



Diogo José da Silva Jerônimo

**Proposta de novos critérios para avaliação de
desempenho de laboratórios clínicos por meio de ensaios
de proficiência**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do título de Mestre pelo Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação) da PUC-Rio.

Orientador: Prof^a. Elisabeth Costa Monteiro

Rio de Janeiro
Maio de 2017



Diogo José da Silva Jerônimo

**Proposta de novos critérios para avaliação de
desempenho de laboratórios clínicos por meio de ensaios
de proficiência**

Dissertação apresentada como requisito parcial
para obtenção do grau de Mestre pelo Programa de
Pós-Graduação em Metrologia da PUC-Rio.
Aprovada pela Comissão Examinadora abaixo
assinada.

Prof^a. Elisabeth Costa Monteiro

Orientador

Programa de Pós-Graduação em Metrologia - PUC-Rio

Prof. Reinaldo Castro Souza

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Élcio Cruz de Oliveira

Programa de Pós-Graduação em Metrologia - PUC-Rio

Prof. Márcio da Silveira Carvalho

Coordenador Setorial de Pós-Graduação do
Centro Técnico Científico - PUC-Rio

Rio de Janeiro, 05 de maio de 2017

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Diogo José da Silva Jerônimo

Bacharel em Ciências Estatísticas pela ENCE - Escola Nacional de Ciências Estatísticas, Rio de Janeiro, Brasil, 2005.

Ficha Catalográfica

Jerônimo, Diogo José da Silva

Proposta de novos critérios para avaliação de desempenho de laboratórios clínicos por meio de ensaios de proficiência / Diogo José da Silva Jerônimo ; orientador: Elisabeth Costa Monteiro. – 2017.

101 f. : il. color. ; 30 cm

Dissertação (mestrado)–Pontifícia Universidade Católica do Rio de Janeiro, Centro Técnico Científico, Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), 2017.

Inclui bibliografia

1. Metrologia – Teses. 2. Ensaio de proficiência. 3. Laboratório clínico. 4. Critérios. 5. Material de referência certificado. 6. Rastreabilidade metrológica. I. Monteiro, Elisabeth Costa. II. Pontifícia Universidade Católica do Rio de Janeiro. Centro Técnico Científico. Programa de Pós-Graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação). III. Título.

CDD: 389.1

*Aos dois pais que tive nessa terrena vida: José Cândido Jerônimo e José Luiz
Virgílio (in memoriam).*

Agradecimentos

A Deus Pai Criador, a Ele Toda Honra e Glória, e a Maria Santíssima, Mãe de Deus e da Santa e Madre Igreja Católica;

A minha família, em especial minha mãe Elisabete da Silva Jerônimo, que sempre foi meu anjo, meu primeiro amor, minha inspiração e meu humano modelo de moral e conduta para toda minha vida, além de ser a âncora viva que ainda nos mantém unidos;

A Diretoria da Control Lab - Controle de Qualidade para Laboratórios (em especial, Dr. Vinícius Almeida Biasoli e Dr. Márcio Mendes Biasoli), que sempre acreditaram em mim ao longo de mais de dez anos de trabalho na empresa, e que me ajudaram a chegar com muito custo e paciência ao fim dessa jornada, sendo sempre modelos para minha vida pessoal, profissional e acadêmica;

A equipe técnica da Control Lab, em especial meu gestor Rafael Monsorez Lopes, cujas orientações e colaboraram demais com minha pesquisa, e cuja paciência com a época de estudos me trouxe segurança na conclusão desse caminho;

A todos os coordenadores e professores do Programa de Pós-MQI da PUC-Rio, pelo apoio, confiança e dedicação presentes no momento de concepção dessa dissertação. Em especial, minha professora, Dr^a Elisabeth Costa Monteiro, que foi misto de competente orientadora, grande amiga e dedicada conselheira na condução desse trabalho, sempre acreditando em minha força de vontade e determinação para encerrar essa dissertação;

Aos meus amigos e colegas de curso que aqui fiz, e que não poderei a todos citar pela extensão. Foram fontes de conhecimento e ouvidos de desabafo nas horas complicadas;

A todos os meus amigos mais chegados, que diretamente ou indiretamente colaboraram com a realização desse trabalho.

Resumo

Jerônimo, Diogo José da Silva; Costa Monteiro, Elisabeth. **Proposta de novos critérios para avaliação de desempenho de laboratórios clínicos por meio de ensaios de proficiência**. Rio de Janeiro, 2017. 101p. Dissertação de Mestrado – Programa de Pós-graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), Pontifícia Universidade Católica do Rio de Janeiro.

A participação de laboratórios clínicos em programas de ensaio de proficiência é uma importante ferramenta para gestão da qualidade. A maior parte das avaliações de desempenho de laboratórios clínicos por ensaio de proficiência se realiza com base no critério Z-Score, cujo valor designado é o consenso de resultados obtidos pelos participantes da rodada de proficiência, que é fortemente dependente da qualidade dos sistemas analíticos dos participantes. Utilizando uma base de dados fornecida por um provedor de ensaios de proficiência acreditado pelo INMETRO, este trabalho compara o desempenho de critérios para a avaliação da proficiência de laboratórios clínicos para medição de glicose. Consideraram-se não só critérios comumente utilizados, mas também novos critérios propostos que independem do consenso, baseados apenas em parâmetros associados ao valor designado pelo material de referência. As discrepâncias de classificação de desempenho dos laboratórios em função dos critérios de avaliação empregados indicaram inadequações da classificação baseada em critérios com parâmetros afetados pelo consenso entre participantes. O estudo também permitiu caracterizar o desempenho dos sistemas analíticos empregados pelos participantes, evidenciando-se uma discrepância significativa entre seus resultados. Os resultados do presente trabalho possibilitam analisar o impacto dos critérios utilizados no ensaio de proficiência para a adequada avaliação do desempenho dos laboratórios, contribuindo, assim, para a garantia da confiabilidade dos diagnósticos clínicos e condutas terapêuticas.

Palavras-chave

Metrologia; Ensaio de proficiência; Laboratório clínico; Critérios; Material de referência certificado; Rastreabilidade metrológica.

Abstract

Jerônimo, Diogo José da Silva; Costa Monteiro, Elisabeth (Advisor). **Proposed new criteria for performance evaluation of clinical laboratories by means of proficiency tests.** Rio de Janeiro, 2017. 101p. MSc. Dissertation - Programa de Pós-graduação em Metrologia (Área de concentração: Metrologia para Qualidade e Inovação), Pontifícia Universidade Católica do Rio de Janeiro Pontifícia Universidade Católica do Rio de Janeiro.

The participation of clinical laboratories in proficiency testing programs is an important tool for quality management. Most evaluations of clinical laboratory performance by proficiency testing are carried out based on the Z-Score criterion; whose assigned value is the consensus of results obtained by the participants of the round, which is heavily dependent on the quality of analytical systems of participants. Using a database provided by a proficiency testing provider accredited by INMETRO, this paper compares the performance of criteria for the assessment of proficiency of clinical laboratories for measuring glucose. The analysis considered not only commonly used criteria, but also new proposed criteria that are independent of the consensus effect, based only on parameters associated with the assigned value defined by the reference material. The observed discrepancies in the performance rating of laboratories, according to the evaluation criteria employed, allowed the evidencing of inadequacies in the classifications based on criteria using parameters affected by the consensus among participants. The study also characterized the performance of analytical systems employed by the laboratories, evidencing a significant discrepancy between the results its. The present work allowed a better understanding of the impact of criteria applied in proficiency testing programs on the adequacy of evaluation of laboratories performance, thus contributing to ensuring the reliability of clinical diagnostics and therapeutical procedures.

Keywords

Metrology; Proficiency Testing; Clinical Laboratory; Criteria; Certified reference material; Metrological traceability.

Sumário

1 Introdução	17
1.1. Organização da Dissertação	19
2 Confiabilidade metrológica em ensaios de proficiência	20
2.1. Contexto Internacional	20
2.1.1. BIPM (Bureau International des Poids et Mesures)	20
2.1.2. OIML (Organisation Internationale de Métrologie Légale)	23
2.1.3. IFCC (International Federation of Clinical Chemistry and Laboratory Medicine)	24
2.1.4. IUPAC (International Union of Pure and Applied Chemistry)	24
2.1.5. WHO (World Health Organization)	24
2.1.6. ILAC (International Laboratory Accreditation Cooperation)	25
2.1.7. EURACHEM	26
2.1.8. ISO (International Organization for Standardization)	26
2.2. Contexto Nacional	28
2.2.1. INMETRO (Instituto Nacional de Metrologia, Qualidade e Tecnologia)	28
2.2.2. ANVISA (Agência Nacional de Vigilância Sanitária)	31
2.2.3. ABNT (Associação Brasileira de Normas Técnicas)	32
3 Ensaio de proficiência para laboratórios clínicos	33
3.1. Etapas para execução do ensaio de proficiência	34
3.2. Avaliação do erro sistemático por meio de EP	36
3.3. Acreditação laboratorial e ensaio de proficiência	37
3.4. Material de Referência e Material de Referência Certificado	39
3.4.1. Homogeneidade e estabilidade do MR	44
3.4.1.1. Homogeneidade	44
3.4.1.2. Estabilidade	45
3.5. Parâmetros que regem o EP	45
3.5.1. Valor designado	46
3.5.2. Desvio padrão para a proficiência	47
3.5.3. Limites fixos empregados em ensaios de proficiência para laboratórios clínicos	49

3.6. Critérios recomendados para avaliação do desempenho em ensaio de proficiência	50
3.7. Análise Clínica da Concentração de glicose segundo métodos analíticos	52
4 Materiais e métodos	54
4.1. Base de dados utilizada	54
4.2. Critérios empregados para determinação do desempenho analítico dos laboratórios	55
4.2.1. Apresentação matemática dos critérios sob teste	55
4.2.2. Regra para determinação do desempenho para os critérios	58
4.2.3. Apresentação matemática dos critérios em forma de faixas de aceitação	59
4.3. Métodos para obtenção dos valores designados	60
4.4. Análise estatística dos parâmetros dos critérios de avaliação de desempenho	61
4.5. Análise gráfica do desempenho dos critérios de avaliação de desempenho dos laboratórios	62
5 Resultados	64
5.1. Análise estatística dos parâmetros	64
5.2. Análise gráfica dos escores obtidos por diferentes critérios para avaliação de desempenho do laboratório	66
5.3. Análise gráfica de desempenho do laboratório segundo as faixas de concentração associadas a cada critério estatístico utilizado para determinação dos escores	69
5.4. Análise gráfica e estatística de desempenho dos laboratórios segundo o sistema analítico utilizado	72
6 Discussão	83
7 Conclusões e Trabalhos Futuros	88
7.1. Conclusões	88
7.2. Trabalhos Futuros	90
8 Referências Bibliográficas	91

Apêndice A - Algoritmo A (ISO 13528, 2015)	96
Apêndice B – Testes estatísticos empregados	98

Lista de Figuras

Figura 1 – Organograma ISO acerca setores e documentos relevantes para EP, MRC e laboratórios clínicos.	27
Figura 2 – Ciclo de realização de um Ensaio de Proficiência (adaptado de Oliveira & Mendes, 2011).	34
Figura 3 – Representação esquemática das componentes aleatória e sistemática do erro de medição (Oliveira & Mendes, 2011).	36
Figura 4 – Reação de Oxidase com Peroxidase.	53
Figura 5 – Reação de Hexoquinase com Desidrogenase.	53
Figura 6 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 1, segundo os diferentes critérios estudados.	67
Figura 7 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 2, segundo os diferentes critérios estudados.	67
Figura 8 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 3, segundo os diferentes critérios estudados.	68
Figura 9 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 1, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.	70
Figura 10 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 2, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.	70
Figura 11 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 3, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.	71
Figura 12– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, segundo os métodos analíticos.	75
Figura 13 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, segundo os métodos analíticos.	75

Figura 14 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, segundo os métodos analíticos.	76
Figura 15 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, segundo equipamentos que empregam o método analítico Oxidase/Peroxidase.	78
Figura 16– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, entre equipamentos que empregam o método analítico Oxidase/Peroxidase.	79
Figura 17– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, entre equipamentos que empregam o método analítico Oxidase/Peroxidase.	79
Figura 18– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.	80
Figura 19– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.	81
Figura 20– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.	81

Lista de Tabelas

Tabela 1 – Parâmetros estimados para o cálculo dos escores dos participantes, segundo os critérios propostos para avaliação de desempenho em EP, para medição de glicose.	65
Tabela 2 – Valores designados e respectivas faixas de avaliação, geradas pelos critérios Z-Score e $Z'_{LF}\text{-Score}_{(Ref)}$, ao longo do período estudado (Jan/15 a Out/15). Em vermelho itálico sublinhado, os resultados correspondentes à rodada Abr/15.	73
Tabela 3– Estatística de teste e respectivos p-valores obtidos pelos testes de Kruskal-Wallis (KW) e Levene, para comparação entre os métodos analíticos (Oxidase/Peroxidase e Hexoquinase / Desidrogenase), nos itens da rodada Abr/15.	74
Tabela 4 – Estatística de teste e respectivos p-valores obtidos pelos testes de Kruskal-Wallis (KW) e Levene, para comparação entre equipamentos, segundo os métodos analíticos, nos itens da rodada Abr/15.	77
Tabela 5 – Correspondência diagnóstica, segundo o Consenso Brasileiro sobre Diabetes, para os valores de glicose plasmática (em mg/dL) com relação às faixas de aceitação dos resultados reportados pelos laboratórios participantes (X_{ref} no valor de 143 mg/dL), considerando os critérios Z-Score e $Z'_{LF}\text{-Score}_{(Ref)}$.	86

Lista de Quadros

Quadro 1 – Documentos orientativos para acreditação de laboratórios, com ênfase no EP, MRC ou laboratórios.	30
Quadro 2 – Documentos normativos para acreditação de laboratórios, com ênfase no EP, MRC ou laboratórios.	30
Quadro 3 – Comitês Técnicos de interesse da ABNT.	32
Quadro 4 – Informação contida no Apêndice C do KCDB sobre a Capacidade de Medição associada a materiais de referência para glicose (BIPM, 2017).	42
Quadro 5 – Métodos e procedimentos de medição de referência para glicose em soro humano, reconhecidos e rastreáveis, listados no JCTLM-DB (BIPM, 2017).	43
Quadro 6 – Resumo dos tipos, vantagens e desvantagens das propostas para definição do valor designado para o EP (ABNT NBR ISO/IEC 17043, 2011; Thompson et al., 2006).	47

Resumo das siglas

Sigla	Instituições
ANVISA	Agência Nacional de Vigilância Sanitária
BIPM	<i>Bureau International des Poids et Mesures</i>
CCQM	<i>Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology</i>
CLIA	<i>Clinical Laboratory Improvement Amendments</i>
IFCC	<i>The International Federation of Clinical Chemistry and Laboratory Medicine</i>
ILAC	<i>International Laboratory Accreditation Cooperation</i>
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
ISO	<i>International Organization for Standardization</i>
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
JCTLM	<i>Joint Committee for Traceability in Laboratory Medicine</i>
NIST	<i>National Institute of Standards and Technology</i>
OIML	<i>Organisation Internationale de Métrologie Légale</i>
WHO	<i>World Health Organization</i>

Sigla	Termos
EP	Ensaio de proficiência
GUM	<i>Guide to the expression of uncertainty in measurement</i>
IVD	<i>In Vitro Diagnostic (Medical Devices)</i>
MR	Material de referência
MRC	Material de referência certificado
SI	Sistema Internacional de Unidades
VIM	Vocabulário Internacional de Metrologia
TC	<i>Technical Committee</i>
LF	Limite fixo para avaliação em ensaios de proficiência
Bias	erro sistemático

*“Mede o que for mensurável, e torna mensurável aquilo que não o é” Galileu
Galilei.*

1

Introdução

O conceito de qualidade evoluiu frente às exigências dos clientes desde o século passado (Chaves, 2010). É ressaltado o papel cada vez maior do laboratório clínico na tomada de decisões médicas, o que conduz a uma avaliação cuidadosa da confiabilidade dos resultados laboratoriais (Plebani et al., 2017). O laboratório clínico deve atender aos usuários, de modo que as informações produzidas reflitam a real situação clínica dos pacientes (Santos & Zanusso Júnior, 2015). Dessa forma, a exatidão e comparabilidade dos resultados das biomedicações realizadas pelos laboratórios clínicos são essenciais para a garantia da confiabilidade dos diagnósticos e tratamentos clínicos (Monteiro, 2007; Delatour et al., 2012; Monteiro & Leon, 2015). É importante destacar que os laboratórios usam uma grande variedade de técnicas para quantificar um grande número de parâmetros, de modo que os resultados podem não ser comparáveis entre laboratórios (Delatour et al., 2012).

Há necessidade de melhorias contínuas, que dependem de um controle eficaz de processos, na identificação e eliminação de falhas recorrentes no laboratório clínico. Esse processo é chamado garantia da qualidade, e exige conhecimento das etapas de realização do exame, conhecidas como fases pré-analítica, analítica e pós-analítica (Chaves, 2010). Ao longo do século XX, foram desenvolvidas ferramentas para avaliar a qualidade do laboratório clínico por meio de desempenho, de modo a atender às especificações de qualidade (Plebani et al., 2017; Oliveira & Mendes, 2011).

Os laboratórios clínicos podem medir, monitorar e melhorar seus desempenhos por meio de ferramentas como o controle interno, os ensaios de proficiência e de indicadores da qualidade dos processos do laboratório (Plebani et al., 2017). Em ensaios de proficiência, a qualidade está relacionada à adequação do resultado frente ao verdadeiro valor do mensurado (De Bièvre, 2016; Oliveira & Mendes, 2011). Tal valor não é de fato conhecido, mas pode ser definido pelo melhor método disponível, ou por análise repetida em diferentes laboratórios, constituindo um valor designado, cuja estimativa é a melhor do valor verdadeiro (Oliveira & Mendes, 2011).

O reconhecimento do potencial de contribuição dos ensaios de proficiência os tornaram uma prática regular. Estudos acadêmicos relatam a associação entre a participação em ensaios de proficiência e a melhoria da qualidade laboratorial (Ceriotti, 2014). A realização do EP é associada a diminuição de erros nos laboratórios médicos, e a melhoria da qualidade analítica e na determinação do desempenho, em vista de melhor diagnosticar o paciente (Bonini et al., 2002).

A ABNT NBR ISO/IEC 17043:2011 (Avaliação de conformidade – Requisitos gerais para ensaios de proficiência) apresenta os parâmetros para a avaliação do desempenho dos laboratórios, dentre eles a determinação do valor designado a ser empregado no ensaio de proficiência. O consenso entre participantes é o método mais empregado para definição desse valor designado (Thompson et al., 2006). Esse método apresenta desvantagens: o valor designado é dependente dos resultados dos participantes, o que pode gerar tendência e classificação errônea de desempenho dos participantes; além disso, a incerteza do valor designado aumenta para um número menor de laboratórios participantes do ensaio de proficiência.

Em normas, regulamentos e diretrizes, não é destacada a importância do emprego de critérios independentes (De Bièvre, 2016), cujos parâmetros podem ser oriundos dos valores de propriedade de um material de referência certificado ou de um método de referência, permitindo assim a rastreabilidade metrológica à unidade do SI e a harmonização de resultados (De Bièvre, 2016). Assim, a possibilidade de aplicar valores designados e medidas de dispersão por métodos independentes do consenso entre laboratórios participantes do EP é essencial para uma adequada avaliação da qualidade analítica do laboratório (De Bièvre, 2016) e, conseqüentemente, para uma apropriada utilização do resultado do EP na melhoria do sistema de medição utilizado e na confiabilidade dos resultados clínicos fornecidos.

O presente trabalho tem por objetivo realizar uma análise comparativa de desempenho na avaliação de laboratórios clínicos por meio de ensaio de proficiência, utilizando tanto critérios amplamente utilizados, recomendados por documentos normativos e regulatórios, quanto novos critérios propostos no âmbito desta pesquisa. Esses últimos buscam evitar as distorções de avaliação que podem ser ocasionadas pelo efeito de parâmetros do protocolo de avaliação que sejam baseados nos valores associados ao consenso entre participantes. Os dados para a realização desse estudo correspondem aos resultados reportados por participantes de um programa de ensaio de proficiência para dosagem de

glicose em soro liofilizado, realizado por um provedor de ensaios de proficiência acreditado pelo INMETRO.

1.1.

Organização da Dissertação

A seguir, está descrito de forma sucinta, o conteúdo dos capítulos do presente documento:

Capítulo 1: é realizada a introdução ao trabalho sendo apresentados o contexto inicial, associado a garantia da qualidade em laboratórios clínicos, as motivações para a realização do estudo, e a estrutura e conteúdo da dissertação.

Capítulo 2: detalha os organismos e documentos normativos e regulatórios que dão suporte à confiabilidade metrológica da ferramenta ensaio de proficiência.

Capítulo 3: descreve o processo, os requisitos gerais, o contexto metrológico, a correlação com o processo de acreditação de laboratórios de ensaio e calibração, os parâmetros para procedimento da avaliação de desempenho, e a apresentação do processo de medição para o analito escolhido para a análise por meio de ensaio de proficiência (glicose em soro humano liofilizado).

Capítulo 4: descreve a apresentação do conjunto de dados que será utilizado e dos critérios de avaliação de desempenho sob teste, a metodologia estatística adotada e os métodos gráficos para apresentação dos resultados obtidos.

Capítulo 5: apresenta os resultados do estudo, onde são analisados os critérios para avaliação de desempenho por meio de ensaio de proficiência.

Capítulo 6: é apresentada uma breve discussão dos resultados obtidos.

Capítulo 7: detalham-se as conclusões do trabalho, e são propostos trabalhos futuros.

2

Confiabilidade metrológica em ensaios de proficiência

Para que seja obtida confiabilidade metrológica nos processos de medição é necessário todo um conjunto de processos organizacionais, conduzido por organismos que fornecem o suporte necessário para o processo envolvido. Nesse capítulo, é apresentada a estrutura de organismos, no contexto internacional e nacional, que podem dar suporte para a confiabilidade dos resultados de avaliação de desempenho de laboratórios de análises clínicas e biológicas por meio da realização do ensaio de proficiência (EP).

2.1.

Contexto Internacional

2.1.1.

BIPM (Bureau International des Poids et Mesures)

O *Bureau International des Poids et Mesures* é uma das organizações estabelecidas para manutenção do Sistema Internacional de Unidades (SI) sob os termos da Convenção do Metro (*Convention du Mètre*). O BIPM zela pela uniformidade e rastreabilidade global dos resultados de medição, em conformidade com o SI.

A organização científica opera por meio de comitês consultivos, gerenciados pelo CIPM (*Comité international des poids et mesures*). Entre as tarefas desses comitês, destacam-se a elaboração de recomendações para debate pelo CIPM, o planejamento e execução de comparação de padrões de medição, e consultorias científicas.

Uma das formas de atuação do CIPM, no que se refere à comparação dos padrões de medição, é a aplicação dos acordos de reconhecimento mútuo (CIPM MRA). Através deles, os institutos nacionais demonstram a equivalência internacional de seus padrões de medição e calibração com os padrões definidos pelo BIPM. Dessa forma, os resultados de calibração e medição certificados (CMCs) dos institutos participantes são reconhecidos internacionalmente, sendo aprovados e disponibilizados ao público, a partir do KCDB (banco de dados do

CIPM MRA). Esse banco de dados disponibiliza, entre outros, os institutos de metrologia que produzem MRC com rastreabilidade ao SI.

Um dos últimos comitês a ser criado, o CCQM (*Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology*) é o comitê consultivo para o desenvolvimento, melhoria e documentação da equivalência dos padrões nacionais (que inclui materiais de referência certificados e métodos de referência) para medições químicas e biológicas. Suas responsabilidades principais (BIPM, 2017):

1ª Estabelecer comparabilidade global de medições através da promoção da rastreabilidade ao SI, e quando essa rastreabilidade não é viável, a outras referências acordadas internacionalmente;

2ª Contribuir para a implementação e manutenção dos acordos de reconhecimento mútuo (CIPM MRA) com relação às medições químicas e biológicas;

3ª Agir como fórum para o intercâmbio de informações acerca de programas de pesquisa, serviços de medição e outras atividades técnicas.

Dentro do CCQM, existem diversos grupos de trabalho, cujas atividades envolvem realização de EP e desenvolvimento de MRC:

- CCQM *Working Group on Cell Analysis* (CAWG) – Responsável pelo parecer consultivo e planejamento (CCQM *strategy document*, acerca de comparações interlaboratoriais);
- CCQM *Working Group on Nucleic Acid Analysis* (NAWG) – Supervisiona as revisões de procedimentos para medição e calibração (inclusive informações para materiais de referência certificados); responsável pelo desenvolvimento, melhoria e documentação da equivalência dos padrões nacionais (materiais de referência certificados e métodos de referência) para medições químicas e biológicas;
- CCQM *Strategic Planning Working Group* (SPWG) – Supervisiona comparações, projetos-piloto e análise crítica dos elementos de medição;

desenvolvimento conjunto de metodologias e investigação para rastreabilidade metrológica;

- CCQM WG *on Key Comparisons and CMC Quality* (KCWG) - Análise crítica de padrões de medição e calibração, e estabelecimento de estudos interlaboratoriais.

Outra forma de atuação do BIPM é pelos seus *Joint Committes*, que são atuações conjuntas do BIPM com outras organizações internacionais, dentro de uma área específica. Atualmente, são quatro, sendo três mais diretamente relacionados às realizações de EP:

- JCGM (*Joint Committee for Guides in Metrology*) – responsável pela manutenção e promoção da GUM (Guide to the Expression of Uncertainty in Measurement) e do VIM (International Vocabulary of Metrology). São membros desse comitê: BIPM, OIML, IUPAC, IUPAP, IFCC, ILAC, ISO e IEC;
- JCTLM (*Joint Committee for Traceability in Laboratory Medicine*) – é uma cooperação para estabelecimento da rastreabilidade metrológica em medicina laboratorial, fornecendo uma plataforma para promoção e orientação sobre a equivalência reconhecida sob padrões de medição adequados. De forma similar ao KCDB, o JCTLM dissemina os organismos produtores de MRC da mais alta ordem, os métodos de medição de referência e os serviços de medição de referência. Essa ação é uma resposta a diretiva europeia IVD 98/79/EC, que determina que os sistemas analíticos forneçam rastreabilidade metrológica a unidade do mensurando. São membros o BIPM, o IFCC e o ILAC.
- JCRB (*Joint Committee of the Regional Metrology Organizations and the BIPM*) – O JCRB coordena as atividades entre as RMOs para o estabelecimento da confiança no reconhecimento mútuo dos Certificados de Calibração e Medição (CMC), em conformidade com o Acordo de Reconhecimento Mutuo (MRA) do CIPM. Os dados do CMC são formalmente aprovados pelo JCRB, antes de sua publicação no KCDB (*key comparison database*), a base de dados que compila as informações

dos estados membros do CIPM MRA acerca dos serviços de medição e calibração disponíveis, com incertezas declaradas.

2.1.2.

OIML (Organisation Internationale de Métrologie Légale)

Organização responsável pela metrologia na legislação de cada país, com o objetivo de proteger fornecedores e consumidores, e mesmo de gerenciar sanções pela presença de irregularidades nas relações comerciais. Sua atuação se dá entre as exigências legais associadas às unidades de medida, métodos e instrumentos de medição, com vista à proteção do consumidor. A OIML tem como objetivo harmonizar regulamentações e controles metrológicos dos seus Estados Membros, e fornecer sistemas de reconhecimento mútuo que reduzem as barreiras comerciais e os custos no mercado. Ela ainda orienta sobre os requisitos de certificação e calibração de produtos, em especial quando há impactos legais em áreas, como o comércio e a saúde. As principais categorias de publicações da OIML:

- Recomendações Internacionais (OIML R) – são os modelos de regulamentação que fixam as características metrológicas dos instrumentos de medição e dos métodos e equipamentos de medição padrão;
- Documentos Internacionais (OIML D) – documentos de natureza informativa, que visam harmonizar a atividade da Metrologia Legal;
- Guias Internacionais (OIML G) – documentos de natureza informativa, que visam fornecer regras gerais para a aplicação de algumas exigências em Metrologia Legal;
- Publicações de Base Internacionais (OIML B) – definem regras de funcionamento de várias estruturas e sistemas da OIML.

Os mais relevantes documentos para a aplicação dentro do EP são: o OIML D30, que é a Guia de aplicação da ISO/IEC 17025 (que inclui os EP), para avaliação de laboratórios envolvidos em metrologia legal, desenvolvido pelo TC3; e o OIML R135, que é a recomendação sobre espectrofotômetros para laboratórios médicos, desenvolvido pelo TC18.

2.1.3.

IFCC (International Federation of Clinical Chemistry and Laboratory Medicine)

Organização voltada às análises clínicas e à medicina laboratorial. Entre suas atribuições, está o estabelecimento de padrões globais, o incentivo aos esforços científicos e educativos, e a colaboração entre especialistas em vista da obtenção de melhores práticas e soluções para o laboratório clínico. A IFCC é membro do JCGM, JCTLM e do CCQM.

2.1.4.

IUPAC (International Union of Pure and Applied Chemistry)

Organização científica internacional, não governamental, sem fins lucrativos, que se propõe a disseminar os avanços da Química, e contribuir para a aplicação correta da mesma. É a autoridade para desenvolvimento de padrões de denominação para os compostos, tendo como membros sociedades científicas, academias nacionais e órgãos de representação. Participou das discussões para definição, junto a IUPAP (voltada a Física), da unidade de base do SI mais recentemente adotada, o mol (para a grandeza quantidade de substância).

A IUPAC é dividida em comitês, conforme as responsabilidades associadas a pesquisa, avaliação, educação, terminologias, nomenclatura, entre outros. O BIPM é representado na IUPAC em seu Comitê Interdivisional sobre Terminologia, Nomenclatura e Símbolos (ICTNS). Por sua vez, a IUPAC é membro dos seguintes comitês consultivos do BIPM: CCQM, CCU e JCGM.

O comitê *Analytical Chemistry Division* é responsável pela elaboração do *INTERNATIONAL HARMONIZED PROTOCOL FOR THE PROFICIENCY TESTING OF ANALYTICAL CHEMISTRY LABORATORIES (IUPAC Technical Report)*, ou simplesmente **Protocolo Harmonizado** (Thompson et al., 2006), orientando a harmonização no emprego da ferramenta EP, desde 1996. Uma nova versão foi elaborada em 2006, com disseminação pelo INMETRO.

2.1.5.

WHO (World Health Organization)

A WHO é uma agência criada pela ONU para o desenvolvimento da saúde global. Sua atuação consiste na coordenação de esforços para controle de doenças, além do patrocínio de programas de prevenção e tratamento. A WHO possui um programa coordenado, o EQAAS (*External quality assurance*

assessment scheme), com o apoio da EDQM (Divisão Europeia de Qualidade para Medicamentos e Cuidados com a Saúde). O objetivo desse programa é avaliar o desempenho técnico da qualidade dos laboratórios farmacêuticos designados pela WHO.

A ECBS (Comitê de Especialistas em Padronização Biológica) (WHO, 2016) desenvolve padrões de medição para a produção e controle de qualidade de produtos. Entre outras atividades, fornecem orientações para órgãos reguladores e fabricantes, servindo como padrão para a aprovação de vacinas. Também são fornecidos padrões para a comparação laboratorial de vacinas no mundo.

No documento “Boas práticas da WHO para laboratórios de controle de qualidade de produtos farmacêuticos” (WHO, 2010) encontra-se um requisito de que os laboratórios farmacêuticos nacionais de controle de qualidade participem de ensaios de proficiência de qualidade para avaliar seus processos analíticos e de materiais de referência. Nos requisitos de teste para monitoramento da qualidade de vacinas para o HIV na África (WHO, 2001), é destacada a necessidade do EP associado a um programa de controle de qualidade com competência para ensaios informada aos laboratórios participantes.

2.1.6. ILAC (International Laboratory Accreditation Cooperation)

Organização internacional de organismos de acreditação de laboratórios que promove o mecanismo de acreditação, fomentando o emprego de resultados de laboratórios acreditados. Através do acordo de reconhecimento mútuo da ILAC, os organismos de acreditação julgados competentes aumentam a aceitação de seus produtos e serviços, fortalecendo sua credibilidade. O objetivo principal é remover as barreiras comerciais, e estimular o uso e aceitação pela indústria de resultados de organismos acreditados no âmbito internacional.

Os organismos membros do ILAC operam em conformidade com a ISO/IEC 17011, participando da acreditação de organismos de avaliação da conformidade através das normas ISO/IEC 17025 (para laboratórios de calibração e ensaio), ISO 15189 (para laboratórios de ensaio médicos e clínicos) e a ISO/IEC 17020 (para órgãos de inspeção).

A política da ILAC reconhece a atividade de EP como fundamental para o processo de harmonização dos acordos multilaterais internacionais. Bons desempenhos em EP devem ser considerados no processo de acreditação de

laboratórios. Atualmente, o ILAC está desenvolvendo a extensão de seu acordo de referência mútuo (MRA ILAC), para acreditação de provedores de EP e produtores de MRC (ILAC, 2017; INMETRO, 2017).

2.1.7. EURACHEM

É uma rede de organizações europeia, criada para estabelecer um sistema de rastreabilidade internacional das medições químicas e de promoção de boas práticas de qualidade laboratorial. Proporciona um fórum para a discussão de problemas comuns ao laboratório clínico (EURACHEM, 2015). É aberta aos Estados membros da UE (União Europeia) ou da EFTA (Associação Europeia de Livre Comércio). Outros países europeus e organizações com interesse na qualidade das medições analíticas podem participar como membros Associados ou Observadores.

Para atender seus objetivos, a EURACHEM disponibiliza algumas publicações:

- *The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics* (EURACHEM, 2014);
- *Selection, Use and Interpretation of Proficiency Testing (PT) Schemes* (EURACHEM, 2011).

2.1.8. ISO (International Organization for Standardization)

Organização internacional independente, não-governamental, que reúne especialistas para compartilhar conhecimentos e desenvolver normas internacionais voluntárias, baseadas em consenso e relevantes para o mercado, que apoiem a inovação e proporcionem soluções para os desafios globais. Tais normas buscam facilitar o comércio e promover boas práticas de gestão e o avanço tecnológico, além de disseminar conhecimentos.

As normas ISO são elaboradas por comitês técnicos de diversas áreas, e por outros corpos analíticos. A Figura 1 relaciona os comitês técnicos, com as respectivas normas destacadas em cinza claro, mais fortemente relacionadas ao EP, ao MRC e às análises no laboratório clínico:

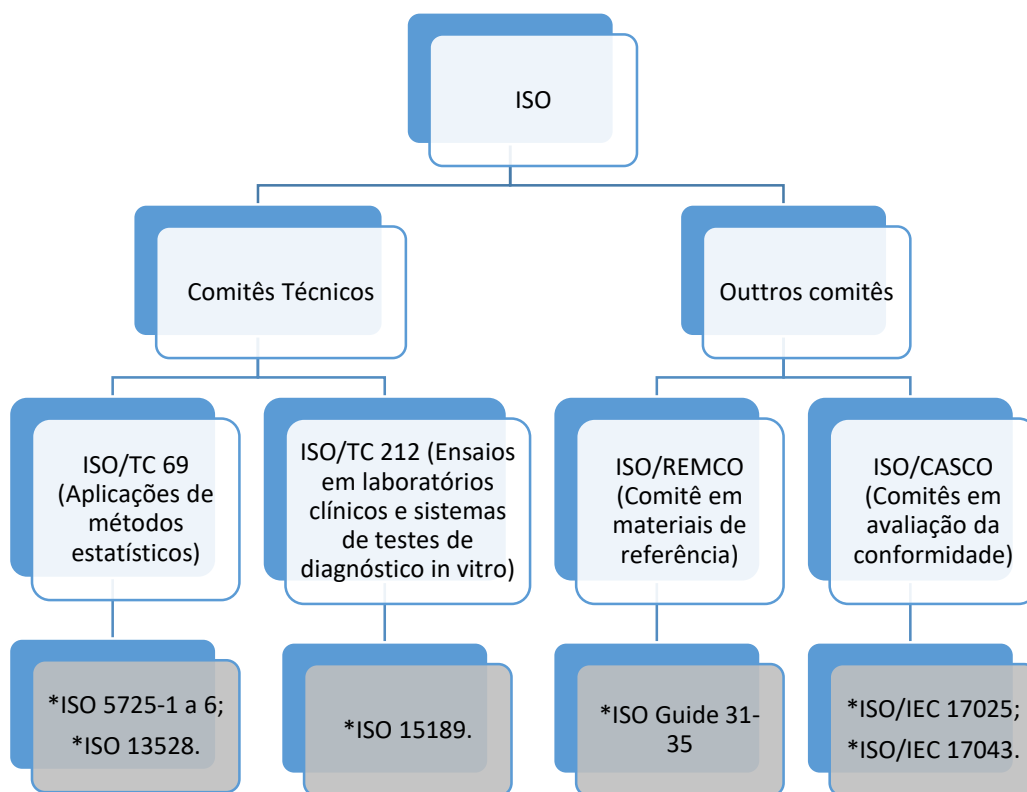


Figura 1 – Organograma ISO acerca setores e documentos relevantes para EP, MRC e laboratórios clínicos.

O comitê técnico (TC) responsável pela aplicação de métodos estatísticos é o TC 69 (Applications of statistical methods), cujo subgrupo técnico TC 69/SC 6 (Measurement methods and results) é responsável por desenvolver os métodos de medição, e incluirá as normas associadas aos ensaios de proficiência, comparações interlaboratoriais, calibrações, cálculos de incerteza e métodos de precisão. Nessa se tem o destaque para a normas ISO 13528, que é a norma que delimita os métodos estatísticos para aplicação em EP.

No TC 212, para testes e diagnósticos In Vitro, para laboratórios da área médica, a ISO 15189 é a norma a ser cumprida para a acreditação pelo INMETRO, em substituição a ISO/IEC 17025, conforme o INMETRO e o ILAC. Essa norma é responsável pelas adequações das investigações a ser feitas em laboratórios clínicos, acerca da competência, rastreabilidade metrológica, reagentes e materiais biológicos.

Para a acreditação de produtores de MRC, são delimitados os passos, que incluem as obrigações técnicas e estatísticas para a caracterização do MRC. Finalmente, para a avaliação da conformidade, têm-se a ISO/IEC 17025 (que

torna obrigatória a participação em EP, para qualquer laboratório de calibração ou ensaio) e a ISO/IEC 17043 (que orienta no processo de acreditação do provedor em EP).

2.2.

Contexto Nacional

2.2.1.

INMETRO (Instituto Nacional de Metrologia, Qualidade e Tecnologia)

É uma autarquia federal brasileira em forma de agência executiva, fundada pela Lei 5.966, de 11 de dezembro de 1973 (INMETRO, 2011). Essa autarquia é vinculada ao Ministério do Desenvolvimento, Indústria e Comércio Exterior, e que atua como Secretaria Executiva do Conselho Nacional de Metrologia, Normalização e Qualidade Industrial (CONMETRO). Além disso, o INMETRO atua no SINMETRO (Sistema Nacional de Metrologia, Normalização e Qualidade Industrial) com as seguintes principais atribuições designadas pelo CONMETRO:

- ✓ Organismo Acreditador.
- ✓ Secretaria Executiva do CONMETRO e dos seus comitês técnicos assessores.
- ✓ Supervisão dos Organismos de Fiscalização.

O INMETRO é o único organismo acreditador reconhecido no SINMETRO e, internacionalmente acreditado, também exerce no campo da avaliação da conformidade. O INMETRO delega as atividades de Metrologia Legal, como verificação, fiscalização e da certificação às entidades da Rede Brasileira de Metrologia Legal e Qualidade – RBMLQ que são os Institutos de Pesos e Medidas (IPEM) dos estados brasileiros. (INMETRO, 2007).

O INMETRO tem como missão promover a qualidade de vida e a competitividade da economia nacional por meio da metrologia e avaliação da conformidade. Entre suas atribuições, estão:

- Verificar a observância de normas técnicas e legais associadas às unidades de medida, instrumentos e métodos de medição;
- Conservar padrões de medição, implantar e manter a cadeia de rastreabilidade metrológica em solo nacional;

- Planejar e executar as atividades de acreditação de laboratórios de calibração e de ensaio, de provedores de ensaios de proficiência, de organismos de certificação e de inspeção.

O INMETRO é um dos signatários do CIPM MRA do BIPM, e dessa forma, tem equivalência internacional de seus padrões de medição, bem como dos certificados de medição e calibração que emite. No que se refere às atividades de acreditação, o INMETRO segue a política do ILAC, que considera o EP ferramenta fundamental para o processo de harmonização dos acordos multilaterais. Dessa forma, organismos como a CGCRE (Coordenação Geral de Acreditação do Inmetro) devem ser capazes de avaliar a participação dos laboratórios acreditados em atividades de EP. São comissões técnicas associadas a CGCRE:

- Comissões técnicas na acreditação de laboratórios;
- Comissões técnicas na acreditação de organismo de certificação;
- Comissões técnicas na acreditação de organismos de inspeção.

O INMETRO realiza a acreditação de laboratórios de calibração e ensaio, de forma a comprovar sua competência e capacidade de medição, atendendo os requisitos da ABNT NBR ISO/IEC 17025. Dessa forma, são formadas duas grandes redes de laboratórios acreditados: a Rede Brasileira de Laboratórios de Ensaio (RBLE) e a Rede Brasileira de Laboratórios de Calibração (RBC). Dentro da RBLE, os laboratórios podem ser ainda divididos conforme a modalidade de acreditação: laboratórios clínicos, laboratórios independentes e laboratórios de fábrica. Para o caso dos laboratórios clínicos, devido às peculiaridades associadas à medicina laboratorial, a norma exigida é a ISO 15189.

Uma das exigências para os laboratórios que desejam obter acreditação do INMETRO, conforme a ABNT NBR ISO/IEC 17025 e a ISO 15189, é a participação do laboratório em programas de EP. Para servir de subsídio, a CGCRE estabelece regulamentos (NIE-CGCRE, NIT-DICLA), que constituem requisitos para a acreditação, e também documentos orientativos (DOQ-CGCRE), que fornecem informações para a implementação dos requisitos de acreditação dentro dos laboratórios. Os Quadros 1 e 2 apresentam os documentos relacionados à acreditação dos laboratórios, e que ressaltam a importância da participação em EP, bem como a acreditação de laboratórios e fornecedores de MR.

Quadro 1 – Documentos orientativos para acreditação de laboratórios, com ênfase no EP, MRC ou laboratórios.

Documento	Data da atualização	Revisão	Descrição
DOQ-CGCRE-001	fev/17	13	Orientação para a Acreditação de Laboratórios, Produtores de Materiais de Referência e Provedores de Ensaios de Proficiência.
DOQ-CGCRE-008	ago/16	5	Orientação sobre validação de métodos analíticos.
DOQ-CGCRE-016	fev/10	2	Orientações para a seleção e uso de materiais de referência.
DOQ-CGCRE-019	ago/13	3	Exemplos de estimativa de incerteza de medição ensaios químicos.
DOQ-CGCRE-020	mai/15	7	Definições de termos utilizados nos documentos relacionados à acreditação de laboratórios, produtores de materiais de referência e provedores de ensaios de proficiência.
DOQ-CGCRE-053	fev/14	0	Exemplos de estimativa de incerteza de medição em ensaios microbiológicos.
DOQ-CGCRE-063	abr/16	3	Orientação para a Elaboração dos Escopos de Acreditação Voltados aos Laboratórios que Atuam na Área de Atividade: Saúde .
DOQ-CGCRE-064	set/14	0	Orientação para a elaboração dos escopos de acreditação voltados aos laboratórios de ensaios que atuam na área de atividade: saúde humana, focando análises clínicas e patológicas .
DOQ-CGCRE-078	jun/16	1	Orientações para Utilização de Materiais de Referência em Ensaios Biológicos.
DOQ-CGCRE-083	jul/16	1	Orientações Gerais sobre a Acreditação de Laboratórios que Realizam Calibração de Equipamentos com o Uso de Materiais de Referência Certificados (MRC).

Quadro 2 – Documentos normativos para acreditação de laboratórios, com ênfase no EP, MRC ou laboratórios.

Documento	Data da atualização	Revisão	Descrição
NIT-DICLA-016	set/14	7	Elaboração de Escopo de Laboratórios de Ensaios e de Provedores de Ensaios de Proficiência.
NIT-DICLA-021	mar/13	9	Expressão da incerteza de medição por laboratórios de calibração.
NIT-DICLA-026	jun/16	10	Requisitos para a Participação de Laboratórios em Ensaios de Proficiência.
NIT-DICLA-030	nov/15	9	Rastreabilidade Metrológica na Acreditação de Organismos de Avaliação da Conformidade e no Reconhecimento da Conformidade aos Princípios das BPL.
NIT-DICLA-031	jan/17	18	Regulamento da Acreditação de Laboratório, de Produtores de Materiais de Referência e de Provedores de Ensaios de Proficiência.
NIT-DICLA-062	ago/15	3	Aplicação da ABNT NBR ISO/IEC 17025 no âmbito da metrologia legal.
NIT-DICLA-063	abr/14	0	Política de Transição para a adoção da versão da norma ISO 15189:2012.

2.2.2.

ANVISA (Agência Nacional de Vigilância Sanitária)

A Agência Nacional de Vigilância Sanitária é uma autarquia sob regime especial, criada pela Lei nº 9.782, de 26 de janeiro 1999, funcionando como agência reguladora. Tem por finalidade promover a proteção da saúde da população por intermédio do controle sanitário de produtos e serviços voltado à saúde, pela aprovação dos mesmos para posterior comercialização e implementação, e ainda pelo monitoramento e fiscalização de ambientes, processos, insumos e tecnologias. A ação da ANVISA é coordenada entre estados e União. As resoluções da ANVISA são tomadas pela diretoria colegiada, composta do diretor-presidente, do diretor de controle e monitoramento sanitário e pelo diretor de autorização e registros sanitários.

Devido a essa estrutura, um dos instrumentos pelos quais a ANVISA realiza sua regulação são as RDC (Resolução da Diretoria Colegiada). Com a RDC 343/2002, o ensaio de proficiência passou a ser exigência para o funcionamento dos serviços de hemoterapia. Em 2003, foi instituído pela Portaria nº. 864, de 30 de setembro 2003, um Grupo de Trabalho para elaborar o Regulamento Técnico de Funcionamento do Laboratório Clínico. Finalmente, tendo como ponto principal esse regulamento, é publicada em 13 de outubro de 2005 a RDC 302/2005 (ANVISA, 2005a), que define os requisitos para o funcionamento dos laboratórios de análises clínicas, patológicas e citológicas.

Através do item 8.1, o laboratório clínico deve assegurar a confiabilidade dos serviços laboratoriais prestados, por meio de, no mínimo:

- a) controle interno da qualidade;
- b) controle externo da qualidade (ensaios de proficiência).

A RDC 302/2005 inclui ainda orientações acerca de controle de qualidade, calibração e materiais. Através de outras publicações, a ANVISA ressalta a importância do EP. No procedimento GGLAS 02/43 (ANVISA, 2002), é proposto um critério de habilitação para provedores de EP, incluindo sugestões de aplicação de limites fixos de aceitação para alguns ensaios. Em parceria com o SENAI, como auxílio a acreditação de laboratórios, é também publicado o Guia para a Qualidade em Química Analítica (ANVISA, 2005b), onde o ensaio de proficiência é abordado como ferramenta para o controle de qualidade.

2.2.3.

ABNT (Associação Brasileira de Normas Técnicas)

A Associação Brasileira de Normas Técnicas é o Foro Nacional de Normalização por reconhecimento da sociedade brasileira desde a sua fundação, em 28 de setembro de 1940, sendo assim confirmado pela Resolução n.º 07 do CONMETRO, de 24.08.1992. É membro da ISO (*International Organization for Standardization*), da *Comisión Panamericana de Normas Técnicas* (Copant), e da *Asociación Mercosur de Normalización* (AMN). A ABNT também é membro da IEC (*International Electrotechnical Commission*).

A ABNT é responsável pela elaboração das Normas Brasileiras (ABNT NBR), elaboradas por seus Comitês Brasileiros (ABNT/CB), Organismos de Normalização Setorial (ABNT/ONS) e Comissões de Estudo Especiais (ABNT/CEE). Desde 1950, a ABNT atua também na avaliação da conformidade e dispõe de programas para certificação de produtos, sistemas e rotulagem ambiental.

A ABNT atua por meio de Comitês Técnicos. No Quadro 3 são destacados aqueles mais relevantes para o tema desta pesquisa:

Quadro 3 – Comitês Técnicos de interesse da ABNT.

Número e título do Comitê Técnico da ABNT	Área de Atuação
ABNT/CB-25 – Qualidade	Normalização no campo de gestão da qualidade; diretrizes para seleção e aplicação das normas da qualidade; estudos de meios para a avaliação da conformidade de produtos, processos, serviços; elaboração de guias e normas.
ABNT/CB-53 – Normalização em Metrologia	Normalização no campo da metrologia científica e industrial: terminologia, generalidades, controle metrológico, confiabilidade laboratorial, métodos e técnicas de calibração e ensaios, comparação interlaboratorial, ensaios de proficiência.
ABNT/CEE-83 – Aplicações de Métodos Estatísticos	Normalização na aplicação de métodos estatísticos: estimação, planos de amostragem, análise, apresentação e interpretação de dados (espelho do ISO/TC 69).
ABNT/CEE-150 – Materiais de Referência	Normalização no campo de materiais de referência: definições, categorias, níveis e classificação de MR; determinar a estrutura de MR; formular critérios para a escolha das fontes de referência nos documentos da ABNT; preparar diretrizes para os Comitês Técnicos; propor ações sobre MR para o trabalho da ABNT; aconselhar o Conselho Técnico da ABNT.

3

Ensaio de proficiência para laboratórios clínicos

Conceitos de ensaio de proficiência (EP) são aplicados desde a década de 1940 (Sunderman, 1987). Mas o estabelecimento de uma política de qualidade abrangente que incluísse a ferramenta foi possibilitado apenas com o início do processo de automação pelos laboratórios (Oliveira & Mendes, 2011). Motivados por divergências em estudos clínicos, que resultaram em diferentes diagnósticos por laboratórios locais (Cerriotti, 2014), um grupo de pesquisadores solicitou o que seria o primeiro estudo por meio de EP em 1946, no estado da Pensylvania, EUA (Sunderman, 1987). A discrepância de resultados desse estudo foi tanta que suscitou uma pesquisa acerca dos parâmetros de exatidão e precisão desses laboratórios. Ainda em 1946, com a fundação do CAP (*College of American Pathologists*), são intensificadas tais pesquisas, que resultariam na realização do primeiro programa de EP (Sunderman, 1987). As pesquisas acerca da eficácia do método em detectar problemas nos laboratórios resultaram na expansão de sua prática nos EUA e Europa (Cerriotti, 2014).

Desde a sua criação, o EP sempre foi identificado como uma rotina para implementação de um programa adequado de garantia da qualidade no âmbito do laboratório (Oliveira & Mendes, 2011). Mas ao contrário do controle interno, que é útil no monitoramento do erro aleatório, e que emprega técnicas igualmente sofisticadas, como as regras de Westgard (Westgard et al., 1997), um dos focos do EP é o monitoramento do erro sistemático, auxiliando o laboratório na eliminação de suas causas. Tais causas podem estar relacionadas à validação do método (avaliação de repetibilidade e reprodutibilidade, linearidade, calibração...), às diferenças entre condições analíticas, ao treinamento ineficaz de pessoal, entre outros (Oliveira & Mendes, 2011). Em meio à complexidade nesse processo, é fato que cada laboratório deve buscar o programa de EP que venha a se adaptar a sua demanda e às especificações da qualidade, o que exige conhecimento profundo da cadeia de processos do laboratório clínico (Oliveira & Mendes, 2011).

3.1.

Etapas para execução do ensaio de proficiência

De acordo com definição da ABNT NBR ISO/IEC 17043 - Avaliação de Conformidade - Requisitos Gerais para Ensaio de Proficiência, o EP consiste na determinação do desempenho de um laboratório por meio da realização de ensaios e comparação interlaboratorial. Ainda segundo essa norma, a realização do programa de EP envolve duas partes distintas: o provedor do EP, responsável pelo desenvolvimento e operação do programa, e o cliente, também dito participante do programa de EP, para o qual esse programa é oferecido por contrato (ABNT NBR ISO/IEC 17043, 2011). O provedor do EP deve ter competência para realizar comparações interlaboratoriais, acesso a um conhecimento especializado e possuir competência na medição das propriedades a serem determinadas.

O esquema de realização do EP pode ser resumido na Figura 2:

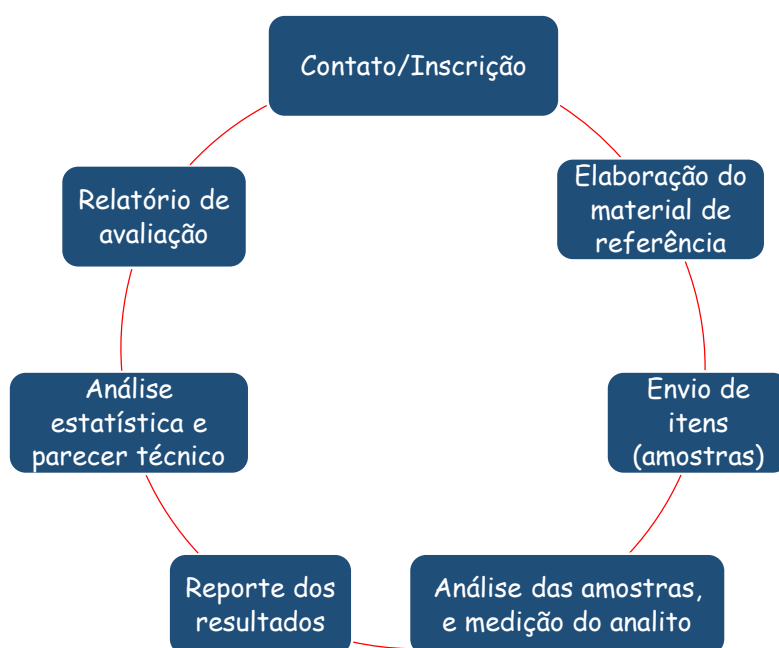


Figura 2 – Ciclo de realização de um Ensaio de Proficiência (adaptado de Oliveira & Mendes, 2011).

Numa primeira etapa do ciclo para realização do ensaio de proficiência, um laboratório se inscreve como participante de um programa de EP organizado pelo provedor definido. Tal provedor pode ser um outro laboratório, uma instituição científica, entre outros. O provedor realiza o preparo de um material apropriado,

que no caso do laboratório clínico, deve reproduzir a rotina do laboratório. Esse processo de preparo de materiais inclui a triagem de materiais, o preparo biológico, ensaios e avaliação de homogeneidade e estabilidade, estoque, liofilização, entre outros. O objetivo é gerar amostras similares, segundo a condição de análise do material. Essas amostras são então enviadas na forma de itens (que podem conter mais de um analito para mensuração dentro de cada item), separadamente para todos os participantes inscritos, com validade dos itens definida e registrada no item. De forma similar ao executado em sua rotina, cada laboratório analisa seu(s) respectivo(s) item(ns), e para um determinado analito, reporta o(s) resultado(s) da quantificação do(s) analito(s) presente no(s) itens enviados pelo provedor (exemplo: quantificação de Colesterol em uma amostra de soro humano). Os resultados dos participantes respondentes são armazenados pelo provedor, e é realizada uma análise técnica e estatística dos dados. Após finalizar essa análise, as conclusões são reportadas para o laboratório, de forma individual, com uma análise do desempenho para cada item da rodada.

Uma linha de exigências deve ser atendida para seu correto funcionamento (ABNT NBR ISO/IEC 17043, 2011):

- **Planejamento:** objetivos do programa definidos; endereço físico, identificação do coordenador e envolvidos; atividades e escopo de ensaios; cronograma; critérios para participação; datas e procedimentos para análise de itens, reporte de dados e compilação; descrição dos métodos de cálculo; análise técnica e estatística, e avaliação;
- **Pessoal:** equipes gerenciais e técnicas treinadas e qualificadas para atividades autorizadas, associadas ao planejamento, amostragem, preparo, logística, processamento de dados, análise estatística, avaliação de desempenho;
- **Área Física e Recursos:** instalações e equipamentos apropriados para a realização do EP, para fins de fabricação, armazenamento e expedição de itens, bem como de processamento, comunicação e recuperação de dados; softwares e planilhas validados; segurança, confidencialidade e integridade de dados brutos;
- **Materiais:** seleção, coleta e aquisição de materiais; normas para manuseio e armazenamento adequado; critérios de descarte; investigação de efeitos-matriz e de concentração do mensurando; recomendações sobre material: tomada da porção, transporte e logística, identificação, armazenamento e manuseio documentados; homogeneidade e estabilidade testados e documentados.

3.2. Avaliação do erro sistemático por meio de EP

Enquanto ferramenta de natureza metrológica, a realização do EP se insere no conceito de validação, sendo esta do tipo interlaboratorial. Para sua realização, o provedor de EP deve ter então definidos uma estimativa do valor verdadeiro (chamada de valor designado) e uma medida de dispersão. Tal medida de dispersão pode ser obtida pelo desvio padrão para a proficiência, ou por um balanço de incertezas associadas ao valor designado. Esse balanço pode ser oriundo de incertezas-padrão ou de incertezas expandidas (oriundas de incertezas combinadas de diversas incertezas-padrão). Os métodos para o cálculo dessas incertezas, modelos estatísticos e requisitos de aplicação são definidos de acordo com a GUM (GUM, 2008). Conforme o objetivo do EP, podem ser enviados aos laboratórios um ou mais itens, e de posse da amostra enviada, o laboratório pode realizar uma ou mais medições. A Figura 3 ilustra as componentes do erro total da medição.

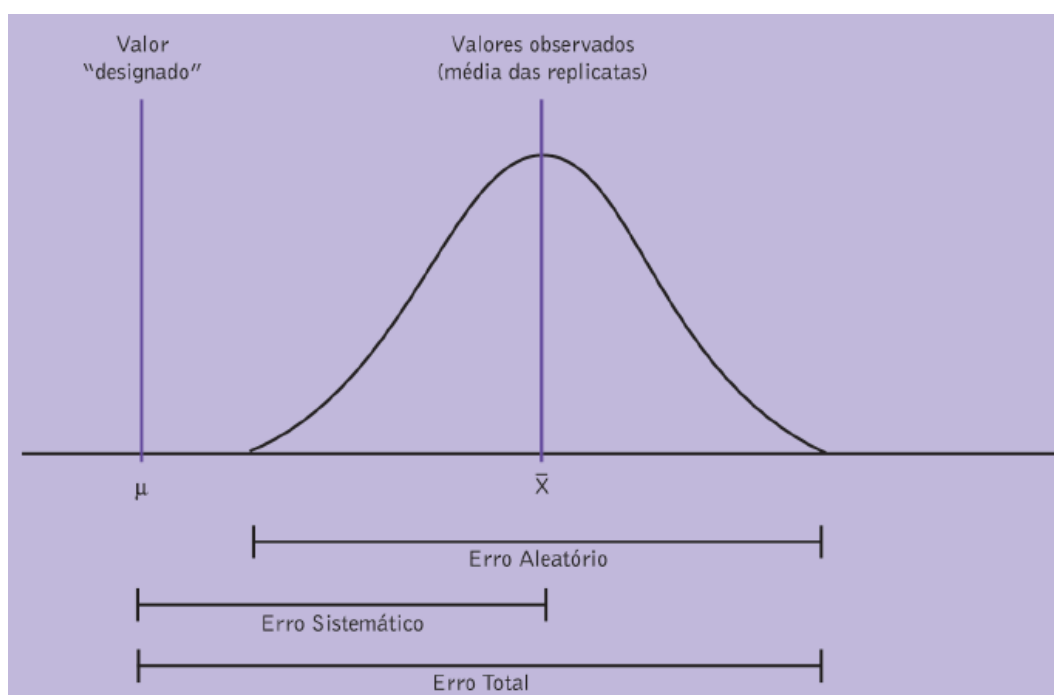


Figura 3 – Representação esquemática das componentes aleatória e sistemática do erro de medição (Oliveira & Mendes, 2011).

O erro total (ou erro de medição) é composto pela soma dos erros sistemático e aleatório (VIM, 2012; GUM, 2008; Oliveira & Mendes, 2011). O erro sistemático é a diferença entre a medida de tendência central dos dados em condições de repetibilidade, para infinitas repetições, e o valor verdadeiro do

mensurando (GUM, 2008). Essa medida de tendência central (GUM, 2008), sob condições de repetibilidade, é associada à média das replicatas geradas pelo laboratório. Em EP, a estimativa do valor verdadeiro é chamada de valor designado (Oliveira & Mendes, 2011), também representado na Figura 3.

Uma das principais vantagens do EP é o monitoramento do erro sistemático desse laboratório. O EP é uma ferramenta para garantia da qualidade que funciona de forma poderosa, quando empregado em conjunto com o controle interno, o uso de materiais de referência certificados e a validação de métodos analíticos (Oliveira & Mendes, 2011; Thompson et al., 2006). Embora não seja o foco do EP, os dados reportados pelos participantes de uma rodada de EP podem também ser empregados para validar métodos analíticos (Thompson et al., 2006). Ferrero & Casaril (2009) empregaram dados de ensaios de proficiência para determinar a repetibilidade, a precisão intermediária e a reprodutibilidade de um método padrão de medição, com o objetivo de melhorar a rastreabilidade em análises químicas. A Rede Metrológica RS (Rede Metrológica, 2017) emprega dados de programas interlaboratoriais para certificar seus materiais de referência.

3.3.

Acreditação laboratorial e ensaio de proficiência

Acreditação é o reconhecimento formal por um organismo competente, de que outro organismo de avaliação de conformidade atende a requisitos previamente definidos, e demonstra ser competente para realizar suas atividades com confiança. A acreditação é uma ferramenta estabelecida em escala internacional para gerar confiança na atuação de organizações que executam atividades de avaliação da conformidade.

Devido ao seu caráter voluntário, o processo de acreditação está associado ao desenvolvimento no seu campo de atuação. A história da acreditação voltada à saúde tem início com a criação nos EUA do PPH (Programa de Padronização Hospitalar), estabelecido pelo Colégio Americano de Cirurgiões para definir os requisitos essenciais à medicina assistencial, nos anos 1920 (Vieira, 2011). O PPH descrevia a organização do corpo de médicos, as normas de atendimento, os exames e condições de realização dos mesmos. Nos anos 1930, um processo similar tem início no Brasil (Vieira, 2011), com a criação da Ficha de Inquérito Hospitalar, para estabelecer os padrões de organização no ambiente do hospital, o que incluía organização do corpo médico, dos serviços e laboratórios, entre outros.

O ILAC (*International Laboratory Accreditation Cooperation*), se inicia em 1977 na forma de conferências que se realizam anualmente para promover a aceitação de testes e resultados de calibração acreditados para laboratórios. Nos anos 1980, os programas de acreditação se disseminam quando da elaboração do manual de padrões de acreditação, realizado pela WHO e pela Organização Pan-Americana de Saúde (OPAS). Em 1996, o ILAC torna-se formalmente uma Cooperação, passando a estabelecer uma rede de ação entre organismos acreditadores em todo o mundo. Em 1997, o Ministério da Saúde no Brasil decidiu estabelecer uma comissão nacional de especialistas para o desenvolvimento do modelo brasileiro de acreditação, o qual foi oficialmente lançado em 1998. Desde os anos 2000, com a criação dos acordos de reconhecimento mútuos (ILAC MRA) para testes e resultados de calibração acreditados, o ILAC incorpora normas da ISO para disseminar o mecanismo de acreditação. Os órgãos membros do ILAC MRA operam em conformidade com a ISO/IEC 17011, participando da acreditação de organismos de avaliação da conformidade, através das normas ISO/IEC 17025 (para laboratórios de calibração e ensaio), ISO 15189 (para laboratórios de ensaio médicos e clínicos) e a ISO/IEC 17020 (para órgãos de inspeção).

A acreditação é vista como um processo dinâmico, de benefícios para todos os envolvidos (Sevgi, 2013). Através da acreditação, os laboratórios de ensaio e calibração demonstram que suas medições apresentam resultados rastreáveis aos padrões internacionais, garantindo aos clientes que os resultados do laboratório são confiáveis (Sevgi, 2013). Por outro lado, a acreditação permite aos fabricantes de sistemas analíticos a capacidade de identificação de melhorias na qualidade dos seus produtos. Finalmente, a acreditação fortalece a credibilidade de resultados emitidos pelas instituições governamentais (Sevgi, 2013).

A acreditação é voluntária, independente, imparcial e funciona de forma não lucrativa, independendo assim de motivações comerciais, e dessa forma, sem competir com organismos de acreditação ou de avaliação da conformidade. O EP é tido como uma das formas de garantia da continuidade do processo dinâmico da acreditação (Sevgi, 2013).

Antes de solicitar a acreditação ou sua extensão, o laboratório de ensaio ou calibração deve participar de forma satisfatória de ao menos uma atividade de ensaios de proficiência para cada grupo de calibração, classe de ensaio ou exame por especialidade, conforme o escopo solicitado pelo INMETRO (NIT-Dicla-026, 2016). Após obter a acreditação, o laboratório deve, segundo a mesma norma, participar de ao menos uma atividade de ensaio de proficiência relacionada a cada

parte significativa do seu escopo de acreditação a cada quatro anos. Para demonstrar sua competência, o laboratório deve requerer a acreditação, sendo essa competência regida pela ABNT NBR ISO/IEC 17025:2005 (para laboratórios de calibração e ensaio) e pela ABNT NBR NM ISO 15189:2015 (específica para os laboratórios de ensaio da área médica) (ILAC, 2017). Por outro lado, a competência dos provedores de MRC é regida pela ABNT ISO Guia 34:2012, e a competência dos provedores de EP pela ABNT NBR ISO/IEC 17043:2011 (ILAC, 2017).

3.4.

Material de Referência e Material de Referência Certificado

Uma das razões essenciais que dificultam a harmonização de resultados em laboratórios clínicos é a ausência de rastreabilidade metrológica à unidade do mensurando, associada aos resultados emitidos pelos laboratórios. Tal rastreabilidade é exigida pela diretiva europeia *IVD 98/79/EC* (Bais et al., 2013), que estabelece que quando disponíveis serviços de medição de referência, os fabricantes de sistemas analíticos devem garantir rastreabilidade aos valores que designa para seus calibradores. Devido aos altos custos e recursos associados, nem sempre a cadeia de rastreabilidade é executada de forma completa (Bais et al., 2013). Dessa forma, ainda há grande dificuldade de se obter equivalência entre os resultados de muitos mensurandos comumente avaliados, mesmo para os casos em que haja MRC ou MR disponível.

É nítido um aumento considerável no número, na variedade e na utilização de MR (ABNT ISO Guia 31, 2015). Esse fato é oriundo da exigência cada vez maior de confiabilidade nos resultados obtidos por técnicas analíticas. Alguns sistemas analíticos não utilizam calibradores com rastreabilidade a um MRC (PESSOA & FERREIRA JR, 2015), gerando impactos na qualidade e confiabilidade dos resultados, na validação de novas metodologias e na comparação e harmonização de resultados. O impacto direto se dá na condução correta do diagnóstico clínico.

Conforme o VIM (2012):

- **Material de referência (MR)** – Material, suficientemente homogêneo e estável em relação a propriedades específicas, preparado para se adequar a uma utilização pretendida numa medição ou num exame de propriedades qualitativas;

- **Material de referência certificado (MRC)** – Material de referência, acompanhado de uma documentação emitida por uma entidade reconhecida, a qual fornece um ou mais valores de propriedades especificadas com as incertezas e as rastreabilidades associadas, utilizando procedimentos válidos.

Pela definição, a diferença entre um MR e um MRC reside no fato de o segundo tipo de MR ser certificado, isso é, de possuir um certificado associado. Tal certificado contém todas as informações essenciais para o uso do MRC (ABNT ISO Guia 34, 2012), sendo que é recomendado que o MRC e seu certificado jamais sejam separados (ABNT ISO Guia 31, 2015). A diferença no certificado conduz a outras três diferenças intrínsecas:

- Caracterização: é uma descrição do material, que informa o tipo, fornecedor, validade, uso pretendido, condições de uso..., que deve ser realizada por meio de procedimento metrologicamente válido (ABNT ISO Guia 34, 2012);
- A realização de um procedimento metrologicamente válido para se caracterizar um MRC torna esse processo associado a incertezas. Assim, no certificado devem estar presentes o(s) valor(es) de propriedade especificada(s), a(s) incerteza(s) associada(s) e o(s) método(s) de cálculo que resultaram nesse(s) valor(es) e incerteza(s);
- Finalmente, no certificado deve estar presente uma declaração de rastreabilidade metrológica. Elementos para confirmar a rastreabilidade metrológica são, além dos valores e incertezas acima citados, a cadeia de rastreabilidade metrológica ininterrupta ao SI, o procedimento de medição documentado, a competência técnica reconhecida e os intervalos entre calibrações (ILAC, 2013; VIM, 2012).

Algumas possibilidades de emprego de MRC (DOQ-CGCRE-016, 2010):

- ✓ Validação de métodos e determinação da incerteza;
- ✓ Verificação de correto uso de método (treinamento de pessoal, adequação de reagentes, padrões, equipamentos...);
- ✓ Calibração, que investiga a contribuição da incerteza do MRC para a incerteza total da medição;

- ✓ Controle e garantia da qualidade (investigação de homogeneidade e estabilidade, determinação do valor designado...).

Materiais de referência devem ser caracterizados com relação à homogeneidade, estabilidade e valor(es) de propriedade certificado(s) (DOQ-CGCRE-016, 2010). Para controle interno de qualidade, o último requisito é dispensável. No caso do ensaio de proficiência, avaliar a homogeneidade é essencial, e a estabilidade da amostra dentro da escala de tempo de execução do ensaio deve ser avaliada e controlada.

Embora desejável, o custo da atribuição de valores de propriedade certificados limita muito a sua utilização em EP (Thompson et al., 2006), o que faz com que resultados consensuais sejam empregados em seu lugar. Assim, é com frequência restar dúvidas sobre a confiabilidade de valores designados e respectivas incertezas em EP, pois os valores carregam algum elemento não revelado de incerteza (DOQ-CGCRE-016, 2010). Quando empregado o consenso, a interpretação dos parâmetros exige cuidado. Os pressupostos se baseiam em dois aspectos: o valor certificado é a melhor estimativa do valor verdadeiro da propriedade do MRC; e que toda variação seja aleatória, oriunda de uma distribuição normal (ABNT ISO Guia 35, 2012). No BIPM, o apêndice C do KCDB contém 14 declarações de CMC com material de referência para glicose ou d-glicose, conforme apresentado no Quadro 4. Dentre os nove países que possuem CMC declarado para esse analito, dois são pertencentes ao continente americano, os Estados Unidos e o México.

Quadro 4 – Informação contida no Apêndice C do KCDB sobre a Capacidade de Medição associada a materiais de referência para glicose (BIPM, 2017).

<u>Analyte or component</u>	<u>Matrix or material</u>	<u>Service Provider</u>		
		<u>Country</u>	<u>NMI</u>	<u>RMO</u>
glucose	high purity glucose	China	NIM	APMP
glucose	native or processed blood serum or plasma	France	LNE	EURAMET
glucose	blood serum	Germany	PTB	EURAMET
d-glucose	human serum (frozen or lyophilized)	Korea, Republic of	KRISS	APMP
d-glucose	human serum (frozen)	Korea, Republic of	KRISS	APMP
d-glucose	human serum (lyophilized)	Korea, Republic of	KRISS	APMP
glucose	serum (frozen)	Korea, Republic of	KRISS	APMP
glucose	glucose	Mexico	CENAM	SIM
glucose	human serum	Mexico	CENAM	SIM
glucose	blood serum	Russian Federation	VNIIM	COOMET
glucose	human serum	Singapore	HSA	APMP
glucose	blood serum	Turkey	UME	EURAMET
d-glucose	high purity d-glucose	United States	NIST	SIM
glucose	human serum	United States	NIST	SIM

Para promover a rastreabilidade metrológica em medicina laboratorial, o JCTLM dissemina uma base de dados (JCTLM-DB), com informações de produtores e serviços de MRC e MetR (métodos e procedimentos de medição de referência). Através dela, é possível discriminar analitos, técnicas e matriz de realização dos ensaios clínicos. No Quadro 5, são disponibilizados os métodos e procedimentos de medição de referência, reconhecidos e rastreáveis, para o analito glicose em soro humano. Para dosagem de glicose em soro humano, podem ser empregados os métodos cromatografia de troca líquida (ID-LC-MS/MS), cromatografia de troca gasosa (ID/GC/MS) e enzimático. Quando empregada uma das cromatografias, é analisada a concentração de quantidade de substância dentro do soluto, e o método é rastreável a uma unidade derivada do mol (Quadro 5). Quando empregado o método enzimático, é analisada a concentração de massa no soluto, e o método é rastreável a uma unidade derivada do kg (Quadro 5).

Quadro 5 – Métodos e procedimentos de medição de referência para glicose em soro humano, reconhecidos e rastreáveis, listados no JCTLM-DB (BIPM, 2017).

Enzymatic method for glucose in blood serum	
▶ JSCC reference method	
Applicable matrice(s)	blood serum
Full description of technique(s)	Enzymatic
Quantity	Mass concentration
Applicable range	50 mg/dl to 300 mg/dl
Expected uncertainty (level of confidence 95%)	0.4 mg/dl to 2 mg/dl
JCTLM DB identification number	C3RMMP7

Isotope dilution mass spectrometry methods for glucose in blood serum	
▶ DGKL reference method for glucose	
Applicable matrice(s)	lyophilized, fresh or frozen blood serum, blood plasma, whole blood, urine, or calibration solution; and, other biological material
Full description of technique(s)	ID/GC/MS
Quantity	Amount-of-substance-concentration
Applicable range	0.5 mmol/l to 20 mmol/l
Expected uncertainty (level of confidence 95%)	1 % to [dependent from the matrix and the special properties of the analyte]
Reference(s)	<i>Biomedical Mass Spectrometry</i> , 1982, 9 , 395-405
JCTLM DB identification number	C3RMMP19
▶ HSA ID/LC/MS/MS reference measurement procedure for glucose in human serum	
Applicable matrice(s)	human serum
Full description of technique(s)	ID/LC/MS/MS
Quantity	Amount-of-substance concentration
Applicable range	1.5 mmol/l to 20 mmol/l
Expected uncertainty (level of confidence 95%)	0.028 mmol/l to 0.25 mmol/l
Reference(s)	High accuracy analysis of glucose in human serum by isotope dilution liquid chromatography-tandem mass spectrometry, Chen Y. et al., <i>Clinica Chimica Acta</i> , 2012, 413 (7-8), 808-813
JCTLM DB identification number	C13RMP4
▶ NIST reference method for serum glucose	
Applicable matrice(s)	lyophilized, fresh, or frozen human serum
Full description of technique(s)	ID/GC/MS
Quantity	Amount-of-substance concentration
Applicable range	2 mmol/l to 20 mmol/l
Expected uncertainty (level of confidence 95%)	0.5 % to 1.5 %
Reference(s)	<i>Biomed. Mass. Spectrom.</i> , 1982, 9 , 395-405
Comparability assessment study(ies)	<i>Metrologia</i> , 2003, 40 , <i>Tech. Suppl.</i> , 08003
JCTLM DB identification number	NRMeth 80
▶ University of Ghent reference method for glucose	
Applicable matrice(s)	lyophilized, fresh, or frozen human serum
Full description of technique(s)	ID/GC/MS
Quantity	Amount-of-substance concentration
Applicable range	1 mmol/l to 20 mmol/l
Expected uncertainty (level of confidence 95%)	1 % to 2 %
Reference(s)	<i>Clin. Chem.</i> , 1993, 39 , 1001-1006 <i>Clin. Chem.</i> , 1993, 39 , 993-1000 <i>Eur. J. Clin. Chem. Clin. Biochem.</i> , 1996, 34 , 853-860
Comparability assessment study(ies)	EUROMET 563
JCTLM DB identification number	NRMeth 4

Uma forma de se aplicar critérios em EP que independam da qualidade dos dados fornecidos pelos participantes do ensaio é empregar os valores de propriedade de um MR ou MRC (De Bièvre, 2016; ISO 13528, 2015). O emprego dos valores de propriedade de um MRC permite não só uma avaliação direta do desempenho, independente da qualidade dos dados, como também fornece a rastreabilidade metrológica à unidade do SI do valor designado (Thompson et al., 2006).

3.4.1. Homogeneidade e estabilidade do MR

Em estudos interlaboratoriais, os materiais enviados aos participantes possuem algum grau de heterogeneidade, o que pode ter impacto na subdivisão do material em itens de EP (Thompson et al., 2006). O objetivo é que essa heterogeneidade não seja significativa, isso é, que todos os laboratórios recebam os itens sem diferença significativa quanto à concentração do analito. Essa condição é denominada de **homogeneidade suficiente** (Thompson et al., 2006). Por outro lado, os itens de EP distribuídos devem ser suficientemente estáveis no período durante a sua validade (Thompson et al., 2006). Qualquer mudança durante o período deve ser irrelevante para a interpretação dos resultados da rodada de EP. Essa segunda condição é denominada **estabilidade suficiente**. Tais averiguações também são descritas pela ISO (ISO 13528, 2015).

Os ensaios de homogeneidade e estabilidade suficientes para fins de EP (Thompson et al., 2006) não devem ser confundidos com os ensaios de homogeneidade e estabilidade para certificação de um MR (ABNT ISO Guia 35, 2012). As condições de certificação de um MR são mais rígidas, operando em períodos maiores de tempo, e com objetivos diferentes (ABNT ISO Guia 35, 2012).

3.4.1.1. Homogeneidade

Se o material possui “homogeneidade suficiente”, então o desvio padrão real da amostra (σ_{sam}) é desprezível em comparação à variação introduzida pelas medições dos participantes em EP (Thompson et al., 2006), também conhecida como desvio padrão para a proficiência (σ_p). É requerido que o desvio padrão estimado da amostra (s_{sam}) seja no máximo, de 30 % do desvio padrão para a proficiência definido (σ_p) (Thompson et al., 2006), ou seja:

$$s_{sam} \leq 0,3\sigma_p$$

O ensaio de homogeneidade suficiente demanda a análise de uma amostra com ao menos 10 resultados em duplicata, sob condições de repetibilidade. Os resultados são testados por meio de uma ANOVA One-Way (um fator), em que esse fator é constituído dos grupos de duplicatas. Testes estatísticos são úteis nesse caso, para identificar *outliers* entre as duplicatas. Nesse caso, uma alternativa é a aplicação do teste de Cochran para variâncias, objetivando identificar duplicatas que sejam passíveis de descarte (Thompson et al., 2006).

3.4.1.2. Estabilidade

Durante o período de validade do item, o material deve ser suficientemente estável para aplicação no EP, ou seja, mudanças que ocorram durante o período pertinente devem ser irrelevantes para a interpretação dos resultados (Thompson et al., 2006). É indicado que modificações de $0,1\sigma_p$ seriam irrelevantes, pois conduziriam a mudança de ± 1 no Z-Score (Thompson et al., 2006). Por outro lado, é sugerido que o material é estável num momento adiante se a diferença em módulo entre a média da homogeneidade e a média da estabilidade não extrapolar $0,3\sigma_p$ (ISO 13528, 2015). Para fins de EP, o período de validade é o intervalo entre a preparação do material e o prazo para o retorno dos resultados. Esse período pode ser maior, em caso de definidos outros fins para o material enviado e empregado (Thompson et al., 2006).

3.5. Parâmetros que regem o EP

Os parâmetros necessários para avaliar o laboratório participante de um EP são o valor designado e uma medida de dispersão associada. O valor designado representa uma estimativa do valor verdadeiro (X), ilustrado na Figura 3. Como estimativa, estará sujeito a uma incerteza, que pode ser obtida pelo desvio padrão para a proficiência (σ_p) ou por um balanço de incertezas associadas ao valor designado. São descritos diversos métodos para o cálculo desses parâmetros (ISO 13528, 2015; Thompson et al., 2006), dependendo dos objetivos do EP, da disponibilidade de MRC's ou MR's e da natureza das informações.

3.5.1. Valor designado

Valor designado é o “valor atribuído a uma propriedade específica de um item de EP” (ABNT NBR ISO/IEC 17043, 2011). Na prática em medições, o valor verdadeiro do mensurando não é possível de ser conhecido (VIM, 2012). Deve ser definida, para aplicação em EP, uma estimativa do que seria o valor que mais se aproximaria do valor verdadeiro. Essa estimativa é conhecida como valor designado (Oliveira & Mendes, 2011). Os métodos para definição do valor designado estão presentes na ISO 13528 (2015), sendo esse o documento recomendado (ABNT NBR ISO/IEC 17043, 2011) para a aplicação dos métodos estatísticos para avaliação do desempenho em EP. Para os protocolos de realização do EP em laboratórios de química analítica, é largamente difundida a aplicação do International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories, publicado pela IUPAC (Thompson et al., 2006).

Uma vez que um dos principais objetivos do ensaio de proficiência é monitorar o erro sistemático dos laboratórios participantes, a correta escolha do valor designado é crucial para esse processo. A ISO (ISO 5725-4, 2004) recomenda que esse valor designado seja definido a partir dos valores de propriedade de um material de referência, que pode ser certificado (MRC), manufaturado com propriedades conhecidas, ou estabelecido por um método alternativo de viés conhecido e desprezível. Quando não é possível definir esse valor pelas propriedades de um material de referência, uma alternativa é o emprego de estatísticas por consenso de participantes, que segundo a ISO (ISO 13528, 2015) é um método onde o erro sistemático poderá ser afetado pela pobreza de repetibilidade e estabilidade.

Em ordem direta do aumento da incerteza associada ao valor designado, são listadas no Quadro 6 as abordagens distintas para a definição do valor designado, bem como as vantagens e desvantagens de aplicação de cada uma dessas abordagens.

Quadro 6 – Resumo dos tipos, vantagens e desvantagens das propostas para definição do valor designado para o EP (ABNT NBR ISO/IEC 17043, 2011; Thompson et al., 2006).

Tipo de Valor Designado	Origem do Valor Designado	Vantagens	Desvantagens
Valores Conhecidos	Obtido por formulação específica do item de EP	<ul style="list-style-type: none"> ▪ Método com menor incerteza potencial, devido à concentração e à homogeneidade; ▪ Funciona melhor em soluções líquidas e homogêneas. 	<ul style="list-style-type: none"> ▪ Concentração deve ser conhecida com precisão; ▪ Difícil obter homogeneidade suficiente em materiais sólidos; ▪ Analito adicionado pode ser menos ligado a matriz que o analito já presente.
Valores de Referência Certificados	Valor de propriedade de um MRC	<ul style="list-style-type: none"> ▪ Rastreabilidade metrológica automaticamente fornecida; ▪ Em quantidade suficiente, valores designados e incertezas do certificado são usadas diretamente. 	<ul style="list-style-type: none"> ▪ MRC's são escassos ou indisponíveis; ▪ Alto custo na produção de MRC's; ▪ Facilmente identificáveis, comprometendo o objetivo do EP.
Valores de Referência	Associado a um MR ou padrão de referência (laboratório, visto por meio de consenso suficientemente confiabilidade para o programa)	<ul style="list-style-type: none"> ▪ Na maioria das vezes, equivale ao uso de um MRC; ▪ Material adaptado aos requisitos do programa. 	<ul style="list-style-type: none"> ▪ Requer custos e esforços para validar a metodologia do material, ou para eliminar possíveis interferências.
Valores de Consenso de Laboratórios Especialistas	Obtido por comparação direta com os valores de consenso obtidos por laboratórios especialistas	<ul style="list-style-type: none"> ▪ Método útil para parâmetros empíricos; ▪ Efetivo na averiguação da consistência de resultados laboratoriais de rotina, com fins regulatórios. 	<ul style="list-style-type: none"> ▪ Esforço equivalente ao de se certificar um material de referência; ▪ Pouco efetivo para avaliar rotina; ▪ A incerteza do valor designado por consenso pode ser suficientemente grande.
Valores de Consenso de Participantes do Programa	É definido por meio de um consenso dos resultados reportados para o item de EP	<ul style="list-style-type: none"> ▪ Baixo custo, o valor designado não requer análises mais aprofundadas; ▪ Boa aceitação de pares; ▪ O cálculo do valor designado é geralmente direto. 	<ul style="list-style-type: none"> ▪ Valores designados dependentes dos resultados dos participantes; risco de tendências (erro sistemático ou presença de subpopulações); ▪ Incerteza alta para poucos resultados de laboratórios (notadamente, para $n < 15$).

3.5.2.

Desvio padrão para a proficiência

A escolha do tipo de valor designado terá impacto na escolha da medida de dispersão associada, para definição do critério a ser empregado. Para esse fim, deve ser definido o desvio padrão para a proficiência, que é uma medida de

dispersão utilizada na avaliação de resultados de EP (ABNT NBR ISO/IEC 17043, 2011), baseada nas informações disponíveis. Esse parâmetro pode ser definido como apresentado pela ISO (ISO 13528, 2015):

- **Por meio de percepção de especialistas.** Nesse caso, o desvio padrão para a proficiência corresponderá a um valor fixado, razoável para a avaliação dos participantes do EP, e definido por autoridades reguladoras, órgãos de acreditação ou especialistas técnicos. Pode ser associado ao erro máximo analítico permitido (item 3.5.3);

- **Por meio de rodadas anteriores do programa de EP.** Nesse caso, o desvio padrão para a proficiência será um valor fixado, invariante ao longo das rodadas do EP, onde os participantes empreguem métodos de medição comparáveis e compatíveis. Métodos estatísticos, como a análise de regressão linear simples, podem ser empregados para sua definição;

- **Como valor definido por um modelo geral.** Nesse caso, o desvio padrão pode ser definido por um modelo matemático específico, como por exemplo, a curva de Horvitz;

- **Como valor definido por um experimento de precisão.** Nesse caso, o método de medição deve ser padrão para todos os laboratórios. Através da repetibilidade e da reprodutibilidade (se estiverem disponíveis), ele pode ser estimado por meio de cálculo apresentado na ISO (ISO 13528, 2015);

- **Como valor obtido pelos dados da rodada de EP.** Nesse caso, a ISO (ISO 13528, 2015) recomenda algumas técnicas, como a aplicação do Algoritmo A, descrito no Anexo C da norma, e cuja aplicação está descrita no Apêndice A desse documento.

É recomendado que todo desvio padrão definido para a rodada de EP seja prático e aceito pelos participantes (ISO 13528, 2015). Para alguns ensaios, é necessário somente que o valor escolhido seja claro o suficiente para discriminar uma situação simples de aceitação ou reprovação (ISO 13528, 2015).

3.5.3.

Limites fixos empregados em ensaios de proficiência para laboratórios clínicos

O conceito de limites fixos surgiu na década de 1980 (Oliveira & Mendes, 2011), como uma alternativa ao emprego dos parâmetros do z-score (valor designado e desvio padrão para a proficiência), quando estes parâmetros são obtidos por consenso de laboratórios participantes. Essa prática foi adotada na área clínica devido a incoerências do emprego do desvio padrão amostral no cálculo do z-score. Era observado que laboratórios dentro de um intervalo clínico útil eram considerados inadequados, em casos de o desvio padrão amostral ser suficientemente pequeno (Ehrlmeyer & Laessig, 1988). Por outro lado, laboratórios que necessitavam de maior compatibilidade com as necessidades médicas não apresentavam melhoria, quando avaliados com um desvio padrão amostral suficientemente grande.

Limites fixos representam uma parcela do valor designado, resultando em um intervalo em torno do mesmo (Ehrlmeyer & Laessig, 1988). Embora esse conceito possa ser associado diretamente com o cálculo do z-score, a vantagem dele é não fazer distinção entre erros devido à inexatidão ou imprecisão dos dados, e não requerer suposição estatística sobre a distribuição dos dados. São usualmente definidos com base: (1) no TEa (“erro total analítico”, também conhecido como “limite de utilidade médica”) (Ehrlmeyer & Laessig, 1988); (2) no erro total derivado da variação biológica (Oliveira & Mendes, 2011); ou (3) no estudo da dispersão média acumulada dos sistemas analíticos. O emprego dessa alternativa ganha relevância entre especialistas, no segmento da saúde e na indústria de sistemas analíticos (Oliveira & Mendes, 2011).

Limites fixos são associados como sendo desvios padrão para a proficiência obtidos por definição de especialistas, ou por análise de rodadas anteriores do EP (ISO 13528, 2015). Esses podem ser adaptados ou revistos, conforme a necessidade do provedor e do programa de EP (Thompson et al., 2006).

Quando definidos por especialistas, esses limites irão variar. A CLIA (*Clinical Laboratory Improvement Amendments*, regida pelo FDA), o RCPA (*Royal College of Pathologists of Australasia*), entre outros institutos, regulam ou recomendam a aplicação desses limites, conforme a necessidade. A ANVISA (Agência Nacional de Vigilância Sanitária) propõe limites fixos, que podem ser empregados pelos provedores para orientar o processo de avaliação dentro do EP

(ANVISA, 2002). Segundo Oliveira & Mendes (2011), limites fixos podem ser considerados parte das especificações da qualidade.

3.6.

Crítérios recomendados para avaliação do desempenho em ensaio de proficiência

De acordo com a ISO (ABNT NBR ISO/IEC 17043, 2011), a ISO 13528 (2015) é o documento que descreve os métodos específicos para determinação dos valores designados e de suas respectivas incertezas, sendo tais métodos também apresentados por Thompson et al. (2006). A forma mais frequente de determinar o desempenho dos laboratórios participantes é padronizar seus resultados na forma de escores. As equações que conduzem a essa padronização correspondem aos critérios de avaliação de desempenho. É de responsabilidade do provedor definir um critério que seja empregado de forma única a todos os participantes.

Os critérios mais comumente empregados para a determinação do desempenho em EP, com suas respectivas equações, são a seguir listadas, numa ordem crescente do grau de transformação dos resultados dos participantes (ABNT NBR ISO/IEC 17043, 2011) e com detalhes acerca de sua aplicação (ISO 13528, 2015):

- Z-Score:

$$z - score = \frac{x - X}{\sigma_p} \quad (1)$$

Onde:

x – valor reportado pelo participante;

X – valor designado;

σ_p – desvio padrão para a proficiência.

- Z'-Score:

$$z' - score = \frac{x - X}{\sqrt{\sigma_p^2 + u_x^2}} \quad (2)$$

Onde:

x – valor reportado pelo participante;

X – valor designado (que não pode ser definido por consenso entre participantes);
 σ_p – desvio padrão para a proficiência;
 u_x – incerteza padrão do valor designado.

- Zeta-Score(ξ):

$$\xi = \frac{x - X}{\sqrt{u_x^2 + u_X^2}} \quad (3)$$

Onde:

x – valor reportado pelo participante;
 X – valor designado (que não pode ser definido por consenso entre participantes);
 u_x – incerteza padrão do resultado do participante;
 u_X – incerteza padrão do valor designado.

- Número E_n :

$$E_n = \frac{x - X}{\sqrt{U_x^2 + U_X^2}} \quad (4)$$

Onde:

x – valor reportado pelo participante;
 X – valor designado por um laboratório de referência;
 U_x – incerteza expandida do resultado do participante;
 U_X – incerteza expandida do valor designado.

Para os critérios definidos nas eq. (1-3):

$ z \leq 2$	Desempenho satisfatório do laboratório;
$2 < z \leq 3$	Desempenho questionável do laboratório (como resultado de EP, geraria um sinal de alerta para o laboratório participante);
$ z > 3$	Desempenho insatisfatório do laboratório (como resultado de EP, geraria um sinal de necessidade de ação por parte do laboratório participante).

Para o critério definido na eq. (4):

$ E_n \leq 1$	Desempenho satisfatório do laboratório, não gera sinal de ação;
$ E_n > 1$	Desempenho insatisfatório do laboratório, gera sinal de ação.

3.7.

Análise Clínica da Concentração de glicose segundo métodos analíticos

A glicose é um dos mais analitos mais comumente empregados na investigação de patologias clínicas (Moodley et al., 2015). Foi somente no século XVIII que se associou o diabetes como sendo causada por hiperglicemia (Moodley et al., 2015). No século seguinte, foram desenvolvidas novas técnicas, que permitiam identificar quantidade de açúcar (rico em glicose) na urina, no líquido e no suor, empregando diversas reações químicas.

Segundo a Mbiolog (Mbiolog, 2017), que produz o kit Cepa, a glicose é formada pela digestão de carboidratos e pela conversão do glicogênio nas células hepáticas. Os dois hormônios que regulam os níveis de glicemia são o glucagon e a insulina. O glucagon acelera a liberação da glicose do glicogênio, aumentando a glicemia, enquanto a insulina aumenta a permeabilidade e o transporte das células com glicose, estimulando a formação de glicogênio, resultando na diminuição dos níveis séricos de glicose. Segundo a Mbiolog (Mbiolog, 2017), a Associação Americana de Diabetes informa ser necessária uma glicemia em jejum igual ou superior a 126 mg/dL. Para averiguação de diabetes gestacional, deve ser realizado exame específico entre a 24ª e 28ª semana de gestação.

Foi identificada em 1928 a primeira aplicação de métodos enzimáticos para dosagem de glicose (Moodley et al., 2015). Desde então, são os métodos enzimáticos os mais empregados para detecção de Glicose e outras substâncias no corpo, como Colesterol e Triglicerídeos. Dois métodos, em resumo, são destacados (Moodley et al., 2015): a Oxidase, com reação da Peroxidase, e a Hexoquinase, com reação de Desidrogenase.

A Oxidase catalisa a oxidação da glicose para ácido glucônico e peróxido de hidrogênio (H_2O_2), na presença de peroxidase adicionada (figura 4). O H_2O_2 reage com fenol e antipirina, gerando um produto colorido (beta-D-Glucose), que é medido por um espectrofotômetro, e cuja conversão resulta na quantidade de glicose na amostra.

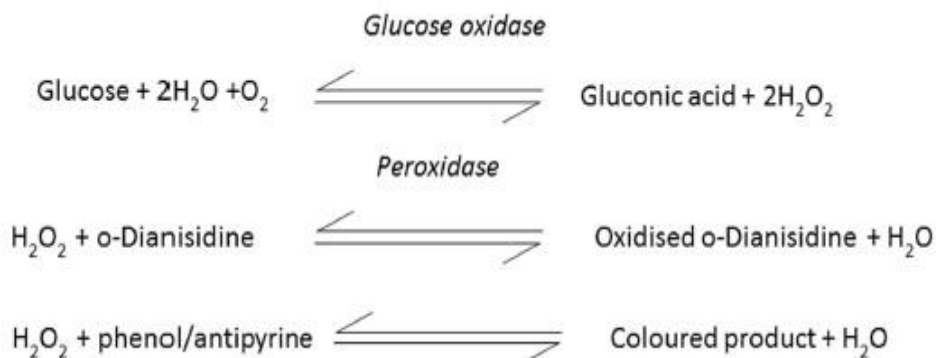


Figura 4 – Reação de Oxidase com Peroxidase.

A Hexoquinase catalisa a fosforilação da glicose por trifosfato de adenosina (ATP), formando Glucose-6-fosfato (Figura 5). A glicose-6-fosfato é oxidada pela Glucose-6-fosfato-desidrogenase (G6PD), na presença de nicotinamida-adenina-dinucleotídeo (NADP+) ou da nicotinamida-adenina-dinucleotídeo (NAD+). É produzida assim a nicotinamida-adenina-dinucleotídeo reduzida (NADPH), que é medida por um espectrofotômetro, cuja medida será proporcional à quantidade de glicose na amostra.

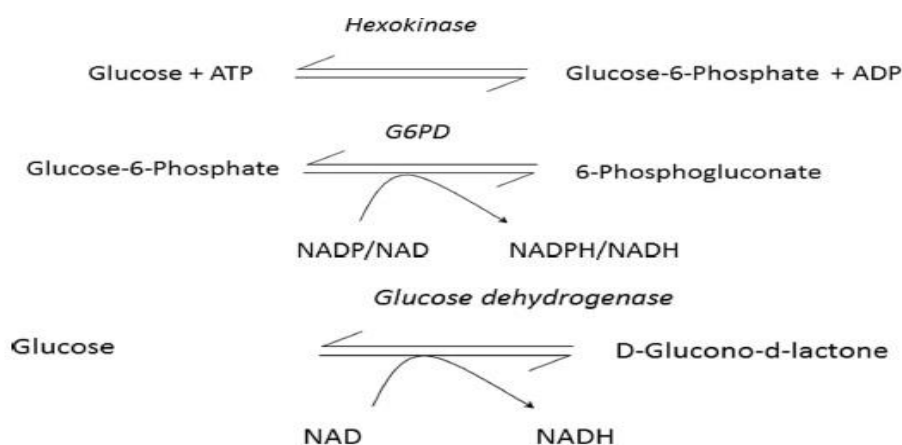


Figura 5 – Reação de Hexoquinase com Desidrogenase.

Os sistemas analíticos para medição da concentração em glicose dependerão dessas duas combinações de métodos enzimáticos. No programa de EP da Control Lab, somente esses dois métodos (Oxidase/Peroxidase e Hexoquinase/Desidrogenase) são atualmente reportados pelos participantes, frente a uma gama variada de sistemas disponíveis no mercado.

4

Materiais e métodos

Neste capítulo são apresentados as bases de dados e os métodos utilizados no desenvolvimento do presente estudo, que realiza uma análise comparativa de diferentes abordagens para avaliação de desempenho de laboratórios clínicos por meio de ensaio de proficiência. Foram empregados critérios recomendados pela ISO (ISO 13528, 2015) e pela IUPAC (Thompson et al., 2006) amplamente utilizados, e também novos critérios, propostos no âmbito deste trabalho. A análise comparativa de impacto dos diferentes critérios empregados para a classificação dos laboratórios foi realizada por diversas abordagens, descritas ao longo das sessões desse capítulo.

4.1.

Base de dados utilizada

Para a análise de desempenho dos critérios de avaliação foram utilizados resultados de medição obtidos e reportados por participantes do programa de EP realizado pela Control Lab, que é um provedor acreditado pelo INMETRO. Os resultados são reportados ao longo de rodadas periódicas, onde a cada rodada são distribuídos itens correspondentes a amostras de soro liofilizado, com diferentes concentrações de glicose. Para cada item de rodada enviado, os laboratórios participantes realizaram uma medição do analito em sua rotina laboratorial, cujo resultado obtido foi reportado na unidade estabelecida pelo provedor (mg/dL), de modo que cada participante neste estudo teve no máximo um número de respostas igual ao número de rodadas empregadas, multiplicado pelo número de itens por rodada. As rodadas para avaliação da proficiência em glicose, cujos dados foram investigados no presente trabalho, são de periodicidade trimestral, e foram realizadas no período de Janeiro de 2015 a Outubro de 2015, com três itens de concentrações de glicose em soro liofilizado por rodada, totalizando 12 itens enviados aos participantes do EP.

4.2.

Critérios empregados para determinação do desempenho analítico dos laboratórios

Com o objetivo de avaliar o impacto dos critérios de avaliação de desempenho dos participantes analisados, foram calculados escores padronizados por meio das equações associadas aos critérios. Tais equações contêm diferentes parâmetros, cuja aplicação determina o valor do escore para o resultado do participante. Um parâmetro de fundamental relevância é o valor designado, que dependendo do critério adotado, pode ser obtido pelo consenso entre participantes, ou com base no valor associado ao material de referência empregado pelo provedor.

Os critérios utilizados no trabalho para avaliação dos resultados reportados por laboratórios clínicos participantes do EP podem ser subdivididos em dois grupos:

- Critérios recomendados: critérios presentes em normas, regulamentos, diretrizes e literatura científica, sendo estes de amplo emprego na determinação do desempenho de laboratórios participantes;
- Novos critérios propostos: critérios que realizam uma alteração na aplicação dos parâmetros, com o objetivo de minimizar ou eliminar possíveis distorções observadas nos critérios recomendados.

A seguir são apresentadas as equações referentes aos diferentes critérios de desempenho, considerando os dois grupos mencionados, e a definição do conjunto de parâmetros que compõem as fórmulas de cada indicador:

4.2.1.

Apresentação matemática dos critérios sob teste

Grupo 1: Critérios recomendados em documentos normativos, regulatórios, diretrizes ou descritos na literatura científica (ISO 13528, 2015; Thompson et. al, 2006).

- Z-Score:

$$z - score = \frac{x - X_{cons}}{\sigma_{cons}} \quad (1)$$

Onde:

x – valor reportado pelo participante;

X_{cons} – valor designado por consenso de participantes;

σ_{cons} – desvio padrão para a proficiência, obtido por meio de consenso de participantes;

O critério **Z-Score** busca identificar possível viés do laboratório, calculado com base no valor designado e no desvio padrão, ambos obtidos por meio de consenso entre participantes do EP.

- **Z_{LF}-Score:**

$$z_{LF} - score = \frac{x - X_{cons}}{\sigma_{consLF}} \quad (2)$$

Onde:

x – valor reportado pelo participante;

X_{cons} – valor designado por consenso de participantes;

σ_{consLF} – é o desvio padrão para a proficiência, obtido por meio de consenso de participantes, considerando o limite fixo em sua obtenção ($\sigma_{consLF} = X_{cons} * (LF/2)$).

O critério **Z_{LF}-Score** tem a mesma natureza do **Z-Score**, mas emprega o conceito de limites fixos (LF) (Ehrmeyer & Laessig, 1988), também conceituados como limites por meio de percepção de especialistas (ISO 13528, 2015) ou legalmente definidos (Thompson et al., 2006), para a definição do desvio padrão para a proficiência. Eventualmente, existirá diferença significativa entre ambas as estimativas de desvio padrão para a proficiência (ou seja, $\sigma_{cons} \neq \sigma_{consLF}$). No presente trabalho, para o EP associado à determinação da concentração de glicose em soro liofilizado, foi adotado o limite fixo (LF) de 13 %, recomendado pelo Anexo B do Procedimento GGLAS 02/43 (ANVISA, 2002).

- **Z'-Score:**

$$z' - score = \frac{x - X_{ref}}{\sqrt{\sigma_{cons}^2 + u_{X_{ref}}^2}} \quad (3)$$

Onde:

x – valor reportado pelo participante;

X_{ref} – valor designado pela referência fornecida pelo provedor;

σ_{cons} – desvio padrão para a proficiência, obtido por meio de consenso de participantes;

u_{Xref} – é a incerteza padrão do valor designado por meio de referência.

A ISO (ISO 13528, 2015) não informa que o valor designado deve ser X_{ref} , mas declara que o valor designado, em caso de aplicação do **Z'-Score**, não pode ser o valor designado por consenso de participantes. No presente trabalho, X_{ref} foi designado com base no material de referência empregado pelo provedor, o mesmo se aplicando a u_{Xref} .

- **Z'_{LF}-Score:**

$$z'_{LF} - score = \frac{x - X_{ref}}{\sqrt{\sigma_{consLF}^2 + u_{Xref}^2}} \quad (4)$$

Onde:

x – valor reportado pelo participante;

X_{ref} – valor designado pela referência fornecida pelo provedor;

σ_{consLF} – é o desvio padrão para a proficiência, obtido por meio de consenso de participantes, considerando o LF em sua obtenção ($\sigma_{consLF} = \sigma_{cons} * (LF/2)$);

u_{Xref} – é a incerteza padrão do valor designado por meio de referência.

Como em **Z_{LF}-Score**, o desvio padrão para a proficiência obtido por meio de consenso de participantes considera em seu cálculo o limite fixo empregado para o analito (Ehrmeyer & Laessig, 1988; ANVISA, 2002), substituindo σ_{cons} por σ_{consLF} .

Grupo 2: Critérios propostos no presente trabalho:

- **Z_{LF}-Score_(Ref):**

$$z_{LF} - score_{(Ref)} = \frac{x - X_{ref}}{\sigma_{refLF}} \quad (5)$$

Onde:

x – valor reportado pelo participante;

X_{ref} – valor designado pela referência fornecida pelo provedor;

σ_{refLF} – é o desvio padrão para a proficiência, obtido do valor designado por meio da referência fornecida pelo provedor, que considera o LF (Ehrmeyer & Laessig, 1988; ANVISA, 2002) em sua obtenção ($\sigma_{refLF} = X_{ref}^* (LF/2)$).

Esse critério é similar ao **Z_{LF} -Score**, mas substituindo X_{cons} por X_{ref} e σ_{consLF} por σ_{refLF} . Dessa forma, o critério não inclui nenhum parâmetro associado ao consenso entre participantes, tornando-o independente da qualidade dos dados do item de rodada.

- **Z'_{LF} -Score_(Ref)**:

$$z'_{LF} - score_{(Ref)} = \frac{x - X_{ref}}{\sqrt{\sigma_{refLF}^2 + u_{X_{ref}}^2}} \quad (6)$$

Onde:

x – valor reportado pelo participante;

X_{ref} – valor designado pela referência fornecida pelo provedor;

σ_{refLF} – é o desvio padrão para a proficiência, obtido do valor designado por meio da referência fornecida pelo provedor, que considera o LF em sua obtenção ($\sigma_{refLF} = X_{ref}^* (LF/2)$);

$u_{X_{ref}}$ – é a incerteza padrão do valor designado por meio de referência.

Esse critério é similar ao **Z'_{LF} -Score**, mas com a substituição do σ_{consLF} por σ_{refLF} , que emprega o LF (Ehrmeyer & Laessig, 1988; ANVISA, 2002), logo não inclui parâmetro associado ao consenso entre participantes, tornando-o igualmente independente da qualidade dos dados do item de rodada. Outro aspecto desse critério consiste na inclusão da incerteza do valor designado por meio da referência fornecida pelo provedor no denominador da expressão.

4.2.2.

Regra para determinação do desempenho para os critérios

Como os critérios propostos são adaptações dos critérios recomendados pela ISO (ISO 13528, 2015; ABNT NBR ISO/IEC 17043, 2011), para qualquer critério presente nesse trabalho, os escores (z) obtidos conforme os resultados declarados pelos participantes, dentro da estratégia formal de EP, são classificados da seguinte forma:

$ z \leq 2$	Desempenho satisfatório do laboratório;
$2 < z \leq 3$	Desempenho questionável do laboratório (como resultado de EP, geraria um sinal de alerta para o laboratório participante);
$ z > 3$	Desempenho insatisfatório do laboratório (como resultado de EP, geraria um sinal de necessidade de ação por parte do laboratório participante).

4.2.3.

Apresentação matemática dos critérios em forma de faixas de aceitação

Para a obtenção das faixas de limites aceitáveis de valores reportados pelos participantes como resultado da medição, adaptou-se as equações dos escores, de forma que sua apresentação fosse também realizada na forma de faixas de adequação. Dessa forma, em torno do valor designado, está estabelecido um erro, que corresponde ao denominador do escore multiplicado pelo valor 2 (que é o escore que discrimina os participantes de desempenho aceitável e não aceitável). Quando o erro é subtraído do valor designado, é obtido o limite inferior da faixa, e quando somado, é obtido o limite superior da faixa. Valores que extrapolem o valor inferior ou superior dessa faixa são declarados como inadequados, caso contrário, são declarados como adequados.

Grupo 1: Critérios recomendados:

$$R_{z-score} = [X_{cons} - 2\sigma_{cons}; X_{cons} + 2\sigma_{cons}] \quad (7)$$

$$R_{z_{LF}-score} = [X_{cons} - 2\sigma_{consLF}; X_{cons} + 2\sigma_{consLF}] \quad (8)$$

$$R_{z'-score} = [X_{ref} - 2\sqrt{\sigma_{cons}^2 + u_{X_{ref}}^2}; X_{ref} + 2\sqrt{\sigma_{cons}^2 + u_{X_{ref}}^2}] \quad (9)$$

$$R_{z'_{LF}-score} = [X_{ref} - 2\sqrt{\sigma_{consLF}^2 + u_{X_{ref}}^2}; X_{ref} + 2\sqrt{\sigma_{consLF}^2 + u_{X_{ref}}^2}] \quad (10)$$

Grupo 2: Critérios propostos no presente trabalho:

$$R_{z_{LF}-score_{(Ref)}} = [X_{ref} - 2\sigma_{refLF}; X_{ref} + 2\sigma_{refLF}] \quad (11)$$

$$R_{z'_{LF-score(Ref)}} = [X_{ref} - 2\sqrt{\sigma_{refLF}^2 + u_{X_{ref}}^2}; X_{ref} + 2\sqrt{\sigma_{refLF}^2 + u_{X_{ref}}^2}] \quad (12)$$

4.3.

Métodos para obtenção dos valores designados

Para a avaliação do desempenho dos laboratórios foram utilizados valores designados obtidos por duas vias: por meio do valor de referência fornecida pelo provedor (X_{ref}) e por consenso de laboratórios participantes das rodadas de EP (X_{cons}). Os valores designados por referência foram definidos com base nos dados empregados para averiguação de homogeneidade suficiente dos itens de rodada (ISO 13528, 2015). Para cada item de EP, o valor designado por referência (X_{ref}) e sua respectiva incerteza padrão ($u_{X_{ref}}$) foram calculados conforme as eq. (13) e eq.(14):

$$X_{ref} = \frac{\sum_{i=1}^{n_h} x_{ih}}{n_h} \quad (13)$$

$$u_{X_{ref}} = \frac{\sqrt{\frac{1}{n_h - 1} \left(\sum_{i=1}^{n_h} x_{ih}^2 - n_h X_{ref}^2 \right)}}{n_h} \quad (14)$$

Onde:

x_{ih} – é o i-ésimo valor da amostra obtida que é analisada para averiguação da homogeneidade suficiente;

n_h – é o tamanho da amostra onde foi efetuada a averiguação da homogeneidade suficiente.

Os valores designados por consenso de participantes foram calculados pela média robusta dos resultados reportados pelos laboratórios utilizando o Algoritmo A, descrito no Anexo C.3 da ISO (ISO 13528, 2015), e cuja aplicação no presente trabalho está descrita no Apêndice A desse documento. Essa abordagem foi proposta para contornar as limitações da média comum (extremamente influenciada por outliers) e da mediana (estatística não suficiente) empregadas como valor designado, sem proceder com a eliminação de outliers. Aplicando-se esse algoritmo, foram geradas médias robustas (μ_{rob}) e desvios

padrão robustos (σ_{rob}). O parâmetro μ_{rob} é uma estimativa para o valor designado por consenso, e σ_{rob} é uma estimativa do desvio padrão para a proficiência por consenso (ISO 13528, 2015). A norma resume a seguir os cálculos para o valor designado por consenso de participantes, e para sua respectiva incerteza:

$$X_{cons} = \hat{\mu}_{rob} \quad (15)$$

$$u_{X_{cons}} = 1.25 \frac{\hat{\sigma}_{rob}}{\sqrt{n}} \quad (16)$$

4.4.

Análise estatística dos parâmetros dos critérios de avaliação de desempenho

Conforme o item 5.3.1 da ISO (ISO 13528, 2015), as técnicas mais comuns para a o EP assumem que os dados sejam normalmente distribuídos, ou ao menos de caráter unimodal e de razoável simetria, sem contaminação pela presença de outliers (valores que aparentam não pertencer a distribuição dos dados). Dessa forma, não é frequentemente necessário investigar a normalidade dos dados, mas a simetria aproximada deve ser analisada (ISO 13528, 2015). Em caso de essa simetria não ser contemplada, a norma recomenda técnicas robustas resistentes a assimetria, e mesmo a necessidade de transformação dos dados.

Para investigar se é significativa a diferença entre o desvio padrão para a proficiência obtido pelo consenso de participantes, calculado diretamente utilizando o Algoritmo A da ISO (ISO 13528, 2015) ($\sigma_p = \sigma_{cons}$), e o desvio padrão para proficiência definido pelo limite fixo associado ao valor designado por consenso ($\sigma_p = \sigma_{consLF}$), foi utilizado o teste Qui-Quadrado para variância única (Bussab & Morettin, 2010), em que a variância amostral (s^2) equivale ao σ^2_{cons} , e a variância populacional sob H_{0v} (σ_0^2) equivale ao σ^2_{consLF} , sendo H_{0v} a hipótese nula formulada conforme o Apêndice B do presente documento, a um nível de significância de 5 %.

Por sua vez, para investigar se é significativa a diferença entre o valor designado por consenso (X_{cons}) e o valor designado por referência (X_{ref}), empregou-se o teste T para amostra única (Bussab & Morettin, 2010), sendo que a média amostral (\bar{y}) é o X_{cons} , e a média populacional sob H_{0m} (μ_0) é o X_{ref} , sendo H_{0m} a hipótese nula, também formulada conforme o mesmo Apêndice B do presente documento, a um nível de significância de 5 %.

Finalmente, para investigar as diferenças entre os resultados reportados segundo os métodos analíticos e equipamentos, foram empregados os testes não paramétricos de Kruskal-Wallis (Fonseca & Martins, 2010) e de Levene (Levene, 1960). O teste de Kruskal-Wallis permite investigar se há diferença entre as médias dos resultados dos métodos analíticos e equipamentos (quando há suspeita de violação das condições de normalidade, para mais de dois grupos de resultados) (Fonseca & Martins, 2010), enquanto o teste de Levene (Levene, 1960) permite investigar a diferença entre as variâncias (para mais de dois grupos de resultados). Quando identificadas quaisquer diferenças, há indícios da falta de harmonização dos resultados, bem como indícios de potenciais fontes de incoerência entre os critérios que empregam X_{cons} e X_{ref} .

4.5.

Análise gráfica do desempenho dos critérios de avaliação de desempenho dos laboratórios

Para analisar a adequação da classificação de desempenho dos laboratórios, com base em valores designados por consenso ou pela referência, foram elaborados histogramas dos escores obtidos, conforme a recomendação da ISO (ISO 13528, 2015). Para evitar distorções na apresentação gráfica, foram excluídos da representação os resultados cujo escore correspondente fosse menor que -6 ou maior que +6.

Alguns resultados reportados pelos participantes foram destacados para exemplificar as potenciais discrepâncias de classificação de desempenho entre os diferentes critérios de avaliação empregados. Esses resultados foram apresentados na forma de gráficos de dispersão, sobre os quais foram adicionadas as faixas de limites de aceitação correspondentes a cada um dos critérios indicados pelas eq.(8) a eq.(12). O objetivo foi identificar casos de inadequação na classificação do desempenho do laboratório. Considerou-se como padrão a classificação segundo critérios que geram escores através do emprego do valor designado por meio de referência (X_{ref}). Também foram inclusos resultados obtidos por laboratórios de desempenho satisfatório esperado, com as seguintes características:

- *Laboratório de referência*: foi definido como referência o laboratório de Soro Controle da Control Lab, que é acreditado pelo INMETRO como laboratório de ensaio, e que é participante do programa de EP;

- *Laboratório acreditado*: laboratório com acreditação pelo INMETRO para ensaios, e que também é participante do programa de EP da Control Lab;

- *Laboratório excelente*: laboratório participante do programa de EP da Control Lab, cujo histórico de desempenho no programa ao longo do período estudado se revelou satisfatório. Esse desempenho inclui uma adequação plena no período (isso é, 100 % de adequação em itens de EP para proficiência em glicose, ao longo do período estudado), e erro sistemático médio de até 2 %, em relação ao valor designado por consenso.

Finalmente, com o objetivo de analisar comparativamente o desempenho alcançado pelos diferentes sistemas analíticos (métodos e equipamentos empregados na medição), foram construídos boxplots dos resultados. Tais gráficos são construídos frente às faixas obtidas por eq.(1) e eq.(12), ou seja, pelos critérios **Z-Score** e $Z'_{LF-score}_{(Ref)}$. O objetivo é comparar esses resultados quando a avaliação é feita, utilizando o critério que emprega o consenso (**Z-Score**) e o critério que emprega a incerteza do valor designado por referência, sem a interferência de nenhum parâmetro obtido por consenso ($Z'_{LF-Score}_{(Ref)}$). Devido à enorme variedade de sistemas analíticos pouco representativos, e à presença de outliers, foram definidos dois procedimentos de filtro para essa análise:

1º - efetuou-se o teste bilateral de detecção de outliers de Grubbs, sendo eliminados todos os detectados por esse teste, a um nível de 5 % de significância. Sua aplicação está descrita no Apêndice B do presente documento;

2º - dentre os dados não eliminados pelo teste de Grubbs, foram selecionados aqueles cuja combinação de métodos e equipamentos apresentasse uma frequência mínima de 10 resultados, sendo esse o mínimo aceitável para proceder com o cálculo da incerteza do valor designado por consenso de participantes do EP, segundo a versão anterior da norma ISO 13528 (2005).

5 Resultados

Neste capítulo, são apresentados os resultados das análises comparativas de desempenho dos diferentes critérios apresentados no item 4.2, para a avaliação do desempenho dos laboratórios clínicos por meio de ensaio de proficiência. A base de dados utilizada, proveniente do programa de Proficiência Clínica da Control Lab, consistiu dos resultados de medição da concentração de glicose em soro liofilizado, reportados pelos laboratórios clínicos participantes de quatro rodadas realizadas ao longo de 2015, para três itens diferentes, com concentrações desconhecidas pelos participantes. O impacto na classificação dos laboratórios participantes do EP resultante das diferenças entre os parâmetros utilizados para cada critério foi realizada por meio de análises gráficas por histogramas dos escores obtidos pelos participantes (sessão 5.2), por desempenho do laboratório segundo as faixas de concentração associadas a cada critério estatístico utilizado na determinação dos escores (sessão 5.3). Na sessão 5.4, por sua vez, é realizado um estudo mais detalhado do desempenho dos laboratórios segundo o sistema analítico utilizado pelo participante.

As equações dos critérios para determinação do desempenho de laboratórios são compostas de parâmetros definidos por meio de consenso ou com base em uma referência fornecida pelo provedor de EP. Os diferentes critérios são divididos em dois grupos: os que são descritos na literatura normativa, regulatória ou científica (eq.(1) a eq.(4)), e os que são propostos neste trabalho (eq. (5) e eq.(6)).

5.1. Análise estatística dos parâmetros

Os riscos de discrepância entre resultados de desempenho podem ser previstos pela análise dos valores de cada parâmetro utilizado nas equações associadas a cada uma das abordagens estatísticas (critérios e faixas) analisadas no presente trabalho e descritas nas eq.(1) a eq.(12). Na Tabela 1, estão resumidos os parâmetros estimados para emprego nas equações dos critérios utilizados para a geração dos escores associados aos resultados de laboratórios

participantes do EP para a quantificação de glicose em soro liofilizado. São informados os valores do tamanho amostral N , correspondente ao número total de resultados reportados pelos laboratórios para o item de rodada; o valor designado obtido por consenso de participantes (X_{cons}); o desvio padrão para a proficiência obtido por consenso de participantes por cálculo direto (σ_{cons}); o desvio padrão para a proficiência obtido por consenso de participantes, considerando um limite de aceitação fixo de 13 % para glicose (ANVISA, 2002) (σ_{consLF}); o valor designado pela concentração de glicose da amostra de referência utilizada no EP pelo provedor (X_{ref}); e a incerteza padrão do valor designado por meio dessa referência (u_{Xref}). Nas duas últimas colunas, estão os p-valores associados aos testes T para amostra única (Bussab & Morettin, 2010), e Qui-quadrado para variância única (Bussab & Morettin, 2010), delimitados pelas hipóteses apresentados no capítulo 4. Uma diferença significativa entre X_{cons} e X_{ref} indica deslocamento na distribuição dos dados, sendo uma potencial fonte de inadequação na avaliação de desempenho do participante. Da mesma forma, ocorre para o caso de uma diferença entre σ_{cons} e σ_{LF} que seja significativa, a qual interfere na análise ponderada da distância entre o valor observado (resposta do participante) e o valor designado, resultando em escores diferentes para um mesmo participante.

Tabela 1 – Parâmetros estimados para o cálculo dos escores dos participantes, segundo os critérios propostos para avaliação de desempenho em EP, para medição de glicose.

Estatísticas									
Rodada	Item	N	X_{cons}	σ_{cons}	Limite Fixo	X_{ref}	u_{Xref}	p-valor ($H_0: X_{cons}=X_{ref}$)	p-valor ($H_0: \sigma_{cons}=\sigma_{LF}$)
jan/15	1	1761	66,09	4,57	13%	66,48	0,19	0,00	0,00
jan/15	2	1761	183,12	10,93	13%	184,88	0,35	0,00	0,00
jan/15	3	1761	393,22	25,24	13%	389,72	1,05	0,00	0,47
abr/15	1	1737	94,05	5,97	13%	91,08	0,49	0,00	0,17
abr/15	2	1733	301,82	18,76	13%	301,96	0,61	0,76	0,01
abr/15	3	1734	140,44	9,13	13%	143,33	0,34	0,00	0,99
jul/15	1	1815	275,42	16,86	13%	268,40	0,88	0,00	0,00
jul/15	2	1813	101,27	6,30	13%	97,98	0,25	0,00	0,01
jul/15	3	1813	172,33	10,72	13%	174,28	0,49	0,00	0,01
out/15	1	1740	93,83	6,10	13%	91,79	0,28	0,00	0,99
out/15	2	1738	228,93	12,71	13%	223,55	0,52	0,00	0,00
out/15	3	1737	354,80	21,69	13%	344,99	1,05	0,00	0,00

Em todos os itens de EP para glicose ao longo do período estudado, exceto o item 2 da rodada de abril de 2015, foi observada diferença estatística significativa entre X_{cons} e X_{ref} , ao nível de 5 %. Por outro lado, nesse item, houve diferença ao nível de 5 % entre os valores de σ_{cons} e σ_{consLF} , o mesmo se dando em mais sete dos 12 itens das 4 rodadas de EP estudadas. Dessa forma, no item 2 de Abr/15, é esperado que uma possível incoerência entre os resultados de desempenho segundo os diferentes critérios seja devida a magnitude do erro que define a faixa de aceitação, ou seja, da magnitude do denominador da equação do critério empregado, utilizando ou não o limite fixo. Ao longo dos itens de EP estudados, quando não houve diferença significativa entre σ_{cons} e σ_{consLF} , foi identificada diferença significativa entre X_{cons} e X_{ref} . Nesse caso, a incoerência de resultados ocorrerá entre os critérios que empregam valores de referência como valor designado (Z' -Score, Z_{LF} -Score_(Ref), Z'_{LF} -Score e Z'_{LF} -Score_(Ref)) e os critérios que utilizam valor designado por meio de consenso (Z -Score e Z_{LF} -Score).

5.2.

Análise gráfica dos escores obtidos por diferentes critérios para avaliação de desempenho do laboratório

Os histogramas das Figuras 6, 7 e 8 apresentam a distribuição dos escores obtidos pelos laboratórios na rodada de Abr/15, promovida pelo provedor de EP, exibidos para cada um dos seis protocolos estatísticos aplicados para avaliar a adequação dos resultados de concentração de glicose (em mg/dL) em soro liofilizado, reportados pelos participantes. Essa abordagem gráfica foi produzida pelo software SPSS 15.

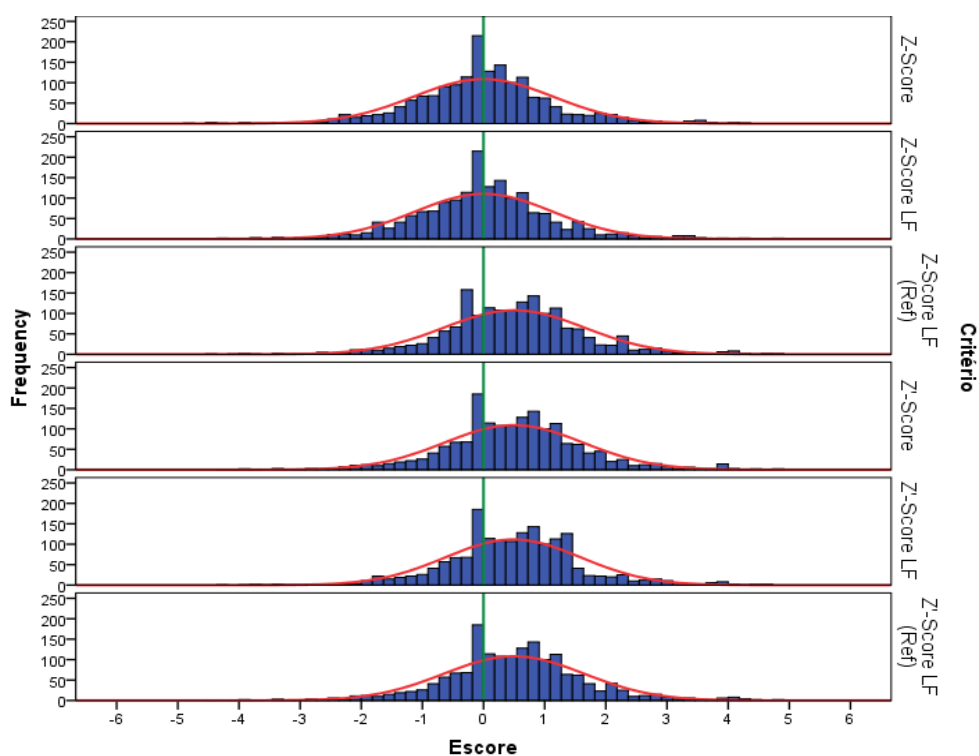


Figura 6 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 1, segundo os diferentes critérios estudados.

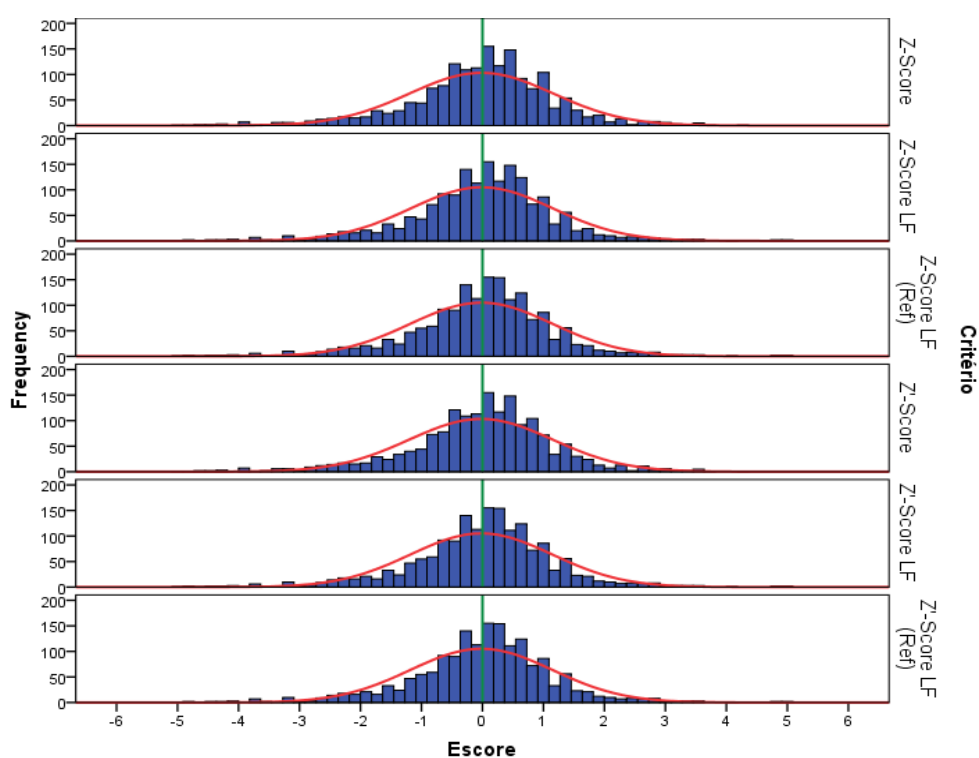


Figura 7 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 2, segundo os diferentes critérios estudados.

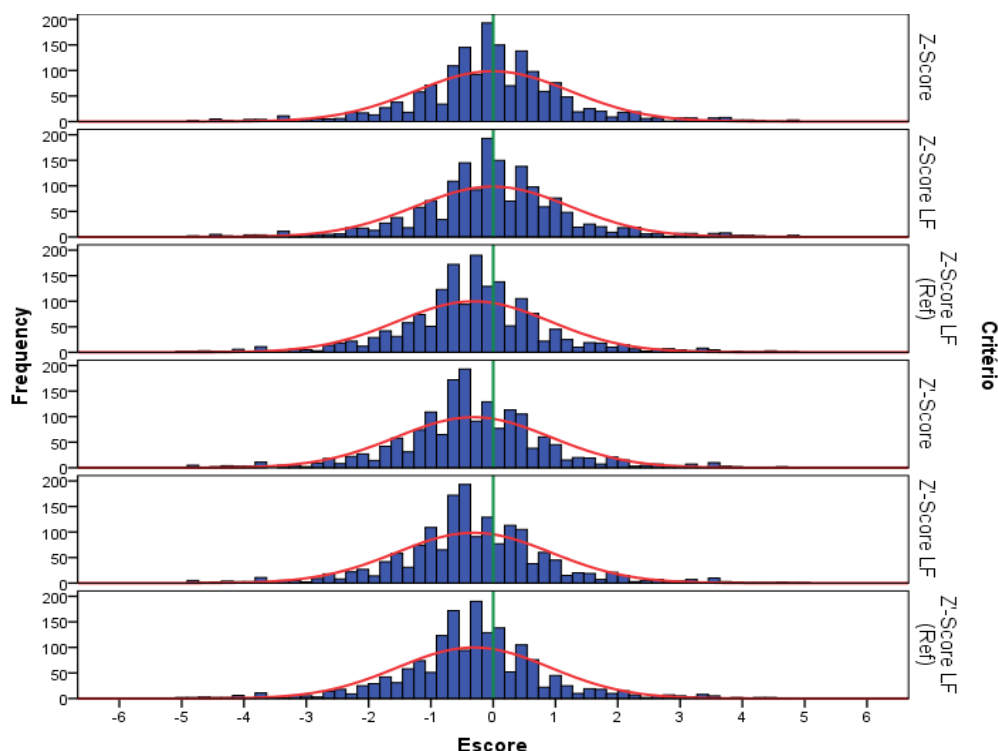


Figura 8 – Distribuição dos escores associados aos resultados dos participantes do EP para glicose, rodada Abr/15, item 3, segundo os diferentes critérios estudados.

Por meio das Figuras 6, 7 e 8, são observados os efeitos das discrepâncias entre valores designados e faixas de aceitação para os diferentes critérios empregados. Para qualquer item da rodada, ao utilizar um protocolo estatístico que aplica o valor designado por consenso (Z-Score ou Z_{LF} -Score), naturalmente os escores se tornam centralizados em torno do escore zero. Para o item 2 (Figura 7), todos os critérios aparentam estar numa mesma convergência ao centro da distribuição (histogramas com frequência simétrica em torno do escore zero), o que é corroborado pelo resultado apresentado pelo teste T apresentado na Tabela 1, que indica não ter sido observada uma diferença significativa entre X_{cons} e X_{ref} , ao nível de 5 %. Para os itens 1 e 3, os critérios baseados em valor designado por meio de referência (Z_{LF} -Score_(Ref), Z'-Score, Z'_{LF}-Score e Z'_{LF}-Score_(Ref)) apresentam deslocamento da distribuição dos escores, o que também é corroborado pela Tabela 1, onde foi observada diferença significativa entre X_{cons} e X_{ref} . Assim, é produzido um deslocamento dos escores, como visto nos histogramas das Figuras 6 e 8. No caso dos histogramas das Figuras 6 e 8, participantes que tenham escore próximo de zero em critérios que empreguem valor designado por referência são prejudicados se avaliados segundo o valor designado por consenso, representando casos de falsos negativos. Por outro lado, participantes que recebam escore próximo de zero em valores designados por consenso são

avaliados baixo desempenho em critérios que empreguem o valor por referência, sendo potenciais falso-aceitáveis quando avaliados por critérios baseados no consenso.

5.3.

Análise gráfica de desempenho do laboratório segundo as faixas de concentração associadas a cada critério estatístico utilizado para determinação dos escores

Uma forma prática de analisar os resultados de um participante no EP, sugerida pela ISO (ABNT NBR ISO/IEC 17043, 2011), é analisar a adequação dos resultados dos participantes frente a uma faixa de valores aceitáveis para o mensurando. A análise dos resultados dos participantes frente a uma faixa de aceitação definida com base nos parâmetros empregados, de acordo com o critério escolhido, permite identificar as incoerências de classificação do desempenho. As discrepâncias de classificação podem ser causadas tanto pela diferença entre os valores designados por consenso (X_{cons}) e por referência (X_{ref}), quanto pela amplitude da faixa de aceitação adotada, dependente dos parâmetros utilizados no denominador da equação do critério empregado.

Nas Figuras 9, 10 e 11 são apresentadas as faixas de adequação associadas a cada protocolo estatístico de escore, com base nos limites aceitáveis de concentração de glicose (mg/dL) correspondentes a cada abordagem. Nessas figuras são destacados os resultados da medição reportados por alguns dos laboratórios participantes do programa de EP do provedor acreditado, para os itens 1, 2 e 3 da rodada de Abr/15, respectivamente nas Figuras 9 (baixa concentração), 10 (alta concentração) e 11 (média concentração). Dentre esses resultados, foram incluídos os identificados pelas legendas A, B, C, D, E e F. Inclusos a estes, três laboratórios de maior destaque, considerados laboratórios de bom desempenho esperado no programa. São o laboratório de referência, o laboratório acreditado e o laboratório dito excelente. Conforme descrito no capítulo 4, o laboratório de referência corresponde ao laboratório de soro controle do provedor Control Lab, que participa das rodadas do EP; o laboratório acreditado possui acreditação pelo INMETRO como laboratório de ensaios; finalmente, o laboratório dito excelente possui histórico de resultados satisfatórios continuados ao longo do ano pesquisado, em rodadas de EP do provedor, para o mesmo analito. Essa abordagem gráfica foi produzida pelo software R.3.3.0.

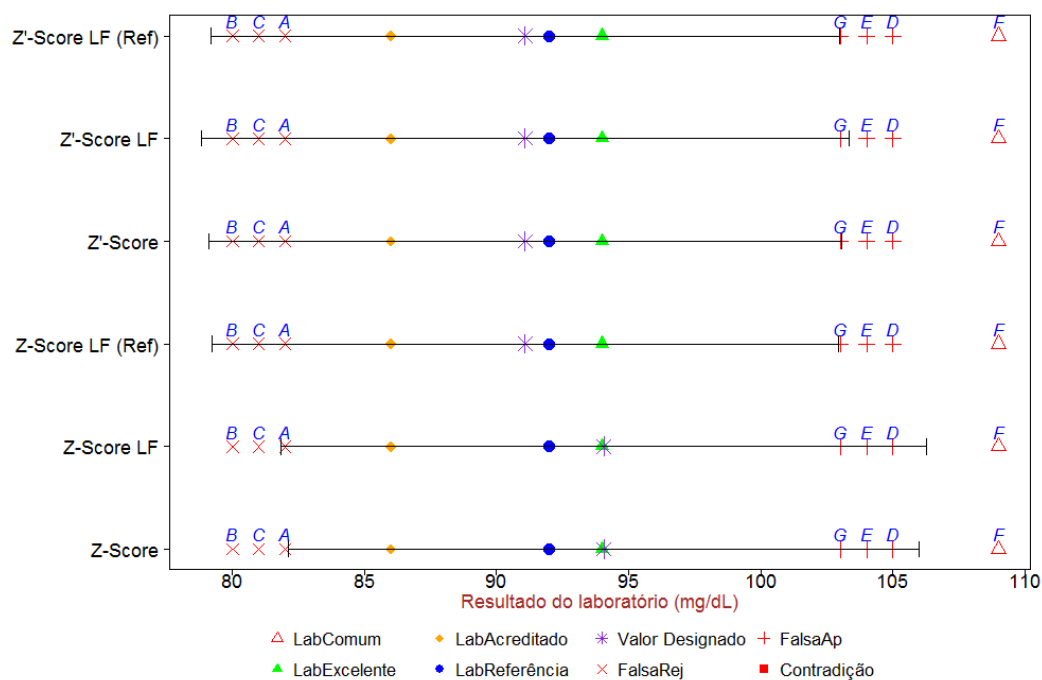


Figura 9 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 1, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.

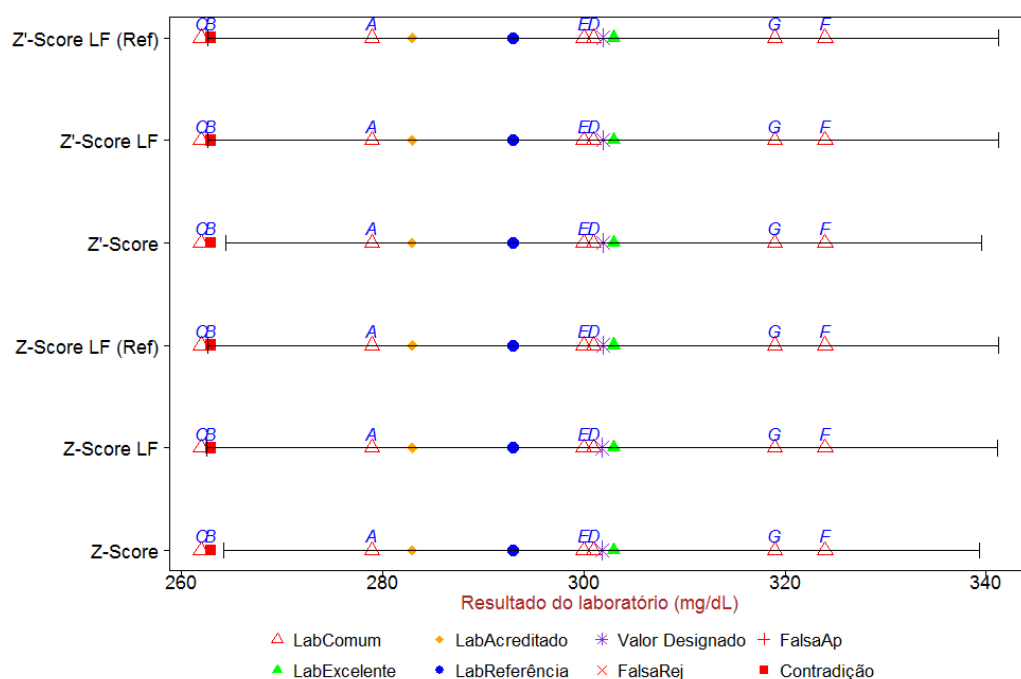


Figura 10 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 2, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.

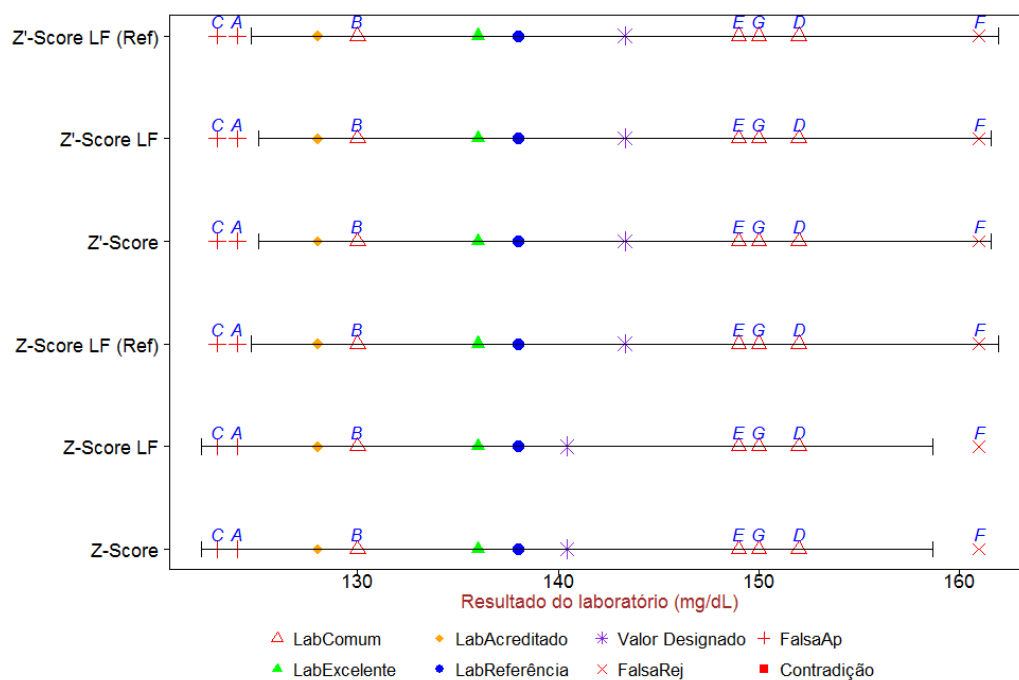


Figura 11 – Faixas de concentração aceitáveis para o EP em glicose, rodada Abr/15, item 3, segundo os protocolos estatísticos utilizados, frente aos resultados de laboratórios de bom desempenho esperado, além de resultados notáveis.

Por meio da análise das Figuras 9 a 11, apresentada para o item de baixa concentração, verifica-se que os laboratórios B e C foram declarados com desempenho inadequado considerando os critérios Z-Score e Z_{LF} -Score, que empregam valor designado pelo consenso entre participantes (X_{cons}); e foram declarados satisfatórios pelos demais critérios, os quais baseiam-se no valor designado por meio de referência (X_{ref}). O resultado caracteriza uma situação de reprovação indevida (falso-negativos). Por outro lado, os laboratórios D e E foram considerados de desempenho satisfatório pelos critérios Z-Score e Z_{LF} -Score (baseados em X_{cons}) e inadequados frente aos demais (baseados em X_{ref}). Assim, esses resultados caracterizam uma situação de adequação indevida (falso-positivos). No item de alta concentração, apresentado na Figura 10, não foi identificada diferença significativa entre o valor designado por consenso e o valor designado por referência. Dessa forma, somente uma discrepância foi identificada entre o Z-Score e o Z'-Score, envolvendo o laboratório B. Notar que B foi bem avaliado nos critérios Z_{LF} -Score e Z'_{LF} -Score, o que caracteriza o impacto do emprego do limite fixo, também indicado pelo teste de variância apresentado na Tabela 1. Finalmente, no item de concentração média, representado na Figura 11, também por motivos de discrepância entre X_{cons} e X_{ref} , os laboratórios A, C e F

foram classificados erroneamente com desempenhos satisfatório, satisfatório e reprovado, respectivamente.

5.4.

Análise gráfica e estatística de desempenho dos laboratórios segundo o sistema analítico utilizado

Nesta sessão é apresentada uma análise comparativa de desempenho obtido pelos diferentes sistemas analíticos (métodos e equipamentos) utilizados pelos participantes, para dosagem de glicose, da rodada de EP de Abr/15, realizada pela Control Lab. Tal análise se divide em duas partes: na primeira, é realizada uma análise estatística, que verificou a compatibilidade entre métodos e equipamentos, por meio dos testes estatísticos descritos no item 4.5. Na segunda, os resultados obtidos pelos diferentes sistemas analíticos foram confrontados graficamente com a faixa de concentração aceitável para classificação de desempenho satisfatório segundo o critério $Z'_{LF}\text{-Score}_{(Ref)}$, cujo cálculo é apresentado pela eq.(12), e que considera unicamente o emprego de parâmetros associados ao material de referência produzido. Para avaliar o impacto do emprego de parâmetros baseados no consenso entre participantes, foi também representada a faixa de concentração aceitável para classificação de desempenho satisfatório segundo o critério Z-Score, cujo cálculo está descrito na eq.(1).

Os valores designados e respectivas faixas de aceitação para os critérios Z-Score e $Z'_{LF}\text{-Score}_{(Ref)}$, para todos os itens de rodada, são apresentados na Tabela 2, onde os parâmetros para cálculo são oriundos da Tabela 1. São destacados, em especial, os resultados para a rodada de Abr/15.

Tabela 2 – Valores designados e respectivas faixas de avaliação, geradas pelos critérios Z-Score e $Z'_{LF-Score(Ref)}$, ao longo do período estudado (Jan/15 a Out/15). Em vermelho itálico sublinhado, os resultados correspondentes à rodada Abr/15.

Rodada.Item	Z-Score			$Z'_{LF-Score(Ref)}$		
	Limite Inferior	Valor Designado	Limite Superior	Limite Inferior	Valor Designado	Limite Superior
Jan15.1	56,94	66,09	75,24	57,83	66,48	75,13
Jan15.2	161,27	183,12	204,98	160,84	184,88	208,92
Jan15.3	342,74	393,22	443,71	339,01	389,72	440,43
<u><i>Abr15.1</i></u>	<u><i>82,11</i></u>	<u><i>94,05</i></u>	<u><i>105,99</i></u>	<u><i>79,20</i></u>	<u><i>91,08</i></u>	<u><i>102,96</i></u>
<u><i>Abr15.2</i></u>	<u><i>264,30</i></u>	<u><i>301,82</i></u>	<u><i>339,35</i></u>	<u><i>262,69</i></u>	<u><i>301,96</i></u>	<u><i>341,23</i></u>
<u><i>Abr15.3</i></u>	<u><i>122,19</i></u>	<u><i>140,44</i></u>	<u><i>158,69</i></u>	<u><i>124,68</i></u>	<u><i>143,33</i></u>	<u><i>161,98</i></u>
Jul15.1	241,69	275,42	309,14	233,46	268,40	303,34
Jul15.2	88,67	101,27	113,87	85,23	97,98	110,73
Jul15.3	150,88	172,33	193,77	151,60	174,28	196,96
Out15.1	81,64	93,83	106,03	79,84	91,79	103,74
Out15.2	203,51	228,93	254,36	194,47	223,55	252,63
Out15.3	311,43	354,80	398,18	300,09	344,99	389,89

Diferenças observáveis entre os diferentes tipos de métodos analíticos e equipamentos podem evidenciar a falta de harmonização de resultados, e potencializar incoerências no emprego dos critérios estudados. Para investigar tais diferenças, foram aplicados os testes de Kruskal-Wallis (para testar a diferença de médias entre os grupos de métodos analíticos) e de Levene (para testar a diferença de variâncias entre os grupos de métodos analíticos). Tais resultados foram submetidos preliminarmente ao teste bilateral de detecção de outliers de Grubbs, seguido pela seleção dos dados mais representativos, como apresentado no item 4.5, considerando as diferentes combinações de métodos.

A Tabela 3 apresenta os resultados da análise dos testes de Kruskal-Wallis (KW) e Levene objetivando investigar as diferenças entre os dois métodos analíticos, Oxidase/Peroxidase e Hexoquinase/Desidrogenase, empregados pelos participantes do EP para medição de glicose.

Tabela 3– Estatística de teste e respectivos p-valores obtidos pelos testes de Kruskal-Wallis (KW) e Levene, para comparação entre os métodos analíticos (Oxidase/Peroxidase e Hexoquinase / Desidrogenase), nos itens da rodada Abr/15.

Item	Teste	Estatística	p-valor
IT1	KW	19,07900	0,00001
	Levene	116,86000	0,00000
IT2	KW	48,76600	0,00000
	Levene	101,89000	0,00000
IT3	KW	0,10081	0,75090
	Levene	97,08300	0,00000

O teste de Levene foi significativo ao nível de 5 % para todos os itens da rodada, o que indica que há diferença entre as variâncias dos dois conjuntos de métodos analíticos. Essas diferenças impactam no desvio padrão para proficiência quando este é obtido por consenso (σ_{cons} ou σ_{consLF}). Se o critério não é independente, a qualidade dos dados disponíveis afetará a decisão do critério empregado, majorando ou minorando o desvio padrão para a proficiência. Por outro lado, o teste de Kruskal-Wallis indicou haver diferença significativa entre as médias dos métodos analíticos nos dois primeiros itens, mas ela foi não significativa no terceiro item. Nos dois primeiros itens é esperado que incoerências dentro do critério que emprega o consenso sejam causadas pelas diferenças entre as médias e variâncias dos métodos analíticos (Oxidase/Peroxidase ou Hexoquinase/Desidrogenase), enquanto no terceiro item, isso é apenas esperado pelas diferenças entre as variâncias. Além disso, a diferença observada no primeiro e segundo item evidencia a falta de harmonização entre resultados.

As Figuras 12, 13 e 14, correspondentes aos itens 1, 2 e 3 da rodada de Abr/15, respectivamente, apresentam uma comparação dos resultados reportados pelos participantes, segundo os métodos analíticos por eles informados (Oxidase/Peroxidase vs. Hexoquinase/Desidrogenase), na forma de boxplots. Esses resultados foram graficamente analisados frente às faixas de concentração que delimitam a aceitação dos resultados, de acordo com os critérios Z-Score (linhas em azul) e $Z'_{LF}\text{-Score}_{(Ref)}$ (linha vermelha e zona sombreada em cinza). Para os resultados apresentados nos gráficos, empregou-se os resultados correspondentes aos respectivos métodos e equipamentos, considerando o tratamento descrito no item 4.5. A abordagem gráfica foi realizada através do software R.3.3.0.

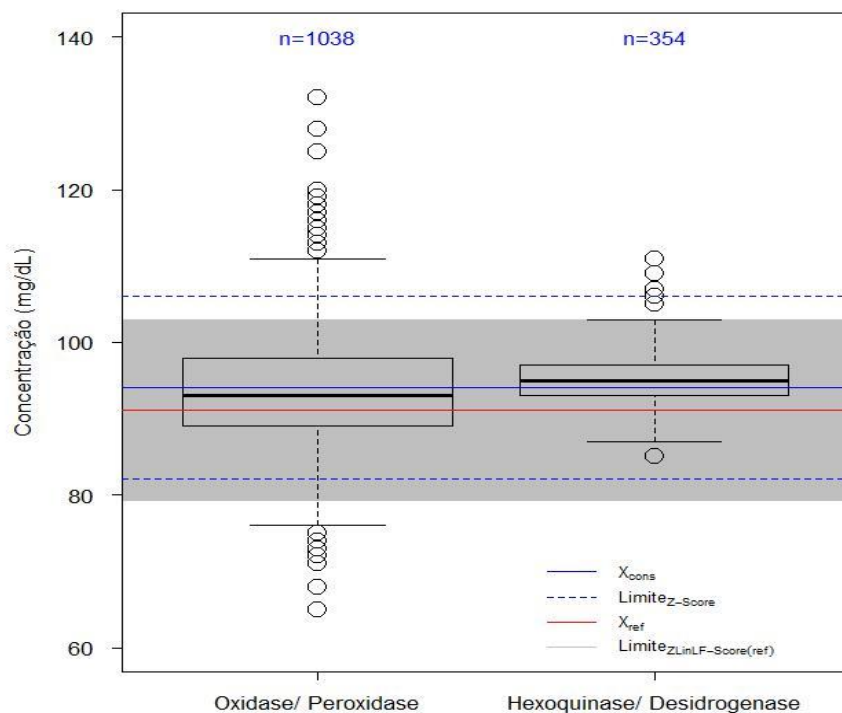


Figura 12– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, segundo os métodos analíticos.

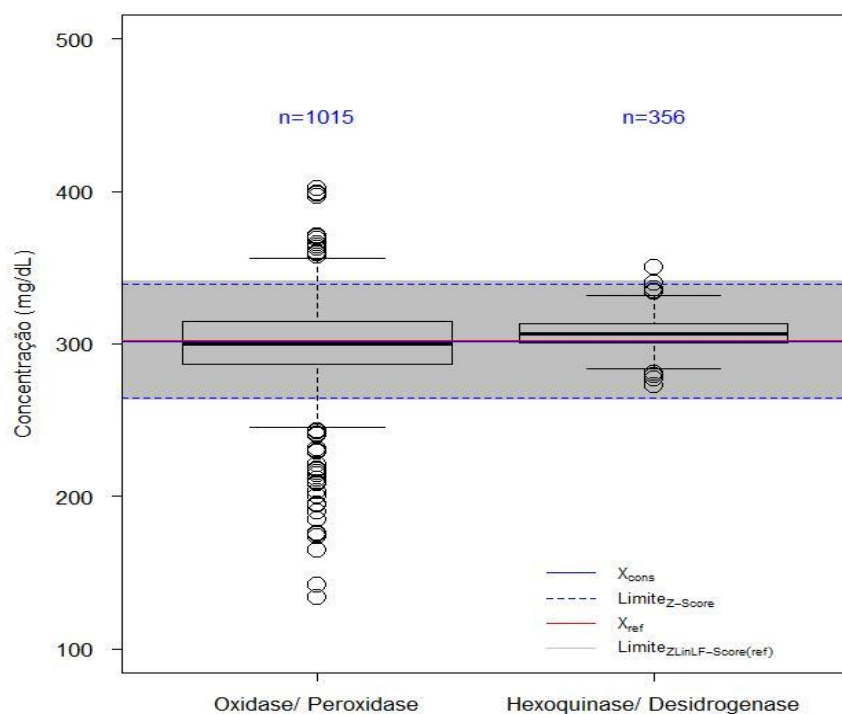


Figura 13 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, segundo os métodos analíticos.

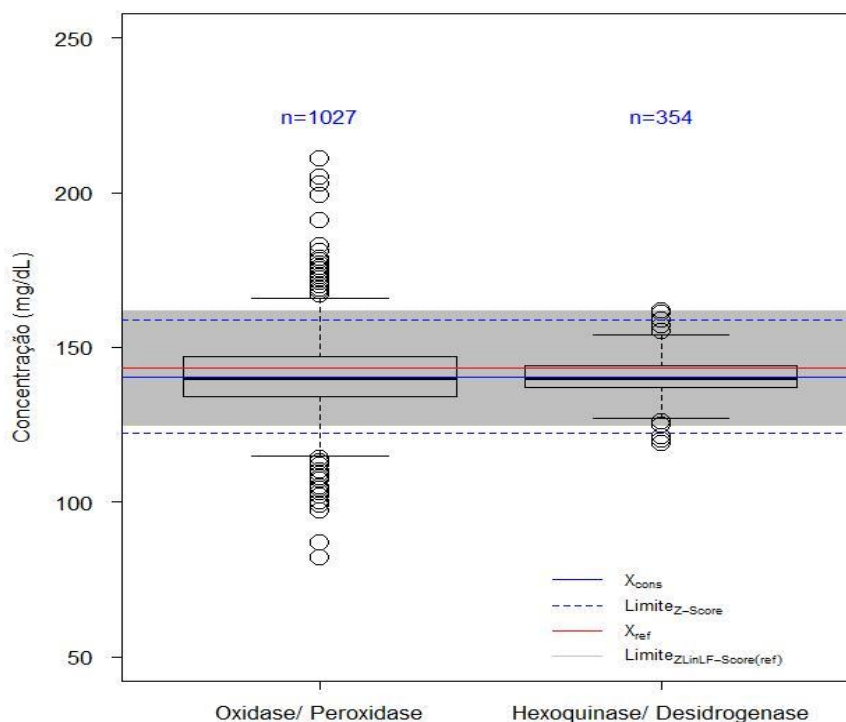


Figura 14 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, segundo os métodos analíticos.

Os resultados apresentados nas Figuras 12 a 14 indicam que o método utilizado pela maioria dos participantes da rodada de EP (Oxidase/Peroxidase) apresenta maior dispersão dos resultados em comparação ao método Hexoquinase/Desidrogenase, para todos os itens da rodada, excedendo os limites de desempenho satisfatório definido pela faixa cinza correspondente ao critério $Z'_{LF-Score(Ref)}$. Nas Figuras 13 e 14 (itens 2 e 3, respectivamente) observa-se uma assimetria da distribuição dos resultados obtidos pelo método Oxidase/Peroxidase em relação ao valor designado por referência. Por outro lado, o conjunto dos dados apresenta um valor da mediana próximo a esse valor, indicando menor erro sistemático, particularmente para o item 2, relativo à maior concentração. Os dados de Hexoquinase/Desidrogenase aparentam possuir menor erro aleatório, com a maior parte dos resultados atendendo à faixa aceitável (em cinza). Por outro lado, aparentam também apresentar erro sistemático maior quando comparados ao conjunto dos resultados obtidos pela Oxidase/Peroxidase.

As diferenças observáveis entre os tipos de método analítico podem não revelar as diferenças segundo os equipamentos utilizados na análise. Assim, para avaliação das diferenças de desempenho entre os equipamentos para determinação de glicose, segundo os métodos analíticos, foram novamente aplicados os testes de Kruskal-Wallis (para testar a diferença de médias entre os

grupos de equipamentos, segundo os métodos analíticos) e de Levene (para testar a diferença de variâncias entre os grupos de equipamentos, segundo os métodos analíticos). Os resultados obtidos estão dispostos na Tabela 4.

Tabela 4 – Estatística de teste e respectivos p-valores obtidos pelos testes de Kruskal-Wallis (KW) e Levene, para comparação entre equipamentos, segundo os métodos analíticos, nos itens da rodada Abr/15.

Item	Método Analítico	Teste	Estatística	p-valor
IT1	Oxidase/Peroxidase	KW	93,50200	0,00000
		Levene	5,91820	0,00000
	Hexoquinase/Desidrogenase	KW	35,95600	0,00017
		Levene	0,84900	0,57530
IT2	Oxidase/Peroxidase	KW	91,39800	0,00000
		Levene	9,98430	0,00000
	Hexoquinase/Desidrogenase	KW	42,70800	0,00001
		Levene	1,05420	0,39810
IT3	Oxidase/Peroxidase	KW	105,38000	0,00000
		Levene	8,51640	0,00000
	Hexoquinase/Desidrogenase	KW	89,12200	0,00000
		Levene	1,22790	0,26670

O teste de Levene foi significativo ao nível de 5 % para as variâncias das medições realizadas pelos equipamentos que utilizam o método Oxidase/Peroxidase, em todos os três itens da rodada (Tabela 4). Dessa forma, há diferença significativa entre as variâncias dos resultados obtidos pelos equipamentos, de forma que ao menos um equipamento possui variância diferente dos demais. Por outro lado, o mesmo teste revelou que não há diferença significativa entre as variâncias dos resultados obtidos pelos diferentes equipamentos que foram empregados com o método Hexoquinase/Desidrogenase (Tabela 4). O teste de Kruskal-Wallis indicou diferenças entre os resultados obtidos pelos equipamentos, independentemente do método analítico utilizado (Tabela 4). Considerando uma análise conjunta com os resultados da Tabela 3, nos itens 1 e 2 há diferença entre as médias dos resultados segundo métodos analíticos e equipamentos, mas no item 3, a diferença ocorre somente entre os equipamentos. Os métodos são diferentes entre si segundo a variância, mas na análise separada por método analítico, indicou homogeneidade de variâncias entre os resultados obtidos pelos diferentes equipamentos que empregam o método analítico Hexoquinase/Desidrogenase. Dessa forma, considerando as diferenças identificadas entre os resultados segundo a média ou variância, o emprego de parâmetros baseados no consenso

entre os resultados dos participantes de um EP pode impactar na avaliação fidedigna do desempenho dos laboratórios participantes.

As Figuras 15, 16 e 17 correspondem aos resultados dos participantes, após o tratamento descrito no item 4.5, segundo os itens 1, 2 e 3 da rodada de Abr/15, respectivamente. Esses resultados são apresentados na forma de boxplots, segundo o equipamento, para participantes que declararam empregar o método Oxidase/Peroxidase na medição de glicose.

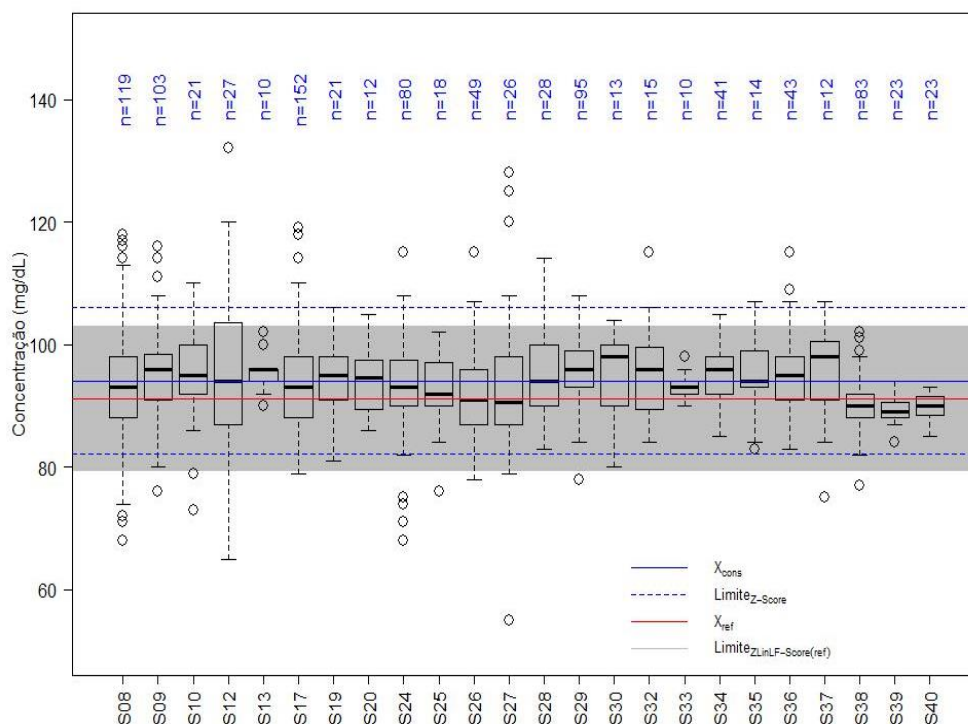


Figura 15 – Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, segundo equipamentos que empregam o método analítico Oxidase/Peroxidase.

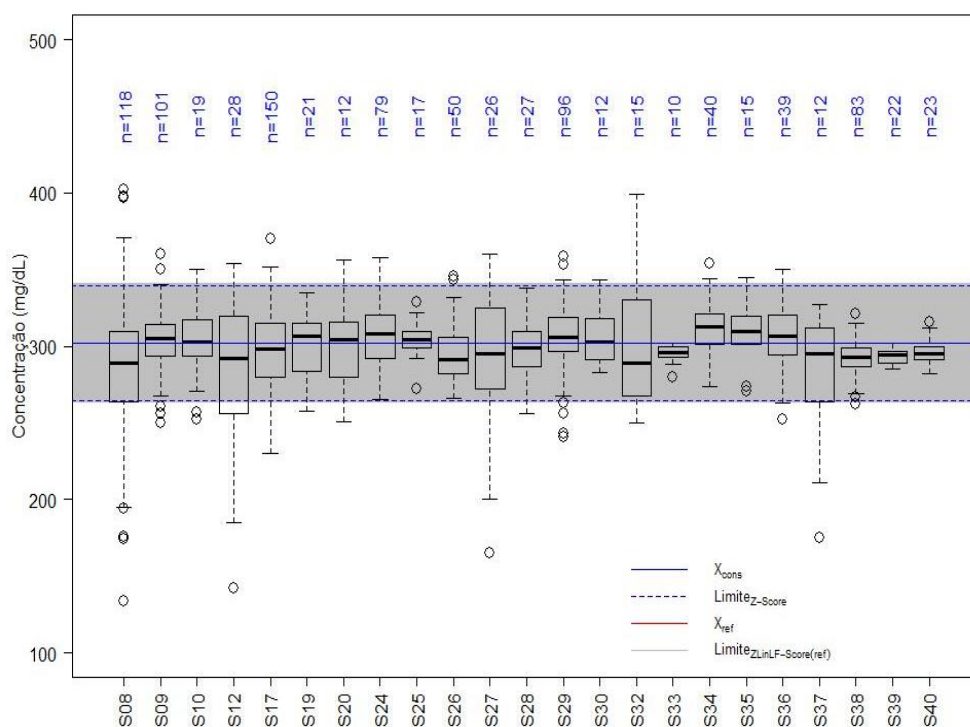


Figura 16– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, entre equipamentos que empregam o método analítico Oxidase/Peroxidase.

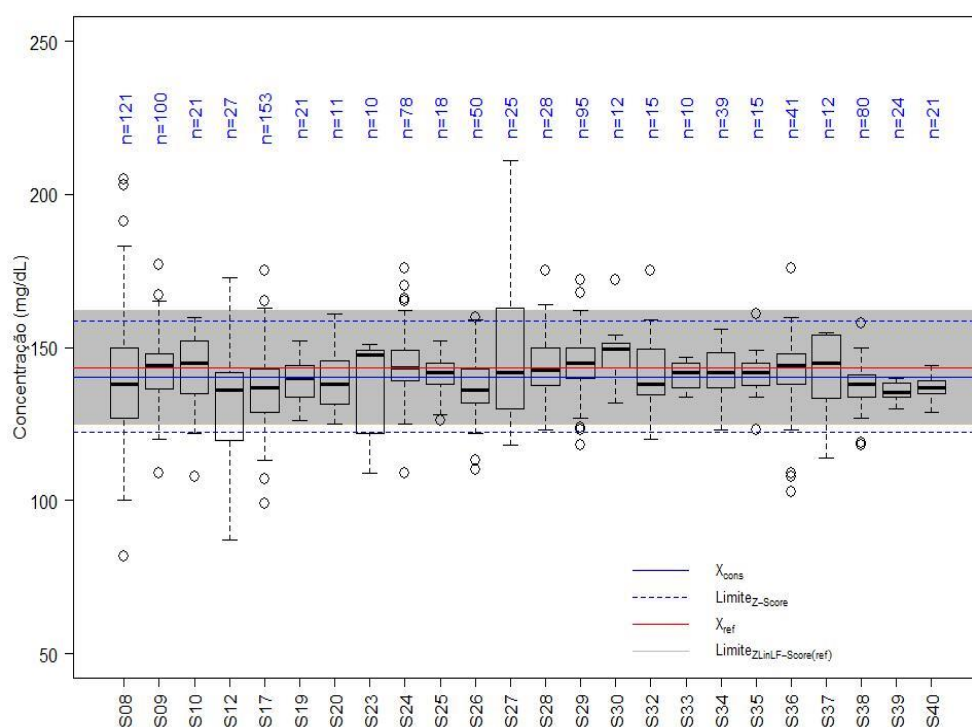


Figura 17– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, entre equipamentos que empregam o método analítico Oxidase/Peroxidase.

Os resultados avaliados para os equipamentos mais empregados pelos participantes que utilizam o método Oxidase/Peroxidase (S17, S08, S09 e S29) apresentaram diferenças de distribuição, entre si e em relação ao valor designado por meio de referência, mesmo empregando o método que gera X_{Ref} . Participantes avaliados por consenso em grupo de equipamentos de mediana diferente do valor designado por referência são potencialmente falsos aceitáveis, ao passo que participantes com resultado próximo ao valor designado por referência, mas inferior ao valor designado por consenso (se esse for maior que o de referência) ou superior a esse valor (se esse for menor que a referência) são potencialmente falsos não-aceitáveis. Notar que participantes cujos resultados se situaram entre a linha inferior pontilhada em azul e a parte inferior da faixa cinza representam participantes erradamente dados como inadequados, enquanto que participantes cujos resultados se situaram entre a linha pontilhada azul superior e a parte superior da faixa cinza são erradamente dados como adequados.

As Figuras 18, 19 e 20 correspondem aos resultados após o tratamento descrito no item 4.5, segundo os itens 1, 2 e 3 da rodada de Abr/15, sendo essa análise realizada entre os participantes que declaram empregar equipamentos que utilizam o método Hexoquinase/Desidrogenase. Novamente, esses resultados são apresentados na forma de boxplots.

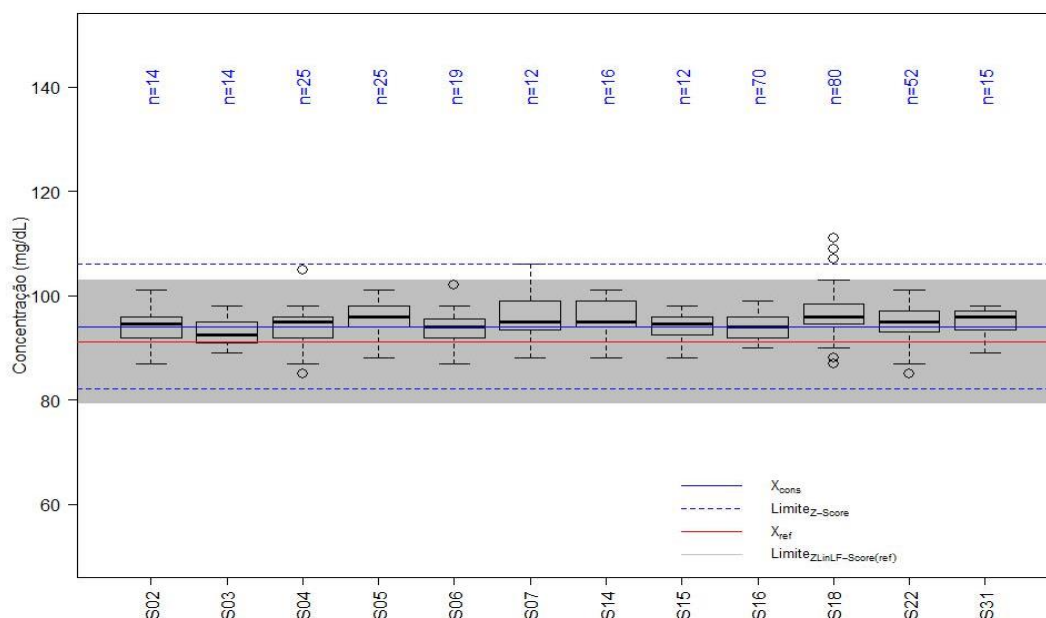


Figura 18– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 1, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.

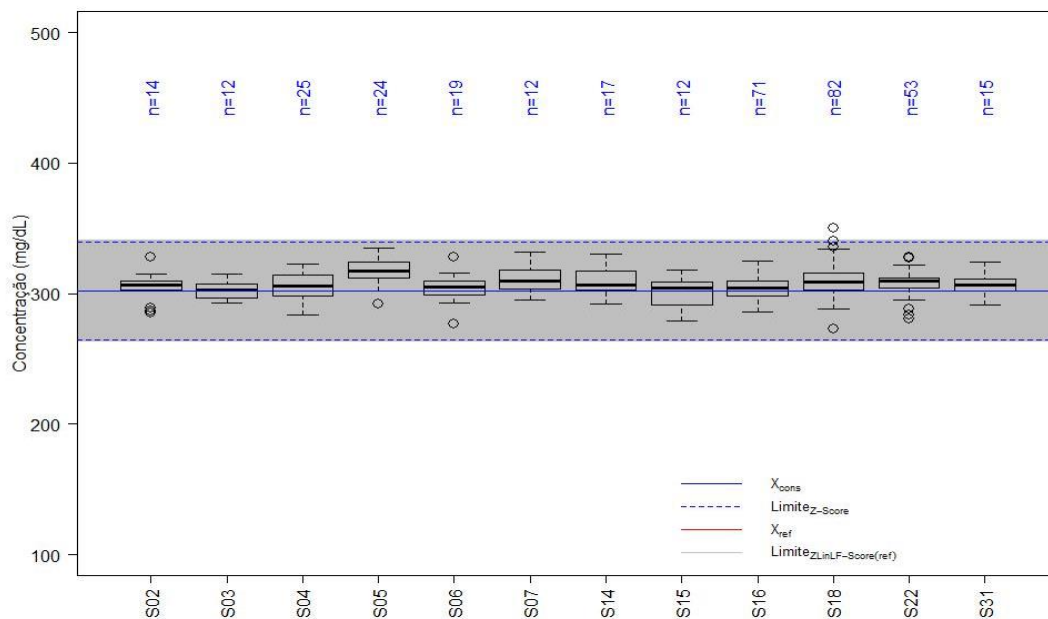


Figura 19– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 2, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.

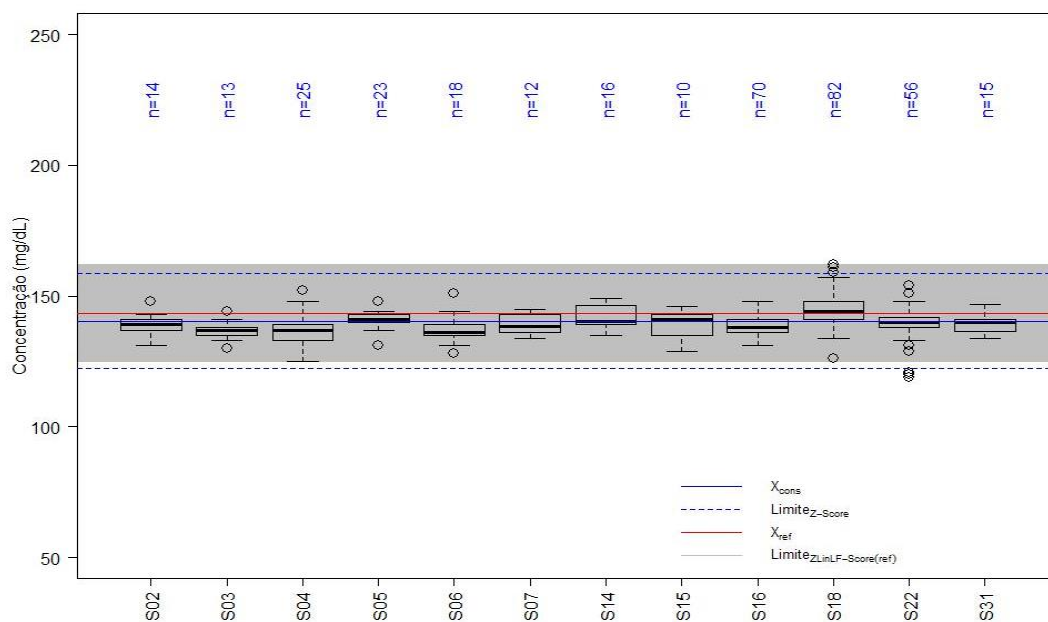


Figura 20– Distribuição dos resultados dos participantes do EP para dosagem em glicose, rodada Abr/15, item 3, entre equipamentos que empregam o método analítico Hexoquinase/Desidrogenase.

Os equipamentos que empregam a Hexoquinase/Desidrogenase apresentam uma evidência de erro sistemático maior, em comparação aos equipamentos que empregam o método Oxidase/Peroxidase (Figuras 15 a 17), quanto ao distanciamento da linha vermelha que representa o valor designado por referência (X_{ref}). Por outro lado, esses equipamentos possuem poucos resultados

que extrapolam a faixa cinza, o que indica um erro aleatório menor para equipamentos que empreguem esse método analítico. Novamente, participantes cujos resultados se situaram entre a linha inferior pontilhada em azul e a parte inferior da faixa cinza representam participantes erradamente dados como inadequados, e participantes cujo resultado se situaram entre a linha pontilhada azul superior e a parte superior da faixa cinza são erradamente dados como adequados.

A análise de resultados, segmentada por informações do sistema analítico, complementa as análises anteriores. Os resultados indicam que tanto para uma avaliação de desempenho baseada no consenso de todos os resultados dos participantes, quanto através de resultados segmentados por equipamentos, ocorre igual risco de classificação errônea dos participantes se não for baseada no valor do material de referência (X_{ref}) empregado no EP. Esse risco se torna mais potencial quando são identificadas diferenças na distribuição dos resultados dos participantes segmentados por sistema analítico, independentemente dessa diferença se manifestar entre as médias ou entre as variâncias.

6 Discussão

O presente trabalho apresenta uma análise comparativa de desempenho de diferentes critérios para avaliação de laboratórios clínicos, incluindo tanto aqueles que vêm sendo comumente empregados, quanto novos critérios propostos no âmbito desta pesquisa. As novas abordagens propostas buscam eliminar possíveis distorções na avaliação dos laboratórios que podem ser ocasionadas pelo efeito de parâmetros baseados nos valores associados ao consenso entre participantes. Dessa forma, os novos critérios de avaliação de desempenho de laboratório clínico são determinados unicamente pelo valor designado pelo material de referência do EP e por limites definidos por documentos que considerem aspectos da aplicação clínica, sem interferência da dispersão associada aos resultados dos participantes do programa.

Considerando que os critérios propostos são calculados com base apenas em parâmetros associados ao valor designado por referência, e que o critério apresentado na eq.(12) emprega a incerteza padrão do valor designado por meio do valor do material de referência utilizado no EP, esse critério foi definido como aquele que resultaria em classificações de desempenho mais justas dos laboratórios clínicos, e contra o qual as classificações resultantes dos outros critérios foi comparada.

Na análise por histogramas (Figura 6), percebe-se que, caso haja diferença entre valores designados, sendo o valor designado por consenso (X_{cons}) maior ou menor que o valor de referência (X_{ref}), há o risco potencial de que participantes com bom desempenho analítico sejam indevidamente avaliados, recebendo falsa classificação de inadequação no EP, assim como participantes com desempenho duvidoso podem potencialmente receber falsa aprovação no EP. Esse aspecto tem um caráter ainda mais crítico quando não há diferença significativa entre os denominadores dos critérios utilizados, pois a fonte de desvio que gera as falsas classificações de aprovação e reprovação estará associada exclusivamente à discrepância entre X_{cons} e X_{ref} , sendo esse último obtido do material utilizado pelo provedor para a realização do EP.

A análise por meio de faixas de aceitação de resultados para a concentração de glicose, definidas para cada um dos diferentes critérios

empregados, segundo os diferentes itens, revela haver clara possibilidade de erros de diversas fontes na determinação do desempenho do laboratório clínico quando utilizam-se critérios baseados que não consideram o valor do material de referência utilizado no EP. As faixas de aceitação podem ser deslocadas pela discrepância entre X_{cons} e X_{Ref} , como também minoradas ou majoradas conforme a abrangência de seus limites de erro aceitáveis, correspondendo aos limites de adequação do desempenho do laboratório (definidos pela magnitude do denominador da equação do critério). Observam-se resultados indevidos de classificação de laboratórios, tanto de aceitação, como o caso dos laboratórios E e D para o item 1 da rodada (Figura 9) e os laboratórios A e C para o item 3 (Figura 11), quanto inaceitação, como no caso dos laboratórios B e C para o item 1 e F para o item 3, que foram considerados com desempenho aceitável pelos critérios em efeito do consenso. Para o item de mais alta concentração, que não apresentou discrepância entre X_{cons} e X_{Ref} , foi evidenciada diferença de classificação apenas para o laboratório B, com base na abrangência maior das faixas de aceitação que utilizam o critério de limite fixo no denominador. Para os critérios baseados no limite fixo, o laboratório B foi considerado com um desempenho aceitável.

A análise dos resultados dos laboratórios fragmentada por sistemas analíticos utilizados pelos participantes da rodada de EP de Abr/15 para dosagem de glicose permitiu melhor caracterizar o desempenho desses sistemas separadamente, além de avaliar as possíveis inadequações de suas classificações (Figuras 12-20; Tabelas 3-4). Verificou-se que o método Oxidase/Peroxidase, o mais utilizado pelos participantes do EP, apresenta maior dispersão dos resultados em comparação ao outro método utilizado (Hexoquinase/Desidrogenase). O método Oxidase/Peroxidase evidencia ainda uma menor possibilidade de erro sistemático somente para o item 2, relativo à maior concentração (Figura 13). Os resultados obtidos pelo método Hexoquinase/Desidrogenase apresentam menor erro aleatório, com a maior parte dos resultados atendendo à faixa aceitável (em cinza), baseada no valor de referência, utilizando o critério proposto $Z'_{LF-Score(Ref)}$. Por outro lado, aparentam maior erro sistemático quando comparados ao conjunto dos resultados obtidos pela Oxidase/Peroxidase, que foi utilizada por número superior de participantes.

O estudo do desempenho com foco nos equipamentos utilizados pelos participantes indicou comportamentos muito diversificados, com sistemas analíticos que apresentaram elevada dispersão e possibilidade de erro sistemático para o usuário, observados especialmente para o método Oxidase/Peroxidase,

por exemplo o S12 (Figuras 15 a 17), e de forma não tão crítica para o método Hexoquinase/Desidrogenase, por exemplo o S18 (Figuras 18 a 20). Apesar de grande parte dos sistemas utilizando o método Oxidase/Peroxidase apresentarem maior dispersão, alguns se destacaram por muito baixa dispersão, como por exemplo S25, S33, S38, S39, S40, corroborado pelo teste de Levene, que indicou diferença entre as variâncias dos diferentes equipamentos (Tabela 4).

Eventualmente, considerando a variação das dispersões apresentadas pelos diferentes sistemas analíticos, uma análise de desempenho baseada no valor designado por consenso e realizada em separado para cada sistema pode ser realizada (Coucke et al., 2015). No entanto, os resultados obtidos no presente trabalho para a rodada analisada para medição de glicose indicam que os sistemas analíticos com dados de menor dispersão, ainda que apresentassem bias, o mesmo não foi suficiente para deslocar os resultados para além dos limites aceitáveis segundo o critério $Z'_{LF-Score(Ref)}$, reduzindo o risco de classificação errônea em uma avaliação por consenso. Por outro lado, os resultados do trabalho indicam que a análise segmentada do EP considerando os resultados de sistemas analíticos com maior dispersão conduziram a distorções de classificação, se avaliados com base no valor designado por consenso entre os laboratórios correspondentes.

Os resultados observados separadamente por sistema analítico permitem visualizar as discrepâncias mencionadas por Delatour et al.(2012), que indica que os resultados laboratoriais são fortemente dependentes do método escolhido para realizar um ensaio, podendo não serem comparáveis (Delatour et al., 2012). Esse fato é confirmado pelo teste de Kruskal-Wallis, que indicou haver diferença significativa entre os resultados obtidos por meio de diferentes métodos analíticos ou equipamentos (Tabela 4). O impacto da classificação de adequação de desempenho aceitável para um laboratório clínico tem repercussões na tomada de decisão do diagnóstico clínico.

Considerando os valores definidos para diagnóstico pré-clínico, apresentados pelo Consenso Brasileiro sobre Diabetes (Conselho Brasileiro sobre Diabetes, 2000), na Tabela 5 é apresentada uma análise da repercussão para o diagnóstico clínico consequente à faixa de limites aceitáveis de cada critério de desempenho do laboratório. Essa análise foi realizada com base nos limites para classificação de resultados laboratoriais dos participantes como aceitáveis, para o item 3 da rodada de EP de abril de 2015, com material de referência com a concentração de glicose de 143 mg/dL (X_{ref}). Segundo o Consenso sobre Diabetes, para uma medição da glicemia, medida em um paciente em jejum, com

valor de 143 mg/dL resultaria em um diagnóstico de Diabetes Mellitus. Se esse resultado, por outro lado, fosse obtido em uma medição de uma glicemia realizada 2 h após a ingestão de 75 g de glicose, levaria ao diagnóstico de uma tolerância diminuída à glicose. Na Tabela 5 é apresentada ainda a correspondência diagnóstica que seria obtida se forem considerados os resultados de um laboratório dentro dos limites das faixas aceitáveis, com base tanto no critério que emprega os parâmetros por meio de consenso entre participantes (Z-Score) quanto no critério proposto ($Z'_{LF}\text{-Score}_{(Ref)}$), cujo parâmetros são independentes do consenso.

Tabela 5 – Correspondência diagnóstica, segundo o Consenso Brasileiro sobre Diabetes, para os valores de glicose plasmática (em mg/dL) com relação às faixas de aceitação dos resultados reportados pelos laboratórios participantes (X_{ref} no valor de 143 mg/dL), considerando os critérios Z-Score e $Z'_{LF}\text{-Score}_{(Ref)}$.

Glicemia real medida em jejum		Valor (mg/dL)	Diagnóstico
Material de Referência (X_{ref})		143,33	Diabetes Mellitus
Z-Score	Limite Inferior	122,19	Glicemia de jejum alterada
	Valor Designado (X_{cons})	140,44	Diabetes Mellitus
	Limite Superior	158,69	Diabetes Mellitus
$Z'_{LF}\text{-Score}_{(Ref)}$	Limite Inferior	124,68	Glicemia de jejum alterada
	Valor Designado (X_{ref})	143,33	Diabetes Mellitus
	Limite Superior	161,98	Diabetes Mellitus
Glicemia real, 2 h após ingestão de 75 g de glicose			
Material de Referência (X_{ref})		143,33	Tolerância diminuída à glicose
Z-Score	Limite Inferior	122,19	Glicemia normal
	Valor Designado (X_{cons})	140,44	Glicemia normal
	Limite Superior	158,69	Tolerância diminuída à glicose
$Z'_{LF}\text{-Score}_{(Ref)}$	Limite Inferior	124,68	Glicemia normal
	Valor Designado (X_{ref})	143,33	Tolerância diminuída à glicose
	Limite Superior	161,98	Tolerância diminuída à glicose

Observa-se que, se considerarmos 143 mg/dL como o valor associado à glicemia de jejum, o que levaria ao diagnóstico de Diabetes Mellitus, os laboratórios que indicassem os resultados dessa medição no limite inferior da faixa de aceitação baseada tanto em critério definido pelo valor designado por consenso (X_{cons}) quanto por referência (X_{ref}) resultariam em diagnóstico diferente daquele correspondente à realidade. Para o caso de 143 mg/dL tratar-se da glicemia realizada 2 h após ingestão de 75 g de glicose, o diagnóstico correspondente seria de tolerância diminuída à glicose, o que está associado a uma conduta médica específica. Considerando a faixa aceitável do critério com valor designado por consenso, os laboratórios que fornecessem resultados em toda sua faixa de aceitação de limite inferior, incluindo o próprio valor designado (X_{cons}) levariam a um diagnóstico de glicemia normal, portanto diferente daquele correspondente à realidade. Para esse caso, o limite inferior de sua faixa de aceitação pelo critério $Z'_{LF-Score(Ref)}$ também apresentaria inadequação de diagnóstico.

Dessa forma, embora o emprego do valor designado por consenso possa resultar em distorções mais significativas, a faixa de resultados aceitáveis dos critérios podem ainda não estar adequadas às demandas clínicas, muito embora os limites aplicados no critério $Z'_{LF-Score(Ref)}$ tomem por base o valor fixo de 13 % do valor do material de referência (X_{ref}) recomendado pela ANVISA, que considera os interesses clínicos (ANVISA, 2002).

7

Conclusões e Trabalhos Futuros

7.1.

Conclusões

O desempenho de diferentes abordagens empregadas em ensaios de proficiência para avaliar a qualidade de laboratórios clínicos foi analisado de forma comparativa no presente trabalho. No estudo, para a avaliação dos laboratórios participantes de EP foram empregados não só critérios já descritos em diretrizes e literatura científica, mas também novas abordagens, propostas no âmbito do presente trabalho.

Os critérios propostos para avaliação de desempenho de laboratórios clínicos por ensaio de proficiência caracterizam-se por independerm da dispersão dos resultados dos participantes do programa, baseando-se no valor designado por meio do material de referência (X_{ref}) e em limites definidos como um percentual de X_{ref} , prescritos por diretrizes que considerem aspectos da aplicação clínica.

Os resultados da análise do desempenho da avaliação dos laboratórios indicaram discrepâncias significativas entre critérios cujo valor designado baseia-se no consenso entre os participantes do ensaio de proficiência (X_{cons}) quando comparado aos resultados obtidos pelo valor designado por meio do material de referência para EP (X_{ref}). As diversas avaliações realizadas por meio da análise de histogramas da distribuição de resultados do escore dos participantes para cada critério de avaliação do EP utilizado; das faixas aceitáveis para os resultados da medição da concentração de glicose associadas a cada um dos diferentes critérios; além da análise dos resultados de desempenho dos laboratórios fragmentada por sistemas analíticos utilizados pelos participantes da rodada de EP, indicaram importantes inadequações de classificação de desempenho. Assim, as avaliações de desempenho baseadas em critérios cujos parâmetros são fortemente afetados pelo consenso entre os valores reportados pelos participantes resultaram em classificações indevidas, tanto de aceitação de laboratórios quanto de rejeição.

O presente estudo permitiu caracterizar as discrepâncias de desempenho entre os sistemas analíticos empregados pelos participantes do EP para medição de glicose em soro liofilizado. Observou-se que embora a maior parte dos sistemas analíticos que empregaram o método Oxidase/Peroxidase, mais utilizado pelos participantes do EP, apresente maior dispersão dos resultados em comparação ao outro método utilizado (Hexoquinase/Desidrogenase), vários equipamentos utilizando o método Oxidase/Peroxidase destacaram-se pela reduzida dispersão de resultados. Todos os participantes que utilizaram esses últimos equipamentos mencionados foram classificados com desempenho aceitável. Por outro lado, a maior parte dos sistemas analíticos que empregaram o método Hexoquinase/Desidrogenase apresentaram resultados de reduzida dispersão, conduzindo à uma classificação de desempenho aceitável.

Considerando as demandas do diagnóstico clínico para glicose plasmática, baseadas no consenso publicado em 2000 pela Sociedade Brasileira de Diabetes, observou-se que as faixas aceitáveis de concentração de glicose abrangem valores que ultrapassam os limites que possibilitam o adequado diagnóstico clínico. Embora os resultados sejam mais críticos para critério de avaliação de desempenho baseado no consenso de participantes, a faixa isenta dessa influência, baseada em recomendação baseada em aspectos biológicos, também apresenta extensão além da desejável (Tabela 5). Esses resultados indicam a necessidade não só de mais estudos e discussão mais aprofundada sobre os parâmetros a serem definidos no denominador dos critérios empregados para o cálculo dos escores obtidos pelos laboratórios clínicos; mas também em relação à definição das faixas de valores de concentração do analito associadas aos diagnósticos clínicos, considerando a real condição dos sistemas analíticos disponíveis até o momento para medição pelos laboratórios clínicos.

Os resultados obtidos no presente trabalho contribuem para uma melhor compreensão da avaliação da qualidade analítica do laboratório clínico. O emprego dos novos critérios possibilita a classificação mais adequada de desempenho do laboratório clínico participante, propiciando a utilização dos resultados da avaliação de desempenho no EP para a melhoria do sistema analítico utilizado pelo participante, e para a maior confiabilidade de seus resultados de medição, com base nos quais se realizam os diagnósticos clínicos.

7.2.

Trabalhos Futuros

- Análise do erro sistemático, associado a sistema analíticos considerando consenso segmentado por sistemas ao longo do tempo (três anos, por exemplo);
- Estudo do impacto dos critérios considerando rodadas de EP com itens mais diretamente associados aos limites de concentração empregados para diagnóstico clínico;
- Proposta de realização de EP com valores de material de referência mais próximos dos limites de concentração associados às decisões do diagnóstico clínico e utilização dos resultados na proposta de novos critérios cujas faixas de concentração aceitáveis melhor atendam às demandas clínicas.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT ISO Guia 30 – Materiais de referência - Termos e definições selecionados**. 3ª Ed., 2015;

_____. **ABNT ISO Guia 31 – Materiais de Referência - Conteúdo de certificados e rótulos**. 3ª Ed., 2015;

_____. **ABNT ISO Guia 33 – Usos de Materiais de Referência Certificados**. 3ª Ed., 2015;

_____. **ABNT ISO Guia 34 – Requisitos gerais para a competência de produtores de material de referência**. 3ª Ed., 2012;

_____. **ABNT ISO Guia 35 – Materiais de referência - Princípios gerais e estatísticos para certificação**. 3ª Ed., 2012;

ABNT NBR ISO/IEC 17025 – **Requisitos gerais para a competência de laboratórios de ensaio e calibração**. 1ª Ed., 2005;

ABNT NBR ISO/IEC 17043 – **Avaliação da conformidade - Requisitos gerais para ensaios de proficiência**. 1ª Ed., 2011;

ABNT NBR NM ISO 15189. **Laboratórios de análises clínicas – Requisitos de qualidade e competência**. 3ª Ed., 2012.

ANVISA. **Guia para Qualidade em Química Analítica – Uma Assistência à Habilitação**. Séries Temáticas, v. 1, 2005;

_____. **Procedimento GGLAS nº. 02/43: Critérios para a Habilitação de Provedores de Ensaios de Proficiência**. 2ª Ed., Brasília, 2002;

_____. **Resolução RDC nº 302, de 13 de outubro de 2005. Dispõe sobre Regulamento Técnico para funcionamento de Laboratórios Clínicos**. Diário Oficial da União da República Federativa do Brasil, Brasília, 14 out. 2005;

BAIS, R et al. **Defining acceptable limits for the metrological traceability of specific measurands**. Clin. Chem. Lab. Med., 51(5), 973–979, 2013;

BONINI, P. et al. **Errors in Laboratory Medicine**. Clinical Chemistry, v. 48, nº 5, p. 691-698, 2002;

BUREAU INTERNATIONAL DES POIDS ET MESURES. **Consultative Committee for Amount of Substance: Metrology in Chemistry and Biology (Mission)**. Disponível em: <<http://www.bipm.org/en/committees/cc/ccqm>>. Acesso em: 28 abr. 17;

BUREAU INTERNATIONAL DES POIDS ET MESURES. **The BIPM key comparison database**. Disponível em: <<http://kcdb.bipm.org>>. Acesso em: 28 abr. 17;

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 6ª Ed. São Paulo: Saraiva, 2010;

CERRIOTTI, F. **The role of External Quality Assessment Schemes in Monitoring and Improving the Standardization Process**. Clinica Chimica Acta v. 432, p. 77–81, 2014;

CHAVES, C. D. **Controle de qualidade no laboratório de análises clínicas**. J Bras Patol Med Lab, v. 46, nº 5, out. 2010;

CONSELHO BRASILEIRO SOBRE DIABETES. **Diagnóstico e classificação do diabetes mellitus e tratamento do diabetes mellitus tipo 2**. Recomendações da Sociedade Brasileira de Diabetes, mai. 2000. Disponível em: <http://bvsmms.saude.gov.br/bvs/publicacoes/consenso_bras_diabetes.pdf>;

COUCKE, W. et al. **An Application of the Characteristic Function to Evaluate and Compare Analytical Variability in an External Quality Assessment Scheme for Serum Ethanol**. Clinical Chemistry, v. 61, nº 7, p. 948-954, 2015;

DE BIÉVRE, P. **Measurement results in a Proficiency Testing programme need to be evaluated against independent criteria**. Accred Qual Assur, 21:167-169, 2016;

DELATOUR, V. et al. **Continuous improvement of medical test reliability using reference methods and matrix-corrected target values in proficiency testing schemes: Application to Glucose assay**. Clinica Chimica Acta 413, 1872–1878, 2012.;

EC Directive 98/79/EC. **In vitro diagnostic medical devices**. 27 out. 1998;

EHRMEYER, S.S.; LAESSIG, R. H. **An Assessment of the Use of Fixed Limits to Characterize Intralaboratory Performance by Proficiency Testing**. Clinical Chemistry, v. 33, nº 10, p. 1901-1902, 1987;

EURACHEM. **Selection, Use and Interpretation of Proficiency Testing (PT) Schemes**. English edition, 2ª Ed., 2011;

_____. **The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics**. English edition, 2ª Ed., 2014;

FELDMAN, L. B. et al. **Hospital quality evolution history: from patterns to accreditation.** Acta Paul Enferm., 18(2):213-9, 2005;

FERRERO, C.; CASARIL, M. **Proficiency testing programs to improve traceability in chemical analysis.** Measurement, nº42, p.1502–1509, ago. 2009;

FRECKMANN, G. et al. **System Accuracy Evaluation of Different Blood Glucose Monitoring Systems Following ISO 15197:2013 by Using Two Different Comparison Methods.** Diabetes Technology & Therapeutics, V. 17, nº 9, 2015;

FONSECA, J.S; MARTINS, G.A. **Curso de Estatística.** 6ª Ed. São Paulo: Editora Atlas, 2010;

HUBER, Peter J. RONCHETTI, Elvezio M. **Robust Statistics.** 2.ed. John Wiley & Sons, Inc. ISBN: 978-0-470-12990-6. 2009;

INTERNATIONAL LABORATORY ACCREDITATION COOPERATION. **Policy on the Traceability of Measurement Results – P10:01,** 2013;

_____. **Policy for Participation in Proficiency Testing Activities - P9,** jun. 2014;

_____. **ABOUT ILAC.** Disponível em: <<http://ilac.org/about-ilac/>>. Acesso em: 28 abr. 17;

INSTITUTO NACIONAL DE METROLOGIA, QUALIDADE E TECNOLOGIA. **Vocabulário Internacional de Metrologia: Conceitos Fundamentais e Gerais e Termos Associados (VIM 2012).** 2013, 1ª Ed. Brasileira. Rio de Janeiro: Inmetro, 92 p;

_____. **Acreditação de Provedores de Ensaio de Proficiência.** Disponível em:<http://www.inmetro.gov.br/credenciamento/acre_prod_ep.asp>. Acesso em: 28 abr. 17;

_____. **DOQ-CGCRE-008 - Orientação sobre validação de métodos de ensaios químicos.** Rev.05, ago. 2016;

_____. **DOQ-CGCRE-016 - Orientação para a seleção e uso de materiais de referência.** Rev.02, fev. 2010;

_____. **Avaliação de dados de medição – Guia para a expressão da incerteza de medição.** JCGM 100:2008. Duque de Caxias/RJ, CICMA/SEPIN, 2012;

_____. **NIT-DICLA-026 - Requisitos sobre a participação dos laboratórios de ensaio e calibração em atividades de ensaio de proficiência.** Rev.10, jun. 2016;

ISO 13528 – **Statistical methods for use in proficiency testing by interlaboratory comparisons**. 2ª Ed., 2015;

ISO 5725-2 – **Accuracy (trueness and precision) of Measurement Methods and Results - Part 2: Basic method for the Determination of Repeatability and Reproducibility of a Standard Measurement Method**. 1ª Ed., 2004;

ISO 5725-4 – **Accuracy (trueness and precision) of Measurement Methods and Results - Part 4: Basic Methods for the Determination of the Trueness of a Standard Measurement Method**. 1ª Ed., 2004;

LAESSIG, R. H. et al. **Limitations of Proficiency Testing under CLIA '67**. Clinical Chemistry, v. 38, nº 7, p. 1237-1244, 1992;

LEVENE, H. "**Robust tests for equality of variances**". In Ingram Olkin; Harold Hotelling; et al. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling. Stanford University Press. pp. 278–292, 1960;

MBIOLOG DIAGNÓSTICOS. **Instruções de Uso -Glicose Cepa**. Disponível em: < http://www.mbiolog.com.br/produtos/Bula_Glicose_C.pdf>. Acesso em 28 abr. 2017;

MOODLEY, N. et al. **Historical perspectives in clinical pathology: a history of Glucose measurement**. J Clin Pathol, p. 1-7, 7 jan. 2015;

MONTEIRO, E. **Biometrologia: confiabilidade nas biomedicações e repercussões éticas**. Metrologia e Instrumentação, v. 6, p. 6-12. 2007;

MONTEIRO, E; LEON, L.F. **Metrological Reliability of Medical Devices**. Journal of Phys. Conf. Series 588 012032. 2015;

NIST – Engineering Statistics Handbook. **Grubbs' Test for Outliers**. Disponível em:<<http://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm>>. Acesso em: 28 abr. 17;

OLIVE, David J. **Applied Robust Statistics**. Department of Mathematics, Southern Illinois University, 23 June, 2008;

OLIVEIRA, C.A; MENDES, M.E. **Gestão da fase analítica do laboratório: como assegurar a qualidade na prática; v.2**. 1ª ed, Rio de Janeiro, ControlLab, 2011;

PESSOA, M.C.F.; FERREIRA JR., O.C. **Metrological traceability in clinical laboratory**. J Bras Patol Med Lab, v. 52, nº 3, p. 157-164, jun. 2016;

PLEBANI, M et al. **Quality Indicators for the Total Testing Process**. Clin Lab Med 37, 187–205, 2017;

REDE METROLÓGICA RS. **Materiais de Referência Certificados**. Disponível em: <<http://www.redemetrologica.com.br/materiais-de-referencia-certificados>>. Acesso em: 28 abr. 17;

RICÓS, C. et al. **Biological variation database, and quality specifications for imprecision, bias and total error (desirable and minimum)**. 1999. Disponível em: <http://www.westgard.com/biodatabase-2010-update.htm>;

SANTOS, A. P.; ZANUSSO JUNIOR, G. **Controle de qualidade em laboratórios clínicos**. Revista UNINGÁ, v. 45, p. 60-67, set. 2015;

SEVGI, L. **Accreditation: Crucial in World Trade, Public Safety, and Human Rights**. IEEE Antennas and Propagation Magazine, V. 56, 2014;

SUNDERMAN, F.W. **The History of Proficiency Testing / Quality Control**. Clinical Chemistry, v. 38, nº 7, p. 1205-1209, 1987;

THOMPSON, M. et al. **The International Harmonized Protocol for the Proficiency Testing of Analytical Chemistry Laboratories**. IUPAC Technical Report., Pure Appl. Chem., V. 78, nº 1, 2006;

VIEIRA, K.F. et al. **A utilidade dos indicadores da qualidade no gerenciamento de laboratórios clínicos**. J Bras Patol Med Lab, V. 47, nº3, p. 201-210, jun. 2011;

WESTGARD, J.O. et al. **QC Validator® 2.0: a computer program for automatic selection of statistical QC procedures for applications in healthcare laboratories**. Computer Methods and Programs in Biomedicine, 53, 175–186, 1997;

WORLD HEALTH ORGANIZATION et al. **Guidelines for Appropriate Evaluations of HIV Testing Technologies in Africa**. 2001;

WORLD HEALTH ORGANIZATION. **Boas práticas da OMS para laboratórios de controle de qualidade de produtos farmacêuticos**. OMS Series de Informes Técnicos, nº 957, 2010;

_____. **WHO Expert Committee on Biological Standardization**. WHO Technical Report Series 987, 2016.

Apêndice A - Algoritmo A (ISO 13528, 2015)

Os critérios de avaliação de desempenho para EP quantitativos propostos pela ISO (ISO 13528, 2015) supõe que os dados reportados pelos participantes sejam minimamente unimodais e simétricos. A presença de outliers pode distorcer essa simetria, bem como a estimação do valor designado e das medidas de incerteza obtidos por consenso, o que acarreta em problemas na avaliação desse desempenho. É realizado um algoritmo iterativo, que efetua o tratamento desses outliers, de forma a obter estimativas confiáveis sem a eliminação de dados. Esse algoritmo está descrito na ISO (ISO 13528, 2015; HUBER, 2009), e é chamado de Algoritmo A, estando seu processo detalhado no Anexo C.3 da norma.

Esse algoritmo pode ser assim resumido:

1º Calcula-se a mediana M de \vec{x} , onde:

$\vec{x} = (x_1, \dots, x_n)$, sendo n o número de observações no conjunto de dados

2º Obtêm-se o seguinte vetor:

$\vec{q} = (q_1, \dots, q_n)$, onde $q_i = |x_i - M|$, $i = 1, \dots, n$

3º Calcula-se a mediana M_q , com os dados de \vec{q} ;

4º Calcula-se as seguintes estatísticas:

$$DAM_n(X) = 1.4826 \cdot M_q$$

$$\phi(X) = 1.5 \cdot DAM_n(X)$$

5º Gera-se um novo vetor \vec{x}^* , com as observações transformadas do vetor \vec{x} , segundo a seguinte regra:

$$x_i^* = \begin{cases} Md(X) - \varphi(X), & \text{se } x_i < Md(X) - \varphi(X) \\ x_i, & \text{se } Md(X) - \varphi(X) \leq x_i \leq Md(X) + \varphi(X) \\ Md(X) + \varphi(X), & \text{se } x_i > Md(X) + \varphi(X) \end{cases}$$

6º Calcula-se com as observações do vetor $x^*=(x_1^*,...,x_n^*)$ a nova média (valor designado por consenso) e o novo desvio padrão (desvio padrão para a proficiência por consenso), empregando as expressões abaixo:

$$\hat{\mu}_{rob} = \frac{\sum_{i=1}^n x_i^*}{n} \quad \hat{\sigma}_{rob} = 1.134 \cdot \sqrt{\frac{\sum_{i=1}^n (x_i^* - \hat{\mu}_{rob})^2}{n-1}}$$

O processo do algoritmo é desenvolvido de forma iterativa, ou seja, efetuando o mesmo cálculo repetitivamente, sendo tomado como $Md(X)$ o $\hat{\mu}_{rob}$, e como novo $DAM_n(X)$ o $\hat{\sigma}_{rob}$. É efetuado novo truncamento, e obtido novos $\hat{\mu}_{rob}$ e $\hat{\sigma}_{rob}$. O processo é finalizado quando um valor de convergência é atingido, e dessa forma, não é observada diferença entre as estimativas do σ_{cons} entre iterações. O Algoritmo A sugere que, se não houver uma diferença ξ entre estimativas do $\hat{\sigma}_{rob}$ que interfira na terceira casa decimal da estimativa obtida ($\xi < 0.001$), o processo pode ser interrompido.

Apêndice B – Testes estatísticos empregados

Teste bilateral de detecção de outliers de Grubbs (ISO 5725-2, 2004; NIST, 2017)

Seja y_1, y_2, \dots, y_n uma amostra de tamanho n de um conjunto de dados. Sejam dadas as hipóteses:

H_0 : A observação y_i ($i=1, \dots, n$) não é um outlier dentro do conjunto de dados;

H_1 : A observação y_i ($i=1, \dots, n$) é um outlier dentro do conjunto de dados.

Seja a estatística de teste:

$$G = \max_i \frac{|y_i - \bar{y}|}{s},$$

Onde:

- \bar{y} é a média amostral;
- s o desvio padrão amostral.

Rejeitaremos H_0 , e assim y_i é um outlier, se $G > G_{tab}$, onde:

$$G_{tab} = \frac{n-1}{\sqrt{n}} \sqrt{\frac{t_{\alpha/2n, n-2}^2}{n-2 + t_{\alpha/2n, n-2}^2}},$$

onde $t_{\alpha/2n, n-2}$ indica o quantil de uma distribuição T-Student, com $n-2$ graus de liberdade, a um nível de significância α .

Teste Qui-Quadrado para variância – amostra única (BUSSAB & MORETTIN, 2010)

Seja y_1, y_2, \dots, y_n uma amostra de uma variável aleatória $N(\mu, \sigma^2)$, sendo μ e σ^2 parâmetros desconhecidos. Sejam dadas as hipóteses:

$$H_{0v}: \sigma^2 = \sigma_0^2$$

$$H_{1v}: \sigma^2 \neq \sigma_0^2$$

Rejeitaremos H_{0v} se:

$$X = \frac{(n-1)s^2}{\sigma_0^2} < \sigma_0^2 \chi_{n-1; \frac{\alpha}{2}}^2 \quad \text{ou} \quad X = \frac{(n-1)s^2}{\sigma_0^2} > \sigma_0^2 \chi_{n-1; 1-\frac{\alpha}{2}}^2$$

Onde:

- σ_0 é o desvio padrão populacional sob H_0 ;
- s é o desvio padrão amostral;
- n é o tamanho amostral;
- α é um nível de significância definido a priori;
- $\chi_{n-1; \frac{\alpha}{2}}^2$ e $\chi_{n-1; 1-\frac{\alpha}{2}}^2$ são quantis de uma distribuição Qui-Quadrado, com $n-1$ graus de liberdade, para um nível de significância α .

Teste T para média - amostra única (BUSSAB & MORETTIN, 2010)

Seja y_1, y_2, \dots, y_n uma amostra de uma variável aleatória $N(\mu, \sigma^2)$, sendo μ e σ^2 parâmetros desconhecidos. Sejam ainda dadas as hipóteses:

$$H_{0m}: \mu = \mu_0$$

$$H_{1m}: \mu \neq \mu_0$$

Rejeitaremos H_{0m} se:

$$\left| \frac{(\bar{y} - \mu_0)\sqrt{n}}{s} \right| > t_{(n-1), \frac{\alpha}{2}}$$

Onde:

- \bar{y} é a média amostral;
- μ_0 é a média populacional sob H_0 ;
- s é o desvio padrão amostral;
- n é o tamanho amostral;
- α é um nível de significância definido a priori;
- $t_{(n-1),\alpha/2}$ é o quantil de uma distribuição T de Student, com $n-1$ graus de liberdade com nível de significância α .

Teste de Kruskal-Wallis por soma de rankings (FONSECA & MARTINS, 2010)

Sejam T_1, T_2, \dots, T_k variáveis aleatórias mutuamente independentes, que chamaremos de **tratamentos**. Sejam dadas as hipóteses:

$$H_{0v}: T_1 = T_2 = \dots = T_k.$$

$$H_{1v}: \exists T_i, T_j / T_i \neq T_j, i, j = 1, \dots, k.$$

Seja ainda estatística H , tal que:

$$H = \frac{\left(\frac{12}{n(n+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} \right) - 3(n+1)}{1 - \frac{\sum_{j=1}^g t_j^3 - t_j}{n^3 - n}}$$

Onde:

- $R_i = \sum_{j=1}^{n_i} r_{ij}$, onde r_{ij} é o posto (ordem) da observação T_{ij} frente a todas as observações do conjunto de dados;
- n_i é o tamanho amostral do i -ésimo tratamento;
- n é o tamanho amostral de todos os resultados;
- t_j é o tamanho do j -ésimo grupo de elementos repetidos.

Sob H_0 , $H \sim \chi_{k-1}^2$, e rejeitaremos H_0 com $(1 - \alpha) \%$ de confiança, se $H > Q_\alpha$, onde $P(H > Q_\alpha) = \alpha$.

Teste de Levene para igualdade de variâncias (LEVENE, 1960)

Sejam T_1, T_2, \dots, T_n variáveis aleatórias independentes, que chamaremos de **grupos**.

Sejam dadas as hipóteses:

$$H_{0v}: \sigma^2_{T1} = \sigma^2_{T2} = \dots = \sigma^2_{Tk}.$$

$$H_{1v}: \exists \sigma^2_{Ti}, \sigma^2_{Tj} / \sigma^2_{Ti} \neq \sigma^2_{Tj}, i, j = 1, \dots, k.$$

Seja ainda estatística W , tal que:

$$W = \frac{n - k}{k - 1} \frac{\sum_{i=1}^k n_i (Z_{i.} - Z_{..})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} n_i (Z_{ij} - Z_{i.})^2}$$

Onde:

- k é o número de diferentes grupos (variáveis) sob teste;
- n_i é o número de casos no i -ésimo grupo;
- n é tamanho amostral total;
- Y_{ij} é o j -ésimo valor do i -ésimo grupo;
- $Z_{ij} = |Y_{ij} - \bar{Y}_{i.}|$, onde $\bar{Y}_{i.}$ é a média do i -ésimo grupo;
- $Z_{i.} = \frac{1}{n_i} \sum_{j=1}^{n_i} Z_{ij}$ é a média de Z_{ij} para o i -ésimo grupo;
- $Z_{..} = \frac{1}{n} \sum_{i=1}^k \sum_{j=1}^{n_i} Z_{ij}$ é a média de todos os Z_{ij} .

Sob H_0 , $W \sim F_{k-1, n-k}$ (distribuição F de Snedecor), e rejeitaremos H_0 com $(1 - \alpha) \%$ de confiança, se $W > F_\alpha$, onde $P(W > F_\alpha) = \alpha$.