which a slight increase was observed. The decrease was especially striking for Phe, C0, C3, C4, C10 and C14, where the values decreased by more than 12 percentage points.

When analyzing the results depending on the concentration range, obtained from the most used methods (Table 4), higher TE APS were obtained in general for lower concentrations, excluding Phe, C0, C5 C10 and C18 that obtained slightly lower TE. At higher concentrations some measurands such as TSH, C5DC, C6, C8 and C14 showed a marked decrease of more than 9 percentage points regarding the TE APS obtained at lower concentration ranges.

### 4. Discussion

The results provided can serve as guidance to establish TE APS for the measurands of the NBS. These TE APS based on the SOTA can be recommended as minimum analytical quality requirements, thus a poor performance is assumed for those laboratories that can not reach these requirements [14].

The TE APS obtained considering all methods were higher than those calculated including only the data of the laboratories that performed the tests with the majority method (Table 3). On the one hand, it must be considered that the good performance of IRT and TSH could be due to the method used by all laboratories (except one not specified), usually on the same analytical platform with a fully automated system.

**Table 4**

Total error analytical performance specifications obtained for the different concentration ranges studied.

Measurand	Units	Concentration	TE APS (%)	Concentration	TE APS (%)
TSH	μU/mL	< 10	24.03	≥ 10	14.89
IRT	ng/mL	< 50	15.55	≥ 50	15.54
Phe	μmol/L	< 110	18.33	≥ 110	19.92
Tyr	μmol/L	< 400	20.56	≥ 400	19.40
C0	μmol/L	< 58	17.27	≥ 58	19.02
C2	μmol/L	< 8	22.18	≥ 8	17.83
C3	μmol/L	< 3.5	15.36	≥ 3.5	13.86
C4	μmol/L	< 1.1	15.79	≥ 1.1	13.45
C5	μmol/L	< 0.9	18.73	≥ 0.9	19.60
C5DC	μmol/L	< 0.27	45.53	≥ 0.27	36.24
C6	μmol/L	< 0.2	38.89	≥ 0.2	14.65
C8	μmol/L	< 0.3	25.00	≥ 0.3	13.34
C10	μmol/L	< 0.4	20.28	≥ 0.4	20.76
C14	μmol/L	< 0.6	25.66	≥ 0.6	13.15
C16	μmol/L	< 7	12.40	≥ 7	12.37
C18	μmol/L	< 2.4	10.63	≥ 2.4	14.73

(TE: total error; APS: analytical performance specifications; TSH: thyroid-stimulating hormone; IRT: immunoreactive trypsinogen; Phe: phenylalanine; Tyr: tyrosine; C0: free carnitina; C2: acetylcarnitine; C3: propionylcarnitine; C4: butyrylcarnitine; C5: isovalerylcarnitine; C5DC: glutaryl carnitine; C6: hexanoylcarnitine; C8: octanoylcarnitine; C10: decanoylcarnitine; C14: myristoylcarnitine; C16: palmitoylcarnitine; C18: stearyl carnitine.)

The other analytes were measured with less homogeneous instruments and, moreover, sample preparation could not be so automated (i.e. derivatization or not), what could be affecting the TE APS obtained when considering all methods. On the other hand, the poor performance of C5DC could be related to the detection of a dicarboxylic acylcarnitine in the positive ion mode, as after ionization the two negative charges of carboxyl groups can enter in conflict with the positive ionization. We could not prove if derivatized methods, that can quantify the isobaric compounds, have a better performance than non-derivatized methods because there were not enough participants that used this methodology.

The results suggest that minority methods such as fluorimetry, thin layer chromatography or high performance liquid chromatography may have a poorer analytical performance and, therefore, TE APS obtained from majority methods would not be suitable for them. We could not make groups of laboratories to calculate a method-specific target for a subset of results due to the small number of participants in the EQA scheme. Therefore, it was not possible to calculate TE APS for each methodology. Hence, in the absence of specific TE APS, the use of the TE APS obtained from all methods would be more appropriate for those laboratories that use minority methods. It is important to note that, since there are methodologies with higher analytical performance available, replacing minority methods should be considered if they do not meet the intended use.

Regarding the results obtained for the different concentration ranges, it is observed that most of the measurands showed a better performance at higher concentrations (Table 4). This fact is important since most of pathological situations respond to high concentrations for the measurands studied. Committing lower ME at the clinical decision levels means a better control of false negatives results. It is important to note that, the ME obtained at low concentration values might exceed the quality specifications when no concentration-dependent TE APS were established (TE APS from majority methods). This might suggest that there is a problem in the methodologies used, when in fact it is due to worse performance at low concentrations. This shows the importance of establishing adequate goals, since they could determine the proper assessment of the analytical performance. Other biological measurands have also shown differences in the analytical performance depending on the concentration [15]. Therefore, for those measurands that present greater differences depending on the concentration (TSH, C5DC, C6, C8

and C14), it might be better to evaluate the analytical performance using different TE APS.

For the measurands included in the NBS, there are no quality specifications neither based on the effect on clinical outcomes, nor based on biological variability. Obtaining data based on components of biological variation can be more complicated than usual when dealing with newborn population. In these cases, the only feasible option is to obtain APS based on the SOTA. EQA schemes constitute an ideal setting to study the analytical performance. Given the necessary centralization of screening programs, the number of NBS laboratories is low and this makes it difficult to analyze the analytical performance in these measurands. Even so, Spain is one of the European countries where a greater number of laboratories perform these tests [16]. Thanks to the participation of the laboratories in the Spanish EQA in NBS, it has been possible to obtain highly useful quality specifications. We cannot know whether APS based on the SOTA satisfy clinical requirements [17,18] but literature reveals that in some cases the current technology allows to achieve TE APS that are more demanding than those based on biological variability [19,20]. In the absence of APS based on Models 1 and 2, specialists should assess whether the TE APS provided are suitable for the intended use.

## 5. Conclusions

The quality specifications obtained can help NBS laboratories to establish acceptance criteria for the analytical performance of their methodologies. Despite the limitations of the study, we consider that the results provided by the present study are of great value given the absence of quality specifications for the NBS measurands based on superior models according to the Milan consensus statement.

## Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## CRedit authorship contribution statement

**Leonor Guiñón:** Conceptualization, Methodology, Formal analysis, Writing - original draft. **Anna Soler:** Writing - review & editing. **Angel Molina:** Formal analysis, Writing - original draft. **José Luis Marín:** Writing - review & editing. **Rosa María López:** Writing - review & editing. **Judit García-Villoria:** Writing - review & editing. **Ángeles Sahuquillo:** Visualization. **Luisa Alvarez:** Visualization, Supervision.

## Acknowledgements

We would like to acknowledge Dr. Elena Dulín, president of the Spanish Society for Newborn Screening (AECNE) during the period of the study, for allowing us to use EQA data to carry out this study.

## References

- [1] E. Martínez-Morillo, B. Prieto, F.V. Álvarez, Challenges for worldwide harmonization of newborn screening programs, *Clin. Chem.* 62 (5) (2016) 689–698, <https://doi.org/10.1373/clinchem.2015.240903>.
- [2] T. Torresani, Quality control requirements in neonatal screening, *Eur. J. Pediatr.* 162 (1) (2003) S54–S56, <https://doi.org/10.1007/s00431-003-1353-1>.
- [3] J.L. Dhondt, Neonatal screening: from the “Guthrie age” to the “genetic age”, *J. Inher. Metab. Dis.* 30 (4) (2007) 418–422, <https://doi.org/10.1007/s10545-007-0624-9>.
- [4] J.J. Pitt, Newborn screening, *Clin. Biochem. Rev.* 31 (2) (2010) 57–68.
- [5] ISO 15189. Medical laboratories. Requirements for quality and competence. International Organization for Standardization, 2012.
- [6] C.G. Fraser, A. Kallner, D. Kenny, P. Hyltoft, Introduction: strategies to set global quality specifications in laboratory medicine, *Scand. J. Clin. Lab. Invest.* 59 (7) (1999) 477–478, <https://doi.org/10.1080/00365519950185184>.
- [7] International vocabulary of metrology – Basic and general concepts and associated terms (VIM), 3rd edition, Joint Committee for Guides in Metrology, 2012.
- [8] G.R.D. Jones, Analytical performance specifications for EQA schemes—need for harmonisation, *Clin. Chem. Lab. Med.* 53 (6) (2015) 919–924, <https://doi.org/10.1515/cclm-2014-1268>.
- [9] S. Sandberg, F.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, 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.* 53 (2015) 833–835, <https://doi.org/10.1515/cclm-2015-0067>.
- [10] F. Ceriotti, P. Fernandez-Calle, G.G. Klee, G. Nordin, S. Sandberg, T. Streichert, et al., Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference, *Clin. Chem. Lab. Med.* 55 (2) (2017) 189–194, <https://doi.org/10.1515/cclm-2016-0091>.
- [11] G.R.D. Jones, S. Albaredo, D. Kessler, F. MacKenzie, J. Mammen, M. Pedersen, A. Stavelin, et al., Analytical performance specifications for external quality assessment—definitions and descriptions, *Clin. Chem. Lab. Med.* 55 (7) (2017) 949–955, <https://doi.org/10.1515/cclm-2017-0151>.
- [12] D.E. Castiñeras, M.L. Couce, J.L. Marín, D. González-Lamuño, H. Rocha, Newborn screening for metabolic disorders in Spain and worldwide, *An. Pediatr.* 91 (2) (2019) 128.e1–128.e14, <https://doi.org/10.1016/j.anpedi.2019.05.007>.
- [13] ISO 13528. Statistical methods for use in proficiency testing by interlaboratory comparison, International Organization for Standardization, 2015.
- [14] A. Salas, C. Ricós, E. Prada, F. Ramón, J. Moráncho, J.M. Jou, R. Blazquez, State-of-the-art Approach to Goal Setting, *Clin. Lab. Med.* 37 (1) (2017) 73–84, <https://doi.org/10.1016/j.cll.2016.09.007>.
- [15] A. Molina, J. Alcaraz, L. Guiñón, A. Pérez, A. Segurana, J.C. Reverter, J.L. Bedini, et al., Study of the analytical performance at different concentrations of hematological parameters using Spanish EQAS data, *Clin. Chem. Lab. Med.* 57 (12) (2019) 1980–1987, <https://doi.org/10.1515/cclm-2019-0108>.
- [16] J.G. Loeber, P. Burgard, M.C. Cornel, T. Rigter, S.S. Weinreich, K. Rupp, G.F. Hoffmann, et al., Newborn screening programmes in Europe; arguments and efforts regarding harmonization. Part 1—From blood spot to screening result, *J. Inher. Metab. Dis.* 35 (4) (2012) 603–611, <https://doi.org/10.1007/s10545-012-9483-0>.
- [17] R. Haecckel, W. Wosniok, T. Streichert, Optimizing the use of the “state-of-the-art” performance criteria, *Clin. Chem. Lab. Med.* 53 (6) (2015) 887–891, <https://doi.org/10.1515/cclm-2014-1201>.
- [18] C.G. Fraser, P.H. Petersen, Analytical performance characteristics should be judged against objective quality specifications, *Clin. Chem.* 45 (3) (1999) 321–323, <https://doi.org/10.1093/clinchem/45.3.321>.
- [19] A. Molina, L. Guiñón, A. Perez, A. Segurana, J.L. Bedini, J.C. Reverter, A. Merino, State of the art vs biological variability: Comparison on hematology parameters using Spanish EQAS data, *Int. J. Lab. Hematol.* 40 (2018) 284–291, <https://doi.org/10.1111/ijlh.12783>.
- [20] J.A. Castilla, J. Moráncho-Zaragoza, J. Aguilar, R. Prats-Gimenez, M.C. Gonzalvo, E. Fernández-Pardo, C. Alvarez, et al., Quality specifications for seminal parameters based on the state of the art, *Hum. Reprod.* 20 (9) (2005) 2573–2578, <https://doi.org/10.1093/humrep/dei061>.

**2. Segundo objetivo: Establecer una metodología basada en el análisis de riesgos y la métrica sigma para realizar una evaluación anual del rendimiento de la fase analítica.**

**2.1. Tercer estudio**

**Evaluación y mejora del rendimiento analítico mediante el Análisis Modal de Fallos y Efectos (AMFE).**

**2.1.1. Presentación**

De todas las fases del proceso analítico, la fase analítica fue hace unas décadas el objetivo inicial para la minimización de errores. Mediante el diseño e implantación del plan de control de calidad interno y la participación en programas de evaluación externa de la calidad, entre otras estrategias, se consiguió disminuir el número de errores analíticos. En consecuencia, actualmente es en las fases extraanalíticas donde ocurren la mayoría de los errores (78).

Desde que este hecho fue conocido, los laboratorios clínicos se enfocaron en analizar las causas de los errores extraanalíticos y en buscar estrategias para asegurar su calidad. Esta mayor vulnerabilidad de las fases extraanalíticas puede generar suposiciones falsas con respecto al aseguramiento de la calidad de la fase analítica. Sin embargo, hay muchas evidencias que demuestran que todavía es necesario mejorar el rendimiento analítico como, por ejemplo, el que existan procedimientos de medida con un nivel sigma no satisfactorio o que no logren alcanzar el nivel mínimo de las especificaciones de calidad basadas en la variabilidad biológica (5,79). En la última década, todavía se ha estimado que el porcentaje de errores analíticos es aproximadamente el 23 % del total de errores que ocurren en el proceso analítico (78).

En el marco de esta Tesis Doctoral se planteó la búsqueda de nuevas herramientas para minimizar los errores todavía frecuentes en la fase analítica, estableciendo una metodología basada en el análisis de riesgos y la métrica sigma para realizar una evaluación anual del rendimiento de la fase analítica.

Esta metodología se aplicó a algunas magnitudes de bioquímica general que se analizaban en el laboratorio core y que no precisaban ningún tratamiento pre-analítico especial. Las 26 magnitudes estudiadas se incluían en el PEEC "Bioquímica suero" organizado por la Sociedad Española de Medicina de Laboratorio (SEQC<sup>ML</sup>). Concretamente fueron: alanina aminotransferasa, albúmina, alfa-amilasa, aspartato aminotransferasa, bilirrubina total, bilirrubina directa, calcio, cloruro, colesterol total, colesterol HDL, creatina quinasa, creatinina, fosfatasa alcalina, fosfato no esterificado, gamma glutamil transferasa, glucosa, hierro (II + III), lactato deshidrogenasa, lipasa, magnesio, potasio, proteínas, sodio, triglicéridos, urato y urea. Excepto los iones sodio, potasio y cloruro que se miden por potenciometría indirecta, el procedimiento de medida del resto de las magnitudes es espectrofotométrico.

El PEEC "Bioquímica suero" dispone de la acreditación según la norma ISO 17043 (68). El número aproximado de laboratorios participantes es de 1.000. El material de control es suero liofilizado de origen animal. Al inicio del programa se reciben 12 viales de material de control. Cada mes del año se analiza un único material de control para todas las magnitudes citadas con anterioridad. La evaluación de los resultados también es mensual con valor asignado por consenso entre laboratorios participantes.

El AMFE es una herramienta de gestión del riesgo empleada con frecuencia en los laboratorios clínicos pero que, tal y como se ha comentado en la introducción, tiene el inconveniente de presentar cierta subjetividad (26). Para inferir objetividad al AMFE, los factores de estimación del riesgo (ocurrencia, severidad y detectabilidad) se asignaron a partir de los datos generados en la participación del laboratorio en el PEEC y de los datos obtenidos del procesamiento de los

controles de calidad internos. Para estandarizar la evaluación de los resultados y la priorización de las acciones se convirtieron los defectos por millón de oportunidades (DPMO) a la escala sigma (80).

Para aquellos procedimientos de medida cuyo nivel sigma fue insatisfactorio ( $\sigma < 3$ ) se llevó a cabo un análisis de las causas y se implementaron las acciones correctivas que se consideraron apropiadas para conseguir una mejora del rendimiento analítico. Por último, se realizó de nuevo un AMFE para evaluar la eficacia de las acciones implementadas y la posible aparición de nuevos riesgos residuales, utilizando los datos del plan de control de la calidad del año 2018.

### **2.1.2. Resultados y discusión**

Tras revisar los errores de medida de todos los resultados correspondientes a los doce envíos que había realizado el PEEC en el año 2017, se decidió incluir en el AMFE las siguientes diez magnitudes: fosfatasa alcalina, bilirrubina directa, cloruro, creatina quinasa, creatinina, colesterol HDL, lactato deshidrogenasa, potasio, sodio y proteínas totales. El motivo de su inclusión fue que, para todas ellas, el error de medida cometido había superado en alguno de los envíos del PEEC la especificación de error total establecida por el laboratorio; es decir, se consideraron en el AMFE aquellas magnitudes que presentaron un valor de ocurrencia superior a 0.

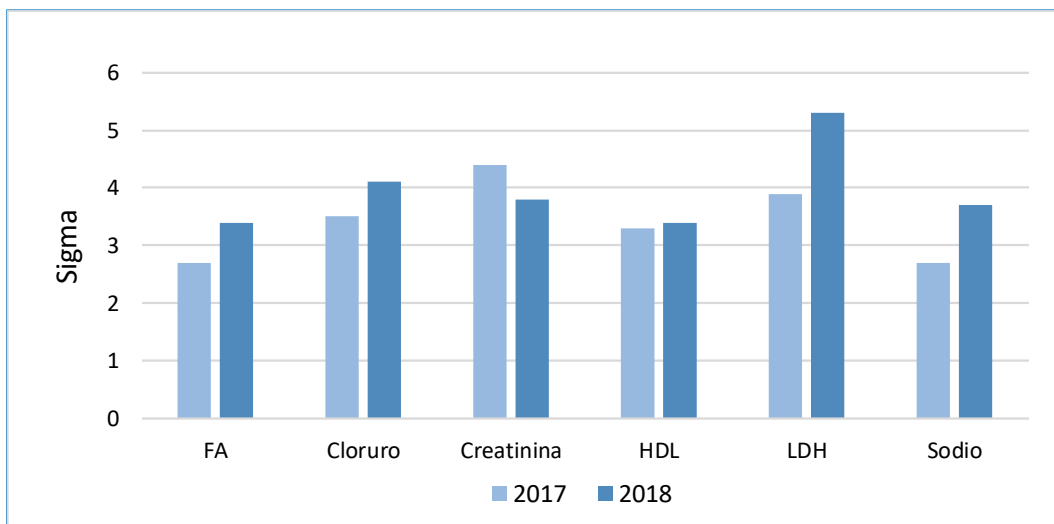
Después de la estimación del número de DPMO, para cada magnitud, éstos se expresaron según la métrica sigma. Dos de las magnitudes, fosfatasa alcalina y sodio, mostraron un nivel sigma no satisfactorio ( $\sigma < 3$ ) y, por tanto, se procedió a realizar un análisis de causa raíz obteniéndose los siguientes resultados:

- El análisis causa raíz realizado para la fosfatasa alcalina puso de manifiesto que el error sistemático observado en los envíos de los últimos 5 meses de 2017 fue debido a un cambio de lote del reactivo. Para corregir el sesgo se aplicó una ecuación a los resultados

de los pacientes obtenidos con el nuevo lote con la finalidad de que fueran intercambiables con los obtenidos con el lote el anterior (81).

- El análisis causa raíz realizado para el sodio puso de manifiesto un bajo rendimiento analítico a concentraciones elevadas. Esto se detectó al realizar una revisión del seguimiento de la imprecisión y el sesgo estimados a partir de los resultados de la medida de los materiales de control de calidad interno. Para optimizar el rendimiento analítico del procedimiento de medida se aumentó la frecuencia de medida del material de control de calidad de alta concentración de dos a tres veces al día.

Tras revisar los errores de medida de todos resultados correspondientes a los envíos que había realizado el PEEC en el año 2018, se incluyeron en el AMFE las siguientes seis magnitudes: fosfatasa alcalina, cloruro, creatinina, colesterol HDL, lactato deshidrogenasa y sodio. Tal y como se muestra en la Figura 6, todas ellas ya se habían incluido en el estudio del año 2017, y en el año 2018 ninguna de las magnitudes obtuvo un nivel sigma insatisfactorio ( $\sigma < 3$ ). Además, no se identificaron riesgos residuales para otras magnitudes.



FA: fosfatasa alcalina; HDL: colesterol HDL; LDH: lactato deshidrogenasa.

Figura 6. Comparación del valor sigma obtenido para las magnitudes incluidas en el AMFE del año 2018 con respecto al año 2017.

En cuanto a la valoración de la eficacia de las acciones correctivas implementadas en el 2017 se observó que tanto para la fosfatasa alcalina como para el sodio habían sido satisfactorias, dado que el nivel sigma pasó de bajo ( $\geq 2 \sigma < 3$ ) a moderado ( $\geq 3 \sigma < 4$ ). Esta mejora del rendimiento analítico fue consecuencia de que:

- Al eliminar el sesgo del procedimiento de medida de la fosfatasa alcalina la probabilidad de detección de errores aumentó y la ocurrencia de los errores de medida en los envíos de 2018 fue menor.
- Al aumentar la frecuencia de medida del material de control de calidad alto del sodio aumentó la probabilidad de detección de errores y los errores de medida en los envíos de 2018 tuvieron una menor severidad.

Por todo ello se consideró que la metodología propuesta resultó útil para evaluar el rendimiento analítico de los procedimientos de medida. Sin embargo, es importante tener en cuenta que todos los factores de estimación del riesgo utilizados en este estudio dependen del error total máximo permitido establecido por el laboratorio. Por lo tanto, es fundamental que al evaluar los resultados se asegure previamente su idoneidad, ya que en función de la especificación establecida podríamos pasar de un nivel sigma insatisfactorio a satisfactorio o viceversa.

A pesar de que el objetivo sea alcanzar las especificaciones de calidad basadas en la variabilidad biológica, se ha demostrado que la tecnología actual no siempre es capaz de conseguirlo (75). El caso más evidente es el del sodio: en la fecha del estudio el error total que era capaz de alcanzar el 90 % de los laboratorios en España era de un 5 % (61), siendo la especificación mínima de variabilidad biológica de un 1,1 % (82). En cualquier caso, consideramos que la mejor opción es ajustar la especificación de calidad teniendo en cuenta el rendimiento de la tecnología disponible siempre y cuando ésta permita dar respuesta al uso previsto. De no ser así podría dar lugar a interpretaciones incorrectas y/o acciones correctivas innecesarias. De acuerdo a esta premisa, para el sodio en el laboratorio se había establecido una especificación de calidad de

error total intermedia, en línea con el requisito de obligado cumplimiento establecido por el CLIA (83). Para la fosfatasa alcalina, también se había fijado la especificación de calidad más realista, en este caso en línea con el nivel deseable basado en la variabilidad biológica (82). En ambos casos, si se hubiesen establecido especificaciones de calidad más permisivas, el nivel sigma hubiese sido satisfactorio y habría dado lugar a una falsa seguridad. En consecuencia, el análisis causa raíz no se habría realizado y por tanto no se habrían realizado las acciones correctivas pertinentes.

El efecto contrario hubiese ocurrido si se hubiesen establecido especificaciones de calidad más estrictas de lo que el laboratorio es capaz de alcanzar. Sería por ejemplo el caso del cloruro. En el laboratorio se había establecido como especificación de calidad el error total máximo permitido por el CLIA (83). Si se hubiese establecido una especificación de calidad demasiado estricta, el nivel sigma hubiese sido insatisfactorio y se habrían generado toda una serie de acciones innecesarias para encontrar la posible causa del bajo rendimiento analítico, con el consecuente coste y tiempo asociado.

A pesar de que otros autores han constatado la utilidad del AMFE para detectar la necesidad de ajustar la estrategia de control de calidad con la finalidad de mejorar el rendimiento analítico (84-87), y de que también han intentado aportar objetividad sus estudios modificando la metodología tradicional, debido a la subjetividad implícita en la aplicación de esta herramienta, no es posible realizar una comparación de los resultados.

### **2.1.3. Conclusión**

La metodología utilizada en este estudio superó los inconvenientes del AMFE. Utilizar los datos del plan de control de calidad aportó objetividad a la valoración de los factores de estimación del riesgo. Por otro lado, utilizar la métrica sigma permitió estandarizar la priorización de los riesgos identificados y evaluar con objetividad la eficacia de las acciones implementadas para minimizar los riesgos. La combinación de ambas herramientas (AMFE y métrica sigma) demostró



ser de gran utilidad para tener una visión general del rendimiento analítico de los procedimientos de medida.

Debemos destacar, por último, que esta metodología podría aplicarse para evaluar el rendimiento analítico de procedimientos de medida cuantitativos de otras áreas de especialización.

#### **2.1.4. Cita al artículo publicado**

Guiñón L, Soler A, Díaz MG, Fernández RM, Rico N, Bedini JL, Mira A, Alvarez L. **Analytical Performance Assessment and Improvement by Means of the Failure Mode and Effect Analysis (FMEA)**. Biochemia medica. 2020;30:020703.

#### **2.1.5. Publicación completa**



# Analytical performance assessment and improvement by means of the Failure mode and effect analysis (FMEA)

Leonor Guiñón<sup>\*1</sup>, Anna Soler<sup>1</sup>, Mónica Gisell Díaz<sup>1</sup>, Rosa María Fernández<sup>2</sup>, Nayra Rico<sup>2</sup>, Josep Lluís Bedini<sup>2</sup>, Aurea Mira<sup>3</sup>, Luisa Alvarez<sup>1</sup>

<sup>1</sup>Quality Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Barcelona, Spain

<sup>2</sup>Biochemistry and Molecular Genetics Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Barcelona, Spain

<sup>3</sup>Direction, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Barcelona, Spain

\*Corresponding author: leonorguinon@yahoo.es

## Abstract

**Introduction:** Laboratories minimize risks through quality control but analytical errors still occur. Risk management can improve the quality of processes and increase patient safety. This study aims to use the failure mode and effect analysis (FMEA) to assess the analytical performance and measure the effectiveness of the risk mitigation actions implemented.

**Materials and methods:** The measurands to be included in the study were selected based on the measurement errors obtained by participating in an External Quality Assessment (EQA) Scheme. These EQA results were used to perform an FMEA of the year 2017, providing a risk priority number that was converted into a Sigma value ( $\sigma^{\text{FMEA}}$ ). A root-cause analysis was done when  $\sigma^{\text{FMEA}}$  was lower than 3. Once the causes were determined, corrective measures were implemented. An FMEA of 2018 was carried out to verify the effectiveness of the actions taken.

**Results:** The FMEA of 2017 showed that alkaline phosphatase (ALP) and sodium (Na) presented a  $\sigma^{\text{FMEA}}$  of less than 3. The FMEA of 2018 revealed that none of the measurands presented a  $\sigma^{\text{FMEA}}$  below 3 and that  $\sigma^{\text{FMEA}}$  for ALP and Na had increased.

**Conclusions:** Failure mode and effect analysis is a useful tool to assess the analytical performance, solve problems and evaluate the effectiveness of the actions taken. Moreover, the proposed methodology allows to standardize the scoring of the scales, as well as the evaluation and prioritization of risks.

**Keywords:** risk management; FMEA; sigma metric; analytical performance; quality control

Submitted: October 01, 2019

Accepted: January 30, 2020

## Introduction

In the field of healthcare, the clinical laboratory is one of the areas that has made many efforts to minimize errors that may affect patient safety. Errors made in laboratories are approximately 0.3%, which is much lower than the percentage of errors in other areas of medicine (1). Despite this, taking into account the high number of laboratory tests that are usually carried out, this percentage of error can total millions of erroneous results *per year*. Although it is unlikely, some of these errors can lead to adverse effects on patients (2).

Errors can occur at any step in the testing process. However, in the last few years it has been found that most of the laboratory errors occur in the ex-

tra-analytical phases. This is because the vast majority of the strategies adopted to minimize risks have been focused on reducing failures in the analytical phase, such as the design of an internal quality control (IQC) plan or the participation in external quality assessment (EQA) schemes (3). Despite these strategies, nowadays the estimated percentage of analytical errors is approximately 23% of the total errors produced in laboratories (4). This indicates that it is necessary to improve analytical performance, therefore efforts to mitigate these errors should not cease.

A requirement of the international standard for accreditation of clinical laboratories, the Internation-

al Organization for Standardization (ISO) 15189:2012, is the incorporation of risk management principles aimed at patient safety (5). The last version of the ISO 9001:2015 standard, frequently implemented in clinical laboratories, has also incorporated risk management to improve the quality of the processes (6). There are ISO directives and Clinical and Laboratory Standards Institute (CLSI) documents that provide guidance on risk management (7-10). Although these standards do not indicate which methodology should be used, the failure mode and effect analysis (FMEA) is the most applied (2).

Failure mode and effect analysis is a risk management tool for identifying possible failures that can occur and for solving known errors, analysing the causes and effects of the failures, and eliminating or reducing the most relevant ones by proposing control measures (11). After identifying and listing all the possible failure modes, the risk must be estimated using different factors. Clinical and Laboratory Standards Institute EP23A "Laboratory Quality Control Based on Risk Management" recommends a 2-factor risk model considering only the occurrence and the severity of a failure. However, to assess the analytical performance in medical laboratories a 3-factor risk model can be more appropriate (12). This model considers also the detectability, as the probability of not detecting a failure. Then, a value is assigned to each risk factor according to a numerical scale. Higher scores are assigned to those risks with a higher occurrence, higher severity and lower probability of detection. To identify the high priorities for action, the values assigned to the occurrence, severity and detectability are multiplied, thus obtaining an index score named risk priority number (RPN). Risk mitigation actions are more urgently needed for those risks with a higher RPN. After the implementation of these actions, the RPN must be calculated again to assess their effectiveness (3,11). Despite being the most used methodology, FMEA has shown some drawbacks such as the lack of standardization of scoring scales, subjectivity when assigning values to each risk factor or the lack of a unified criterion to assess the RPN value obtained (13).

It has been postulated that the Sigma metric, a tool that clinical laboratories have been using for some years to measure the effectiveness of its quality control procedures, could also be useful for the risk management of the analytical phase (14). One of the advantages of the Sigma metric is that it allows an objective assessment of the process performance and can also be used as a benchmark to compare the results obtained universally. The objective of the Six Sigma model is to reduce the variability of the processes, in such a way that six standard deviations can fit within the established tolerance limits of the process. Reducing variation, fewer defects (the results that fail to meet the specified tolerance limits) are generated so a performance improvement of the process is achieved. In order to use Sigma metric as a benchmarking scale, the number of defects produced must be expressed in defects *per* million opportunities (DPMO), and this must be converted into a Sigma value. The higher the Sigma value is, the lower the variation of the process. Six Sigma levels have been described (15). A Sigma value of 6, recognized as the world-class quality, corresponds to 3.4 DPMO. Sigma values between 5 and 6 are considered of excellent quality. Sigma levels of 4, 3 and 2 correspond to a good, moderate and poor quality, respectively. Sigma values below 2 indicate an unacceptable quality. It has been described that the number of defects in the analytical phase due to quality control failures (IQC and EQA) is around 3 to 4-Sigma (15). Outside of health care, a Sigma value of 3, which corresponds to 66,807 DPMO, has been described as the minimum acceptable quality for routine operation (14). Currently, as clinical laboratories perform a high number of tests, the same standards might be applied (16). Obtaining lower Sigma values would indicate the need to implement improvement actions.

The objective of this study was to carry out an FMEA of the analytical phase to assess the analytical performance of the measurement procedures, implement actions to mitigate the risks and evaluate their effectiveness.

## Materials and methods

### Materials

The study was performed in the Biomedical Diagnostic Center of the Hospital Clinic of Barcelona, Spain, in the period between 1st of January 2017 and 31st of December 2018. In the study were included all measurands whose measurement errors (ME) exceeded the allowable total error (TEa) in any of the surveys of the EQA scheme "Serum biochemistry" organized by the Spanish Society of Laboratory Medicine (SEQC). The following measurands were evaluated: alkaline phosphatase (ALP), direct bilirubin (DBIL), chloride (Cl), creatine kinase (CK), creatinine (CREA), high density lipoprotein (HDL), lactate dehydrogenase (LD), potassium (K), sodium (Na) and total protein (TP). The materials analysed were: i) quality control materials provided by the EQA scheme and ii) IQC materials: Liquid Assayed Multiquel level 2 and 3 (Bio-Rad Laboratories, Marnes-la-Coquette, France).

### Methods

All tests were run on ADVIA Chemistry XPT System analyser (Siemens Healthcare GmbH, Erlangen, Germany). From each EQA survey, the ME were obtained. It was calculated (Equation (Eq.) 1) as the distance in percentage from the value reported by the laboratory to the target value (instrument group mean):

$$ME (\%) = \frac{[(\text{Result} - \text{Target value})]}{\text{Target value}} \quad (\text{Eq. 1}).$$

The selection of the allowable total error was preferably based on the biological variability (BV) (minimum, desirable or optimal values) (17). When the BV model was so demanding that the laboratory was not possible to achieve the minimum values, the criteria of the Clinical Laboratory Improvement Amendments (CLIA) was adopted (18). Table 1 shows the TEa set by the laboratory for the measurands included in this study.

Failure mode and effect analysis of the years 2017 and 2018 were performed using a 3-factor model. In order to assign objective values to the occurrence, severity and detectability, they were calcu-

**TABLE 1.** Total error quality specifications established by the laboratory

Measurand	Source of Quality Specification	TEa (%)
ALP	Desirable BV	12.04
DBIL	Optimal BV	22.30
Cl	CLIA	5.00
CK	Optimal BV	15.20
CREA	Desirable BV	8.87
HDL	Desirable BV	11.63
LD	Desirable BV	11.35
K	Minimum BV	8.40
Na	CLIA	2.79
TP	Minimum BV	5.40

BV – biological variability. CLIA – Clinical Laboratory Improvement Amendments. TEa – allowable total error. ALP – alkaline phosphatase. DBIL – direct bilirubin. Cl – chloride. CK – creatine kinase. CREA – creatinine. HDL – high density lipoprotein. LD – lactate dehydrogenase. K – potassium. Na – sodium. TP – total protein.

lated using data from the EQA and the IQC, as follows:

1) Occurrence (O) was calculated (Eq. 2) as the percentage of errors (results whose ME exceeded the TEa) with respect to the number of surveys of the EQA scheme:

$$O (\%) = \frac{\text{Number of errors} \times 100}{\text{Number of surveys}} \quad (\text{Eq. 2}).$$

Those measurands that did not exceed the TEa in any of the EQA surveys were not included in the FMEA, since the occurrence value would have been zero.

2) Severity (S) was calculated (Eq. 3) as the difference expressed as a percentage between the ME and the TEa. A ME higher than TEa means a significant deviation that could lead to a false diagnosis or inadequate treatment, so the risk for the patient safety would rise as that difference increases.

$$S (\%) = \frac{(ME - TEa) \times 100}{TEa} \quad (\text{Eq. 3}).$$

When for a certain measurand there was more than one ME higher than the TEa, the average of ME was calculated.

3) Detectability (D) was calculated (Eq. 4) from the probability of error detection (Pde) provided by the Westgard Advisor module available in Unity Real Time v.2.0, a quality control data management software from Bio-Rad Laboratories (19). With this purpose, we obtained the Pde for both concentration levels of IQC and calculated the average. Since Pde is the probability of detecting an error, the detectability was calculated as the complementary value of the Pde (1 - Pde), thus obtaining the probability of not detecting an error. When the Pde value was 1, it was transformed into 0.999 in order to obtain data for the calculation. Results were expressed as a percentage.

$$D (\%) = (1 - Pde) \times 100 \quad (\text{Eq. 4})$$

Following the methodology published by Sten Westgard, risk factors were expressed as a percentage to obtain the RPN on a scale of 1 to 1 million. Therefore, an estimation of the number of DPMO was obtained (15). Then, the DPMO was converted into a Sigma value ( $\sigma^{\text{FMEA}}$ ) to standardize the evaluation of the results and the prioritization of actions, providing a better assessment of the risks that are acceptable than the one provided by the RPN. The DPMO value (described in the Six Sigma conversion tables) closest to the DPMO obtained provides the  $\sigma^{\text{FMEA}}$  (20).

The laboratory investigated the possible causes of failure for those measurands with a  $\sigma^{\text{FMEA}}$  lower than 3. Thus, a root cause analysis by means of the repeated "whys" query tool was performed (21). After the investigation, some actions were taken to decrease the number of DPMO produced, such as applying an equation to correct systematic error or increasing the frequency of IQC concentration level 3 testing (3 times a day).

An FMEA of the year 2018 was performed to evaluate the effectiveness of these measures and to identify new residual risks. A  $\sigma^{\text{FMEA}}$  between 3 and 4-Sigma or higher would indicate the success of the corrective measures.

## Results

Tables 2 and 3 show the ME obtained from the EQA surveys in the years 2017 and 2018 respectively.

The results of the FMEA of the year 2017 are shown in Table 4. Ten measurands were evaluated. Alkaline phosphatase and Na presented a  $\sigma^{\text{FMEA}}$  of less than 3 and only CK was close to the world-class quality. The results of the FMEA of the year 2018 are shown in Table 5. Only six measurands were included. None of the measurands presented a  $\sigma^{\text{FMEA}}$  below 3. The results showed that  $\sigma^{\text{FMEA}}$  for ALP and Na increased.

**TABLE 2.** Measurement errors expressed as a percentage obtained in the EQA surveys of the year 2017

Measurand	ME (%)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ALP	8.91	3.96	2.43	5.20	1.31	10.16	11.79	18.85	11.08	14.93	17.66	15.84
DBIL	14.60	30.47	24.02	1.44	4.49	4.63	4.02	0.39	5.81	0.85	7.29	0.23
Cl	1.08	3.39	1.21	2.17	1.54	0.26	1.96	0.70	0.63	2.10	1.75	7.81
CK	16.12	9.29	11.54	1.70	2.90	4.81	2.21	3.00	4.16	2.62	1.80	7.13
CREA	8.45	2.49	0.95	1.93	4.62	0.39	2.54	9.46	0.13	5.80	2.92	3.49
HDL	17.47	3.52	9.28	10.61	9.87	0.52	14.75	9.22	3.05	9.09	1.39	5.55
LD	4.79	6.03	5.39	6.15	7.33	3.68	5.52	4.37	5.49	3.95	5.42	12.61
K	1.86	1.98	0.73	3.23	0.74	1.19	1.15	0.21	1.01	0.21	1.56	11.62
Na	1.37	1.24	0.01	0.98	0.95	0.54	0.22	0.48	1.46	3.84	2.59	6.87
TP	0.35	2.34	0.97	1.38	0.65	7.07	0.31	4.65	1.84	2.93	1.00	1.78

ME – measurement error. ALP – alkaline phosphatase. DBIL – direct bilirubin. Cl – chloride. CK – creatine kinase. CREA – creatinine. HDL – high density lipoprotein. LD – lactate dehydrogenase. K – potassium. Na – sodium. TP – total protein.

**TABLE 3.** Measurement errors expressed as a percentage obtained in the EQA surveys of the year 2018

Measurand	ME (%)											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
ALP	6.01	10.55	13.45	5.14	5.85	4.61	8.66	8.25	5.38	9.38	21.09	4.35
DBIL	2.09	10.02	4.57	0.59	2.61	1.92	6.37	0.26	2.59	3.89	16.66	10.74
Cl	2.71	4.00	1.17	2.62	0.64	5.82	0.07	1.16	0.45	0.61	0.51	0.28
CK	3.64	1.63	3.21	0.54	2.94	2.29	8.27	7.40	1.59	4.76	2.34	1.52
CREA	0.74	1.90	6.92	3.13	2.68	1.22	4.49	7.47	1.90	3.98	11.65	5.85
HDL	0.27	2.37	0.88	0.39	3.35	0.53	26.50	0.57	0.13	12.37	0.90	2.53
LD	4.12	22.50	2.95	0.61	1.44	1.19	1.95	1.24	0.01	0.69	2.26	1.34
K	0.81	4.40	0.06	4.48	2.73	4.19	0.49	0.71	0.25	1.92	0.23	2.65
Na	0.46	3.80	1.61	1.31	0.65	2.84	2.36	0.53	0.09	1.19	0.05	1.55
TP	3.85	2.54	0.47	3.29	3.45	2.87	1.48	0.47	1.81	1.57	2.16	3.52

ME – measurement error. ALP – alkaline phosphatase. DBIL – direct bilirubin. Cl – chloride. CK – creatine kinase. CREA – creatinine. HDL – high density lipoprotein. LD – lactate dehydrogenase. K – potassium. Na – sodium. TP – total protein.

**TABLE 4.** FMEA of the year 2017

Measurand	Occurrence (%)	Severity (%)	Pde	Detectability (%)	DPMO	$\sigma^{FMEA}$
ALP	33.33	39.70	0.200	80.00	105,856	2.7
DBIL	16.67	22.17	0.948	5.25	1940	4.4
Cl	8.33	56.20	0.500	50.00	23,407	3.5
CK	8.33	6.05	0.999	0.10	5	5.9
CREA	8.33	6.65	0.610	39.05	2163	4.4
HDL	16.67	38.52	0.431	56.90	36,537	3.3
LD	8.33	11.10	0.200	80.00	7397	3.9
K	8.33	38.33	0.904	9.65	3081	4.2
Na	16.67	91.94	0.274	72.65	111,346	2.7
TP	8.33	30.93	0.426	57.45	14,802	3.7

Pde – probability of error detection. DPMO – defects per million opportunities.  $\sigma^{FMEA}$  – Sigma value. ALP – alkaline phosphatase. DBIL – direct bilirubin. Cl – chloride. CK – creatine kinase. CREA – creatinine. HDL – high density lipoprotein. LD – lactate dehydrogenase. K – potassium. Na – sodium. TP – total protein.

**TABLE 5.** FMEA of the year 2018

Measurand	Occurrence (%)	Severity (%)	Pde	Detectability (%)	DPMO	$\sigma^{FMEA}$
ALP	16.67	43.44	0.569	43.15	31,247	3.4
Cl	8.33	16.40	0.625	37.50	5122	4.1
CREA	8.33	31.34	0.595	40.50	10,607	3.8
HDL	8.33	67.11	0.449	55.10	30,802	3.4
LD	8.33	98.24	0.999	0.10	82	5.3
Na	16.67	19.00	0.582	41.80	13,239	3.7

Pde – probability of error detection. DPMO – defects per million opportunities.  $\sigma^{FMEA}$  – Sigma value. ALP – alkaline phosphatase. Cl – chloride. CREA – creatinine. HDL – high density lipoprotein. LD – lactate dehydrogenase. Na – sodium.



## Discussion

In the present study, a 3-factor risk model was used to perform the FMEA. The proposed methodology allowed us to assess the analytical performance of the laboratory methodologies, detecting poor analytical performance of ALP and Na, which led us to implement improvement actions. The actions taken in order to reduce the number of DPMO were different for the two measurands according to the results of the root-cause analysis. Therefore, for the bias on the EQA results of ALP from August until December 2017, that was due to a lot-to-lot reagent variation, the action taken by the laboratory was to apply an equation to correct the systematic deviation until the reagent lot changed. The elimination of the bias led to a higher probability of detection in 2018 (from 0.200 to 0.569) and a reduction of the occurrence of errors in the EQA surveys. In the case of Na, a poor performance of IQC concentration level 3 was observed. The action taken was to increase the frequency of IQC concentration level 3 testing. The modification of the quality control procedure in 2018 increased the probability of error detection (from 0.274 to 0.582) and the errors committed in the EQA surveys had a lower severity. Because of the actions carried out in 2017,  $\sigma^{FMEA}$  obtained for these measurands in the FMEA of the year 2018 were between 3 and 4-Sigma, indicating an improvement of the analytical performance.

Other authors have applied the FMEA methodology to assess the analytical performance of clinical chemistry measurands (22-25). As a conclusion, all of them stated the usefulness of the risk assessment to detect if it was necessary to adjust the quality control procedures in order to improve the analytical performance. However, due to the lack of standardization of the FMEA methodology, it is not feasible to compare the results of the present study with those obtained in their studies. They also used a modified FMEA and the Sigma metric in order to overcome the drawbacks of the FMEA. However, their methodology was neither entirely objective nor easy to standardize mainly because the value assigned to the severity was based on personal experience (22-24). Furthermore, those studies do not explain the risk mitigation actions

proposed nor even verify the effectiveness of actions implemented. Only Capunzo *et al.*, that performed an FMEA of the analytical process of three measurands (glucose, total cholesterol, and total bilirubin), listed the improvement actions performed and reassessed the risks once these actions were applied (25). As in the present study, they achieved a risk reduction (measured as RPN) for all measurands after implementing actions such as increasing the IQC frequency analysis or changing the procedure of preparation, storage and use of the calibrators.

It is important to remark that all the factors used to perform the FMEA depend on the TEa established by the laboratory. Therefore, in order to analyse correctly the results obtained, the laboratory must also assess whether the established tolerance limits are the right ones. Depending on the TEa established by the laboratory, it is possible to move from a  $\sigma^{FMEA}$  of less than 3 to a  $\sigma^{FMEA}$  of more than 3 or *vice versa*, as it occurs in other studies in which they have assessed the importance of selecting a proper TEa to obtain a reliable Sigma value that represents analytical performance (26). As it has been shown in other studies the current technology does not always allow to reach quality specifications based on BV, even though these specifications should be the objective (27). Therefore, to assess the results correctly, it should be considered whether the goals set by the laboratory are realistic or not. The evaluation of the analytical performance of the measurement procedures may not be adequate if the quality specifications established by the laboratory are too demanding. According to the authors, laboratory quality departments and responsible laboratory personnel should set together adequate goals.

The proposed methodology allowed us to identify those magnitudes with a less satisfactory analytical performance and, once the causes were detected and the improvement actions implemented, verify their effectiveness. Performing an FMEA using EQA and IQC data allows to standardize how the scoring scales are made. This means a reduction in the time needed by the staff to calculate the risk score and ensures the treatment of risks under the same criteria. Furthermore, this study



shows that the Sigma metric can be used in the FMEA to standardize the prioritization of risks, as well as to measure the change in risk score after having made risk mitigation actions.

## References

- Lippi G, Plebani M. A Six-Sigma approach for comparing diagnostic errors in healthcare—where does laboratory medicine stand? *Ann Transl Med.* 2018;6:180. <https://doi.org/10.21037/atm.2018.04.02>
- Aita A, Padoan A, Antonelli G, Sciacovelli L, Plebani M. Patient safety and risk management in medical laboratories: theory and practical application. *J Lab Precis Med.* 2017;2:75. <https://doi.org/10.21037/jlpm.2017.08.14>
- Jiang Y, Jiang H, Ding S, Liu Q. Application of failure mode and effects analysis in a clinical chemistry laboratory. *Clin Chim Acta.* 2015;448:80-5. <https://doi.org/10.1016/j.cca.2015.06.016>
- Plebani M. The journey toward quality and patient safety in laboratory medicine continues. *North Am J Med Sci.* 2014;6:229-30. <https://doi.org/10.4103/1947-2714.132942>
- International Organization for Standardization (ISO). ISO 15189:2012. Medical laboratories – requirements for quality and competence. Geneva: ISO, 2012.
- International Organization for Standardization (ISO). ISO 9001:2015. Quality management systems -requirements. Geneva: ISO, 2015.
- International Organization for Standardization (ISO). IEC 31010:2019. Risk management – risk assessment techniques. Geneva: ISO, 2019.
- International Organization for Standardization (ISO). ISO/TS 22367:2008. Medical laboratories – reduction of error through risk management and continual improvement. Geneva: ISO, 2008.
- Clinical and Laboratory Standards Institute (CLSI). Laboratory quality control based on risk management; Approved Guideline. CLSI document EP23-A. Wayne, PA: CLSI, 2011.
- Clinical and Laboratory Standards Institute (CLSI). Risk management techniques to identify and control laboratory error sources; Approved guideline, 2nd ed. CLSI document EP18-A2. Wayne, PA: CLSI, 2009.
- Liu HC. FMEA for Proactive Healthcare Risk Analysis: A Systematic Literature Review. In: *Improved FMEA Methods for Proactive Healthcare Risk Analysis.* Singapore:Springer; 2019. p.15-45. <https://doi.org/10.1007/978-981-13-6366-5>
- Westgard JO. Perspectives on quality control, risk management, and analytical quality management. *Clin Lab Med.* 2013;33:1-14. <https://doi.org/10.1016/j.cll.2012.10.003>
- Dean Franklin B, Shebl NA, Barber N. Failure mode and effects analysis: too little for too much? *BMJ Qual Saf.* 2012;21:607-11. <https://doi.org/10.1136/bmjqs-2011-000723>
- Westgard S. Prioritizing risk analysis quality control plans based on Sigma metrics. *Clin Lab Med.* 2013;33:41-53. <https://doi.org/10.1016/j.cll.2012.11.008>
- Westgard JO, Armbruster D, Westgard SA. Risk, Error and Uncertainty: Laboratory Quality Management in the Age of Metrology. An Issue of the Clinics in Laboratory Medicine, Vol. 37-1. New York: Elsevier, 2017. [https://doi.org/10.1016/s0272-2712\(16\)30120-2](https://doi.org/10.1016/s0272-2712(16)30120-2)
- Westgard S, Bayat H, Westgard JO. Analytical Sigma metrics: A review of Six Sigma implementation tools for medical laboratories. *Biochem Med (Zagreb).* 2018;28:020502. <https://doi.org/10.11613/BM.2018.020502>
- Ricos C, Alvarez V, Cava F, Garcia-Lario JV, Hernandez A, Jimenez CV, et al. Desirable specifications for total error, imprecision, and bias, derived from intra- and inter-individual biological variation. The 2014 update. Available at: <http://www.westgard.com/biodatabase1.htm>. Accessed: January 2nd 2019.
- CLIA Requirements for Analytical Quality. Available at: <https://www.westgard.com/clia.htm>. Accessed October 2nd 2018.
- Westgard JO. Basic Quality Management Systems. Chapter 12. Designing Statistical QC procedures. Madison WI: Westgard QC; 2014. p. 171-88.
- Six Sigma Table. Available at: <https://www.westgard.com/sixsigtable.htm>. Accessed February 21th 2019.
- Clinical and Laboratory Standards Institute (CLSI). Nonconforming Event Management; Approved guideline, 2nd ed. CLSI guideline QMS11. Wayne, PA: CLSI, 2015.
- Xia Y, Xue H, Yan C, Li B, Zhang S, Li M, Ji L. Risk analysis and assessment based on Sigma metrics and intended use. *Biochem Med (Zagreb).* 2018;28:020707. <https://doi.org/10.11613/BM.2018.020707>
- Ceriotti F, Cappelletti P, Caputo M, Di Serio F, Messeri G, Ottomano C, et al. A risk-analysis approach to the evaluation of analytical quality. *Clin Chem Lab Med.* 2011;50:67-71. <https://doi.org/10.1515/cclm.2011.740>
- Westgard SA, Salcedo G. A Six Sigma risk analysis example: defining QC frequency with FMEA. July 2012. Available at: <https://www.westgard.com/endoclab-risk.htm>. Accessed October 17th 2018.
- Capunzo M, Cavallo P, Boccia G, Brunetti L, Pizzuti S. A FMEA clinical laboratory case study: how to make problems and improvements measurable. *Clin Leadersh Manag Rev.* 2004;18: 37-41.
- Hens K, Berth M, Armbruster D, Westgard S. Sigma metrics used to assess analytical quality of clinical chemistry assays: importance of the allowable total error (TEa) target. *Clin Chem Lab Med.* 2014;52:973-80. <https://doi.org/10.1515/cclm-2013-1090>
- Molina A, Guiñón L, Perez A, Segurana A, Bedini JL, Reverter JC, Merino A. State of the art vs biological variability: Comparison on hematology parameters using Spanish EQAS data. *Int J Lab Hematol.* 2018;40:284-91. <https://doi.org/10.1111/ijlh.12783>

## Potential conflict of interest

None declared.



### **3. Tercer objetivo: Establecer una metodología para verificar periódicamente la intercambiabilidad de resultados.**

#### **3.1. Cuarto estudio**

**Estudio para la verificación periódica de la intercambiabilidad de resultados de las magnitudes bioquímicas que se miden indistintamente en cualquiera de los analizadores disponibles en el laboratorio Core.**

##### **3.1.1. Introducción**

Es habitual que los laboratorios clínicos dispongan de más de un analizador para poder dar respuesta tanto a las necesidades de los solicitantes de las prestaciones analíticas como a las de los pacientes, siendo tales necesidades la emisión rápida y fiable de resultados. Normalmente se opta por disponer por duplicado de aquellas magnitudes que tienen una mayor demanda o aquellas que por su carácter urgente precisan que la emisión del resultado se realice en un corto margen de tiempo. De ese modo, se evita la sobrecarga de trabajo en un único analizador, y se asegura la emisión de resultados a pesar de que un analizador presente alguna incidencia. En laboratorios altamente automatizados, como suele ser el caso de los laboratorios que dan servicio a hospitales de tercer nivel, en los que se trabaja en modo de carga continua y con la actividad urgente integrada con la actividad de rutina, el número de sistemas analíticos disponibles suele ser incluso mayor.

Ante la incorporación de un nuevo analizador en el laboratorio, ya sea para sustituir al existente o para disponer de un analizador adicional, debe realizarse, previamente a su uso asistencial, un estudio de intercambiabilidad de resultados para poder asegurar que los resultados emitidos no van a presentar diferencias con respecto a los generados hasta la fecha. El objetivo de estos estudios iniciales es evaluar si las diferencias observadas entre los analizadores pueden conllevar

a la modificación de los intervalos de referencia biológicos o niveles de decisión clínica y actuar en consecuencia.

Al ser una situación bastante frecuente, los laboratorios, por lo general, están familiarizados con la metodología para llevar a cabo este tipo de estudios y los criterios a aplicar para evaluar los resultados obtenidos. A este respecto, puede resultar útil consultar la metodología propuesta en la guía EP09-A3 publicada por el CLSI en el año 2013 (88). En base a esta guía, un grupo de trabajo interdisciplinar de calidad en el laboratorio clínico, en colaboración con las asociaciones y sociedades científicas nacionales, elaboró en el año 2014 un documento de consenso sobre la realización de estudios de intercambiabilidad (89). En este documento los autores indicaban que, cuando se disponga de más de un analizador para medir la misma magnitud, además de realizar el estudio inicial de intercambiabilidad de resultados, era recomendable realizar verificaciones de forma periódica con la finalidad de asegurar que la intercambiabilidad de resultados se mantenía a lo largo del tiempo. Con la nueva edición de la norma ISO 15189 publicada a finales del año 2022 (67) esta recomendación ha pasado a ser un requisito (apartado 7.3.7.4.) para aquellos laboratorios que estén acreditados o que quieran conseguir la acreditación.

Antes de realizar un estudio de intercambiabilidad de resultados se debe partir de la premisa de que los procedimientos de medida se encuentren en un estado estable en todos los sistemas analíticos que se desean comparar. Ahora bien, que así sea no exime al laboratorio de la realización de verificaciones de la intercambiabilidad de forma periódica, debido principalmente a la falta de conmutabilidad de los materiales de control de calidad, lo que hace que el comportamiento analítico de éstos pueda ser diferente al de las muestras de los pacientes (90). Aun así, a pesar de la falta de conmutabilidad, si mediante la estrategia de control de calidad se detecta un bajo rendimiento analítico (mayor imprecisión o sesgo de lo deseable) de un analizador respecto a los demás, puede ser necesario realizar un estudio de intercambiabilidad

de resultados con muestras de pacientes para identificar si es preciso realizar acciones correctivas en alguno de los analizadores.

A diferencia de lo que ocurre con el estudio inicial de intercambiabilidad antes del uso asistencial, no existe consenso entre los laboratorios en la forma de proceder para verificar periódicamente la intercambiabilidad, ni en cuanto a la metodología, ni en cuanto al criterio para evaluar los resultados; quizás en parte debido a que son pocas las fuentes bibliográficas disponibles. A este respecto, el grupo interdisciplinar antes mencionado propuso utilizar el protocolo EP31-A-IR publicado por el CLSI (91). Esta guía propone un protocolo para la verificación periódica de resultados que permite realizar la comparación de hasta 10 analizadores dentro de una misma institución. Ante la ausencia de más recomendaciones, esta guía podría servir de ayuda a los laboratorios clínicos que se planteen implementar la realización de estudios de verificación periódica de la intercambiabilidad como una herramienta más, dirigida al aseguramiento de la calidad.

### **3.1.2. Objetivo**

En el marco de esta Tesis Doctoral se ha llevado a cabo un estudio para establecer una metodología basada en la guía EP31-A-IR del CLSI (91) para llevar a cabo la verificación periódica de la intercambiabilidad de resultados entre varios analizadores.

### **3.1.3. Material y métodos**

Se incluyeron en el estudio de intercambiabilidad las 24 magnitudes de bioquímica general que se podían medir en muestras de suero en más de un sistema Alinity c (Abbott Laboratories, IL, USA) de los 4 analizadores disponibles en el laboratorio Core: albúmina, ácido úrico, alanina aminotransferasa (ALT), aspartato aminotransferasa (AST), amilasa, bilirrubina total, bilirrubina directa, calcio, cloruro, colesterol total, creatina quinasa (CK), creatinina, fosfatasa alcalina, fosfato, gamma glutamil transferasa (GGT), glucosa, lactato deshidrogenasa (LDH), magnesio,

potasio, proteína C reactiva (PCR), proteínas totales, sodio, triglicéridos y urea. Excepto los iones sodio, potasio y cloruro que se miden por potenciometría indirecta, el procedimiento de medida del resto de las magnitudes es espectrofotométrico. Para la medición de las magnitudes de este estudio no es preciso realizar ningún tipo de tratamiento pre-analítico de las muestras.

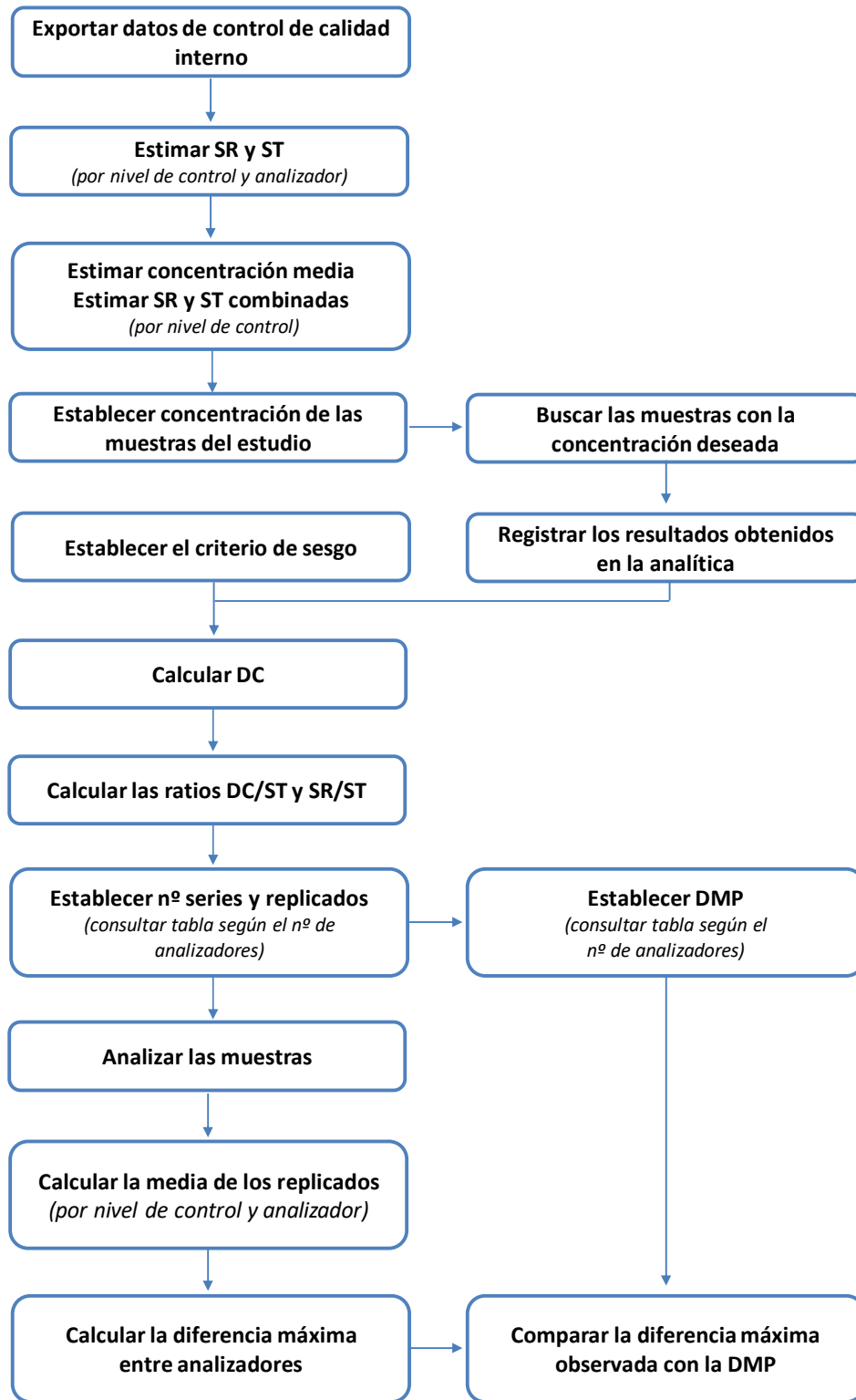
La distribución de las magnitudes en función del número de analizadores en los que era posible realizar la medición se muestra en la Tabla 10.

Tabla 10. Distribución de los procedimientos de medida en los analizadores Alinity c del laboratorio Core.

Dos analizadores	Tres analizadores	Cuatro analizadores
Ácido úrico	AST	Albúmina
Amilasa	Bilirrubina directa	ALT
LDH	Cloruro	Bilirrubina total
	CK	Calcio
	Colesterol total	Creatinina
	Magnesio	Fosfatasa alcalina
	Potasio	Fosfato
	Sodio	GGT
	Triglicéridos	Glucosa
		PCR
		Proteínas totales
		Urea

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa; PCR: proteína C reactiva.

En la Figura 7 se muestran las fases o etapas del protocolo indicadas en la guía EP31-A-IR para llevar a cabo los estudios de intercambiabilidad de resultados.



DC: diferencia crítica; SR: desviación estándar a corto plazo; ST: desviación estándar a largo plazo.

Figura 7. Diagrama de flujo del protocolo de verificación periódica de la intercambiabilidad de resultados descrito en la guía EP31-A-IR del CLSI.

De acuerdo al protocolo, para identificar la concentración o actividad aproximada que debían tener las diferentes magnitudes en las muestras seleccionadas para el estudio, se realizaron los siguientes cálculos:

- A partir de los resultados de los materiales de control interno M S Plus Alinity (Abbott Laboratories, IL, USA) a dos niveles de concentración, niveles 2 (medio) y 3 (alto), se estimó la media y la imprecisión de los procedimientos de medida en cada analizador. Para obtener la imprecisión a corto plazo (SR) se utilizaron los resultados acumulados durante 20 días y para obtener la imprecisión a largo plazo (ST) se utilizaron los resultados acumulados durante 6 meses.
- A continuación, se estimó la concentración o actividad media de todos los analizadores y la SR y la ST combinada de acuerdo a la siguiente expresión:

$$\text{Media} = (\text{media}_1 + \text{media}_2 + \dots + \text{media}_n) / n$$

$$\text{SR} = ([\text{SR}_1^2 + \text{SR}_2^2 + \dots + \text{SR}_n^2]) / n)^{1/2}$$

$$\text{ST} = ([\text{ST}_1^2 + \text{ST}_2^2 + \dots + \text{ST}_n^2]) / n)^{1/2}$$

Siendo n el número de analizadores.

- Por último, se calculó la concentración o actividad aproximada que debían tener las muestras del estudio como un  $\pm 20\%$  de la media obtenida teniendo en cuenta todos los analizadores. La Tabla 11 muestra los datos de SR y ST combinada y los intervalos de concentración o actividad aproximada que debían tener las muestras seleccionadas.



Tabla 11. Estimación del rango de concentración (o actividad) de las diferentes magnitudes para la selección de las muestras del estudio a partir de los datos acumulados de control de calidad.

Magnitud	Unidades	Material de control nivel 2				Material de control nivel 3				
		Concentración o actividad media	SR	ST	Concentración o actividad aproximada	Concentración o actividad media	SR	ST	Concentración o actividad aproximada	
<b>Ácido úrico</b>	mmol/L	0,32	0,012	0,012	0,26	0,39	0,017	0,019	0,39	0,58
<b>Albumina</b>	g/L	41,2	0,56	0,91	33,0	49,5	0,58	0,84	44,7	67,0
<b>ALT</b>	U/L	80	3,6	3,3	64	96	9,4	9,3	156	233
<b>Amilasa</b>	U/L	114	1,9	2,5	91	137	4,6	8,5	300	450
<b>AST</b>	U/L	101	2,3	2,4	81	121	2,1	3,6	186	279
<b>Bilirrubina directa</b>	µmol/L	19	0,7	0,6	15	23	2,6	2,0	33	50
<b>Bilirrubina total</b>	µmol/L	49	2,2	2,2	39	59	5,8	4,8	87	131
<b>Calcio</b>	mmol/L	2,39	0,043	0,047	1,91	2,87	0,046	0,055	2,64	3,96
<b>Cloruro</b>	mmol/L	99	1,1	1,5	79	119	1,3	1,5	91	137
<b>Colesterol</b>	mmol/L	4,54	0,097	0,110	3,63	5,45	0,148	0,151	5,11	7,67
<b>CK</b>	U/L	217	2,9	3,5	173	260	7,6	8,1	350	526
<b>Creatinina</b>	µmol/L	173	6,8	8,3	139	208	21,2	22,5	443	664
<b>Fosfatasa alcalina</b>	U/L	219	10,1	9,1	175	262	15,1	15,4	318	476
<b>Fosfato</b>	mmol/L	1,33	0,032	0,034	1,06	1,59	0,048	0,044	1,84	2,77
<b>GGT</b>	U/L	75	2,7	2,9	60	90	5,5	5,9	135	202
<b>Glucosa</b>	mmol/L	8,2	0,15	0,17	6,5	9,8	0,25	0,29	13,6	20,4
<b>LDH</b>	U/L	239	4,2	6,7	191	287	6,9	10,5	404	606
<b>Magnesio</b>	mmol/L	0,91	0,032	0,028	0,73	1,10	0,060	0,046	1,43	2,15
<b>PCR</b>	mg/L	9,3	0,27	0,27	7,4	11,1	0,70	0,78	24,2	36,4
<b>Potasio</b>	mmol/L	4,07	0,066	0,059	3,25	4,88	0,099	0,101	5,70	8,55
<b>Proteínas totales</b>	g/L	62,6	1,10	1,40	50,1	75,1	1,21	1,66	69,7	104,5
<b>Sodio</b>	mmol/L	150	2,0	1,8	120	180	2,1	2,0	135	203
<b>Triglicéridos</b>	mmol/L	1,64	0,039	0,046	1,31	1,97	0,074	0,079	2,20	3,30
<b>Urea</b>	mmol/L	13,0	0,48	0,48	10,4	15,6	0,81	0,78	18,2	27,2

SR: desviación estándar a corto plazo; ST: desviación estándar a largo plazo.

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa; PCR: proteína C reactiva.

A continuación, se estableció el criterio de sesgo de forma específica para cada magnitud y nivel de concentración teniendo en cuenta dos premisas:

- Que el rendimiento analítico de los procedimientos de medida en estado estable permitiese alcanzar la especificación de calidad.
- Que la especificación de calidad seleccionada no diese lugar a un número de repeticiones de la medida que resultase impracticable para el laboratorio. Así se limitó el número de repeticiones a un máximo de 10.

Así, se seleccionó un criterio de sesgo en el siguiente orden: si al aplicar el criterio de sesgo basado en la variabilidad biológica intraindividual el número de repeticiones resultaba impracticable, entonces se aplicó la especificación de calidad de sesgo basada en la variabilidad biológica; y, si aun así el estudio era inviable, se utilizaron como criterio las especificaciones mínimas de consenso.

1) Basado en la variabilidad biológica:

- a) Calculando el sesgo que sería biológicamente significativo para un procedimiento de medida con una imprecisión analítica mínima (92,93):

$$\text{Sesgo} \leq 0,7 \cdot CV_i$$

Siendo  $CV_i$  el coeficiente de variación basado en la variabilidad biológica intraindividual.

- b) Utilizando las especificaciones de calidad de sesgo publicadas en la base de datos de variabilidad biológica de la EFLM (94) o en su defecto las obtenidas por el grupo de *Ricós* (82). Estas especificaciones son calculadas teniendo en cuenta la variabilidad biológica intraindividual e interindividual.

2) Basado en las especificaciones mínimas de consenso propuestas por el Comité de Expertos Interdisciplinar Especificaciones de la Calidad (62).

Para determinar el número de series, el número de replicados y la diferencia máxima permitida (DMP) entre los diferentes analizadores se calcularon los siguientes parámetros que se recogen en las Tablas 12 y 13, para el material de control de nivel 2 (medio) y 3 (alto), respectivamente:

- La diferencia crítica (DC) multiplicando la concentración o actividad obtenidas en las analíticas originales de las muestras seleccionas por el criterio de sesgo establecido.
- La ratio entre la diferencia crítica y la desviación estándar a largo plazo (DC/ST) y la ratio entre la desviación estándar a corto y a largo plazo (SR/ST). Con el valor de estas ratios y utilizando las tablas del apéndice B para 2, 3 y 4 analizadores en función de la magnitud, se dedujo el número de series, el número de replicados a realizar y la DMP. En aquellos casos en los que no se disponía en las tablas el valor exacto de alguna de las ratios, se escogió el valor más próximo. A modo de ejemplo, se muestra cómo se dedujo el número de series, el número de replicados y la DMP para el sodio en un nivel de concentración cercana al nivel 3 del material de control:

- Las tablas del apéndice B son diferentes en función del número de analizadores que se quieren comparar. Puesto que el sodio podía medirse en tres analizadores, se consultó la tabla B2 de la guía.
- Se buscaron en la tabla los valores más próximos de las ratios obtenidas (DC/ST=2,20 y SR/ST=1,04) que son los siguientes:

Nº de series	DC/ST	SR/ST	Nº de replicados	Diferencia máxima permitida
1	2,00	1,00	4	0,83 * DC
2	2,00	1,00	2	0,83 * DC

- Se seleccionó realizar 4 replicados en una sola serie, tal y como se indica en la primera fila de la tabla.
- Se calculó la DMP como  $0,83 * 4,448 = 3,692$ .

Tabla 12. Criterios de sesgo establecidos, número de series y replicados a realizar, y diferencias máximas permitidas entre analizadores para las diferentes magnitudes a concentraciones cercanas al nivel 2 del material de control.

Magnitud	Unidades	Concentración	Sesgo (%)	Fuente	Sesgo	DC	DC/ST	SR/ST	Nº de series	Nº de replicados	Diferencia máxima permitida
Acido úrico	mmol/L	0,28	5,81	0,7.CvI	0,058	0,016	1,40	1,00	1	7	0,011
Albumina	g/L	34,3	11	EMC	0,110	3,773	4,17	0,62	1	1	3,433
ALT	U/L	87	7,07	0,7.CvI	0,071	6,151	1,86	1,09	1	5	5,720
Amilasa	U/L	107	7,7	VB <sub>des</sub>	0,077	8,239	3,28	0,75	1	5	5,685
AST	U/L	94	6,72	0,7.CvI	0,067	6,317	2,64	2,88	1	3	5,307
Bilirrubina directa	µmol/L	21	25,76	0,7.CvI	0,258	5,410	8,39	1,10	1	1	4,490
Bilirrubina total	µmol/L	43	15,26	0,7.CvI	0,153	6,562	2,99	0,98	1	2	5,906
Calcio	mmol/L	2,06	9,1	EMC	0,091	0,187	3,97	0,90	1	1	0,171
Cloruro	mmol/L	108	6	EMC	0,060	6,480	4,35	0,76	1	1	5,378
Colesterol	mmol/L	3,72	6,4	VB <sub>min</sub>	0,064	0,238	2,17	0,88	1	10	0,207
CK	U/L	200	10,5	0,7.CvI	0,105	21,000	6,01	0,84	1	1	17,430
Creatinina	µmol/L	198	22	EMC	0,220	43,560	5,25	0,82	1	1	39,640
Fosfatasa alcalina	U/L	215	9,1	VB <sub>des</sub>	0,091	19,565	2,15	1,11	1	4	17,804
Fosfato	mmol/L	1,32	5,46	0,7.CvI	0,055	0,072	2,14	0,95	1	6	0,066
GGT	U/L	76	11,4	VB <sub>des</sub>	0,114	8,664	3,02	0,96	1	2	7,798
Glucosa	mmol/L	7,8	8,4	EMC	0,084	0,655	3,91	0,91	1	2	0,563
LDH	U/L	208	21	EMC	0,210	43,680	6,51	0,63	1	1	30,576
Magnesio	mmol/L	0,94	9,4	EMC	0,094	0,088	3,15	1,14	1	2	0,070
PCR	mg/L	8,6	23,87	0,7.CvI	0,239	2,053	7,59	0,99	1	1	1,868
Potasio	mmol/L	4,00	2,87	0,7.CvI	0,029	0,115	1,95	1,12	1	4	0,095
Proteínas totales	g/L	50,6	9	EMC	0,090	4,554	3,26	0,79	1	3	4,190
Sodio	mmol/L	140	3,2	EMC	0,032	4,480	2,54	1,15	1	4	3,718
Triglicéridos	mmol/L	1,51	14	0,7.CvI	0,140	0,211	4,63	0,85	1	1	0,175
Urea	mmol/L	13,9	9,73	0,7.CvI	0,097	1,352	2,81	1,00	1	3	1,136

DC: diferencia crítica; SR: desviación estándar a corto plazo; ST: desviación estándar a largo plazo.

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa;

PCR: proteína C reactiva.

Tabla 13. Criterios de sesgo establecidos, número de series y replicados a realizar, y diferencias máximas permitidas entre analizadores para las diferentes magnitudes a concentraciones cercanas al nivel 3 del material de control.

Magnitud	Unidades	Concentración	Sesgo (%)	Fuente	Sesgo	DC	DC/ST	SR/ST	Nº de series	Nº de replicados	Diferencia máxima permitida
Ácido úrico	mmol/L	0,48	5,81	VB <sub>min</sub>	0,058	0,028	2,24	0,90	2	3	0,029
Albúmina	g/L	46,2	11	EMC	0,110	5,082	6,03	0,68	1	1	4,625
ALT	U/L	159	7,7	VB <sub>des</sub>	0,077	12,243	1,32	1,01	1	10	11,264
Amilasa	U/L	438	7,7	VB <sub>des</sub>	0,077	33,726	3,95	0,54	1	1	23,608
AST	U/L	231	5,7	0,7.Cvi	0,057	15,523	4,30	0,58	1	1	12,884
Bilirrubina directa	µmol/L	40	25,76	0,7.Cvi	0,258	10,304	5,24	1,34	1	1	8,552
Bilirrubina total	µmol/L	115	15,26	0,7.Cvi	0,153	17,549	3,64	1,21	1	2	15,092
Calcio	mmol/L	2,77	9,1	EMC	0,091	0,252	4,57	0,83	1	3	0,229
Cloruro	mmol/L	110	6	EMC	0,060	6,600	4,43	0,86	1	1	5,478
Colesterol	mmol/L	6,29	3,71	0,7.Cvi	0,037	0,233	1,55	0,98	1	7	0,196
CK	U/L	452	10,5	0,7.Cvi	0,105	47,460	5,84	0,93	1	1	39,392
Creatinina	µmol/L	484	22	EMC	0,220	106,480	4,73	0,94	1	1	96,897
Fosfatasa alcalina	U/L	360	9,1	VB <sub>des</sub>	0,091	32,760	2,13	0,98	1	4	29,812
Fosfato	mmol/L	1,86	5,46	0,7.Cvi	0,055	0,102	2,30	1,08	1	4	0,092
GGT	U/L	166	6,37	0,7.Cvi	0,064	10,574	1,80	0,94	1	9	9,834
Glucosa	mmol/L	13,8	8,4	EMC	0,084	1,159	4,00	0,86	1	1	1,055
LDH	U/L	591	21	EMC	0,210	124,110	11,87	0,66	1	1	86,877
Magnesio	mmol/L	1,51	9,4	EMC	0,094	0,142	3,07	1,29	1	2	0,112
PCR	mg/L	24,6	23,87	0,7.Cvi	0,239	5,872	7,49	0,89	1	1	5,344
Potasio	mmol/L	5,72	2,87	0,7.Cvi	0,029	0,164	1,62	0,98	1	7	0,138
Proteínas totales	g/L	71,9	9	EMC	0,090	6,471	3,90	0,73	1	5	6,147
Sodio	mmol/L	139	3,2	EMC	0,032	4,448	2,20	1,04	1	4	3,692
Triglicéridos	mmol/L	2,45	14	0,7.Cvi	0,140	0,343	4,34	0,94	1	1	0,285
Urea	mmol/L	26,8	9,73	0,7.Cvi	0,097	2,608	3,34	1,04	1	2	2,243

DC: diferencia crítica; SR: desviación estándar a corto plazo; ST: desviación estándar a largo plazo.

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa; PCR: proteína C reactiva.

Por último, se seleccionó una muestra que tuviese la concentración deseada para cada una de las magnitudes. Se analizaron según el número de series y replicados establecidos y se calculó la diferencia máxima entre las medias (o los resultados cuando no se requerían replicados) obtenidas en los diferentes analizadores. Los resultados se consideraron intercambiables cuando la máxima diferencia entre analizadores fue inferior o igual a la DMP. En caso de obtenerse resultados no intercambiables, se evaluó la significación clínica de las diferencias observadas para determinar si era preciso realizar un análisis de las causas, corregir las desviaciones detectadas y repetir el estudio de intercambiabilidad.

#### **3.1.4. Resultados y discusión**

Los resultados obtenidos para cada uno de los analizadores se muestran en las Tablas 14 y 15 para el material de control de nivel 2 (medio) y 3 (alto), respectivamente. Para las 24 magnitudes incluidas en el estudio, las diferencias obtenidas entre analizadores fueron inferiores a las DMPs en los dos niveles de concentración estudiados.

Tabla 14. Diferencias máximas observadas entre analizadores y evaluación de la intercambiabilidad de resultados a concentraciones (o actividad) cercanas al nivel 2 del material de control.

Magnitud	Unidades	Concentración o actividad media				Diferencia máxima observada	Diferencia máxima permitida	Intercambiabilidad
		Alinity 1	Alinity 2	Alinity 3	Alinity 4			
Ácido úrico	mmol/L	-	0,27	0,28	-	0,01	0,011	Si
Albúmina	g/L	33,3	34,2	33,0	33,0	1,2	3,433	Si
ALT	U/L	85	83	81	82	4	5,720	Si
Amilasa	U/L	-	107	110	-	3	5,685	Si
AST	U/L	96	-	91	92	5	5,307	Si
Bilirrubina directa	µmol/L	20	-	20	-	0	4,490	Si
Bilirrubina total	µmol/L	40	39	40	39	1	5,906	Si
Calcio	mmol/L	2,08	2,05	2,09	2,09	0,04	0,171	Si
Cloruro	mmol/L	11,1	11,0	109	-	2	5,378	Si
Colesterol	mmol/L	-	3,76	3,75	3,72	0,04	0,207	Si
CK	U/L	-	202	199	199	3	17,430	Si
Creatinina	µmol/L	194	196	187	175	21	39,640	Si
Fosfatasa alcalina	U/L	21,1	21,0	209	195	16	17,804	Si
Fosfato	mmol/L	1,34	1,36	1,36	1,35	0,02	0,066	Si
GGT	U/L	75	74	73	75	2	7,798	Si
Glucosa	mmol/L	7,5	7,4	7,6	7,6	0,215	0,563	Si
LDH	U/L	203	-	199	-	4	30,576	Si
Magnesio	mmol/L	0,92	-	0,91	0,89	0,03	0,070	Si
PCR	mg/L	8,8	8,7	8,6	8,5	0,3	1,868	Si
Potasio	mmol/L	4,01	4,02	3,98	-	0,04	0,095	Si
Proteínas totales	g/L	48,3	48,3	47,0	48,3	1,3	4,190	Si
Sodio	mmol/L	144	143	142	-	2	3,718	Si
Triglicéridos	mmol/L	1,49	-	1,51	1,49	0,02	0,175	Si
Urea	mmol/L	13,9	13,7	14,0	13,7	0,3	1,136	Si

Se indica “-” cuando una magnitud no está disponible en el analizador.

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa; PCR: proteína C reactiva.

Tabla 15. Diferencias máximas observadas entre analizadores y evaluación de la intercambiabilidad de resultados a concentraciones (o actividad) cercanas al nivel 3 del material de control.

Magnitud	Unidades	Concentración o actividad media				Diferencia máxima observada	Diferencia máxima permitida	Intercambiabilidad
		Aliinity 1	Aliinity 2	Aliinity 3	Aliinity 4			
Ácido úrico	mmol/L	-	0,48	0,48	-	0	0,029	SI
Albumina	g/L	45,6	46,2	45,3	44,8	1,4	4,625	SI
ALT	U/L	155	153	150	147	8	11,264	SI
Amlasa	U/L	-	425	437	-	12	23,608	SI
AST	U/L	237	-	230	229	8	12,884	SI
Bilirrubina directa	µmol/L	39	-	39	-	0	8,552	SI
Bilirrubina total	µmol/L	108	105	108	109	4	15,092	SI
Calcio	mmol/L	2,76	2,80	2,75	2,72	0,08	0,229	SI
Cloruro	mmol/L	112	111	111	-	1	5,478	SI
Colesterol	mmol/L	-	6,40	6,45	6,40	0,05	0,196	SI
CK	U/L	-	461	457	457	4	39,392	SI
Creatinina	µmol/L	507	497	483	458	49	96,897	SI
Fosfatasa alcalina	U/L	360	361	356	332	29	29,812	SI
Fosfato	mmol/L	1,90	1,89	1,87	1,89	0,03	0,092	SI
GGT	U/L	165	163	164	164	2	9,834	SI
Glucosa	mmol/L	14,2	14,1	14,3	14,1	0,2	1,055	SI
LDH	U/L	563	-	536	-	27	86,877	SI
Magnesio	mmol/L	1,48	-	1,46	1,48	0,02	0,112	SI
PCR	mg/L	24,7	24,8	25,0	25,2	0,5	5,344	SI
Potasio	mmol/L	5,98	5,98	5,90	-	0,08	0,138	SI
Proteínas totales	g/L	71,4	72,5	70,8	72,4	1,7	6,147	SI
Sodio	mmol/L	141	141	139	-	2	3,692	SI
Triglicéridos	mmol/L	2,60	-	2,57	2,57	0,03	0,285	SI
Urea	mmol/L	27,3	26,6	26,5	26,7	0,8	2,243	SI

Se indica “-” cuando una magnitud no está disponible en el analizador.

ALT: alanina aminotransferasa; AST: aspartato aminotransferasa; CK: creatina quinasa; GGT: gamma glutamil transferasa; LDH: lactato deshidrogenasa; PCR: proteína C reactiva.



La finalidad de los estudios de intercambiabilidad de resultados entre analizadores de forma periódica es detectar errores analíticos que podrían pasar desapercibidos si únicamente se analizan en paralelo los materiales de control de calidad. Deben por tanto formar parte de la estrategia de control de calidad de los laboratorios para asegurar la calidad de los resultados.

Se consideró que uno de los aspectos clave a la hora de diseñar el estudio de intercambiabilidad era establecer el criterio de sesgo teniendo en cuenta el rendimiento analítico de los procedimientos de medida al nivel de concentración estudiado. Si bien el objetivo es poder establecer requisitos que permitan realizar una correcta interpretación de los resultados según la situación clínica del paciente, o en su defecto, que permitan asegurar que un cambio en el resultado de un paciente no es debido a un error analítico, al establecer el criterio de aceptación de las diferencias entre resultados basándose en la variabilidad biológica, la guía resultó impracticable para varias magnitudes.

De hecho, ésta resultó ser la mayor limitación de este estudio ya que, tal y como se ha observado también en otras áreas de especialización (75), para muchas magnitudes el rendimiento analítico de la mejor tecnología disponible en la actualidad no es capaz de alcanzar los objetivos derivados de la variabilidad biológica. Así, para algunas magnitudes, como por ejemplo albúmina, calcio, cloruro, creatinina, magnesio y sodio se observó que, al establecer el criterio de sesgo a partir del  $CV_i$ , el valor de la ST superaba la DC en los dos niveles de concentración estudiados, poniendo de manifiesto que el rendimiento analítico de los analizadores estaba muy por debajo del objetivo deseable. Otros autores que han utilizado la guía para verificar la intercambiabilidad de resultados entre dos analizadores tuvieron esta misma limitación (95).

Una de las mayores ventajas de la metodología propuesta en esta Tesis Doctoral para el estudio periódico de intercambiabilidad es que aporta objetividad en la evaluación de los resultados. Sin embargo, igual que se ha demostrado en los tres primeros estudios, la adecuada selección de las especificaciones de calidad es clave para evaluar los resultados correctamente. Así, en

función del valor de sesgo fijado como objetivo, los resultados podrían pasar de considerarse intercambiables a no intercambiables, y dar lugar a un análisis de causas y/o acciones derivadas innecesarias.

Dado que la finalidad del estudio de intercambiabilidad de resultados es detectar diferencias que puedan dar lugar a un error en la interpretación clínica, para aquellas magnitudes con una amplia variabilidad biológica, se decidió no ajustar el criterio de sesgo en función de lo que el rendimiento analítico de la tecnología permitía alcanzar, ya que hubiese dado lugar a unas especificaciones más exigentes de lo necesario.

También podría ocurrir lo contrario y asumirse que los resultados son intercambiables cuando en realidad presentan diferencias significativas. A este respecto, se tuvo que prestar especial atención a la evaluación de los resultados en aquellas magnitudes en las que se estableció la diferencia crítica a partir del estado del arte, puesto que este criterio no guarda relación alguna con el estado clínico del paciente.

Otro aspecto que resultó fundamental a la hora de diseñar el estudio de intercambiabilidad fue adaptarlo a la realidad del laboratorio. Al aplicar la metodología propuesta por la guía nos encontramos que para algunas magnitudes como, por ejemplo, amilasa y glucosa en los dos niveles de actividad / concentración estudiados, a pesar de ser la DC superior a la ST, el valor de la ratio SR/ST no se encontraba representado en las tablas de la guía. En estos casos también fue necesario establecer un criterio de sesgo menos estricto para conseguir una ratio DC/ST más elevada de forma que la ratio SR/ST no fuera condicionante. Y, por otro lado, para otras magnitudes como, por ejemplo, GGT y ALT a una actividad cercana a la de los niveles 2 y 3 de los materiales de control, el protocolo establecido indicaba realizar un elevado número de replicados, pudiendo superar incluso los 100 replicados. En este caso, el responsable del laboratorio, atendiendo a los recursos disponibles, decidió limitar el número de replicados. Se

ha de señalar que la guía contempla la posibilidad de realizar las modificaciones que han sido realizadas en la metodología propuesta para subsanar sus propias limitaciones.

Dada la complejidad del protocolo, se puede considerar que su aplicación de forma frecuente para todas las magnitudes disponibles en varios analizadores puede resultar poco práctica, si bien es cierto que, una vez se ha implementado en el laboratorio, su aplicación resulta más ágil y que, además, puede ser útil para estudiar la intercambiabilidad de resultados en situaciones puntuales como por ejemplo un cambio de lote del reactivo, actividades de mantenimiento, etc.

En cualquier caso, se ha puesto de manifiesto que lo más conveniente es establecer la periodicidad de las verificaciones de forma individual para cada magnitud en función del impacto que tendría la entrega de resultados no intercambiables en la seguridad del paciente y teniendo en cuenta los recursos disponibles; como, por ejemplo, el tiempo de dedicación del personal del laboratorio en la selección de las muestras para el estudio, en la realización de las mediciones, y el coste asociado al consumo de reactivos.

### **3.1.5. Conclusión**

La guía EP31-A-IR ha sido útil para diseñar un protocolo para el estudio de verificación periódica de la intercambiabilidad de resultados entre varios analizadores en nuestro laboratorio. Debe tenerse en cuenta que para que la guía fuese practicable, se tuvieron que utilizar criterios de sesgo que fuesen técnicamente alcanzables, así como ajustar el número de mediciones analíticas a los recursos del laboratorio. Las modificaciones se realizaron de acuerdo a lo que indica la guía para que los laboratorios puedan superar las limitaciones que se observen.

A pesar de las modificaciones introducidas, la complejidad en la realización del estudio continuó siendo considerable. Sin embargo, ésta queda compensada por la objetividad que proporciona en la evaluación de los resultados y por la seguridad que puede aportar en respuesta al requisito de aseguramiento de la validez de los resultados al enfrentarse a una auditoría de acreditación.

Es importante resaltar que a pesar de que el estudio ha sido realizado para magnitudes de bioquímica, sería posible aplicarlo para procedimientos cuantitativos de cualquier otra área de especialización.

**4. Cuarto objetivo: Establecer una metodología para verificar la fiabilidad de los resultados ante la ausencia de un programa de intercomparación.**

**4.1. Quinto estudio**

**Evaluación externa de la calidad cuando no existe un programa de intercomparación: experiencia en la comparación entre dos laboratorios.**

**4.1.1. Presentación**

La participación en PEECs es una herramienta esencial para asegurar la fiabilidad de los resultados. Algunas de las recomendaciones de la norma ISO 15189 (67) para la selección adecuada de PEECs son que debe darse preferencia a aquellos programas que dispongan de la acreditación según la norma ISO/IEC 17043 (68) y que, en la medida de lo posible, cubran todas las fases del proceso analítico. Sin embargo, no existen PEECs acreditados para todas las magnitudes que se miden en los laboratorios clínicos. Entre las magnitudes que no disponen de ningún tipo de PEEC en el mercado destacan aquellas que se miden sólo en unos pocos laboratorios muy especializados. En estos casos, así como en aquellos en los que existiendo un PEEC, se considere que por sus características es inadecuado, el personal del laboratorio debe mostrar que es competente para buscar una estrategia alternativa que le permita garantizar la validez de los resultados (96). Esto implica que, además de establecer la metodología que se va a utilizar, deberá evaluar la efectividad de la misma y, evidentemente, definir el criterio que se va a utilizar para evaluar los resultados del estudio realizado.

El presente estudio surgió a raíz de la necesidad de acreditar la determinación de la concentración de heparán sulfato. La cuantificación de heparán sulfato se utiliza para el diagnóstico y el seguimiento de la enfermedad de Sanfilippo de tipo A, una enfermedad rara

también conocida como mucopolisacaridosis de tipo III. Se trata de un trastorno metabólico de herencia autosómica recesiva que provoca el déficit de la enzima heparan sulfatasa, dando lugar a una acumulación de heparán sulfato principalmente en el cerebro, lo que conlleva a una grave degeneración del sistema nervioso central. Así, las personas con esta enfermedad presentan concentraciones de heparán sulfato más elevadas que la población sana. El percentil 95 de la concentración de heparán sulfato en muestras de sangre en población sana estimado por el laboratorio fue de 0,880 mmol/L.

Debido a que el uso de la cuantificación de esta magnitud va dirigido a una población muy específica y que, al mismo tiempo, el procedimiento analítico es complejo (97), hay pocos laboratorios que realicen esta medición; y, en consecuencia, no existe ningún PEEC disponible. Entre los posibles enfoques alternativos a los PEECs que propone la norma ISO 15189, se optó por realizar la comparación de resultados mediante el envío de muestras a otro laboratorio, procedimiento definido como split-sample test.

El objetivo de este estudio fue establecer una metodología para verificar la fiabilidad de los resultados ante la ausencia de un programa de intercomparación formal. Para ello, se diseñó un estudio siguiendo la metodología indicada en el apéndice E de la guía QMS24 del CLSI (98). Concretamente se enviaron cinco muestras, dos muestras de pacientes y tres muestras de control de calidad interno, que se analizaron por duplicado en cada laboratorio utilizando el mismo procedimiento de medida (cromatografía líquida de ultra alto rendimiento-espectrometría de masas en tándem (UPLC-MS/MS)). A partir de los resultados obtenidos se calcularon, por un lado, las diferencias interlaboratorio entre las concentraciones medias obtenidas para cada muestra; y, por otro lado, las diferencias máximas permitidas para las concentraciones de las muestras del estudio. La evaluación del ejercicio de intercomparación entre los dos laboratorios se consideró satisfactoria cuando las diferencias interlaboratorio no excedían las diferencias máximas permitidas. En caso contrario, sería necesario realizar un

análisis causa raíz para identificar las posibles causas y proceder a implementar las medidas correctivas oportunas.

#### **4.1.2. Resultados y discusión**

Las diferencias observadas entre los dos laboratorios no excedieron en ningún caso las diferencias máximas permitidas, por lo tanto, la evaluación del estudio fue satisfactoria.

Observando detenidamente los resultados llamó la atención la mayor variabilidad entre laboratorios obtenida en las muestras de los pacientes en comparación con las muestras de control de calidad interno. Este hecho se relacionó con la diferente naturaleza de las muestras del estudio y con el procedimiento preanalítico previo que requiere la medición de heparán sulfato.

Para cuantificar la concentración de heparán sulfato en plasma es preciso realizar una digestión enzimática previa. El procedimiento de medida cuantifica cada uno de los disacáridos obtenidos tras la digestión. Tras ello, la concentración de heparán sulfato en plasma se calcula como la suma de la concentración de los disacáridos. En el caso de las muestras de los pacientes todo el heparán sulfato presente debía sufrir el proceso de digestión. En las muestras de control de calidad se añadieron directamente los diferentes disacáridos a un pool de plasma (de personas sanas) utilizado como matriz, así que el proceso de digestión afectaba únicamente al heparán sulfato presente en el pool de plasma. Por este motivo, a pesar de que las condiciones para llevar a cabo la digestión fueron previamente acordadas, el rendimiento de la digestión en cada laboratorio pudo inferir variabilidad en los resultados. Fue útil por tanto el haber incluido en el envío ambos tipos de muestra, porque así se pudo constatar que la mayor variabilidad interlaboratorios observada en las muestras de los pacientes no era debida a un bajo rendimiento analítico del procedimiento de medida de uno de los dos laboratorios.

Otro de los aspectos que llama la atención fue el valor obtenido para las diferencias máximas permitidas. Al realizar la evaluación de las diferencias obtenidas entre los laboratorios, parece que las diferencias máximas permitidas eran muy permisivas (aproximadamente de un 40 % en todas las concentraciones del estudio).

Al analizar detalladamente el ejemplo que se muestra en el apéndice E de la guía QMS24, se observa que las diferencias máximas permitidas podían ser más o menos permisivas en función de los datos de partida. Al calcularlas a partir de los datos derivados del intercambio de muestras que se proporcionan en el ejemplo, todas las muestras se encontraban dentro del rango permitido. En cambio, al calcularlas utilizando los datos de imprecisión derivados de los resultados de control interno, dos de las muestras excedían la diferencia máxima permitida. Parece por tanto que la segunda opción podría ser más adecuada, ya que resultó menos permisiva que la primera. Sin embargo, es posible que, como ocurrió en nuestro caso, no se disponga de los datos de imprecisión del otro laboratorio y que solamente sea posible utilizar la primera opción. En tal caso, hasta que no se disponga de más datos derivados de nuevos estudios de intercomparación, las diferencias observadas entre laboratorios deben evaluarse con precaución.

Es importante tener en cuenta que, debido a las limitaciones que presentan las herramientas estadísticas para la interpretación de los resultados cuando el número de participantes o de muestras comparadas es muy bajo (99), y a que no se dispone de especificaciones de calidad basadas en ninguno de los modelos resultantes de la conferencia de consenso de Milán, deben ser los profesionales los que en último término determinen la validez de la evaluación en base a su experiencia y al uso previsto de los resultados.

A este respecto, hay que destacar que, dadas las peculiaridades de esta magnitud, buscar la intercambiabilidad de resultados no tenía sentido más allá de identificar diferencias que nos pudiesen hacer pensar que se había cometido un error en alguna etapa del proceso analítico. El



principal motivo es que el diagnóstico y tratamiento de estos pacientes se lleva a cabo en un único centro, y por lo tanto las mediciones se realizan siempre en el mismo laboratorio. Así, las decisiones clínicas se toman teniendo en cuenta la evolución de la concentración de heparán sulfato en las muestras medidas a lo largo del tratamiento del paciente.

#### **4.1.3. Conclusión**

A pesar de que los laboratorios que estén acreditados, o que quieran acreditarse según la norma ISO 15189, deban buscar herramientas alternativas para realizar el seguimiento del rendimiento analítico cuando no exista un PEEC (o cuando exista, pero no se considere adecuado), no se conocen publicaciones que presenten la estrategia llevada a cabo en el marco de esta Tesis Doctoral. Por ello, este estudio puede ser de gran utilidad para otros laboratorios que, para dar respuesta a este requisito, decidan realizar un intercambio de muestras con otro laboratorio, ya que, a pesar de tratarse de una metodología compleja y con algunas limitaciones, aporta objetividad a la evaluación de los resultados.

#### **4.1.4. Cita al artículo publicado**

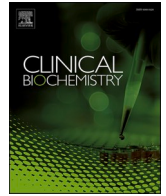
Guiñón L, García-Villoria J, Ribes A, Gort L, Molina A, Soler A, Sahuquillo A, Alvarez L. **External quality assessment in the absence of proficiency testing: a split-sample testing program experience.** Clinical Biochemistry. 2021;97:78-81.

#### **4.1.5. Publicación completa**



Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Clinical Biochemistry

journal homepage: [www.elsevier.com/locate/clinbiochem](http://www.elsevier.com/locate/clinbiochem)

Short Communication

## External quality assessment in the absence of proficiency testing: A split-sample testing program experience

Leonor Guinón<sup>a,b,c,\*</sup>, Judit García-Villoria<sup>d</sup>, Antonia Ribes<sup>d</sup>, Laura Gort<sup>d</sup>, Angel Molina<sup>e</sup>, Anna Soler<sup>a</sup>, Angeles Sahuquillo<sup>f</sup>, Luisa Alvarez<sup>a</sup>

<sup>a</sup> Quality Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Spain

<sup>b</sup> Quality Department, Laboratories, Hospital de la Santa Creu i Sant Pau of Barcelona, Spain

<sup>c</sup> Biochemistry Department, Hospital de la Santa Creu i Sant Pau of Barcelona-IIB Sant Pau, Spain

<sup>d</sup> Division of Inborn Errors of Metabolism, Biochemistry and Molecular Genetics Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, CIBERER, IDIBAPS, Spain

<sup>e</sup> Core Laboratory, Biochemistry and Molecular Genetics Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Spain

<sup>f</sup> Section of Analytical Chemistry, Faculty of Chemistry, University of Barcelona, Spain



### ARTICLE INFO

#### Keywords:

Interlaboratory comparison  
Proficiency testing  
Split-sample  
QMS24

### ABSTRACT

In the field of laboratory medicine, proficiency testing is a vehicle used to improve the reliability of reported results. When proficiency tests are unavailable for a given analyte, an alternative approach is required to ensure adherence to the International Organization for Standardization (ISO) 15189:2012 standard. In this study, we report the results of a split-sample testing program performed as an alternative to a formal PT. This testing method was based on recommendations provided in the Clinical and Laboratory Standards Institute (CLSI) QMS24 guideline. Two different laboratories measured, in duplicate, the heparan sulfate concentration in five samples using ultra-performance liquid chromatography and tandem mass spectrometry. The data analysis to determine the criterion used for the comparability assessment between the two laboratories was based on Appendix E of the QMS24 guideline. Mean interlaboratory differences fell within the maximum allowable differences calculated from the application of the QMS24 guideline, indicating that the results obtained by the two laboratories were comparable across the concentrations tested. Application of the QMS24 split-sample testing procedure allows laboratories to objectively assess test results, thus providing the evidence needed to face an accreditation audit with confidence. However, due to the limitations of statistical analyses in small samples (participants and/or materials), laboratory specialists should assess whether the maximum allowable differences obtained are suitable for the intended use, and make adjustments if necessary.

### 1. Introduction

Proficiency testing (PT) is a type of interlaboratory comparison to monitor laboratory performance and identify potential analytical problems [1]. PT is commonly required by accreditation bodies and regulatory agencies since PT performance has been shown to reflect the quality of patient specimen testing [2]. The International Organization for Standardization (ISO) 15189:2012 standard requires PT in order to ensure testing reliability and quality [3]. According to that standard, PT that meets the ISO/IEC 17043 standard [1] should be performed whenever possible. However, some tests are performed only in a few highly specialized clinical laboratories for which no accredited (or non-

accredited) PTs are available. In this case, the ISO 15189:2012 standard (subclause 5.6.3.2) specifies that an alternative approach—such as sending samples to another laboratory for confirmation—should be used to verify the reliability of the test results. This method is referred to as split-sample testing [4].

When planning split-sample testing, it is essential to establish a criterion for the comparability assessment of the interlaboratory results. The Clinical and Laboratory Standards Institute (CLSI) QMS24 guideline provides methods to assess test performance when commercial PTs are not available [4]. One of the proposals given in those guidelines (Appendix E) is a procedure to determine the maximum allowable differences (MAD) between two laboratories in a split-sample testing

\* Corresponding author at: Sant Quintí 89, 08041 Barcelona, Spain.

E-mail address: [leonorguion@yahoo.es](mailto:leonorguion@yahoo.es) (L. Guinón).

<https://doi.org/10.1016/j.clinbiochem.2021.07.011>

Received 14 April 2021; Received in revised form 21 July 2021; Accepted 23 July 2021

Available online 27 July 2021

0009-9120/© 2021 The Canadian Society of Clinical Chemists. Published by Elsevier Inc. All rights reserved.

program.

To comply with the ISO 15189:2012 requirements, our laboratory performed a split-sample testing program to verify the reliability of the heparan sulfate (HS) measurement. Due to the particularities of the methodology [5], only a few laboratories in the world perform this measurement and no PT is available for this test.

In this context, the aim of the present study is to demonstrate the value of the QMS24 guideline to assess a split-sample testing program. This study may provide valuable information to assist other clinical laboratories in evaluating laboratory results sourced from only two laboratories.

## 2. Materials and methods

Split-sample testing was performed in two laboratories (A and B). Laboratory A sent five samples for testing to Laboratory B, as follows:

- Two plasma samples obtained from patients with Sanfilippo disease type A, also known as mucopolysaccharidosis type III (Patients 1 and 2). Blood samples were collected in lithium-heparin tubes.
- Three internal quality control (IQC) samples were prepared by adding the following disaccharides to a human plasma pool:
  - ΔUA – GlcNAc
  - ΔUA – GlcNS
  - ΔUA - GlcNAc,6S

where ΔUA: 4,5-unsaturated uronic acid; GlcNAc: N-acetyl-D-glucosamine; GlcNS: N-sulfo-D-glucosamine; GlcNAc,6S: N-acetyl-6-O-sulfo-D-glucosamine. All disaccharides were provided by AMSBIO (AMS Biotechnology; Abingdon, United Kingdom [UK]).

The final concentrations of the IQC samples were as follows: 0.664 mmol/L (C1), 2.658 mmol/L (C2) and 10.634 mmol/L (C3).

This study was approved by the Ethics Committee of our Hospital. Informed consent was obtained from all patients. All samples used were exclusively anonymized leftover samples. Chondrosine (GlycoSyn; Lower Hutt, New Zealand) was added as the internal standard to all samples at a concentration of 0.013 mmol/L. The digestion of HS, carried out on all samples, was performed at both laboratories using heparitinase I at a concentration of 0.833 μkat/L (AMS Biotechnology; Abingdon, UK). Each sample was placed on a hot plate at 37 °C for 3 h and then centrifuged at 13,000 r.p.m for 20 min. The digestion products were the disaccharides ΔUA-GlcNAc, ΔUA-GlcNS and ΔUA-GlcNAc,6S.

Both laboratories used ultra-performance liquid chromatography (UPLC) and tandem mass spectrometry (MS/MS). Laboratory A ran the tests on an Acquity UPLC system coupled to Xevo TQ-STM (Waters Chromatography; The Netherlands) and Laboratory B on an Agilent 1260 LC System coupled to an Agilent 6460 Mass Spectrometer (Agilent Technologies; Santa Clara, USA). The HS concentration was obtained by adding the concentration of the three disaccharides.

Each sample was tested in duplicate. The comparability assessment was based on Appendix E of the QMS24 guideline [4]. MADs were calculated with a 95% confidence interval (CI) using the following equation:

$$MAD = z_{1-\alpha/2} \sqrt{\sigma_I^2 + (\sigma_A^2/n_A) + (\sigma_B^2/n_B)}$$

MAD: maximum allowable difference between any pair of results

α: confidence level (0.05)

$z_{1-\alpha/2}$ : z-value corresponding to the level  $1-\alpha/2$

$\sigma_I^2$ : interlaboratory variance

$\sigma_A^2$ : variance for laboratory A

$\sigma_B^2$ : variance for laboratory B

$n_A$ : number of replicates from laboratory A

$n_B$ : number of replicates from laboratory B

Intra- and inter-laboratory variance for laboratories A and B were calculated as the square of the multiplication of the corresponding coefficient of variation (CV) by the mean results obtained by laboratory A, which were defined as the reference values given the greater experience with this type of measurement at laboratory A:

$$\sigma^2 = (CV \cdot X_A)^2$$

$\sigma^2$ : variance

CV: coefficient of variation

$X_A$ : mean of the results of the replicates at laboratory A

CVs were estimated as the square root of the sum of squared proportional differences. For laboratories A and B, these were calculated using the formula:

$$CV = \sqrt{\frac{\sum diff^2}{n-1}}$$

$$diff = \frac{(rep_1 - rep_2)}{(rep_1 + rep_2)/2}$$

CV: coefficient of variation

diff: proportional difference between replicates

n: number of replicates

rep: result of each replicate

The interlaboratory CV was calculated as the square root of the difference between the variance of the proportional differences and the combined repeatability, as follows:

$$CV_I = \sqrt{\left(\frac{\sum diff X_{A,B}^2}{n-1}\right) - \left(\frac{CV_A^2 + CV_B^2}{2}\right)}$$

$$diff X_{A,B} = \frac{X_A - X_B}{(X_A + X_B)/2}$$

$CV_I$ : interlaboratory coefficient of variation

diff  $X_{A,B}$ : proportional difference between the mean values obtained in laboratories A and B

$X_A$ : mean of the results of the replicates at laboratory A

$X_B$ : mean of the results of the replicates at laboratory B

n: number of samples

Finally, the mean interlaboratory differences were compared with the MADs for each concentration. These results were considered equivalent when the differences between the laboratories fell within the CI ( $\pm$ MAD).

## 3. Results

Table 1 shows the results of the replicates and the mean values from the HS measurement obtained by laboratories A and B, as well as the CVs. The mean interlaboratory differences and the interlaboratory CV are also shown.

Calculations to estimate the MADs between each pair of mean values are shown in Table 2. The mean interlaboratory differences fell within the MADs, indicating that the two laboratories obtained comparable results. The MADs obtained by applying the QMS24 guideline showed an acceptability range of approximately 40% across the concentrations tested.

## 4. Discussion

The present study shows our experience in applying the QMS24

**Table 1**

Results of the split-sample testing program and estimation of the coefficients of variation (laboratory A, laboratory B and interlaboratory). All heparan sulfate measurements are expressed in mmol/L.

Laboratory A						
Sample	Replicate 1	Replicate 2	Mean	Difference between replicates	Proportional difference	Squared proportional difference
C1	0.662	0.613	0.637	0.049	0.076	0.00585
C2	2.557	2.633	2.595	−0.076	−0.029	0.00087
C3	10.986	11.292	11.139	−0.306	−0.027	0.00075
Patient 1	3.957	3.957	3.957	0.000	0.000	0.00000
Patient 2	4.146	4.651	4.398	−0.504	−0.115	0.01314
Sum of squared proportional differences						0.021
Coefficient of variation						0.072
Laboratory B						
Sample	Replicate 1	Replicate 2	Mean	Difference between replicates	Proportional difference	Squared proportional difference
C1	0.610	0.673	0.641	−0.063	−0.098	0.00958
C2	2.594	2.603	2.599	−0.009	−0.004	0.00001
C3	13.073	11.312	12.192	1.761	0.144	0.02087
Patient 1	3.162	3.268	3.215	−0.107	−0.033	0.00110
Patient 2	3.094	3.028	3.061	0.066	0.022	0.00047
Sum of squared proportional differences						0.032
Coefficient of variation						0.090
Interlaboratory						
Sample	Laboratory A, mean	Laboratory B, mean	Mean interlaboratory difference	Interlaboratory proportional difference between means	Squared proportional difference	
C1	0.637	0.641	−0.004	−0.007	0.000042	
C2	2.595	2.599	−0.004	−0.001	0.000002	
C3	11.139	12.192	−1.054	−0.090	0.008156	
Patient 1	3.957	3.215	0.742	0.207	0.042855	
Patient 2	4.398	3.061	1.337	0.359	0.128535	
Sum of squared proportional differences					0.180	
Variance of the proportional differences					0.045	
Combined repeatability variance					0.007	
Interlaboratory repeatability variance					0.038	
Interlaboratory coefficient of variation					0.196	

**Table 2**

Maximum allowable differences between laboratories for each sample.

Sample	$\sigma_A^2$	$\sigma_B^2$	$\sigma_I^2$	MAD (mmol/L)	Mean interlaboratory difference (mmol/L)
C1	0.002	0.003	0.016	± 0.265	−0.004
C2	0.035	0.054	0.258	± 1.078	−0.004
C3	0.639	0.994	4.754	± 4.626	−1.054
Patient 1	0.081	0.125	0.600	± 1.644	0.742
Patient 2	0.100	0.155	0.741	± 1.827	1.337

( $\sigma_A^2$ : variance for laboratory A;  $\sigma_B^2$ : variance for laboratory B;  $\sigma_I^2$ : interlaboratory variance; MAD: maximum allowable difference)

guideline to determine the MADs to assess the comparability between tests performed in two laboratories in a split-sample testing program.

The mean interlaboratory differences obtained in this study led us to question whether the MADs were too wide. Indeed, as observed by other authors, when the number of participants in an interlaboratory comparison is so limited, statistical analyses can increase the risk of false acceptance of the results [6]. Ideally, MADs should be based on any of the models included in the Milan hierarchy [7]; but no quality specifications have been published to date for HS. Therefore, the only objective criterion currently available is the one obtained by the application of the QMS24 guideline. In these cases, individual laboratories must assess the suitability of using this criterion based on their previous experience.

Several aspects of the split-sample testing should be highlighted. On the one hand, despite the differences observed between replicates (e.g., sample C3 from Laboratory B), the variability in results fell within the maximum variability established by our laboratory (15%). On the other

hand, the patient samples showed much greater variation between laboratories than the control materials. This is mainly because the HS concentration in plasma depends on enzymatic digestion carried out in the preanalytical phase to obtain the corresponding disaccharides. Control materials were prepared by adding disaccharides to a human plasma pool. Therefore, a higher difference was expected in patient sample results since HS digestion was an additional source of variation compared to control materials.

The QMS24 document proposes two methodologies to calculate the MADs in a split-sample program. One option is to estimate the variances using the split-sample testing data. The other option is to estimate the variances using the CVs derived from IQC data. However, results vary depending on the methodology used to estimate the variances. In the example provided in Appendix E of the QMS24 guideline, if variances are calculated from the split-sample data, no sample exceeds the MAD. However, when variances are calculated from the IQC data, two samples exceed the MAD. Therefore, the MAD depends on the methodology used and is less stringent when split-sample data are used for the calculation. These results suggest that the MADs obtained from IQC data seem to be more realistic. Even so, laboratories involved in a split-sample testing may not know the CV from the IQC data of each laboratory. In these cases, when the first option is the only possible one, caution must be taken to avoid accepting differences between results that may not be comparable. Over time, as more data accumulate from other split-sample testing, the MAD will be adjusted and it may be as appropriate as the one obtained from the IQC data.

To our knowledge, no previous studies have evaluated the alternative assessment approaches applied by other clinical laboratories. Although some authors have described the statistical treatment of data in accredited PTs involving a limited number of participants, those

studies are all intercomparison exercises involving more than two laboratories (mostly in the area of analytical chemistry) [8,9].

A split-sample program is valuable because it allows the laboratory specialists to identify interlaboratory differences. However, when no differences are identified, it cannot be assumed that results of both laboratories are correct, since they can be doing the testing incorrectly and yet achieve similar albeit incorrect results. In either case, this approach does not truly assess accuracy unless a certified reference material or a reference measurement procedure is used. Neither of these options was available for the measurand in our study, but the IQC material used can serve as a form of accuracy benchmark. Consequently, if interlaboratory differences had exceeded the MADs, it would have been possible to determine with reasonable certainty which laboratory had a bias. It is important to keep in mind that the HS measurement is always performed in the same laboratory and medical decisions are based on relative changes in results obtained over time. Therefore, the main value of this interlaboratory comparison is not to ensure that both laboratories provide comparable results, but rather to detect errors that may affect any phase of the analytical procedure. In fact, in some areas of laboratory medicine, such as molecular genetics, an alternative assessment modality is frequently used to supplement PT in order to assess pre- and post-analytical factors that are generally not considered in formal PTs [10].

## 5. Conclusion

The application of the QMS24 guideline in a split-sample testing program allows a laboratory to objectively verify the reliability of the results, thus providing a solid basis of support during the audit process. However, due to the limitations of split-sample testing in only two laboratories, laboratory specialists should ultimately determine whether the criterion provided by this methodology is suitable for the intended use.

## Declarations of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Funding

This research project did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## References

- [1] ISO/IEC 17043, Conformity Assessment—General Requirements for Proficiency Testing, International Organization for Standardization, Geneva, Switzerland, 2010.
- [2] L. Sciacovelli, S. Secchiero, L. Zardo, M. Plebani, The role of the external quality assessment, *Biochem Med* 20 (2) (2010) 160–164, <https://doi.org/10.11613/BM.2010.019>.
- [3] ISO 15189, Medical Laboratories—Requirements for Quality and Competence, International Organization for Standardization, Geneva, Switzerland, 2012.
- [4] CLSI, Using Proficiency Testing and Alternative Assessment to Improve Medical Laboratory Quality. 3rd ed. CLSI Guideline QMS24, Clinical and Laboratory Standards Institute, Wayne, PA, 2016.
- [5] T. Oguma, S. Tomatsu, A.M. Montano, O. Okazaki, Analytical method for the determination of disaccharides derived from keratan, heparan, and dermatan sulfates in human serum and plasma by high-performance liquid chromatography/turbo ionspray ionization tandem mass spectrometry, *Anal Biochem* 368 (1) (2007) 79–86, <https://doi.org/10.1016/j.ab.2007.05.016>.
- [6] E. Szewczak, A. Bondarzewski, Is the assessment of interlaboratory comparison results for a small number of tests and limited number of participants reliable and rational? *Accred Qual Assur* 21 (2) (2016) 91–100, <https://doi.org/10.1007/s00769-016-1195-y>.
- [7] S. Sandberg, F.G. Fraser, A.R. Horvath, R. Jansen, G. Jones, W. Oosterhuis, 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.* 53 (2015) 833–835, <https://doi.org/10.1515/cclm-2015-0067>.
- [8] J.B. Leyva-Morales, P.J. Bastidas-Bastidas, R. Muñoz-Valencia, S.G. Ceballos Magaña, G. Ponce-Vélez, D. Aguilera-Márquez, P. Grajeda-Cota, M.S. Navidad-Murrieta, M.E. Flores-Munguía, J.A. Ramírez-Castillo, C.A. Romero-Bañuelos, R. Ruíz-Ramos, G. Aguilar-Zarate, G. Huerta-Beristain, Measurement of organochlorine pesticides in drinking water: laboratory technical proficiency testing in Mexico, *Accred. Qual. Assur.* 24 (6) (2019) 451–461, <https://doi.org/10.1007/s00769-019-01403-w>.
- [9] M. Kotyczka-Moranska, M. Mastalerz, A. Plis, M. Sciazko, Inter-laboratory proficiency testing of the measurement of gypsum parameters with small numbers of participants, *Accred. Qual. Assur.* 25 (5) (2020) 373–381, <https://doi.org/10.1007/s00769-020-01451-7>.
- [10] D.A. Payne, G. Russomando, M.W. Linder, K. Baluchova, T. Ashavaid, W. Steimer, P. Ahmad-Nejad, External quality assessment (EQA) and alternative assessment procedures (AAPs) in molecular diagnostics: findings of an international survey, *Clin. Chem. Lab. Med.* 59 (2) (2021) 301–306, <https://doi.org/10.1515/cclm-2020-0101>.

**5. Quinto objetivo: Establecer una metodología para evaluar la concordancia entre observadores en aquellas actividades que se consideren subjetivas.**

### **5.1. Sexto estudio**

**Evaluación de la concordancia entre observadores para asegurar la calidad del cribado neonatal.**

#### **5.1.1. Presentación**

A pesar de haberse demostrado que el mayor porcentaje de errores ocurre en actividades de las fases extra-analíticas (12), son pocos los programas de evaluación externa de la calidad que incorporan actividades de estas fases en sus intercomparaciones. Por lo tanto, el laboratorio debe llevar a cabo una revisión de las actividades que se realizan en las fases extra-analíticas con la finalidad de identificar si alguna de ellas puede representar un riesgo para el paciente; y en caso de identificarse alguna, revisar el plan de control de calidad para cerciorarse de que disponga de alguna herramienta de control.

Uno de los aspectos más destacables de la norma ISO 15189 desde su primera edición (66), por el gran impacto que podía tener en la seguridad del paciente, fue la evaluación de la competencia técnica del personal. Este aspecto adquiere especial relevancia en aquellas actividades del proceso analítico que requieren una interpretación. A este respecto, la edición vigente de la norma ISO 15189 (67) en el requisito 7.3.7.2. añade que, además de tener la competencia, debe realizarse una revisión periódica de la adecuación de las interpretaciones con la finalidad de asegurar que hay concordancia entre las opiniones del personal que, en un momento u otro, puede realizar una actividad que presente subjetividad.

El presente estudio es el resultado de la implementación de ejercicios interobservador en el plan de control de la calidad del laboratorio de cribado neonatal. Tal y como se indicó en el segundo estudio de esta Tesis Doctoral, el programa de cribado neonatal está orientado a la detección, diagnóstico y tratamiento de enfermedades endocrino-metabólicas antes de que se manifiesten los síntomas de la enfermedad con la finalidad de reducir la morbimortalidad en los recién nacidos (100).

Una revisión detallada del proceso analítico permitió detectar dos actividades que requerían una interpretación del personal y que, por tanto, podían afectar a la seguridad del paciente.

Concretamente:

- La clasificación de la calidad de las muestras neonatales realizada mediante inspección visual. Las muestras de sangre se recogen en unas tarjetas específicas (DBS – dried blood spot) que contienen un soporte de papel de filtro (validado para el diagnóstico in vitro), tal y como se muestra en la Figura 8. Esta actividad forma parte de la fase pre-analítica. En función de la calidad de la muestra, ésta es aceptada o no para su posterior análisis. Una clasificación incorrecta podría dar lugar a un resultado falso positivo o falso negativo de algunas de las magnitudes que conforman el cribado neonatal (101).



Figura 8. Tarjeta de cribado neonatal con una muestra de calidad aceptable.

- La asignación del patrón de hemoglobinas a partir de la visualización de los electroferogramas obtenidos, un ejemplo de los cuales se muestra en la Figura 9. Esta actividad forma parte de la fase post-analítica. Una interpretación incorrecta del patrón de hemoglobinas podría dar lugar a un resultado incorrecto del cribado de hemoglobinopatías (102).



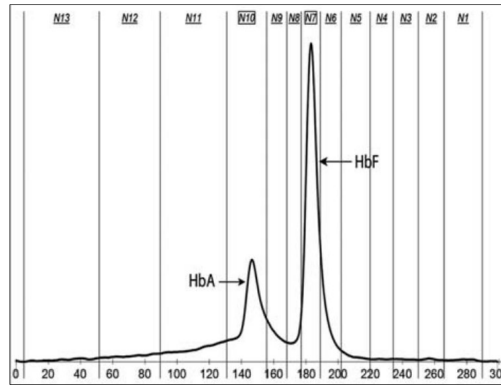


Figura 9. Electroferograma de hemoglobinas en un recién nacido a término con un patrón normal. Adaptada de (102).

Entre los años 2015 y 2020 se realizaron en el laboratorio de cribado neonatal 10 ejercicios interobservador sobre la clasificación de la calidad de las muestras. Entre los años 2017 y 2020 se realizaron 4 ejercicios interobservador sobre la asignación del patrón de hemoglobinas.

El objetivo de este estudio fue establecer una metodología para estimar la concordancia entre observadores para ambas actividades, así como evaluar la eficacia de las acciones implementadas tras haber obtenido un grado de acuerdo insuficiente. Para ello se utilizó el coeficiente alfa de *krippendorff* (103).

El coeficiente alfa de *krippendorff* permite medir el grado de acuerdo entre observadores. La fórmula básica para su cálculo es la ratio entre el desacuerdo observado y el desacuerdo esperado. Sin embargo, a nivel computacional su cálculo es complejo debido a que se realiza un remuestreo (*bootstrap*). Se seleccionó este test estadístico por presentar las siguientes características:

- Puede aplicarse a cualquier número de observadores.
- Puede aplicarse a cualquier número de categorías, escalas de valores o medidas.
- Puede aplicarse a diferentes métricas: nominal, ordinal, intervalos o ratios.
- Puede aplicarse a pesar de que algún observador no proporcione un resultado para alguna de las observaciones.

- Puede aplicarse a cualquier tamaño de muestra.

Para la interpretación de los valores del coeficiente alfa se siguieron las sugerencias de *Krippendorff* (104). En la Tabla 16 se muestra la correspondencia entre el valor del coeficiente alfa y el grado de concordancia entre observadores. Además, el bootstrap permitió obtener la probabilidad de que el valor verdadero de alfa fuese inferior a 0,8.

Tabla 16. Interpretación del coeficiente alfa de *Krippendorff*.

Valor de alfa	Grado de concordancia
1	Perfecto
> 0,80	Bueno
0,67 - 0,80	Bajo
> 0,67	Muy bajo
0	Nulo

### 5.1.2. Resultados y discusión

Para los primeros estudios interobservador realizados para ambas actividades, el grado de concordancia obtenido fue inadecuado. El valor alfa para la clasificación de la calidad de las muestras fue destacadamente inferior (0,59) que el obtenido para la asignación del patrón de hemoglobinas (0,79). El análisis causa raíz realizado puso de manifiesto lo siguiente:

- Clasificación de la calidad de las muestras neonatales: El desacuerdo ocurrió únicamente en la clasificación de una muestra como escasa o como insuficiente, demostrando que faltaba unificar este criterio. Rechazar una muestra que es apta para realizar una parte del cribado (muestra escasa) podría retrasar la detección de una enfermedad. Por otro lado, aceptar una muestra que no es apta para realizar el cribado (muestra insuficiente) podría dar lugar a un resultado falso negativo para algunas magnitudes (101).
- Asignación del patrón de hemoglobinas: El desacuerdo se observó solamente en la decisión de informar o no la presencia de hemoglobina de Bart's. Algunos facultativos

siempre informaban de su presencia y otros únicamente la informaban cuando la cuantificación de la fracción era superior al 15 %.

Como consecuencia, en cuanto a la clasificación de la calidad de las muestras neonatales, se realizó una formación específica a todo el personal para afianzar el criterio de clasificación de una muestra como escasa o insuficiente. En cuanto a la asignación del patrón de hemoglobinas, el personal facultativo consensuó que, dado que la presencia de una hemoglobina de Bart's inferior al 15 % no conllevaba ninguna acción clínica posterior, solamente se informaría cuando la cuantificación fuese superior al 15 % (105).

En los estudios que se realizaron después de haber establecido un consenso y tras las actividades formativas, se consiguió una buena concordancia entre observadores, indicando que las acciones implementadas habían resultado eficaces.

Al cabo de un par de años, en el séptimo estudio interobservador sobre la clasificación de la calidad de las muestras, se obtuvo de nuevo un valor de alfa ligeramente inferior al aceptable ( $\alpha=0,77$ ). Al realizar el análisis causa raíz se observó que el bajo grado de acuerdo correspondía a las interpretaciones realizadas por personal de nueva incorporación que calificaron algunas muestras aceptables como no óptimas. Esto dio lugar a un cambio en la formación planificada del personal de nueva incorporación, ampliando el tiempo de dedicación para adquirir la competencia en la clasificación de la calidad de las muestras.

En los estudios interobservador realizados para la asignación del patrón de hemoglobinas el buen grado de concordancia se mantuvo durante todo el periodo del estudio.

### **5.1.3. Conclusión**

La realización de este tipo de estudios interobservador en el laboratorio de cribado neonatal nos permitió detectar los aspectos más controvertidos en la toma de decisiones e implementar las acciones necesarias para minimizar los riesgos que podrían dar lugar a un retraso en la entrega

del resultado o a un resultado incorrecto suponiendo, en consecuencia, una demora en el diagnóstico o un diagnóstico incorrecto.

La evaluación de la concordancia interobservador fue una herramienta útil tanto para evaluar con objetividad cada uno de los estudios como para verificar la eficacia de las acciones de mejora implementadas. Cabe señalar que el mantenimiento de un buen grado de concordancia fue posible gracias a la estandarización de los procedimientos, al consenso en el criterio a aplicar en cada caso y a la realización de formación continuada para todo el personal.

Por último, debe tenerse en cuenta que la metodología propuesta podría utilizarse para evaluar la concordancia entre observadores de cualquier actividad subjetiva realizada tanto fuera como dentro del ámbito del laboratorio clínico.

#### **5.1.4. Cita al artículo publicado**

Guiñón L, Soler A, López RM, Pajares S, de Aledo JMG, Argudo-Ramírez A, Marín JL, García-Villoria J, Sahuquillo Á, Alvarez L. **Inter-rater reliability assessment for the new-born screening quality assurance**. *Biochemia medica*. 2022;32(3):030901.

#### **5.1.5. Publicación completa**

### Inter-rater reliability assessment for the new-born screening quality assurance

Leonor Guiñón<sup>\*1,2,3</sup>, Anna Soler<sup>1</sup>, Rosa María López<sup>4</sup>, Sonia Pajares<sup>4,5</sup>, José Manuel González de Aledo<sup>4</sup>, Ana Argudo-Ramírez<sup>4</sup>, José Luis Marín<sup>4</sup>, Judit García-Villoria<sup>4,5</sup>, Ángeles Sahuquillo<sup>6</sup>, Luisa Alvarez<sup>1</sup>

<sup>1</sup>Quality Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, Barcelona, Spain

<sup>2</sup>Quality Department, Laboratories, Hospital de la Santa Creu i Sant Pau of Barcelona, Barcelona, Spain

<sup>3</sup>Biochemistry Department, Hospital de la Santa Creu i Sant Pau of Barcelona, Barcelona, Spain

<sup>4</sup>Division of Inborn Errors of Metabolism, Biochemistry and Molecular Genetics Department, Biomedical Diagnostic Center, Hospital Clínic of Barcelona, IDIBAPS, Barcelona Spain

<sup>5</sup>Spain Center for Biomedical Research Network on Rare Diseases (CIBERER), Madrid, Spain

<sup>6</sup>Section of Analytical Chemistry, Faculty of Chemistry, University of Barcelona, Barcelona, Spain

\*Corresponding author: leonorguinon@yahoo.es

#### Abstract

**Introduction:** To ensure the quality of the new-born screening (NBS), our laboratory reviewed the analytical procedure to detect subjective steps that may represent a risk to the patient. Two subjective activities were identified in the extra-analytical phases: the classification of dried blood spots (DBS) according to their quality and the assignment of haemoglobin patterns. To keep these activities under control, inter-rater studies were implemented. This study aimed to evaluate the inter-rater reliability and the effectiveness of the measures taken to improve the agreement between observers, to assure NBS results' quality.

**Materials and methods:** Dried blood spots specimens were used for the inter-rater studies. Ten studies were performed to assess DBS quality classification, and four to assess the assignment of haemoglobin patterns. Krippendorff's alpha test was used to estimate inter-rater reliability. Causes were investigated when alpha values were below 0.80.

**Results:** For both activities, the reliability obtained in the first studies was inadequate. After investigation, we detected that the criterion to classify a DBS as scant was not consolidated, and also a lack of consensus on whether or not to report Bart's haemoglobin depending on its percentage. Alpha estimates became higher once the training was reinforced and a consensus about the appropriate criteria to be applied was reached.

**Conclusion:** Inter-rater reliability assessment helped us to ensure the quality of subjective activities that could add variability to NBS results. Furthermore, the evolution of the alpha value over time allowed us to verify the effectiveness of the measures adopted.

**Keywords:** accreditation; extra-analytical phase; new-born screening; quality control

Submitted: February 24, 2022

Accepted: May 20, 2022

#### Introduction

It is estimated that 70 percent of clinical decisions are based on laboratory tests' results (1). For this reason, laboratories must ensure the reliability of results, since they could directly affect patient safety. Clinical laboratories follow different strategies to assure the quality of reported results. One of the essential tools is participation in external quality assessment and proficiency testing schemes. However, these programs generally do

not address all aspects of the analytical procedure that can affect results' reliability. Extra-analytical steps are usually not included, whereas the majority of errors occur in the pre-analytical (68.2%) or post-analytical phases (18.5%) (2). In this regard, the International Organization for Standardization (ISO) 15189:2012 standard (subclause 5.6.1) requires the implementation of appropriate pre- and post-examination processes for the quality assurance of

examination results (3). Therefore, to add appropriate control measures a thorough revision of the extra-analytical phases must be performed to detect those aspects that may represent a risk to the patient.

New-born screening (NBS) allows the early detection of congenital diseases and it is intended to initiate early treatment to reduce morbidity and mortality. Laboratories receive dried blood spots (DBS) for testing, that is a specimen of blood collected onto a particular filter paper on which printed circles indicate the area to be filled. The panel of screened diseases varies depending on the country. The Catalan NBS program (in Spain) includes the screening of amino acidemias, organic acidemias, fatty acid oxidation disorders, congenital hypothyroidism, cystic fibrosis, sickle cell disease, and severe combined immunodeficiency disease. As in any other screening program, the existence of false positive and negative results is assumed. These false results may be due not only to the biological variability of the measurands, the analytical interferences (*e.g.* caused by drugs or body creams among others), or the analytical performance of technology, but also to failures in other aspects of the NBS program that depend on the decisions taken by the personnel involved in the different phases of the procedure.

Our laboratory reviewed the whole procedure of the NBS to ensure that we provided the highest-quality results. We performed a risk assessment to detect if it was necessary to add control measures to the laboratory's quality control plan. As a result, we identified two potential failures that may represent a risk to the patient and that were not addressed by the quality control plan established. The first one was the classification of the DBS specimen's quality performed in the pre-analytical phase. Despite the existence of sampling guidelines, it is common for NBS laboratories to receive DBS under non-optimal conditions (4). Dried blood spots quality is evaluated by visual inspection; therefore, the decision of acceptance or rejection of specimens can differ between observers, and this decision can lead to false positive or negative results (5). The second one was the assignment of haemoglobin (Hb) patterns carried out in the post-

analytical phase from the visualization of electrophoretograms. In Europe, only a few countries screen for haemoglobinopathies, and only sickle cell disease is included in the screening panels (6). However, the detection of other haemoglobinopathies of interest is also possible. In this regard, the European Consensus issued the recommendation of reporting beta thalassaemia in 2017 (7). In any case, observers' subjectivity may lead to a wrong result in the screening for haemoglobinopathies. Consequently, as both activities could present subjectivity between observers, inter-rater studies were implemented to keep these steps under control.

In this context, this study aimed to evaluate the inter-rater reliability to assure the quality of NBS results, as well as to assess the effectiveness of the measures taken to improve the agreement between observers.

## Materials and methods

### Materials

Blood from a heel prick was collected onto Whatman 903 paper (Whatman International Ltd, Chalfont St Giles, United Kingdom) for NBS. Fifty DBS were used to perform ten inter-rater studies to assess the agreement on the classification of the specimen's quality in the period between 2015 and 2020. Of the selected DBS, 22 had an acceptable quality and 28 were unacceptable (6 scant and 22 non-optimal specimens). Twenty-five DBS were used to conduct four inter-rater studies to assess the agreement on the assignment of Hb patterns between 2017 and 2020. For these studies, the following cases were selected: 2 unaffected, 15 affected (3 beta thalassaemia major, 7 sickle cell disease, 3 alpha thalassaemia, 1 Hb E disease and 1 Hb D disease) and 8 carriers of Hb variants.

Specific data about the samples included in each study and the number of observers who participated are shown in Table 1.

### Methods

The inter-rater studies performed to assess the agreement on the classification of DBS specimen's quality were carried out by visual inspection be-

**TABLE 1.** Number of participants and samples included in the inter-rater studies

<b>Dried blood spots' quality classification</b>				
Study no./Year	Observers	Samples	Samples included	
			N	Classification
1/2015	8	5	2	Acceptable
			2	Scant
			1	Non-optimal
2/2015	7	5	3	Acceptable
			2	Non-optimal
3/2016	6	5	3	Acceptable
			2	Non-optimal
4/2017	7	5	4	Acceptable
			1	Non-optimal
5/2017	7	5	3	Acceptable
			2	Non-optimal
6/2018	7	5	1	Acceptable
			1	Scant
			3	Non-optimal
7/2018	7	5	2	Acceptable
			1	Scant
			2	Non-optimal
8/2019	7	5	1	Acceptable
			1	Scant
			3	Non-optimal
9/2020	7	5	2	Acceptable
			1	Scant
10/2020	7	5	2	Non-optimal
			1	Acceptable
			4	Non-optimal
<b>Haemoglobin pattern assignment</b>				
Study no./Year	Observers	Samples	Samples included	
			N	Hb pattern
1/2017	4	10	1	FA*
			1	F
			1	FS
			1	FCS
			1	FA Bart's
			5	FAX
2/2018	4	5	1	FA*
			1	F
			1	FS
			1	FAC
			1	FAX

**TABLE 1.** Continued.

Study no./Year	Observers	Samples	Samples included	
			N	Hb pattern
3/2019	5	5	1	FS
			1	FE
			1	FA Bart's
			1	FAX
			1	FSX
4/2020	5	5	1	F
			1	FD
			1	FA Bart's
			2	FSA

\*Normal haemoglobin pattern in new-borns. F - foetal haemoglobin. A - adult haemoglobin. Bart's - Bart's haemoglobin. S, C, D, E -  $\beta$ -chain haemoglobin variants S, C, D, and E. X - rare haemoglobin variants.

fore the punching of the spot. Specimens were classified as acceptable, scant, or non-optimal according to the following criteria:

- Acceptable: Blood filled all spots and they were evenly saturated.

- Unacceptable:

1. Scant: Blood transferred to the back side of the paper but it was not enough to perform all tests.

2. Non-optimal: a) Insufficient: Blood did not transfer to the back side of the paper; b) Compressed: Blood filled the spots but they were not saturated; c) Over-impregnated: Blood applied to both sides of the paper; d) Multispotted: Multiple small bloodspots to make one larger.

The inter-rater studies conducted to assess the agreement on the assignment of Hb patterns were carried out by visual inspection of the electrophoretograms obtained by capillary electrophoresis. All tests were run on Capillars 2 Neonat Fast (Serbia, Lisses, France). The criteria followed for reporting the pattern was:

- Hb fractions were sorted by decreasing concentration (*i.e.* a report of FA pattern indicates that the concentration of HbF is higher than that of HbA, which is the normal pattern in new-borns).

- Rare Hb variants (*i.e.* other than S, C, D, E, and Bart's) were coded as X.

### Statistical analysis

Krippendorff's alpha test was used to estimate the inter-rater reliability. A 10,000 bootstrap samples were taken to produce confidence intervals at the 95% level. Estimates of the probability that the true alpha value was less than 0.80 were also obtained (8). Statistical analysis was performed using SPSS v19.

The alpha estimates were interpreted following Krippendorff's suggestions (Table 2) (9). Alpha values close to 1 indicated perfect agreement. Below 0.80 but above 0.67 indicated low reliability. An alpha below 0.67 indicated a really low inter-rater reliability. When alpha values were below 0.80, causes were investigated in order to implement corrective measures.

**TABLE 2.** Interpretation of Krippendorff's alpha

Value of alpha	Level of agreement
1	Perfect
> 0.80	Good
0.67-0.80	Low
< 0.67	Really low
0	None



## Results

For both activities, the results showed that the inter-rater reliability of the first studies performed was inadequate (Table 3). The first alpha estimate for the classification of DBS quality was far lower (0.59) than the one obtained for the assignment of Hb patterns, which was close to achieving a good level of agreement (0.79).

When investigating the causes behind these low alpha values, the lack of unified criteria between the personnel in certain aspects was revealed. Firstly, in the classification of the DBS quality, disagreement between observers in classifying two types of unacceptable specimens was observed: a DBS that had insufficient blood on the spots (*i.e.* the blood did not transfer to the back side of the paper); and, a DBS that was acceptable to perform some tests but was not enough to complete all screening tests (*i.e.* few spots filled with blood and evenly saturated). No differences between observers in the classification of other unacceptable DBS specimens (*i.e.* multispotted, over-impregnated or compressed

bloodspots) were identified. Secondly, in the assignment of Hb patterns, disagreement among the observers in deciding whether or not to report the presence of Bart's Hb depending on the quantity detected was observed. No discrepancies in the assignment of other pathological Hb patterns or the normal Hb pattern (*i.e.* FA) were found.

Once the technicians performed a specific training with insufficient and scant DBS, and the laboratory specialists met a consensus about Bart's Hb, the alpha estimates became higher than 0.80 in the majority of the studies performed, showing strong inter-rater reliability. Only in one study carried out to assess the DBS's quality classification in 2018 an alpha value a bit lower than 0.80 was obtained, with a maximum estimated 64.96 percent chance of alpha failing to exceed 0.80. When investigating the causes, we observed that most of the disagreements belonged to recently hired technical staff, which classified some acceptable DBS specimens as non-optimal.

**TABLE 3.** Inter-rater reliability of the studies performed to assess the classification of dried blood spot's quality and the assignment of haemoglobin patterns

<b>Dried blood spots' quality classification</b>			
<b>Study no. / Year</b>	<b>Alpha</b>	<b>95% CI</b>	<b>q (%)</b>
1/2015	0.59	0.47-0.70	99.99
2/2015	0.88	0.78-0.96	3.76
3/2016	0.86	0.72-0.97	12.23
4/2017	1.00	nc	nc
5/2017	1.00	nc	nc
6/2018	1.00	nc	nc
7/2018	0.77	0.67-0.87	64.96
8/2019	0.85	0.75-0.92	19.66
9/2020	1.00	nc	nc
10/2020	1.00	nc	nc
<b>Haemoglobin pattern assignment</b>			
1/2017	0.79	0.67-0.90	55.39
2/2018	0.88	0.76-1.00	7.37
3/2019	0.91	0.81-0.98	1.63
4/2020	0.90	0.79-0.97	4.30

95% CI – 95% confidence interval. q - probability that the true alpha value was below 0.8. nc - not calculable.

## Discussion

The interpretation of qualitative results presents greater subjectivity than that of quantitative results. Therefore, laboratories must apply the tools at their disposal to make interpretations as objective and homogeneous as possible. In our particular case, the assessment of the inter-rater reliability helped us to ensure the quality of those extra-analytical activities that depended, in whole or in part, on visual examination and that could affect the NBS results.

The root-cause analysis performed to identify the causes of the poor reliability in the first inter-rater study in the classification of DBS quality revealed the need to re-educate the technical staff to clarify the differences between a scant sample and an insufficient sample. Based on this classification, the decision of accepting or rejecting the specimen is taken, leading to different consequences. On the one hand, the rejection of a DBS specimen that is acceptable to perform a portion of tests can lead to a delay in the detection of disease, since it implies requesting a new sample for testing. On the other hand, acceptance of a DBS specimen with insufficient blood may lead to a false negative result for some measurands (5). In both cases, early detection of the disease is not carried out, thus endangering the life of the new-born. In this regard, it should be noted that our laboratory performs testing on DBS specimens with unacceptable quality in order to detect possible positive results, assuming that false positive (e.g. multispotted samples) or false negative results (e.g. compressed bloodspots) can be obtained. However, the final results report is not released until a repeat screening on an additional requested specimen is performed, as long as the new specimen has an acceptable quality. The root-cause analysis performed after the seventh inter-rater study showed the need to enhance training in the classification of DBS specimens for the new staff, to provide them with the expertise required.

The root-cause analysis performed after the first inter-rater study for the assignment of Hb patterns revealed discrepancies between the laboratory specialists when reporting Bart's Hb. Some partici-

pants reported Bart's Hb only when the fraction detected was higher than 15%. Other participants reported always Bart's Hb regardless of the percentage. Based on the fact that low percentages do not require further testing or referral to a haematologist, for the next inter-rater studies, a consensus on reporting Bart's Hb only when the fraction was higher than 15% was reached (10).

The evolution of the alpha value over time allowed us to verify the effectiveness of the corrective measures implemented. For both activities, a better level of agreement between observers was achieved in the subsequent studies. Despite the rapid increase of the alpha value observed in the second inter-rater studies performed, continuous training and education of the laboratory personnel was required to sustain a high level of agreement. Thus, we established to conduct inter-rater studies at least once a year; and, additionally, when recruiting new staff or after a long-term leave.

Other authors have assessed the inter-rater reliability in order to implement actions to minimize the effect of the human factor in laboratory results and have even informed clinicians about the degree of reliability attained (11,12). Some of them have also used Krippendorff's alpha test for the assessment; but, to our knowledge, no previous studies have assessed whether inter-rater variability could affect NBS results (13).

In conclusion we can say that conducting inter-rater exercises is essential to control subjective activities that can lead to an error or delay in diagnosis. Their implementation in our laboratory helped us to detect controversial points in the decision making and to focus corrective measures to standardize important criteria. The assessment of the inter-rater reliability allowed us to verify if the measures taken were and remain effective over time, thus boosting the quality assurance of NBS. Moreover, standardization of subjective activities helped us to accomplish the quality assurance requirements of test results described in the ISO 15189:2012 standard.

### Potential conflict of interest

None declared.

## References

1. Badrick T. Evidence-based laboratory medicine. *Clin Biochem Rev.* 2013;34:43.
2. Plebani M. Quality in laboratory medicine: 50 years on. *Clin Biochem.* 2017;50:101-4. <https://doi.org/10.1016/j.clinbiochem.2016.10.007>
3. International Organization for Standardization (ISO). ISO15189:2012. Medical laboratories – Requirements for quality and competence. Geneva: ISO, 2012.
4. Clinical and Laboratory Standards Institute (CLSI). Blood Collection on Filter Paper for Newborn Screening Programs; Approved Standard, 6th ed. CLSI document NBS01-A6. Wayne, PA: CLSI, 2013.
5. George RS, Moat SJ. Effect of dried blood spot quality on newborn screening analyte concentrations and recommendations for minimum acceptance criteria for sample analysis. *Clin Chem.* 2016;62:466-75. <https://doi.org/10.1373/clinchem.2015.247668>
6. Loeber JG, Platis D, Zetterström RH, Almashanu S, Boemer F, Bonham JR, et al. Neonatal screening in Europe revisited: an ISNS perspective on the current state and developments since 2010. *Int J Neonatal Screen.* 2021;7:15. <https://doi.org/10.3390/ijns7010015>
7. Lobitz S, Telfer P, Cela E, Allaf B, Angastiniotis M, Backman Johansson C, et al. Newborn screening for sickle cell disease in Europe: recommendations from a Pan-European Consensus Conference. *Br J Haematol.* 2018;183:648-60. <https://doi.org/10.1111/bjh.15600>
8. Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas.* 2007;1:77-89. <https://doi.org/10.1080/19312450709336664>
9. Krippendorff K. *Content Analysis: An Introduction to its Methodology.* 2nd ed. Thousand Oaks, CA: Sage; 2004.
10. Tamary H, Dgany O. Alpha-Thalassemia. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Mirzaa GM, Amemiya A, eds. *GeneReviews* [Internet]. 2005 Nov 1 [updated 2020 Oct 1]. Available from: <https://pubmed.ncbi.nlm.nih.gov/20301608/>. Accessed May 6th 2022.
11. Palsson R, Colona MR, Hoenig MP, Lundquist AL, Novak JE, Perazella MA, et al. Assessment of interobserver reliability of nephrologist examination of urine sediment. *JAMA Netw Open.* 2020;3:e2013959. <https://doi.org/10.1001/jama-networkopen.2020.13959>
12. Simundic AM, Nikolac N, Ivankovic V, Ferenec-Ruzic D, Magdic B, Kvaternik M. Comparison of visual vs. automated detection of lipemic, icteric and hemolyzed specimens: can we rely on a human eye? *Clin Chem Lab Med.* 2009;47:1361-5. <https://doi.org/10.1515/CCLM.2009.306>
13. Lee AY, Hudspeth AR, Adelstein S. The concordance of serial ANA tests in an Australian tertiary hospital pathology laboratory. *Pathology.* 2016;48:597-601. <https://doi.org/10.1016/j.pathol.2016.06.003>



## ***IV. Conclusiones***





De los estudios llevados a cabo en el marco de esta Tesis Doctoral se pueden concluir los siguientes aspectos generales:

- Las especificaciones de calidad de error total basadas en el estado del arte, que se han propuesto para las magnitudes incluidas en el cribado neonatal, son la única fuente disponible actualmente para que los laboratorios puedan realizar una evaluación del error de medida.
- Para algunas magnitudes de hematología, de coagulación y del cribado neonatal se ha observado que el error de medida cometido por los laboratorios depende de la concentración. Sería conveniente que los PEEC informasen de las especificaciones de error total para diferentes rangos de concentración. Por otro lado, este hallazgo demuestra la importancia de realizar una selección consciente de especificaciones de calidad considerando las limitaciones de cada uno de los modelos.
- La aplicación de un AMFE, utilizando datos objetivos del plan de control de calidad en combinación con el uso de la métrica sigma para la estandarización de los resultados, ha resultado útil para tener una visión general del rendimiento analítico de los procedimientos de medida. Sin embargo, la correcta interpretación de los resultados está condicionada de nuevo a haber realizado una adecuada selección de la especificación de calidad de error total. Es esencial, por tanto, verificar su idoneidad antes de iniciar estudios cuyos resultados dependan de la especificación establecida.
- Es conocido por todos que la aplicación estricta de las guías o las recomendaciones de las sociedades científicas puede resultar muy compleja o incluso impracticable. Esto se puso de manifiesto al comprobar la utilidad de las recomendaciones disponibles para la realización de verificaciones de la intercambiabilidad de resultados entre diferentes analizadores de forma periódica. Es por tanto indispensable disponer de personal en el laboratorio que conozca en profundidad las guías y que tenga la habilidad de llevar a la

práctica las mismas, encontrando el equilibrio entre la validez de la aproximación propuesta y los recursos necesarios para su implantación; sin olvidar la sobrecarga de trabajo que supone la realización de este tipo de estudios en laboratorios clínicos dedicados a la práctica asistencial.

- Utilizar guías o recomendaciones aprobadas por las sociedades científicas para realizar estudios dirigidos al aseguramiento de la calidad de los resultados analíticos proporciona una objetividad que aporta mucha seguridad ante una auditoría de acreditación. Sin embargo, tal y como se ha observado al aplicar las recomendaciones disponibles en la bibliografía para evaluar los resultados de un intercambio de muestras con otro laboratorio, es posible que el criterio utilizado no se adecue a las necesidades según la magnitud en cuestión y/o su uso previsto. A este respecto, es fundamental que sea el facultativo experto en el área de conocimiento correspondiente quien valore la utilidad de la metodología aplicada.
- La competencia del personal del laboratorio es clave para el aseguramiento de la calidad de los resultados de los pacientes. Este aspecto adquiere especial relevancia cuando en el proceso analítico se realizan actividades que requieren una interpretación por parte del observador, ya que en tal caso debe existir una buena concordancia entre las opiniones de los posibles observadores. Para conseguirlo es fundamental que cada actividad esté respaldada por un procedimiento armonizado, que los criterios de decisión estén consensuados y que se realice un seguimiento del grado de concordancia a lo largo del tiempo.



## ***V. Bibliografía***





- 1) Forsman RW. Why is the laboratory an afterthought for managed care organizations?. Clin Chem. 1996;42(5):813-6.
- 2) MSSSI EPV. Laboratorio Clínico Central estándares y recomendaciones de calidad y seguridad. Informes, estudios e investigación. Ministerio de sanidad, Servicios Sociales e Igualdad. 2013.
- 3) Lundberg GD. Acting on significant laboratory results. JAMA. 1981;245(17):1762-3.
- 4) Laposata M, Dighe A. "Pre-pre" and "post-post" analytical error: high-incidence patient safety hazards involving the clinical laboratory. Clin Chem Lab Med. 2007;45(6):712-9.
- 5) Plebani M. Towards a new paradigm in laboratory medicine: the five rights. Clin Chem Lab Med. 2016;54(12):1881-91.
- 6) Plebani M. Errors in clinical laboratories or errors in laboratory medicine?. Clin Chem Lab Med. 2006;44(6):750-9.
- 7) Plebani M. The detection and prevention of errors in laboratory medicine. Ann Clin Biochem. 2010;47:101-10.
- 8) Bonini P, Plebani M, Ceriotti F, Rubboli F. Errors in laboratory medicine. Clin Chem. 2002;48(5):691-8.
- 9) Plebani M, Carraro P. Mistakes in a stat laboratory: types and frequency. Clin Chem. 1997;43:1348-51.
- 10) Carraro P, Plebani M. Errors in a stat laboratory: types and frequencies 10 years later. Clin Chem. 2007;53:1338-42.
- 11) Lippi G, Plebani M. A Six-Sigma approach for comparing diagnostic errors in healthcare-where does laboratory medicine stand?. Ann Transl Med. 2018;6(10):180.
- 12) Plebani M. Quality in laboratory medicine: 50 years on. Clin Biochem. 2017;50(3):101-4.

- 13) Perich C, Ricós C, Marqués F, Minchinela J, Salas A, Martínez-Bru C, *et al.* Programas de garantía externa de la calidad SEQCML. Evolución de las prestaciones analíticas de los laboratorios clínicos a lo largo de 30 años y comparación con otros programas. *Adv Lab Med.* 2020;1(2):20190024.
- 14) Plebani M. Exploring the iceberg of errors in laboratory medicine. *Clin Chim Acta.* 2009;404(1):16-23.
- 15) Westgard JO. A Total Quality-Control Plan with Right-Sized Statistical Quality-Control. *Clin Lab Med.* 2017;37(1):137-50.
- 16) Jiang Y, Jiang H, Ding S, Liu Q. Application of failure mode and effects analysis in a clinical chemistry laboratory. *Clin Chim Acta.* 2015;448:80-5.
- 17) Sciacovelli L, Panteghini M, Lippi G, Sumarac Z, Cadamuro J, Galoro CAO, *et al.* Defining a roadmap for harmonizing quality indicators in Laboratory Medicine: a consensus statement on behalf of the IFCC Working Group "Laboratory Error and Patient Safety" and EFLM Task and Finish Group "Performance specifications for the extra-analytical phases". *Clin Chem Lab Med.* 2017;55(10):1478-88.
- 18) IEC 31010:2019. Risk management - risk assessment techniques. Geneva.
- 19) ISO 22367:2008. Medical laboratories - reduction of error through risk management and continual improvement. Geneva.
- 20) Clinical and Laboratory Standards Institute (CLSI). Laboratory quality control based on risk management; Approved Guideline. CLSI document EP23-A. Wayne, PA: CLSI, 2011.
- 21) Clinical and Laboratory Standards Institute (CLSI). Risk management techniques to identify and control laboratory error sources; Approved guideline, 2nd ed. CLSI document EP18-A2. Wayne, PA: CLSI, 2009.

- 22) Liu HC. FMEA for Proactive Healthcare Risk Analysis: A Systematic Literature Review. In: Improved FMEA Methods for Proactive Healthcare Risk Analysis. Singapore:Springer. 2019. p.15-45.
- 23) Westgard JO. Perspectives on quality control, risk management, and analytical quality management. Clin Lab Med. 2013;33(1):1-14.
- 24) De la Torre M, Campoy M. Fundamentos de Seguridad del Paciente. Análisis y estrategias en el laboratorio clínico. Editorial Médica Panamericana. 2012. p.143.
- 25) Martinez-Marquez D, Terhaer K, Scheinemann P, Mirnajafizadeh A, Carty CP, Stewart RA. Quality by Design for industry translation: Three-dimensional risk assessment failure mode, effects, and criticality analysis for additively manufactured patient-specific implants. Engineering Reports. 2020;2(1):e12113.
- 26) Dean Franklin B, Shebl NA, Barber N. Failure mode and effects analysis: too little for too much?. BMJ Qual Saf. 2012;21(7):607-11.
- 27) Castro-Castro MJ, Sánchez-Navarro L. Estimación de los límites de cambio (deltacheck) en el laboratorio clínico. Adv Lab Med. 2021;2(3):424-31.
- 28) Clinical and Laboratory Standards Institute (CLSI). Statistical Quality Control for Quantitative Measurement Procedures: Principles and Definitions. 4th ed. CLSI guideline C24. Wayne, PA: CLSI; 2016.
- 29) JCGM, BIPM. "JCGM 200: 2012 International Vocabulary of Metrology: Basic and General Concepts and Associated Terms (VIM)." The Joint Committee for Guides in Metrology and The Bureau International des Poids et Mesures: Paris, France. 2012.

- 30) ISO 17511:2003. In vitro diagnostics medical services - measurement of quantities in biological samples—metrological traceability of values assigned to calibrators and control materials. Geneva.
- 31) Westgard JO. Glossary of QC terms. Disponible en: <http://www.westgard.com/glossary.htm>. [último acceso 19 febrero 2023].
- 32) Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol.* 1950;20(11):1059-66.
- 33) Westgard JO, Bayat H, Westgard SA. Planning risk-based SQC schedules for bracketed operation of continuous production analyzers. *Clin Chem.* 2018;64(2):289-96.
- 34) Westgard JO. Internal quality control: planning and implementation strategies. *Ann Clin Biochem.* 2003;40(6):593-611.
- 35) 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.
- 36) Westgard S, Bayat H, Westgard JO. Analytical Sigma metrics: A review of Six Sigma implementation tools for medical laboratories. *Biochem Med (Zagreb).* 2018;28(2):020502.
- 37) Westgard JO. The Six Sigma Calculators. Disponible en: <https://www.westgard.com/six-sigma-calculators.htm>. [último acceso 19 febrero 2023].
- 38) Westgard JO, Armbruster D, Westgard SA. Risk, Error and Uncertainty: Laboratory Quality Management in the Age of Metrology. An Issue of the Clinics in Laboratory Medicine, Vol. 37-1. New York: Elsevier, 2017.
- 39) Westgard JO, Westgard SA. Establishing Evidence-Based Statistical Quality Control Practices. *Am J Clin Pathol.* 2019;151(4):364-70.

- 40) Prada E, Blazquez R, Gutiérrez-Bassini G, Morancho J, Jou JM, Ramón F, *et al.* Control interno de la calidad vs control externo de la calidad. *Rev Lab Clin.* 2016;9(2):54-9.
- 41) Bayat H, Westgard SA, Westgard JO. Multirule procedures vs moving average algorithms for IQC: an appropriate comparison reveals how best to combine their strengths. *Clin Biochem.* 2022;102:50-5.
- 42) Ricós C, Fernandez-Calle P, Perich C, Westgard JO. Internal quality control—past, present and future trends. *Adv Lab Med.* 2022; 3(3):243-52.
- 43) Vives-Corróns JL, Jou JM, Pastor C, Reverter JC, Jou C. Characteristics of the external quality assessment (EQA) scheme for haematology in Spain. *Qual Assur Health Care.* 1991;3(2):75-83.
- 44) Miller WG, Jones GR, Horowitz GL, Weykamp C. Proficiency testing/external quality assessment: current challenges and future directions. *Clin Chem.* 2011;57(12):1670-80.
- 45) JCTLM Reference measurement laboratories (WG2). Disponible en: <https://www.bipm.org/jctlm/>. [último acceso 22 abril 2023].
- 46) JCTLM Reference materials and reference procedures (WG1). Disponible en: <https://www.bipm.org/jctlm/>. [último acceso 22 abril 2023].
- 47) Jones GR, Jackson C. The Joint Committee for Traceability in Laboratory Medicine (JCTLM)—its history and operation. *Clin Chim Acta.* 2016;453:86-94.
- 48) Sciacovelli L, Zardo L, Secchiero S, Plebani M. Quality specifications in EQA schemes: from theory to practice. *Clin Chim Acta.* 2004;346(1):87-97.
- 49) Ricos C, Baadenhuijsen H, Libeer JC, Hyltoft Petersen P, Stockl D, Thienpont L, *et al.* External quality assesment: currently used criteria for evaluating performance in European countries, and criteria for future harmonization. *Eur J Clin Chem Clin Biochem.* 1996;34(2):159-65.

- 50) Sandberg S, Thue G. External quality assurance of test requesting and test interpretation. *Accredit Qual Assur.* 1999;4:414–5.
- 51) Vasikaran SD, Penberthy L, Gill J, Scott S, Sikaris KA. Review of a pilot quality-assessment program for interpretative comments. *Ann Clin Biochem.* 2002;39:250–60.
- 52) Sciacovelli L, Secchiero S, Padoan A, Plebani M. External quality assessment programs in the context of ISO 15189 accreditation. *Clin Chem Lab Med.* 2018; 56(10):1644-54.
- 53) Caballero A, Gómez-Rioja R, Ventura M, Llopis MA, Bauça JM, Gómez-Gómez C, *et al.* Evaluation of 18 quality indicators from the external quality assurance preanalytical programme of the Spanish Society of Laboratory Medicine (SEQCML). *Adv Lab Med.* 2022;3(2):175-87.
- 54) Badrick T, Gay S, McCaughey EJ, Georgiou A. External Quality Assessment beyond the analytical phase: an Australian perspective. *Biochem Med (Zagreb).* 2017;27(1):73-80.
- 55) Kallner A, McQueen M, Heuck C. The Stockholm Consensus Conference on quality specifications in laboratory medicine, 25-26 April 1999. *Scand J Clin Lab Invest.* 1999;59(7):475-6.
- 56) 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):833-5.
- 57) Badrick T. Biological variation: understanding why it is so important?. *Pract Lab Med.* 2021; 23:e00199.
- 58) Soto AB, Clar RC, Zaragoza JM, Bauzá FR, Aguilá CR, García AS. Consenso sobre especificaciones mínimas de la calidad analítica. *Rev Lab Clin.* 2008; 1(1):35-9.



- 59) Clar RC, Bassini GG, Turallas JM, Zaragoza JM, Bauzá FR, Aguilá CR, *et al.* Consenso sobre especificaciones mínimas de la calidad analítica para magnitudes hematológicas y de bioquímica especial. *Rev Lab Clin.* 2010;3(2), 87-93.
- 60) Salas A, Ricós C, Prada E, Ramón F, Morancho J, Jou JM, *et al.* State-of-the-art Approach to Goal Setting. *Clin Lab Med.* 2017;37(1):73-84.
- 61) Morancho J, Prada E, Gutierrez-Bassini G, Salas A, Blázquez R, Jou JM, *et al.* Actualización de las especificaciones de la calidad analítica 2014. Consenso de las Sociedades Científicas nacionales. *Rev Lab Clin.* 2014;7(1):3-8.
- 62) de Medio EP, Borrás AM, Quiles JA, Castellví JB, Bauzá FR, García AS, *et al.* Especificaciones mínimas de consenso de calidad analítica en España. Comparación con valores previos, preceptivos y de variabilidad biológica. *Rev Med Lab.* 2020;1(3): 93-107.
- 63) Ceriotti F, Fernandez-Calle P, Klee GG, Nordin G, Sandberg S, Streichert T, *et al.* Criteria for assigning laboratory measurands to models for analytical performance specifications defined in the 1st EFLM Strategic Conference. *Clin Chem Lab Med.* 2017;55(2):189-94.
- 64) Berge Kristensen GB, Meijer P. Interpretation of EQA results and EQA-based trouble shooting. *Biochem Med (Zagreb).* 2017; 27(1):49-62.
- 65) Roelofsen-de Beer R, Wielders J, Boursier G, Vodnik T, Vanstapel F, Huisman W, *et al.* Validation and verification of examination procedures in medical laboratories: opinion of the EFLM Working Group Accreditation and ISO/CEN standards (WG-A/ISO) on dealing with ISO 15189: 2012 demands for method verification and validation. *Clin Chem Lab Med.* 2020;58(3):361-7.
- 66) ISO 15189:2003. Laboratorios clínicos - Requisitos particulares relativos a la calidad y la competencia. Ginebra

- 67) ISO 15189:2022. Laboratorios clínicos - Requisitos particulares para la calidad y la competencia. Ginebra
- 68) ISO/IEC 17043: 2010. Evaluación de la conformidad. Requisitos generales para los ensayos de aptitud. Ginebra
- 69) German Medical Association. Revision of the "Guideline of the German Medical Association on Quality Assurance in Medical Laboratory Examinations—RiliBAEK". J Lab Med. 2015;39(1):26–69.
- 70) Jones GR, Sikaris K, Gill J. 'Allowable Limits of Performance' for External Quality Assurance Programs - an Approach to Application of the Stockholm Criteria by the RCPA Quality Assurance Programs. Clin Biochem Rev. 2012;33(4):133-9.
- 71) Horsti J, Uppa H, Vilpo JA. Poor agreement among prothrombin time international normalized ratio methods: comparison of seven commercial reagents. Clin Chem. 2005;51(3):553-60.
- 72) Soumali MR, Van Blerk M, Akharif A, Albarède S, Kessler D, Gutierrez G, *et al.* A new approach to define acceptance limits for hematology in external quality assessment schemes. Clin Chem Lab Med. 2017;55(12):1936-42.
- 73) Pitt JJ. Newborn screening. Clin Biochem Rev. 2010;31(2):57-68.
- 74) Marín JL, López RM, Argudo A, González de Aledo JM, Pajares S, Navarro A, *et al.* 50 años del Programa de Cribado Neonatal en Cataluña. Rev Esp Salud Pública. 2020; 94:e202012177.
- 75) Molina A, Guiñon L, Perez A, Segurana A, Bedini JL, Reverter JC, *et al.* State of the art vs biological variability: Comparison on hematology parameters using Spanish EQAS data. Int J Lab Hematol. 2018;40(3):284-91.

76) Castilla JA, Morancho-Zaragoza J, Aguilar J, Prats-Gimenez R, Gonzalvo MC, *et al.* Quality specifications for seminal parameters based on the state of the art. *Hum Reprod.* 2005;20(9):2573-8.

77) Jones GRD, Albarede S, Kessler D, MacKenzie F, Mammen J, Pedersen M, *et al.*; EFLM Task Finish Group – Analytical Performance Specifications for EQAS (TFG-APSEQA). Analytical performance specifications for external quality assessment - definitions and descriptions. *Clin Chem Lab Med.* 2017;55(7):949-55.

78) Plebani M. The journey toward quality and patient safety in laboratory medicine continues. *N Am J Med Sci.* 2014;6(5):229-30.

79) Weykamp C, Secchiero S, Plebani M, Thelen M, Cobbaert C, Thomas A, *et al.* Analytical performance of 17 general chemistry analytes across countries and across manufacturers in the INPUTS project of EQA organizers in Italy, the Netherlands, Portugal, United Kingdom and Spain. *Clin Chem Lab Med.* 2017;55(2):203-11.

80) Six Sigma Table. Disponible en: <https://www.westgard.com/sixsigtable.htm>. [último acceso 18 febrero 2023].

81) Thompson S, Chesher D. Lot-to-Lot Variation. *Clin Biochem Rev.* 2018;39(2):51-60.

82) Minchinela J, Ricós C, Perich C, Fernández-Calle P, Alvarez V, Domenech M, *et al.* Biological variation database and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update. Disponible en: <https://www.westgard.com/biodatabase-2014-update.htm>. [último acceso 1 febrero 2023].

83) CLIA Requirements for Analytical Quality. Disponible en: <https://www.westgard.com/clia.htm>. [último acceso 1 febrero 2023].

- 84) Xia Y, Xue H, Yan C, Li B, Zhang S, Li M, *et al.* Risk analysis and assessment based on Sigma metrics and intended use. *Biochem Med (Zagreb)*. 2018;28(2):020707.
- 85) Ceriotti F, Cappelletti P, Caputo M, Di Serio F, Messeri G, Ottomano C, *et al.* A risk-analysis approach to the evaluation of analytical quality. *Clin Chem Lab Med*. 2011;50(1):67-71.
- 86) Westgard SA, Salcedo G. A Six Sigma risk analysis example: defining QC frequency with FMEA. July 2012. Disponible en: <https://www.westgard.com/endoclab-risk.htm>. [último acceso 10 mayo 2023].
- 87) Capunzo M, Cavallo P, Boccia G, Brunetti L, Pizzuti S. A FMEA clinical laboratory case study: how to make problems and improvements measurable. *Clin Leadersh Manag Rev*. 2004;18(1):37-41.
- 88) Clinical and Laboratory Standards Institute (CLSI). Measurement procedure comparison and bias estimation using patient samples. Approved guideline - CLSI document EP09-A3. Wayne, PA: CLSI, 2013.
- 89) de Medio EP, Sánchez RB, Alsina CP, Bassini GG, Tenor DP, Ríos AIA, *et al.* Verificación de la intercambiabilidad de resultados entre equipos duplicados. *Rev Lab Clin*. 2014;7(1):17-24.
- 90) Braga F, Panteghini M. Commutability of reference and control materials: an essential factor for assuring the quality of measurements in Laboratory Medicine. *Clin Chem Lab Med*. 2019;57(7):967-73.
- 91) Clinical and Laboratory Standards Institute (CLSI). Verification of Comparability of Patient Results Within One Health Care System. CLSI document EP31-A-IR. Wayne, PA: CLSI, 2012.
- 92) Harris EK. Statistical principles underlying analytic goal-setting in clinical chemistry. *Am J Clin Pathol*. 1979;72(2 Suppl):374-82.

- 93) Fraser CG, Petersen PH. The importance of imprecision. *Ann Clin Biochem.* 1991;28(Pt 3):207-11.
- 94) Aarsand AK, Fernandez-Calle P, Webster C, Coskun A, Gonzales-Lao E, Diaz-Garzon J, *et al.* The EFLM Biological Variation Database. <https://biologicalvariation.eu/>. [último acceso 10 mayo 2023].
- 95) Lee EJ, Lee E, Kim M, Kim HS, Lee YK, Kang HJ. Verification of the comparability of laboratory results from two instruments within one health care system according to Clinical and Laboratory Standard Institute EP31-A-IR. *J Lab Med Qual Assur.* 2016;38(3):129-36.
- 96) Sciacovelli L, Secchiero S, Padoan A, Plebani M. External quality assessment programs in the context of ISO 15189 accreditation. *Clin Chem Lab Med.* 2018;56(10):1644-54.
- 97) Oguma T, Tomatsu S, Montano AM, Okazaki O. Analytical method for the determination of disaccharides derived from keratan, heparan, and dermatan sulfates in human serum and plasma by high-performance liquid chromatography/ turbo ionspray ionization tandem mass spectrometry. *Anal Biochem.* 2007;368(1):79–86.
- 98) Clinical and Laboratory Standards Institute (CLSI). *Using Proficiency Testing and Alternative Assessment to Improve Medical Laboratory Quality.* 3rd ed. CLSI Guideline QMS24, Wayne, PA: CLSI, 2016.
- 99) Szewczak E, Bondarzewski A. Is the assessment of interlaboratory comparison results for a small number of tests and limited number of participants reliable and rational?. *Accred Qual Assur.* 2016;21(2):91–100.
- 100) Grupo de trabajo del Sistema de Información del Programa de Cribado Neonatal del SNS. Programa de Cribado Neonatal del Sistema Nacional de Salud. Informe de Evaluación. Año 2019. Ministerio de Sanidad, 2021.

101) George RS, Moat SJ. Effect of dried blood spot quality on newborn screening analyte concentrations and recommendations for minimum acceptance criteria for sample analysis. *Clin Chem*. 2016;62:466-75.

102) Renom G, Mereau C, Maboudou P, Perini JM. Potential of the Sebia Capillarys® neonat fast automated system for neonatal screening of sickle cell disease. *Clin Chem Lab Med*. 2009;47(11):1423-32.

103) Hayes AF, Krippendorff K. Answering the call for a standard reliability measure for coding data. *Commun Methods Meas*. 2007;1:77-89.

104) Krippendorff K. *Content Analysis: An Introduction to its Methodology*. 2nd ed. Thousand Oaks, CA: Sage; 2004. p.241.

105) Tamary H, Dgany O. Alpha-Thalassemia. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Gripp KW, Mirzaa GM, Amemiya A, eds. *GeneReviews*. 2005 Nov 1 [updated 2020 Oct 1]. Disponible en: <https://pubmed.ncbi.nlm.nih.gov/20301608/>. [último acceso 10 mayo 2023].