



**Universidade de São Paulo**  
Instituto de Química de São Carlos  
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA  
Química Analítica e Inorgânica

**Crise de reprodutibilidade científica em química analítica:  
avaliação com foco na validação dos métodos e na incerteza dos  
resultados**

Autora: Bruna Drielen Ferreira

São Carlos, 2021

Bruna Drielen Ferreira

Crise de reprodutibilidade científica em química analítica: avaliação com foco na validação dos métodos e na incerteza dos resultados

Dissertação apresentada ao Instituto de Química de São Carlos da Universidade de São Paulo como parte dos requisitos para a obtenção do título de mestre em ciências.

Área de concentração: Química Analítica e Inorgânica

Orientador: Prof. Dr. Vitor Hugo Polisél Paccès

Coorientador: Prof. Dr. Emanuel Carrilho

São Carlos, 2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico para fins de estudo e pesquisa, desde que citada a fonte.

**Assinatura:** Bruna Drielen Ferreira Gonçalves  
**Data:** 27.09.2021

*Ficha Catalográfica elaborada pela Seção de Referência e Atendimento ao Usuário do SBI/IQSC*

Ferreira, Bruna Drielen

Crise de reprodutibilidade científica em química analítica: avaliação com foco na validação dos métodos e na incerteza dos resultados / Bruna Drielen Ferreira. — São Carlos, 2021.

140 f.

Dissertação (Mestrado em Química Analítica e Inorgânica) — Instituto de Química de São Carlos / Universidade de São Paulo, 2021.

Orientador: Prof. Dr. Vitor Hugo Polisél Pacces

Coorientador: Prof. Dr. Emanuel Carrilho

1. Reprodutibilidade. 2. Validação. 3. Incerteza. 4. Software. 5. Estatística.  
I. Título.

Sonia Alves - CRB: 4280/8



## DEDICATÓRIA

Desde o início, com base nas sugestões do Prof. Dr. Vitor, pensei e executei este projeto pensando nas pessoas: em poder deixar para elas pistas de como seria possível reproduzir bom trabalho, ou realizar um bom trabalho possível de ser reproduzível. Porém, ao decorrer do período em que o trabalho se desenvolveu, percebi que este viés de "ajudar pessoas a reproduzir um bom trabalho" ou "deixá-lo possível reproduzir" era "de família".

Ao observar e lembrar meu pai executando qualquer trabalho com muito esforço e dedicação, minha mãe realizando tudo com muito afeto e carinho, percebi que essas qualidades eram consequências de lutas, esforços, dedicação, afeto e carinho de duas pessoas principais: minhas Avós.

Por isso, dedico este trabalho à "Vó Edna" e a "Vó Rosa", além de todos aqueles nos quais este estudo possa ajudar de alguma forma.

Apesar de ambas estudarem até a 4.<sup>a</sup> série do ensino fundamental, e nem estarem próximas de obter um título de mestre como estou prestes a receber agora, passaram para meus pais e para mim, valores fundamentais para execução deste trabalho: a vontade de realizar, bem como ajudar as pessoas a reproduzir um bom trabalho.

Um dos exemplos disso, são as receitas culinárias "desenvolvidas" por elas: sempre me ajudaram a ajustar os equívocos a cada receita de pão, balas de coco, torta, arroz, bolo, etc. E até eu conseguir deixar o resultado igual, ou melhor, do que o delas, elas não desistiam! E não desistem!

Portanto, fica aqui minha gratidão e reconhecimento à todas lutas e esforços realizados pelas minhas avós, permitindo que eu estar, hoje, concluindo mais esta etapa!

## AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Vitor Hugo Polisél Pacces, sem o qual não teria conseguido concluir esta gratificante tarefa. Pela sua impecável conduta profissional. Pela paciência e seu aspecto compassivo que compreendeu todas as minhas limitações quando elas apareceram. Por me orientar, não apenas na área acadêmica, mas também em como ser um ser humano melhor em todos os lugares que eu estiver.

Aos professores, Prof. Dr. Emanuel Carrilho e Prof. Dr. Igor Renato Bertoni Olivares, correções e ensinamentos que me permitiram apresentar um melhor desempenho no meu processo de formação profissional ao longo do curso.

Aos meus pais, Marco Aurélio Ferreira e Nivânia Caparelli Ferreira, que me incentivaram e apoiaram em todos os momentos difíceis ou não, por SEMPRE apoiarem meus estudos e compreenderam a minha ausência enquanto eu me dedicava à realização deste trabalho.

Ao meu irmão e melhor amigo, Vitor Hugo Ferreira, que também compreendeu a minha ausência enquanto eu me dedicava à realização deste trabalho, mas também ajudou com sua juventude, bom humor e muita música boa!

Ao meu esposo, Leonam Gouvêa Gonçalves, pelo suporte aos meus estudos e pela caminhada do dia a dia.

As amigas, que estiveram sempre ao meu lado, Bruna Correia da Silva, Rhaissa Mecca Bontempi, Pamela Grizotto, Cássia Watanabe Kock e Paula Souza da Silva Gomes Lima, pela amizade e pelo apoio demonstrado ao longo de todo o período em que me dediquei a este trabalho.

Aos professores, Prof. Dr. Vitor Hugo Polisél Pacces e Prof. Dr. Igor Renato Bertoni Olivares, também por conceder as licenças gratuitas do software ConfLab, pois sem elas o trabalho não teria os resultados de qualidade que forneceu. Ao Grupo de Pesquisa *BioMics* pelas contribuições em relação à visualização aplicável deste trabalho.

À Universidade de São Paulo (USP), corpo docente e discente, a quem fico lisonjeada por dela ter feito parte. Bem como ao Instituto de Química de São Carlos (IQSC) e aos serviços da Pós-Graduação, as agências de fomento pela bolsa e recursos fornecidos.

# EPÍGRAFE

“É mais simples do que você pensa”

*Vitor Hugo Poliséel Pacces*

## RESUMO

Os dados apresentados em 2016 pelo artigo “1500 scientists lift the lid on reproducibility” de Monya Baker, trouxeram uma discussão sobre a existência de significativa crise de reprodutibilidade em diversas áreas da pesquisa científica, e a química foi a área de conhecimento que os cientistas relataram grande dificuldade em reproduzir resultados publicados. Além disso, as principais causas apontadas, de acordo com a experiência dos cientistas consultados, referem-se ao baixo entendimento e aplicação das ferramentas estatísticas, além de replicações insuficientes no laboratório original. Neste trabalho, desenvolveu-se uma pesquisa baseada na busca pela verificação da confirmação ou não de uma potencial crise de reprodutibilidade na área de química analítica, devido a um baixo entendimento e aplicação das ferramentas estatísticas. Primeiramente, o objetivo foi avaliar se a percepção da existência de uma crise de reprodutibilidade era passível de ser confirmada e quantificada. O segundo objetivo foi avaliar a relação entre a potencial existência da crise de reprodutibilidade com as fontes descritas pelos entrevistados com foco em validação métodos e incerteza dos resultados. Com base nos critérios constantes em protocolos e legislações vigentes, foram realizadas avaliações de conformidade e verificações dos cálculos estatísticos dos dados analíticos publicados. Para avaliação dos dados e resultados, utilizou-se como referência os softwares validados voltados à laboratórios de ensaio, denominados ConfLab Validação e ConfLab Incerteza, os quais foram projetados para atender os principais Sistemas de Gestão da Qualidade. As discussões ocorreram a partir da comparação dos resultados obtidos da forma como se apresentaram nos documentos originais com os resultados obtidos pelos softwares. Os resultados, apontaram que 28% dos trabalhos apresentaram valores de incerteza maiores que 100% e apenas 19% dos autores realizaram os estudos dos parâmetros de validação de acordo com os protocolos e em 17% das validações todos os parâmetros atenderam aos critérios de aceitação de resultados dos protocolos, valores que permitiram evidenciar com probabilidade considerável a existência de uma crise de reprodutibilidade causada pelo uso equivocado da validação de métodos, sustentada pelos elevados valores de incerteza não avaliados nos documentos científicos estudados nas diferentes áreas de conhecimento.

Palavras – chaves: reprodutibilidade, validação, incerteza, software, estatística.

## ABSTRACT

The data provided in 2016 by the article “1500 scientists raise the reproducibility limit” by Monya Baker, brought a discussion about the existence of reproducibility crisis in several areas of scientific research, and chemistry was the area of knowledge that scientists reported great difficulty in reproducing published results. Furthermore, the main causes identified, according to the experience of the scientists consulted, are specific to the low understanding and application of statistical tools, in addition to insufficient replication in the original laboratory. In this work, research based on the search for verification of verification or not of a potential reproducibility crisis in the field of analytical chemistry is specialized, due to a low understanding and application of statistical tools. First, the objective was to assess whether the perception of the existence of a reproducibility crisis could be confirmed and quantified. The second objective was to assess the relationship between the existence of a potential existence of the reproducibility crisis with the sources of origin by the interviewees, focusing on validation methods and uncertainty of the results. Based on the criteria contained in current protocols and legislation, conformity assessments and verifications of statistical calculations of published analytical data were carried out. To evaluate the data and results, the validated software aimed at testing laboratories, the ConfLab Validation and ConfLab Uncertainty, which were assigned to meet the main Quality Management Systems, used as a reference. How the cost occurred from the comparison of the results obtained in the form of original documents with the results obtained by the software. The results showed that 28% of the general works with uncertainty values greater than 100% and only 19% of the authors carried out the studies of the validation parameters according to the protocols and in 17% of the validations all parameters met the acceptance criteria of protocol results, values that make it possible to evidence with probability the existence of a reproducibility crisis caused by the mistaken use of the validation of methods, supported by uncertainty values not appreciated in the scientific documents studied in different areas of knowledge.

Keywords: reproducibility, validation, uncertainty, software, statistics..

## LISTA DE ILUSTRAÇÕES

Figura 1 - Participante representativo do estudo pulando de uma aeronave com uma mochila vazia. Este indivíduo não sofreu morte ou ferimentos graves após o impacto com o solo.....	20
Figura 2 - Relações entre a química analítica, outras áreas da química e outras ciências .....	24
Figura 3 - Conteúdo dos comprimidos de amoxicilina em 45 lotes com base nos resultados dos ensaios de HPLC e suas incertezas. ....	48
Figura 4 - Diferentes conceitos e o significado estendido de rastreabilidade .....	49
Figura 5 - Ciclo De Garantia de Qualidade Analítica (Ciclo AQAC) .....	52
Figura 8 - Representação do procedimento de avaliação da conformidade da validação do método com os de critérios de aplicação para caracterização do estudo segundo os protocolos utilizados.....	67
Figura 9 - Representação do procedimento de avaliação da conformidade da validação do método para caracterização do estudo segundo a aplicação dos ensaios de desempenho.....	67
Figura 10 - Esquema de avaliação dos critérios de aceitação de resultados quando o protocolo estabeleceu critério de aceitação de resultado para o parâmetro de validação .....	71
Figura 11 - Esquema de avaliação dos critérios de aceitação de resultados quando o protocolo deixou a critério do autor estabelecer um critério de aceitação para o parâmetro de validação.....	72
Figura 12 - Representação do esquema de avaliação da conformidade dos cálculos descrito no documento científico aos critérios de aceitação de resultados do protocolo e dos cálculos.....	75
Figura 13 - Representação do esquema utilizado na busca informações dos tipos de características que possibilitassem associações dos resultados a um valor de referência no documento.....	81
Figura 14 - Possível relação para investigar o uso equivocado e insuficiente da validação de métodos.....	83
Figura 15 - Possível conclusão sobre a relação entre a crise de reprodutibilidade e incerteza e medições, com base no ciclo AQAC .....	85
Figura 16 - Quantidade de materiais estudados e gerados pelo trabalho .....	87
Figura 17 - Caracterização do estudo em relação ao tipo, ano e área de conhecimento .....	88

Figura 18 - Características do uso dos protocolos de validação (continuação).....	90
Figura 19 - Frequência que os parâmetros de validação foram estudados dentro das validações de métodos.....	91
Figura 20 - Comparação entre as frequências que os parâmetros de validação foram estudados dentro de condições adequadas segundo o protocolo mais apropriado para o objetivo do método.....	92
Figura 21 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos de linearidade .....	93
Figura 22 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos do limite de detecção (LD) .....	93
Figura 23 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos do limite de quantificação (LQ) .....	94
Figura 24 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos da precisão intermediária.....	94
Figura 25 - Equívocos em relação à condução do ensaio e obtenção de dados nos da precisão.....	95
Figura 26 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos da exatidão .....	95
Figura 27 - Resultados sobre a avaliação das características que levam a confiabilidade no resultado no documento científico .....	96
Figura 28 - Frequência que os cálculos dos parâmetros de validação foram realizados corretamente quando foram estudados.....	98
Figura 29 - Características das condições em relação a regressão linear nos estudos da linearidade.....	99
Figura 30 - Características cálculos relacionados estudos de linearidade .....	100
Figura 33 - Comportamento das variâncias dos resíduos, segundo os testes de Cochran realizados pelo software ConfLab Validação .....	103
Figura 34 - Porcentagem de validações que, dos documentos nos quais foi possível identificar o modelo de regressão empregado, foram corretas e incorretas .....	103
Figura 35 - Frequência dos equívocos em relação ao modelo de regressão linear utilizado no documento, comparados ao valor referência (software), dentre os modelos de regressão linear avaliados como incorretos. ....	104
Figura 36 - Dentre as validações que não foi possível identificar qual modelo de regressão linear empregado, de acordo com os resultados do software, quais os modelos mais adequados para realizar a regressão.....	105

Figura 37 - Das validações nas quais foi possível avaliar o modelo de regressão linear utilizado, e dentre essas, as avaliações foram corretas, ou seja, as que os modelos de regressão do software estavam iguais ao do documento, quantas das validações realizaram os testes para verificar o modelo de regressão linear adequado.....	106
Figura 38 - Porcentagem que os resultados atenderam aos critérios de aceitação do protocolo.....	109
Figura 39 - Frequência que utilizaram o mesmo critério de aceitação de resultados do protocolo.....	110
Figura 40 - Quando não utilizou o mesmo critério de aceitação do protocolo, ou o protocolo deixou a critério do pesquisador estabelecer o critério, estão dispostas as frequências que foram utilizados outros critérios.....	111
Figura 41 - Valores de incerteza expandida obtidas dos resultados dos parâmetros de validação, separados no gráfico pelos por três pontos diferentes de concentrações de cada curva de calibração estudada. São eles: primeiros pontos das curvas de calibração, os pontos de média concentração e últimos pontos. ....	113
Figura 42 - Parâmetros que mais contribuíram para a estimativa a incerteza expandida .....	114
Figura 43 - Resumo dos resultados nos quais os autores realizaram os estudos conforme as recomendações do protocolo, os resultados dos parâmetros de validação atenderam aos critérios de aceitação de resultados do protocolo, no primeiro ponto e que os valores de incerteza associadas aos resultados estavam dentro dos valores aceitáveis, ou seja, menores do que 40%, também no primeiro ponto. ....	117
Figura 45 - Forma de apresentação e reporte de resultados das validações de métodos estudadas.....	122
Figura 46 - Valores das estimativas de incerteza expandida associadas as validações de métodos estudadas no primeiro ponto da curva de calibração .....	123
Figura 47 - Relação entre uso equivocado de validação de métodos e incerteza de medição.....	124
Figura 48 - Relação entre incerteza de medição e reprodutibilidade de métodos analíticos .....	126
Figura 49 - Distribuição dos equívocos de cálculos mais comuns .....	140

## LISTA DE TABELAS

Tabela 1 - Protocolos de validação e seus campos de aplicação .....	34
Tabela 2 - Fatores importantes para estudo de linearidade ou curva de calibração .	38
Tabela 3 - Critérios para avaliar e caracterizar os equívocos ocorridos em cada parâmetro de validação .....	69
Tabela 5 - Critérios para a avaliação dos equívocos de cálculos dos parâmetros de validação .....	76
Tabela 6 - Representação da disposição do material suplementar onde estão disponíveis os dados e discussões em materiais suplementares independentes.....	87

## LISTA DE ABREVIATURAS E SIGLAS

ANOVA	<i>Analysis of variance</i> ; Análise de variância
ANVISA	Agência Nacional de Vigilância Sanitária
CMHP	<i>Committee for Medicinal Products for Human Use</i> ; Comitê de Produtos Medicinais para Consumo Humano
CV	Coeficiente de Variação
DICLA	Divisão de Acreditação de Laboratórios
DOQ-CGCRE	Documentos Orientativos da Coordenação Geral de Acreditação
DPR	Desvio padrão relativo
EMA	<i>European Medicine Agency</i> ; Agência Europeia de Medicina
FDA	<i>Food and Drug Administration</i> ; Administração de Alimentos e Fármacos
FMN	Fator de Matriz Normalizado
GUM	<i>Guide to the Expression of Uncertainty in Measurement</i> ; Guia para a Expressão da Incerteza de Medição
ICH	<i>International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use</i> ; Conselho Internacional de Harmonização de Requisitos Técnicos para Produtos Farmacêuticos para Uso Humano
INMETRO	Instituto Nacional de Metrologia, Qualidade e Tecnologia
IUPAC	<i>International Union of Pure and Applied Chemistry</i> ; União Internacional de Química Pura e Aplicada
JCGM	<i>Join Committee for Guides in Metrology</i> ; Comitê Conjunto para Guias em Metrologia
LLOD/LD	<i>Lower Limit of Detection</i> ; Limite de Detecção

LLOQ/LQ	<i>Lower Limit of Quantification</i> ; Limite de Quantificação
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
RDC	Resolução da Diretoria Colegiada
RE	Resolução
REMESP	Rede Metrológica do Estado de São Paulo
VIM	Vocabulário Internacional de Metrologia

## SUMÁRIO

<b>RESUMO .....</b>	<b>7</b>
<b>1</b>	<b>Introdução .....</b>
	<b>18</b>
1.1	Proeminência.....
	18
1.2	A crise de reprodutibilidade e a química analítica.....
	21
1.3	Reprodutibilidade e qualidade de resultados.....
	25
1.4	Crise de reprodutibilidade, custos e carreira .....
	28
1.5	Ferramentas e base teórica.....
	30
1.5.1	Como desenvolver um método analítico.....
	30
1.5.2	Validação de métodos como ferramenta estatística .....
	31
1.5.2.1	Protocolos de validação .....
	32
1.5.2.2	Parâmetros de validação e testes estatísticos .....
	36
1.5.2.3	Linearidade .....
	37
1.5.2.4	Precisão .....
	39
1.5.2.5	Exatidão de medição.....
	40
1.5.2.6	Limite de Quantificação (LQ) .....
	42
1.5.2.7	Limite de Detecção (LD) .....
	42
1.5.2.8	Dados de Validação de Métodos e Apresentação de Resultados .....
	43
1.5.2.9	Incerteza de medição.....
	45
1.5.2.10	Relatórios de Validação .....
	49
1.5.3	Ciclo de Garantia de Qualidade Analítica ( <i>Analytical Quality Assurance Cycle (AQAC)</i> ).....
	51
1.5.4	Softwares ConfLab .....
	53
1.5.4.1	ConfLab Validação ( <i>Versão 1.0.1.0</i> ).....
	54
1.5.4.2	ConfLab Incerteza ( <i>Versão 1.0.1.0</i> ).....
	55
<b>2</b>	<b>Objetivos .....</b>
	<b>57</b>
<b>3</b>	<b>Desenvolvimento.....</b>
	<b>58</b>
3.1	Informações gerais .....
	58
3.2	Objetos do estudo.....
	59
3.2.1	Seleção dos documentos avaliados .....
	59

3.2.2	Seleção dos parâmetros de avaliação.....	60
3.2.3	Ciclo de Garantia de Qualidade Analítica (Analytical Quality Assurance Cycle (AQAC)).....	60
3.2.4	Aplicação dos Softwares ConfLab .....	60
3.2.4.1	Relatórios de Validação e Incerteza.....	61
3.3	Avaliação da percepção da existência de uma crise de reprodutibilidade em relação a validação de métodos analíticos .....	62
3.3.1	Caracterização e avaliação da aplicação da validação de métodos nos documentos analisados .....	63
3.3.1.1	Avaliação do uso dos protocolos de validação .....	64
3.3.1.1.1	Verificação da compreensão do contexto e objetivo de aplicação do método pelo autor .....	64
3.3.1.1.2	Verificação dos critérios e condições de parâmetro de validação e obtenção de dados.....	67
3.3.1.2	Avaliação dos cálculos e testes estatísticos quanto a adequação e suficiência .....	70
3.3.2	Avaliação dos critérios de aceitação de resultado .....	71
3.3.2.1	Verificação dos Cálculos.....	74
3.3.2.2	Verificação de características de rastreabilidade .....	80
3.3.2.3	Verificação da apresentação e qualidade dos resultados .....	82
3.3.3	Investigação da existência de relação entre os itens de uso dos protocolos de avaliação, cálculos e testes estatísticos e os erros mais comuns cometidos durante a validação;.....	82
3.4	Investigar a existência de relação entre as conclusões sobre o uso de validação com a estimativa de incerteza de medição; .....	84
3.5	Investigar a existência de relação da incerteza de medição com a crise de reprodutibilidade .....	84
<b>4</b>	<b>Resultados e discussão .....</b>	<b>87</b>
4.1	Caracterização do Estudo .....	88
4.2	Resultados gerais.....	89
4.2.1	Uso dos protocolos de validação.....	89
4.2.2	Características de rastreabilidade adequada .....	95
4.2.3	Cálculos e testes estatísticos .....	97

4.2.3.1	Cálculos .....	97
4.2.3.2	Critérios de aceitação de resultado.....	108
4.3	Valores de incerteza de medição.....	112
4.4	Discussões dos resultados .....	115
<b>5</b>	<b>Conclusão .....</b>	<b>117</b>
5.1	Avaliação da percepção da existência da crise de reprodutibilidade e o uso equivocado de validação de métodos.....	118
5.2	Características da apresentação e da qualidade dos resultados 121	
5.3	Relação entre incerteza de medição e crise de reprodutibilidade 125	
5.4	Considerações a respeito de medidas na tentativa de amenizar a crise 127	
	<b>Referências.....</b>	<b>129</b>
	<b>Glossário .....</b>	<b>134</b>
	<b>APÊNDICE A - Distribuição dos equívocos de cálculos mais comuns .....</b>	<b>140</b>

# 1 Introdução

## 1.1 Proeminência

A reprodutibilidade na pesquisa científica tem atraído atenção crescente tanto da comunidade científica quanto do público em geral (BAKER, 2015, 2016; BEGLEY; BUCHAN; DIRNAGL, 2015; GOODMAN; FANELLI; IOANNIDIS, 2018; LEEKA; PENG, 2015; LUND-JOHANSEN; TRAN; MEHTA, 2021; NATIONAL INSTITUTES OF HEALTH (EUA), 2021; PRINZ; SCHLANGE; ASADULLAH, 2011; STEPHENS, 2016; ZHAO; SAMPSON; WEN, 2020). A reprodutibilidade diz respeito à produção de resultados pouco dispersos em diferentes experimentos que visam abordar a mesma questão científica. (GOODMAN; FANELLI; IOANNIDIS, 2018; HARRIS, 2010; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, 2020; ZHAO; SAMPSON; WEN, 2020).

Em maio de 2016, um artigo de Monya Baker publicado na revista *Nature* (BAKER, 2016) entrevistou 1576 cientistas de diversas áreas com a seguinte questão: “Existe uma crise de reprodutibilidade?”. Dos entrevistados, 52% deles responderam que “Sim, existe uma crise significativa” e 38% disseram “Sim, existe uma ligeira crise”, ou seja, cerca de 90% dos entrevistados concordaram com a existência de uma crise, sendo a Química a área do conhecimento na qual os cientistas relataram maior dificuldade em reproduzir resultados publicados. Os motivos que contribuíram para a falta de reprodutibilidade, segundo a pesquisa, mais de 50% dos entrevistados apontaram “Replicação insuficiente no laboratório”, “Supervisão deficiente” ou “Baixo poder estatístico”, bem como 90% (mais de 1000 pessoas), assinalaram “Projeto experimental mais robusto”, “melhores estatísticas” e 80% “Validação nos próprios laboratórios”.

Dos entrevistados, 106 eram pesquisadores da área Química. Esta área apresentou um maior número de relatos sobre dificuldades em reproduzir resultados. 90% dos pesquisadores relataram que encontraram problemas ao reproduzir os resultados de outro pesquisador e quase 70% encontraram dificuldades ao reproduzir a sua própria pesquisa.

Um dado crítico e contraditório levantado no artigo (BAKER, 2016) é o fato de que, mesmo com as afirmações das dificuldades de reproduzir os resultados, menos de 31% dos pesquisadores não acreditam que o problema esteja relacionado a

resultados errados. Além disso, a maioria dos entrevistados afirmaram que ainda confiam na literatura publicada.

Em complemento à questão de que um melhor entendimento sobre estatística e um projeto experimental mais robusto seriam a solução do problema da crise de reprodutibilidade (BAKER, 2016) foi discutido no artigo *“Parachute use to prevent death and major trauma when jumping from aircraft: Randomized controlled trial”* (Uso de paraquedas para prevenir morte e traumas graves ao saltar de uma aeronave: ensaio controlado randomizado) (YEH et al., 2018). Este artigo possui foco em pesquisas clínicas, utilizando-se da ironia para chamar atenção que a estatística enquanto ciência é uma ferramenta a ser utilizada de maneira adequada partir de perguntas claramente definidas. Em resumo, a conclusão do artigo é:

“O uso de paraquedas não reduziu a morte ou lesão traumática grave ao saltar de uma aeronave na primeira avaliação randomizada desta intervenção. No entanto, o ensaio só conseguiu inscrever participantes em pequenas aeronaves estacionárias no solo, sugerindo extrapolação cautelosa para saltos de alta altitude.”(YEH et al., 2018, p. 6)

Ou seja, ironicamente, o artigo (YEH et al., 2018) quer tratar do paradigma consensual de que o arcabouço para a construção dos resultados e conclusão de estudos, resultados publicados como “estudos randomizados”, “estatisticamente significativos”, “estatisticamente validados”, dentre outros, são considerados verdadeiros, suficientes e razoáveis ou, pelo menos, como “bem justificados”, independente do contexto. Por mais que a falácia do estudo (YEH et al., 2018) obtivesse um planejamento experimental e um tratamento estatístico durante a análise dos dados, estes não foram realizados de forma suficiente e eficaz, fornecendo uma conclusão de caráter relativamente dissimulado para os leitores. A Figura 1 traz a foto de como “a pesquisa” foi realizada.

Figura 1 - Participante representativo do estudo pulando de uma aeronave com uma mochila vazia. Este indivíduo não sofreu morte ou ferimentos graves após o impacto com o solo



Fonte: YEH et al., Parachute use to prevent death and major trauma when jumping from aircraft: Randomized controlled trial, BMJ (Online) p. 5, 2018

Ainda neste contexto, em julho de 2021, o relatório do comitê de pesquisas animais do Departamento de Saúde e Serviços Humanos dos Estados Unidos (*NIH*) (NATIONAL INSTITUTES OF HEALTH (EUA), 2021), recomendou 5 temas de melhoria no treinamento e na prática estatística dos pesquisadores para melhorar o rigor e a reprodutibilidade das pesquisa animais:

1. melhorar o planejamento do estudo;
2. abordar relatórios incompletos e práticas de pesquisa questionáveis;
3. melhorar a seleção, *design* e relevância dos modelos animais;
4. melhorar a documentação metodológica e o relatório de resultados;
5. um tema transversal voltado para medir e avaliar os custos e a eficácia desses esforços.

Em 19 de agosto de 2021, o artigo “*Three questions to address rigour and reproducibility concerns in your grant proposal*” (WILSON; BOTHAM, 2021), com o objetivo de abordar as fraquezas e limitações na ciência e tranquilizar os financiadores em potencial, também o Instituto Nacional de Saúde dos EUA (NIH) trouxe uma estrutura de três perguntas para ajudar a enquadrar as escolhas científicas ao escrever propostas de pesquisa. O processo orienta os escritores a explicar suas escolhas experimentais, fazendo perguntas que abordam as limitações de sua proposta, como segue:

Quais são as fraquezas ou limitações essenciais em sua pesquisa? - Todo método científico tem limites. Frequentemente, esses limites são metodológicos ou específicos do campo.

Quais métodos você usará, ou usará, para lidar com essas limitações? - Esses podem ser métodos padrão em seu grupo de pesquisa, mas é importante destacá-los como tal para um revisor.

O que torna esses métodos adequados? - Justifique suas escolhas confirmando como o campo testou ou aceitou esses métodos para superar as limitações. ((WILSON; BOTHAM, 2021, p. 610)

É possível perceber que o foco em garantir uma pesquisa científica reprodutível tem aumentado e que soluções pautadas em um melhor preparo e capacidade do pesquisador em relação às suas próprias pesquisas e os estudos estatísticos adequados e suficientes empregados, para publicação de resultados completos e satisfatórios, vêm sendo frequentemente estudadas.

## **1.2 A crise de reprodutibilidade e a química analítica**

Como apresentado anteriormente, não só Monya Baker (BAKER, 2016) tem discutido sobre a dificuldade de se reproduzir resultados e sua relação com estudos insuficientes, mas sim, desde 2011 essa dificuldade de reprodução na pesquisa científica, vem sendo abordada (BAKER, 2016; BEGLEY; BUCHAN; DIRNAGL, 2015; LEEKA; PENG, 2015; PRINZ; SCHLANGE; ASADULLAH, 2011; YEH et al., 2018). Além disso, a busca por soluções pautadas em melhorias na competência do pesquisador em relação aos seus estudos.

Para entender melhor o que de fato significa reproduzir resultados na área de pesquisa científica em química, talvez sejam necessárias algumas interpretações

além da semântica da palavra em si. O Vocabulário Internacional de Metrologia (VIM) traz a palavra “reprodutibilidade” dentro da definição de “precisão de medição”, e separa o sentido da palavra em “reprodutibilidade de medição” e “condição de reprodutibilidade” que seguem:

“Reprodutibilidade de medição: precisão de medição conforme um conjunto de condições de reprodutibilidade (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 23).

Condições de reprodutibilidade: condição de medição em um conjunto de condições, as quais incluem diferentes locais, diferentes operadores, diferentes sistemas de medição e medições repetidas no mesmo objeto ou em objetos similares (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 23).”

Segundo o artigo *“What does research reproducibility mean?”* (GOODMAN; FANELLI; IOANNIDIS, 2018), para um melhor esclarecimento sobre o termo “reprodutibilidade de pesquisa”, propôs separá-lo em três termos: reprodutibilidade dos métodos, reprodutibilidade dos resultados e reprodutibilidade inferencial. Apesar do artigo ter um enfoque biomédico, é possível aproveitar suas definições em relação aos desenvolvimentos de métodos. A reprodutibilidade dos resultados e a reprodutibilidade inferencial serão apresentados no item 1.3 para melhor compreensão.

Para os autores (GOODMAN; FANELLI; IOANNIDIS, 2018), a reprodutibilidade dos métodos pretende captar o significado original de reprodutibilidade, ou seja, a capacidade de implementar, da forma mais exata possível, os procedimentos experimentais, computacionais e analíticos com os mesmos dados e ferramentas, para obter os mesmos resultados, por meio do fornecimento de detalhes suficientes sobre os procedimentos e dados do estudo para que os mesmos procedimentos possam, em teoria ou na realidade, ser repetidos com exatidão. O artigo de (GOODMAN; FANELLI; IOANNIDIS, 2018) cita em relação a este ponto que operacionalmente, isso pode significar aspectos diferentes em ciências diferentes, e por isso, no mínimo, um relatório de estudos detalhado, uma descrição dos procedimentos de medição, os dados coletados, o software e código de análise e os resultados analíticos finais devem estar disponibilizados (GOODMAN; FANELLI; IOANNIDIS, 2018).

Os autores (GOODMAN; FANELLI; IOANNIDIS, 2018), citam que a definição de

quais dados precisam ser examinados para garantir a reprodutibilidade ainda pode ser controversa, na área biomédica. Para o desenvolvimento de métodos em química analítica, não é diferente; visto a infinidade de outras áreas que estão relacionadas e a diversidade de contextos, objetivos e aplicações. O fornecimento de dados brutos (cromatogramas, espectrogramas, planilhas, etc.), para a interpretação e avaliação desses dados, informações sobre otimizações, limitações, julgamentos e escolhas que são feitos ao longo dos processos de limpeza, transformação e análise crítica de dados que podem ser decisivos na determinação dos resultados analíticos (GOODMAN; FANELLI; IOANNIDIS, 2018). Estas são informações que muitas vezes não estão disponíveis para os leitores. E os autores ainda adicionam que:

“mesmo que haja consenso sobre o conjunto de dados analíticos apropriado, a reprodutibilidade metodológica requer uma compreensão de quais e quantas análises foram realizadas e como as análises específicas relatadas em um artigo publicado foram escolhidas.”

Ou seja, mesmo que se estabeleça um consenso sobre o conjunto de dados analíticos apropriados, ainda são necessárias a disponibilidade de informações sobre quais as condições analíticas, parâmetros de validação, os critérios de aceitação de resultados e as respectivas justificativas da escolha de cada um deles.

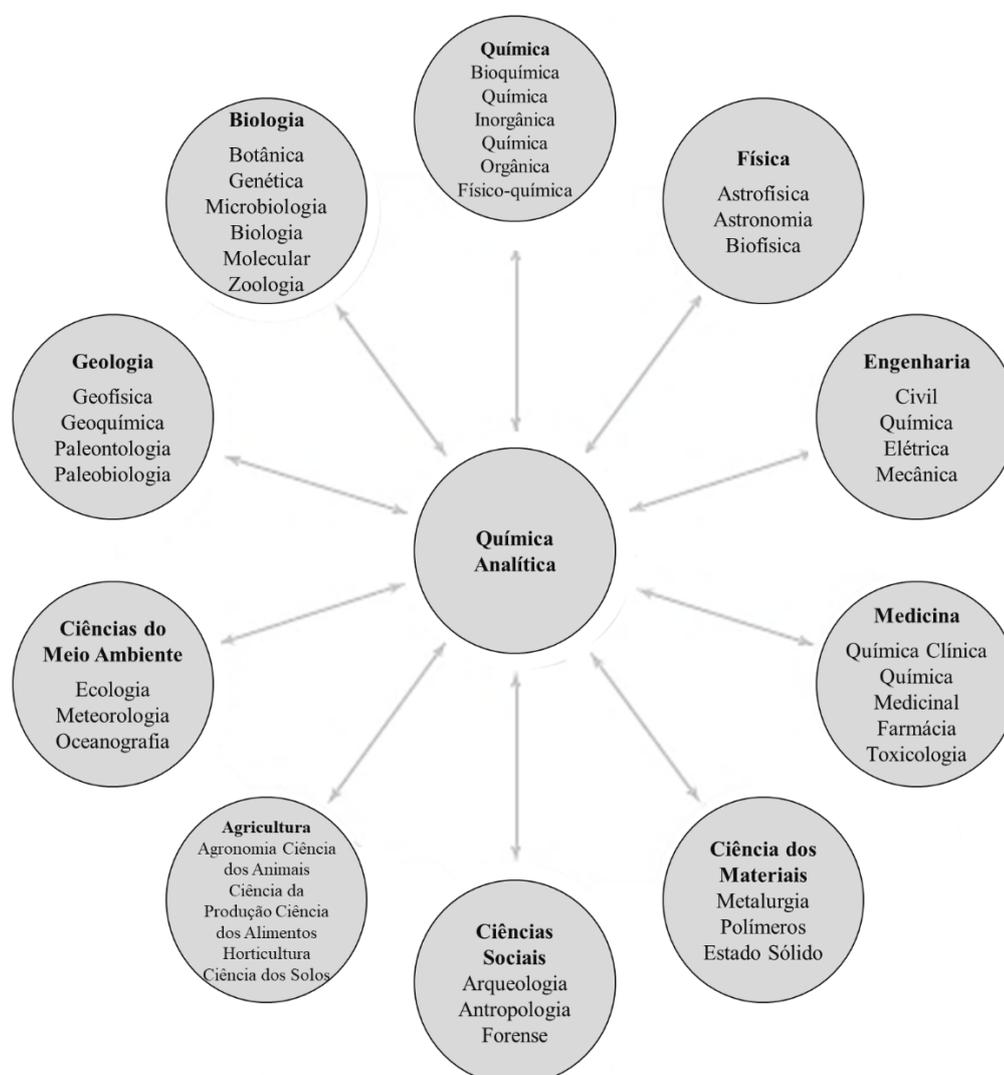
Em relação a avaliação da reprodutibilidade de métodos, o documento do Instituto Nacional de Metrologia do Brasil (Inmetro), intitulado “DOQ-CGCRE-008”, em sua revisão número 9 de junho 2020, onde estão orientações sobre validação de métodos analíticos, traz a seguinte recomendação para a reprodutibilidade dentro da análise do parâmetro de validação de precisão:

“Embora a reprodutibilidade não seja um componente de validação de método executado por um único laboratório, é considerada importante quando um laboratório busca a verificação do desempenho dos seus métodos em relação aos dados de validação obtidos por meio de comparação interlaboratorial.”

Com esta definição, embora a reprodutibilidade não seja um componente da validação de um método executado por uma única instituição que esteja desenvolvendo um método analítico, é considerada importante quando essa instituição busca a verificação do desempenho dos seus métodos em relação aos resultados obtidos por meio de comparação interlaboratorial, ou seja, o mesmo aplica-se, não apenas laboratórios de rotina, mas na pesquisa científica também.

Desta forma, considerando que o baixo conhecimento e aplicação das ferramentas estatísticas foi um dos fatores mais apontados pelos pesquisadores do referido artigo (BAKER, 2016), e que quase 90% dos pesquisadores relataram que não conseguiram reproduzir resultados de outros autores na área de química, é evidente a proeminência do assunto na área, que, segundo Skoog (SKOOG, 2014), a química é frequentemente denominada a ciência central, que por sua vez tem como posição central a química analítica. A Figura 2 mostra um diagrama na qual a localização central da química analítica representa sua importância e a abrangência de sua interação com muitas outras áreas.

Figura 2 - Relações entre a química analítica, outras áreas da química e outras ciências



Fonte: Skoog et. al., **Fundamentos De Química Analítica**, Ed. 9, p. 3, 2014. (Adaptado)

### 1.3 Reprodutibilidade e qualidade de resultados

O artigo “*What does research reproducibility mean?*” (GOODMAN; FANELLI; IOANNIDIS, 2018) mencionado anteriormente, discute vários aspectos dos significados semânticos e operacionais da reprodutibilidade na pesquisa científica e na introdução deste trabalho esses aspectos são retomados e discutidos relacionando-os a química analítica e ao desenvolvimento de métodos.

Portanto, ainda sobre as definições do artigo mencionado (GOODMAN; FANELLI; IOANNIDIS, 2018), a reprodutibilidade de resultados “refere-se à obtenção dos resultados semelhantes da aplicação de um estudo independente cujos procedimentos sejam o mais próximos possível do experimento original” (GOODMAN; FANELLI; IOANNIDIS, 2018). Apesar da definição, o artigo (GOODMAN; FANELLI; IOANNIDIS, 2018) cita que, assim como acontece com a reprodutibilidade dos métodos, “reproduzir resultados é operacionalmente ilusório”, devido aos erros aleatórios substanciais em qualquer resultado, tornando pouco claros os critérios para considerar os resultados “iguais”. Os autores (GOODMAN; FANELLI; IOANNIDIS, 2018) associam essa visão determinística de “resultados idênticos” à intuição e a lógica da reprodutibilidade dos resultados, que levam a negligência em aceitar resultados como verdadeiros quando a relação *signal ± erro* é excessivamente alta, e traz no artigo um comentário de Francis Collins, diretor do *US National Institutes of Health* (NIH), sobre os planos para aumentar a reprodutibilidade na pesquisa (COLLINS; TABAK, 2014):

“... Um conjunto complexo de outros fatores parece ter contribuído para a falta de reprodutibilidade. Os fatores incluem treinamento insuficiente de pesquisadores em projeto experimental, maior ênfase em fazer declarações provocativas em vez de apresentar detalhes técnicos e publicações que não relatam elementos básicos de projeto experimental. Alguns relatórios irreprodutíveis são provavelmente o resultado de descobertas coincidentes que atingem significância estatística, juntamente com tendência de publicação. Outra armadilha é a “superinterpretação” de experimentos criativos de “geração de hipóteses”, que são projetados para descobrir novos caminhos de investigação, em vez de fornecer provas definitivas para qualquer questão isolada. Ainda assim, permanece uma frequência preocupante de relatórios publicados que afirmam um resultado significativo, mas não podem ser reproduzidos.”

Uma outra discussão importante que o artigo traz em relação à reprodutibilidade de resultados na área biomédica é que, muitas vezes, a lógica linear dos estudos determinísticos está “superficialmente limitada por meio de testes de significância estatística” (GOODMAN; FANELLI; IOANNIDIS, 2018), e isto é problemático pois

“descobertas que são ‘estatisticamente significativas’ são frequentemente consideradas verdadeiras ou, pelo menos, como justificativas de uma afirmação de conhecimento” (GOODMAN; FANELLI; IOANNIDIS, 2018, p. 3)

Em relação a este ponto, ressalta o fato de que testes de hipótese-significância nula, não sejam o tipo apropriado e eficaz para avaliar satisfatoriamente aspectos na área biomédica, fato que também se aplica à química analítica, principalmente relacionado ao desenvolvimento de novos métodos. Alguns testes estatísticos isolados (não apenas de significância) não são suficientes para avaliar e garantir um “bom método” ou um “método reproduzível”. Segundo o artigo (GOODMAN; FANELLI; IOANNIDIS, 2018), “a significância estatística por si só diz muito pouco sobre se um estudo ‘replicou’ os resultados de outro”. Este ponto vai ao encontro de possível crise de reprodutibilidade estar associada ao uso inadequado e insuficiente de ferramentas estatísticas (BAKER, 2016).

Uma outra definição de “reprodutibilidade” discutida no artigo “*What does research reproducibility mean?*” (GOODMAN; FANELLI; IOANNIDIS, 2018) é a “reprodutibilidade inferencial”, que para os autores “refere-se à obtenção de conclusões qualitativamente semelhantes a partir de uma replicação independente de um estudo ou de uma reanálise do estudo original”. Eles diferenciam essa definição em relação à reprodutibilidade dos resultados, pois justificam que “nem todos os pesquisadores chegariam as mesmas conclusões dos mesmos resultados ou poderiam fazer diferentes escolhas analíticas que levam a diferentes inferências a partir dos mesmos dados” (GOODMAN; FANELLI; IOANNIDIS, 2018). Pelo fato de existirem muitos contribuintes para essas diferenças de inferências e conclusões, o artigo (GOODMAN; FANELLI; IOANNIDIS, 2018) divide a reprodutibilidade inferencial em duas perspectivas bayesianas e multivariada, e que os motivos relacionados a biomedicina que contribuem para cada uma delas, podem surgir de várias maneiras, porém como mencionado pelo próprio artigo “a definição de quais dados precisam ser examinados para garantir a reprodutibilidade pode ser controversa” (GOODMAN; FANELLI; IOANNIDIS, 2018), esses aspectos em relação a química analítica, ainda

precisam ser discutidos. Enfim, os autores (GOODMAN; FANELLI; IOANNIDIS, 2018) acabam concluindo que as causas e soluções para o que intitularam de “reprodutibilidade insuficiente”, em qualquer campo científico,

“...ainda requerem uma especificação clara do tipo de reprodutibilidade que está sendo discutido (métodos, resultados ou inferências), além de uma compreensão adequada de como isso afeta as alegações de conhecimento, investigação científica de suas causas e uma melhor compreensão das limitações de significância estatística como um critério de aceitação”.

Desta forma, com base nas discussões mencionadas (BAKER, 2016; GOODMAN; FANELLI; IOANNIDIS, 2018; OLIVARES, 2019; OLIVARES; LOPES, 2012; TENNY; ABDELGAWAD, 2021), sobre a importância da qualidade de resultados na pesquisa científica, assegurada por um uso adequado e suficiente de ferramentas estatísticas, e o quanto isso pode estar relacionado a reprodutibilidade de métodos na área de química analítica, os conceitos sobre Sistemas de Gestão de Qualidade Analítica, são importantes de serem mencionados e discutidos.

Nas indústrias e laboratórios de análises químicas, farmacêutica e áreas afins, muitas decisões importantes são baseadas nos resultados analíticos obtidos do departamento de controle e gestão da qualidade. E, uma vez que decisões importantes são tomadas com base nesses resultados, é primordial a indicação da qualidade desses resultados (TRAPLE et al., 2014). Compreender os Sistemas de Garantia de Qualidade dentro do contexto de reprodutibilidade analítica, requer entender como e por que esses sistemas surgiram.

Com avanço da tecnologia, as fabricações de produtos e serviços começaram a avançar em quantidade, porém sem o cuidado com a qualidade desse avanço. Nesse contexto, exigências de confiabilidade começaram a existir e um crescente receio de se produzir produtos e serviços rejeitados ou reprovados, gerou a necessidade de se garantir confiabilidade. Isso fez com que a organização em processos e métodos se tornasse obrigatória, surgindo o conceito de Qualidade Assegurada (LEITE, 2008).

Além disso, o conceito de Qualidade se tornou mais aparente nas últimas décadas. A grande exigência e competitividade de mercado trazida pela globalização em termos de tecnologia, serviços e produtos, trouxe a necessidade de uma forma de se diferenciar e se destacar no mercado agressivo e em expansão (CAMARGO; OLIVARES, 2016; OLIVARES, 2019). O Sistema de Gestão da Qualidade, então, entra como um sistema de gerência para garantir essa diferenciação e destaque pela

qualidade e confiabilidade de resultados de produtos e serviços (OLIVARES, 2019).

Mas como um sistema de gestão e controle de qualidade poderia se relacionar com uma possível crise de reprodutibilidade de métodos analíticos na pesquisa científica?

Como mencionado, decisões importantes são tomadas com base em resultados obtidos a partir dados analíticos, e é importante a indicação da qualidade desses resultados (TRAPLE et al., 2014). Esse ponto, não é importante apenas para indústrias e laboratórios, mas para a pesquisa acadêmica e científica também. É primordial que os resultados publicados em dissertações, teses, artigos etc. obtenham uma indicação do seu respectivo nível de qualidade.

Um indicador de qualidade que pode ser associado a qualidade de um método analítico é a incerteza de medição. Neste contexto de laboratórios temos a ISO 17025:2017 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017) que dispõe de requisitos gerais de sistemas de gestão da qualidade para a competência de laboratórios de ensaio e calibração. Na sua revisão no ano de 2017, ficou mais evidente a importância da incerteza de medição para avaliação de conformidade de produtos, principalmente pelo fato de que a incerteza pode indicar a probabilidade estatística para aprovar ou reprovar produtos e serviços. A ISO 17025:2017 menciona que, ao se avaliar a conformidade de um produto ou serviço, deve-se estabelecer uma regra de decisão a qual determina como a incerteza de medição irá afetar as tomadas de decisão nas avaliações realizadas.

#### **1.4 Crise de reprodutibilidade, custos e carreira**

Como citado anteriormente, algumas soluções do Departamento de Saúde e Serviços Humanos dos Estados Unidos (NIH) (NATIONAL INSTITUTES OF HEALTH (EUA), 2021; WILSON; BOTHAM, 2021), dentre os temas abordados sobre a melhoria no rigor e a reprodutibilidade das pesquisas animais, recomendou medir e avaliar os custos e a eficácia desses esforços. Com o objetivo de tranquilizar os financiadores em potencial, recomendou, ainda, uma melhor conduta do pesquisador em relação aos entendimentos de seus próprios estudos focando em reconhecer as respectivas fraquezas e limitações em sua pesquisa.

Não apenas nas pesquisas animais ou nas pesquisas da área de saúde, mas

também nas pesquisas de maneira geral, há a necessidade da análise de custos de estudos realizados de forma insatisfatória, com um destaque especial para a área de química analítica, uma vez que os custos de pesquisas para o desenvolvimento de métodos analíticos não são triviais e que a incapacidade de se reproduzir a pesquisa, representa um fardo enorme atrasando os tratamentos, desperdiçando o tempo dos cientistas e bilhões de dólares em pesquisa, como mencionado em 2015 (BEGLEY; BUCHAN; DIRNAGL, 2015).

As investigações de soluções pautadas em melhorias na conduta do pesquisador em relação aos seus estudos, para evitar a pesquisa não reproduzível, que causa desperdício de tempo dos cientistas e dinheiro faz sentido, visto que em 14 de dezembro de 2018, a China introduziu punições “sociais” para má conduta científica (CYRANOSKI, 2018). Os líderes chineses estavam focados na má conduta científica, seguindo relatórios contínuos de pesquisadores usando dados fraudulentos, falsificando currículos e falsificando avaliações de pares. Na nova política, dezenas de agências de fomento governamentais teriam o poder de aplicar penalidades aos pesquisadores que fossem pegos cometendo má conduta científica grave. Pesquisadores que fossem identificados aplicando uma má conduta também poderiam enfrentar punições que nada têm a ver com pesquisa, como restrições a empregos fora da academia, bem como penalidades existentes por má conduta, como perda de bolsas e prêmios.

Desta forma, com base na *survey* de Mônia Baker (BAKER, 2016), e considerando a definição de reprodutibilidade de um método em pesquisa científica como a capacidade de um pesquisador de encontrar os resultados de um estudo anterior usando os mesmos materiais que foram usados pelo investigador original (GOODMAN; FANELLI; IOANNIDIS, 2018), e que a reprodutibilidade aborda principalmente questões de confiança de que os dados e análises são representativos, é importante investigar soluções pautadas em melhorias em relação ao uso adequado e suficiente da estatística no intuito de se evitar o desperdício de tempo dos cientistas e custos em pesquisas que não poderão ser aproveitadas após sua conclusão.

## 1.5 Ferramentas e base teórica

### 1.5.1 Como desenvolver um método analítico

Segundo a ISO/IEC 17025 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017) um procedimento analítico é uma “forma especificada de executar uma atividade ou um processo”. Para laboratórios de rotina, que possuem sistemas de gestão de qualidade acreditados, para ser utilizado um procedimento analítico deve estar padronizado e documentado. Este documento é chamado de “POP” (Procedimento Operacional Padrão) e:

“...tem o objetivo de padronizar e minimizar a ocorrência de desvios na execução de tarefas fundamentais para a qualidade do ensaio, independente de quem as faça. Ou seja, um procedimento coerente garante ao usuário que a qualquer momento que ele se dirija ao laboratório, as ações tomadas na fase pré-analítica, analítica e pós-analítica críticas para garantir a qualidade de seus ensaios sejam as mesmas, de uma rodada para a outra, de um turno para outro, de um dia para outro” (PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020, p. 79).

Para que um método analítico (ou procedimento analítico) seja desenvolvido com o objetivo de ser posteriormente aplicado, utilizado ou reproduzido com níveis suficientes e adequados de confiança, são necessárias evidências de que esse método de fato é coerente com seu objetivo e escopo de aplicação, ou seja, é adequado para ao uso pretendido (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; THOMPSON; ELLISON; WOOD, 2002).

O conjunto de processos que traz a confirmação de que os requisitos específicos de um procedimento analítico desenvolvido para um determinado objetivo são atendidos, mediante exame e fornecimento de evidência objetiva (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017) é chamado de “validação de métodos”. Desta forma, para fornecer evidências e obter confirmação de que um procedimento analítico é adequado para ao uso pretendido e poderá ser aplicado, utilizado ou reproduzido com níveis suficientes e adequados de confiança, deve-se realizar um processo de validação desse método.

O documento do Inmetro *DOQ-CGCRE-008* que dispõe de orientações sobre validação de métodos analíticos, em seu item “10.1 Planejamento e execução da validação” sugere o seguinte planejamento para realizar uma validação de método:

- ✓ Definir objetivo e escopo do método;

- ✓ Definir os parâmetros de validação;
- ✓ Definir os critérios de aceitação para cada parâmetro de validação;
- ✓ Verificar se as características de desempenho dos equipamentos estão compatíveis com o exigido pelo método em estudo;
- ✓ Qualificar os materiais, por exemplo, padrões e reagentes;
- ✓ Planejar os experimentos de validação, incluindo o tratamento estatístico a ser utilizado;
- ✓ Realizar os experimentos de validação;
- ✓ Realizar a análise crítica dos resultados obtidos, considerando os critérios de aceitação;
- ✓ Concluir se o método é adequado ao uso pretendido.
- ✓ Os resultados devem estar documentados e registrados de modo organizado e facilmente acessíveis.

### **1.5.2 Validação de métodos como ferramenta estatística**

Segundo Vocabulário Internacional de Metrologia (VIM) (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a), a partir da Abordagem de Erro, o valor verdadeiro de uma grandeza, é considerado único e, na prática, impossível de ser conhecido. E, a partir da Abordagem de Incerteza (EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, 2012b), o valor verdadeiro de uma grandeza consiste no reconhecimento de que, devido à quantidade intrinsecamente incompleta de detalhes na definição de uma grandeza, não existe um valor verdadeiro único, mas sim um conjunto de valores verdadeiros consistentes com a definição.

Neste contexto, a validação do método é uma ferramenta que faz uso de um conjunto de parâmetros de validação, que estudados e avaliados por testes estatísticos investigam quaisquer suposições nas quais o método analítico se baseia estabelecendo as características de desempenho para documentar e, assim, demonstrar se este é adequado para um propósito analítico específico (THOMPSON; ELLISON; WOOD, 2002). Ou seja, a validação é uma ferramenta de análise utilizada para estimar os equívocos e desvios relacionados à uma proposta de método

analítico, bem como de avaliar se amplitude desses desvios é adequada aos critérios propostos para aquela análise. Pode, ainda, encontrar o conjunto de valores verdadeiros consistentes que podem ser atribuídos a uma medição realizada a partir do método desenvolvido (informações para estimar a incerteza da medição).

Os erros que a validação de métodos objetiva avaliar são, normalmente, os erros sistemáticos e os erros aleatórios, que segundo (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a) são:

**Erro sistemático:** Componente do erro que, em medições repetidas, permanece constante ou varia de maneira previsível. Para um erro sistemático, um valor de referência (utilizado como base para comparação com valores de grandezas da mesma natureza) é um valor verdadeiro (o valor verdadeiro de uma grandeza é considerado único e, na prática, impossível de ser conhecido), ou um valor medido de um padrão com incerteza de medição desprezável ou um valor convencional. As causas podem ser conhecidas ou não, e é possível aplicar uma correção para compensar um erro sistemático conhecido. Matematicamente é igual à diferença entre o erro de medição e o erro aleatório (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 22 adaptado).

**Erro aleatório:** Componente do erro de medição que, em medições repetidas, varia de maneira imprevisível. O valor de referência para um erro aleatório é a média que resultaria de número infinito de medições repetidas de um mesmo mensurando. Os erros aleatórios de um conjunto de medições repetidas formam uma distribuição que pode ser resumida por sua esperança matemática ou valor esperado, o qual é geralmente assumido como sendo zero, e por sua variância. Matematicamente, é igual à diferença entre o erro de medição e o erro sistemático (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 22 adaptado).

Para uma validação de método ser consistente, existem documentos orientativos e normativos (neste trabalho, por convenção chamado “protocolos”) que dispõe de informações sobre validações de métodos com focos em diferentes escopos de aplicação.

### 1.5.2.1 Protocolos de validação

Os parâmetros de validação, e seus respectivos critérios de aceitação devem ser definidos de acordo com as características do analito e da natureza do método (BRASIL, 2017). Ou seja, para cada objetivo e escopo de um método, o estudo e condução da validação podem ser diferentes. Isto ocorre devido à grande variedade de necessidades diferentes que podem surgir oriundas da combinação dos tipos de

matriz, analitos, nível de concentração dos analitos, nível de concentração dos interferentes, métodos de preparo de amostra, métodos de separação, extração, detecção, quantificação, equipamentos etc. Por este motivo, existem vários protocolos de validação diferentes, cada qual com seu respectivo objetivo, que dispõe de orientações e critérios de validação como protocolos, guias, normas, resoluções etc.

Esses diversos protocolos para validação de métodos, incluem informações sobre diferentes parâmetros de validação como exatidão, limite de detecção, precisão, recuperação e linearidade (RAPOSO; IBELLI-BIANCO, 2020), mas também existem em cada um deles condições de ensaios específicas e critérios de aceitação de resultados, bem como recomendações de parâmetros de validação que não são comuns a todos os protocolos, como “Efeito matriz” e “Robustez”, por exemplo. A Tabela 1 traz informações de alguns protocolos e suas aplicações, utilizados neste trabalho.

Tabela 1 - Protocolos de validação e seus campos de aplicação

Protocolo	Ano	Órgão responsável	Aplicação
RDC Nº 27 (BRASIL, 2012)	2012	ANVISA	Dispõe sobre os requisitos mínimos para a validação de métodos bioanalíticos empregados em estudos com fins de registro e pós-registro de medicamentos.
RDC nº 166 (BRASIL, 2017)	2017	ANVISA	Dispõe sobre a validação de métodos analíticos e é aplicável a métodos analíticos empregados em insumos farmacêuticos, medicamentos e produtos biológicos em todas as suas fases de produção.
DOQ-CGCRE-008 (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020)	2020	Inmetro	Este documento tem como objetivo auxiliar os laboratórios de ensaio na tarefa de demonstrar que um método analítico, nas condições em que é praticado, tem as características necessárias para a obtenção de resultados com a qualidade exigida.

Protocolo	Ano	Órgão responsável	Aplicação
Manual de Garantia da Qualidade Analítica - Resíduos e Contaminantes em Alimentos (BRASIL, 2011)	2011	MAPA	Este Manual foi elaborado com o objetivo de estabelecer parâmetros e critérios de aceitação aplicáveis à validação de métodos analíticos e à rotina analítica dos laboratórios públicos e privados que pertencem ou desejam pertencer à Rede Nacional de Laboratórios Agropecuários, credenciados ou autorizados a realizar análises para o Plano Nacional de Controle de Resíduos e Contaminantes.
ICH Q2 (R1) (ICH, 2005)	2005	EMA/FDA	Este documento apresenta uma discussão sobre as características a serem consideradas durante a validação dos procedimentos analíticos incluídos como parte dos pedidos de registro apresentados na Europa, Japão e EUA.

Protocolo	Ano	Órgão responsável	Aplicação
Guia de Validação de Métodos Bioanalíticos (EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012)	2012	EMA/FDA	Esta diretriz fornece recomendações para a validação de métodos bioanalíticos aplicados para mensurar concentrações de fármacos em matrizes biológicas de origem animal para estudos toxicocinéticos e em todas as fases dos ensaios clínicos.

Fonte: Própria do autor

### 1.5.2.2 Parâmetros de validação e testes estatísticos

É possível observar que os pontos-chaves para investigar as melhores condições analíticas e realizar um bom planejamento experimental é a definição dos parâmetros de validação dentro de suas condições de ensaio e critérios de aceitação de resultados. Para isto, o objetivo e o escopo do método devem estar definidos, pois é partir dessas características que definem os parâmetros e critérios, bem como qual o protocolo correto para aplicação (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017; HARRIS, 2010; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020; SKOOG, 2014).

Como citado, existem vários protocolos de validação, que dispõem das condições nas quais os parâmetros de validação são relevantes para o objetivo daquele método, quais as condições que devem ser realizados esses ensaios, quais critérios de análise de dados e testes estatísticos a serem utilizados, e quais os critérios de aceitação de resultados dos testes estatísticos.

Os parâmetros na validação de métodos são os indicadores da confirmação (ou não) e de que o método pode ser utilizado com confiabilidade (ou não) para a determinação de uma concentração de analito em uma matriz, ou seja, que ele é

adequado para a finalidade pretendida (ou não) (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017). Os resultados dos estudos desses parâmetros declarados em um documento (um relatório de validação por exemplo) fornecem evidência objetiva de que o método é ou não adequado ao uso pretendido. Para cada parâmetro de validação, ensaios laboratoriais e testes estatísticos específicos são realizados e avaliados conforme o objetivo do método analítico. A seguir estão alguns exemplos de parâmetros de validação, quais são seus objetivos:

### 1.5.2.3 Linearidade

O estudo da linearidade tem como objetivo avaliar a capacidade de um método analítico (dentro de uma dada faixa de concentração) em obter resultados os quais são diretamente proporcionais à concentração do analito na amostra (ICH, 2005). A linearidade é uma propriedade importante dos métodos, usada para fazer medições em uma gama de concentrações. Pode-se determinar a linearidade da resposta a padrões puros e a amostras reais. A linearidade não é, geralmente, quantificada, mas é verificada por inspeção ou usando testes de significância para não-linearidade (EURACHEM, 2012).

A avaliação da linearidade de um método não pode ser realizada apenas por meio da observação do gráfico dos resultados de resposta em função da concentração do analito. Para o estudo deste parâmetro, é necessário investigar a melhor função  $f(x)$  que representa o comportamento linear esperado. Neste contexto, o estudo desta equação deve ser realizado a partir de vários testes e avaliações estatísticas (SKOOG, 2014). A Tabela 2 traz quais pontos são importantes de ser avaliado no estudo da linearidade (BRASIL, 2017; JURADO et al., 2017; OLIVARES; LOPES, 2012; PACCES; OLIVARES, 2020; RAPOSO, 2016; SANAGI et al., 2010; SANCHEZ, 2018), bem como exemplos de testes estatísticos que podem ser utilizados para tal:

Tabela 2 - Fatores importantes para estudo de linearidade ou curva de calibração

O que é necessário avaliar	Testes estatísticos ou matemáticos necessários para avaliação
1. O grau de correlação entre as variáveis da curva construída;	Coefficiente de correlação ( $r$ ) (THOMPSON; PHILIP J, 2011)
2. A proporção de variabilidade nas respostas;	Coefficiente de determinação ( $r^2$ ) (THOMPSON; PHILIP J, 2011)
3. Como as concentrações e as respostas obtidas pela análise se relacionam;	Equação da reta (THOMPSON; PHILIP J, 2011) (MILLER; MILLER, 2010)
4. Qual o melhor modelo de regressão a ser utilizado;	Teste de homogeneidade (ou não) das variâncias dos resultados experimentais entre os diferentes níveis de concentração (Ex. Teste de Cochran) (MILLER; MILLER, 2010)
5. Se realmente há variação significativa da resposta do método em decorrência da variação da concentração;	Análise da variância dos resíduos (ANOVA) (Ex. Teste F) (MILLER; MILLER, 2010)
6. Se há diferença estatística significativa entre o valor médio do coeficiente linear (intercepto) e o valor "0"	Teste de Significância para Intercepto (Ex. Teste T) (MILLER; MILLER, 2010)
7. Exatidão	Tendência de medição ( <i>measurement bias; bias</i> ): estimativa de um erro sistemático ou recuperação: Desvio do valor obtido experimentalmente com relação ao valor obtido por meio da equação que define a regressão linear (admitido como verdadeiro). (EURACHEM, 2017; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a; KONIECZKA; NAMIESNIK, 2018; SKOOG, 2014)

O que é necessário avaliar	Testes estatísticos ou matemáticos necessários para avaliação
<b>8. O grau de dispersão das replicatas por nível de concentração.</b>	CV - Coeficiente de variação (desvio padrão relativo) (SKOOG, 2014)
<b>9. Os valores absolutos e relativos da diferença entre o valor experimental e o valor teórico.</b>	Análise de Resíduos - obtido por meio de da respectiva equação da regressão linear (MILLER; MILLER, 2010)
<b>10. Se existem valores discrepantes de resíduos (<i>Outliers</i>)</b>	Considerando como universo amostral todo o conjunto de dados, avaliar os resíduos. (Ex. Teste Q, Teste de Grubbs, Teste G) (LEITE, 2008)
<b>11. A presença de autocorrelação nos erros de um modelo de regressão.</b>	Teste de Independência para Resíduos (Ex. Teste Durbin-Watson) (PETER C. M.; RICHARD, 2000)
<b>12. Verificar se a distribuição dos resíduos é normal</b>	Teste de Normalidade para Resíduos (Ex. Teste de Shapiro-Wilk) (SHAPIRO; WILK, 1965)

Fonte: Própria do autor

#### 1.5.2.4 Precisão

O estudo de precisão dentro da validação tem como objetivo avaliar o grau de concordância entre indicações ou valores medidos, obtidos por medições repetidas, no mesmo objeto ou em objetos similares, sob condições especificadas (VIM, 2012). Normalmente é expressa em porcentagem, pois é obtida a partir do cálculo do Coeficiente de Variação (CV), ou seja, do desvio padrão relativo (DPR). Ela pode ser obtida e recomendada de acordo com três condições diferentes por alguns protocolos de validação: repetibilidade, intermediária e reprodutibilidade.

- **Repetibilidade de medição:** é a precisão de medição sob um conjunto de condições de repetibilidade (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Condições de repetibilidade: Condição de medição em um conjunto de condições, as quais incluem o mesmo procedimento de medição, os mesmos operadores, o mesmo sistema de medição, as mesmas condições

de operação e o mesmo local, assim como medições repetidas no mesmo objeto ou em objetos similares durante um curto período (VIM, 2012).

- **Precisão intermediária de medição:** precisão de medição sob um conjunto de condições de precisão intermediária (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Condições de precisão intermediária: Condição de medição em um conjunto de condições, as quais compreendem o mesmo procedimento de medição, o mesmo local e medições repetidas no mesmo objeto ou em objetos similares, ao longo de um período extenso de tempo, mas pode incluir outras condições submetidas a mudanças (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).
- **Reprodutibilidade de medição:** precisão de medição conforme um conjunto de condições de reprodutibilidade (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Condição de reprodutibilidade de medição: Condição de medição em um conjunto de condições, as quais incluem diferentes locais, diferentes operadores, diferentes sistemas de medição e medições repetidas no mesmo objeto ou em objetos similares (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

#### 1.5.2.5 Exatidão de medição

Realizar o ensaio de exatidão de medição dentro da validação, tem como objetivo avaliar o grau de concordância entre um valor medido e um valor verdadeiro de um mensurando (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Pode ser definida como a proximidade entre o valor determinado obtido pelo método e a concentração nominal do analito (expressa em porcentagem) (EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012). A exatidão pode ser chamada de “acurácia” (do inglês *accuracy*), e muitos protocolos de validação internacionais consideram tal nomenclatura (EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012; ICH, 2005). Em alguns casos, a exatidão é avaliada numericamente por meio da porcentagem de recuperação ou da tendência (*Bias* do inglês), dependendo do objetivo do protocolo de validação utilizado. A tendência de medição é a estimativa

de um erro sistemático (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a), e pode ser calculada como o valor de recuperação, utilizando-se de materiais de referência certificados (MRC), por exemplo, e quando aplicada a uma série de resultados de ensaio, implica em uma combinação de componentes de erros aleatórios e sistemáticos (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020). A tendência expressa como recuperação analítica, definida como:

$$\frac{\text{valor observado}}{\text{valor esperado}} \times 100\% \quad (1)$$

Esta tendência deve ser corrigida ou demonstrada ser desprezível, mas, em ambos os casos, a incerteza associada à determinação da tendência permanece como um componente essencial da incerteza global.

Outro modo de avaliar a exatidão é mediante o uso correto dos materiais de referência certificados (MRC) que consiste em avaliar o desempenho do método, por meio do cálculo do erro relativo (ER), expresso em porcentagem, e erro normalizado (EN):

$$ER = \frac{X_{lab} - X_v}{X_v} \cdot 100 \quad (2)$$

Sendo:

$X_{lab}$  = valor obtido ou média aritmética de valores obtidos experimentalmente

$X_v$  = valor admitido como verdadeiro (valor certificado do MRC)

$$E_n = \frac{(X_{lab} - X_v)}{\sqrt{U_{lab}^2 + U_{ref}^2}} \quad (3)$$

Sendo:

$U_{ref}$ : incerteza associada ao MRC.

Se  $|E_n| \leq 1$ , considera-se que o resultado do laboratório é adequado

### **1.5.2.6 Limite de Quantificação (LQ)**

Estudar e avaliar o limite inferior de quantificação ou limite de quantificação, consiste em encontrar a menor concentração do analito na curva de calibração preparada na matriz, com precisão e exatidão adequadas (BRASIL, 2012; EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012). Também pode ser a menor quantidade do analito na amostra que pode ser quantitativamente determinada com precisão e exatidão aceitáveis (ICH, 2005). Neste contexto, “aceitáveis” são os resultados de precisão e exatidão que estariam de acordo com critérios de aceitação de resultados de um protocolo de validação adequado ao objetivo daquele método, ou outro critério científica e estatisticamente justificado pelo autor. Para se determinar o LQ, existem várias condições de ensaio, cálculos e testes estatísticos recomendados que podem estimar este parâmetro (mas não determinar, uma vez que o valor deve ser confirmado experimentalmente se este se apresenta como verdadeiro e passível de ser determinado).

Independentemente do método utilizado para se realizar a estimativa do LQ, normalmente, o valor encontrado deve ser posteriormente confirmado pela análise de um número adequado de amostras conhecidas no limite de quantificação, para verificar e conformar a precisão e exatidão do valor estimado (BRASIL, 2017; ICH, 2005).

### **1.5.2.7 Limite de Detecção (LD)**

Estimar e avaliar o Limite inferior de detecção, ou limite de detecção, consiste em encontrar a menor quantidade de analito que pode ser detectada na amostra utilizando-se o método analítico em questão, mas esse valor de concentração, não necessariamente quantificada sob as condições estabelecidas para o ensaio (ICH, 2005). O limite de detecção para um procedimento analítico pode variar em função do tipo da amostra. É fundamental assegurar-se de que todas as etapas de processamento do método analítico sejam incluídas na determinação LD, também é importante para ensaios qualitativos (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020). Assim como o LQ, existem várias condições de ensaio, cálculos e

testes estatísticos recomendados que podem estimar este parâmetro (mas não determinar, uma vez que o valor de LD deve ser confirmado experimentalmente, se este apresentar, pelo menos, diferença entre o sinal do analito e do ruído do sistema).

Em procedimentos analíticos que apresentem ruído de linha de base, a relação sinal/ruído pode ser utilizada e determinada pela comparação dos sinais medidos de amostras com baixas concentrações conhecidas do analito e dos ruídos dos brancos de amostras, definindo-se a concentração mínima em que o analito pode ser detectado com confiança. Uma relação sinal/ruído de 3:1 ou 2:1 é geralmente considerada aceitável para a estimativa do limite de detecção. É importante ressaltar que a região do ruído do branco deve ser a mesma do sinal medido (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020).

Nos casos em que um valor estimado para o limite de detecção é obtido por cálculo ou extrapolação, esta estimativa pode ser subsequentemente validada pela análise independente de um número adequado de amostras conhecidas por estarem próximas ou preparadas no limite de detecção (ICH, 2005). Se o LD for determinado com base na avaliação visual ou na relação sinal/ruído, a apresentação dos cromatogramas relevantes é considerada aceitável para justificativa (ICH, 2005).

É importante destacar que a estimativa deste parâmetro de validação por meio da relação sinal/ruído, faz sentido apenas para os métodos que utilizem equipamentos de medição que permitam este tipo de estimativa. Para outras abordagens, estimativas com base nos desvios padrão das curvas de calibração podem ser abordados (BRASIL, 2012, 2017; EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012; ICH, 2005),

#### **1.5.2.8 Dados de Validação de Métodos e Apresentação de Resultados**

Dentro da validação de método, a avaliação dos dados obtidos é a última etapa antes de apresentar os resultados. Esse processo tem início com a avaliação das matrizes a serem estudadas no processo de validação, dos métodos utilizados nos planejamentos experimentais e definição das condições analíticas (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017; BARWICK, 2019; EURACHEM, 2012, 2017; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; LEITE, 2008; SKOOG, 2014) e se estende até a obtenção das repostas analíticas. Desta forma,

para concluir se o método é adequado ao uso pretendido, é preciso realizar a análise crítica das respostas analíticas obtidas, considerando os critérios de aceitação estabelecidos (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017; EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020).

Segundo o VIM ((INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a), o resultado de uma medição é:

“Conjunto de valores atribuídos a um mensurando, juntamente com toda outra informação pertinente disponível.

NOTA 1 Um resultado de medição geralmente contém “informação pertinente” sobre o conjunto de valores, alguns dos quais podem ser mais representativos do mensurando do que outros. Isto pode ser expresso na forma de uma função densidade de probabilidade (FDP).

NOTA 2 Um resultado de medição é geralmente expresso por um único valor medido e uma incerteza de medição. Caso a incerteza de medição seja considerada desprezável para alguma finalidade, o resultado de medição pode ser expresso como um único valor medido. Em muitos domínios, esta é a maneira mais comum de expressar um resultado de medição.

NOTA 3 Na literatura tradicional e na edição anterior do VIM, o resultado de medição era definido como um valor atribuído a um mensurando obtido por medição, que podia se referir a uma indicação, ou um resultado não corrigido, ou um resultado corrigido, de acordo com o contexto.”

De forma geral, uma vez registrado os resultados obtidos, cabe ao desenvolvedor do método, de acordo com regras definidas, realizar a análise crítica desses resultados para ser possível concluir se o método é ou não adequado ao uso pretendido. (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020). Ou seja, testes estatísticos adequados e suficientes devem aplicados às respostas analíticas a fim de avaliar os parâmetros de validação escolhidos. Então, de acordo com os resultados destes testes estatísticos é que se determina se os parâmetros de validação atendem ou não aos critérios estabelecidos previamente, e por intermédio de uma avaliação crítica geral do processo de validação, é possível concluir se o método é ou não adequado ao uso (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020).

Em relação a obtenção de resultados, valores verdadeiros são impossíveis de ser conhecidos (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Na verdade, o que é possível obter é um conjunto de valores que têm alta probabilidade

de serem verdadeiros e consistentes. Desta forma, após uma avaliação dos dados obtidos da validação do método, os resultados devem ser reportados com a possível dispersão dos valores que poderiam ser razoavelmente atribuídos aos seus resultados, ou seja, com a incerteza de medição associada ao mensurando (BETTENCOURT DA SILVA; WILLIAMS, 2015; EURACHEM, 2012, 2017; OLIVARES; LOPES, 2012; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020).

Resumidamente, os resultados da validação de um método analítico devem ser apresentados com o respectivo intervalo de confiança do método, ou seja, com limites de incerteza válidos especificados e associados ao mensurando (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017; EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2020; MILLER; MILLER, 2010; SKOOG, 2014).

#### **1.5.2.9 Incerteza de medição**

Segundo a IUPAC, na validação de métodos, as características quantitativas de interesse estão relacionadas à acurácia do resultado provável de ser obtido. Portanto, é geralmente verdade dizer que a validação do método equivale à tarefa de fornecer insumos para estimar a incerteza da medição (THOMPSON; ELLISON; WOOD, 2002).

A incerteza de medição é:

“Parâmetro não negativo associado ao resultado de uma medição, que caracteriza a dispersão dos valores que poderiam ser razoavelmente atribuídos ao mensurando, com base nas informações utilizadas (EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Ela inclui componentes provenientes de efeitos sistemáticos, tais como componentes associadas a correções e a valores atribuídos a padrões, assim como a incerteza definicional (ver glossário). Algumas vezes, não são corrigidos efeitos sistemáticos estimados, mas, em vez disso, são incorporadas componentes de incerteza de medição associadas (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Geralmente para um dado conjunto de informações, subentende-se que a incerteza de medição está associada a um determinado valor atribuído ao mensurando. Uma modificação deste valor resulta em uma modificação da incerteza associada (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

A incerteza de medição geralmente engloba muitas componentes. Algumas

delas podem ser estimadas por uma avaliação do Tipo A da incerteza de medição, a partir da distribuição estatística dos valores provenientes de séries de medições e podem ser caracterizadas por desvios-padrão. As outras componentes, as quais podem ser estimadas por uma avaliação do Tipo B da incerteza de medição, podem também ser caracterizadas por desvios-padrão estimados a partir de funções de densidade de probabilidade baseadas na experiência ou em outras informações (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).”

Para entender melhor a importância de apresentar os resultados de uma validação de métodos com seu respectivo intervalo de confiança associado a situação a seguir traz um exemplo prático:

- A incerteza de medição associada a um método de quantificação de mercúrio em água é igual a 15,73%;
- Os limites deste metal no sangue humano devem ser inferiores a 10,00 µg/L;
- Os resultados obtidos em uma estação de tratamento de água de uma cidade qualquer, utilizando este método foram de 9,34 µg/L, 9,68 µg/L e 9,86 µg/L.
- Pergunta: esta água está própria para consumo?

Para responder à questão, é preciso associar a incerteza da medição do mensurando a estes resultados. Portanto, verificando o resultado de menor valor:

$$[9,34 \mu\text{g/L} + ((15,73/100) * 9,34 \mu\text{g/L})] = 10,81 \mu\text{g/L}$$

$$[9,34 \mu\text{g/L} - ((15,73/100) * 9,34 \mu\text{g/L})] = 7,87 \mu\text{g/L}$$

A dispersão dos valores que poderiam ser razoavelmente atribuídos ao resultado de 9,34 µg/L seria o intervalo de valores entre 7,87 µg/L e 10,81 µg/L. E aplicando o mesmo conceito para os outros resultados:

$$[9,68 \mu\text{g/L} + ((15,73/100) * 9,68 \mu\text{g/L})] = 11,20 \mu\text{g/L}$$

$$[9,68 \mu\text{g/L} - ((15,73/100) * 9,68 \mu\text{g/L})] = 8,16 \mu\text{g/L}$$

$$[9,86 \mu\text{g/L} + ((15,73/100) * 9,86 \mu\text{g/L})] = 11,41 \mu\text{g/L}$$

$$[9,86 \mu\text{g/L} - ((15,73/100) * 9,86 \mu\text{g/L})] = 8,31 \mu\text{g/L}$$

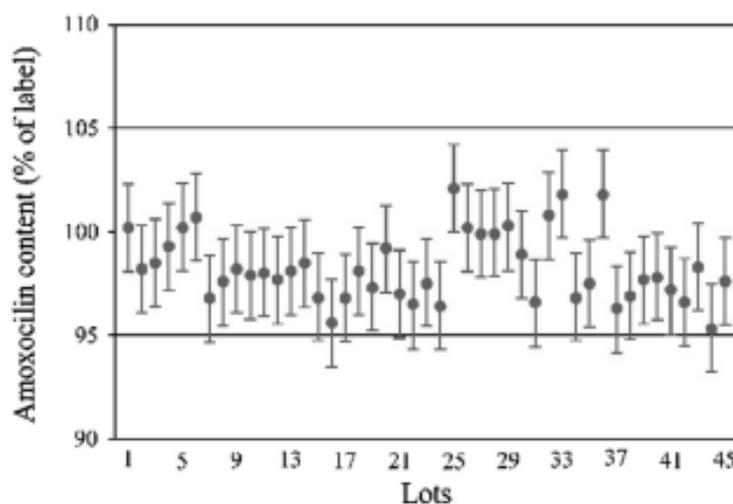
Têm-se que o intervalo de confiança atribuído ao resultado 9,68 µg/L está entre 8,16 µg/L e 11,20 µg/L, e ao resultado 9,86 µg/L está entre 8,31 µg/L e 11,41 µg/L.

Desta forma, os resultados de confiança que esta estação de tratamento possui para tomar uma decisão em relação a toxicidade da água estão no intervalo entre 7,87 µg/L e 11,41 µg/L. Sem a associação da incerteza de medição aos resultados, a água estaria própria para consumo humano, porém ao verificar o intervalo de confiança desses resultados, existe a possibilidade de que esta água não seja própria para consumo humano, o que exigiria uma tomada de decisão importante da parte dos responsáveis pela distribuição de água da cidade.

Os critérios de aceitação para os valores de estimativa de incerteza, assim como a própria validação, dependem do objetivo do método e do desenvolvedor do método, que pode tomar como base, por exemplo, os guias da EURACHEM, GUM (Guia para a Expressão da Incerteza de Medição) e JCGM (Comitê Conjunto para Guias em Metrologia) (BIPM, 2012; EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012b).

Em um artigo de 2018 (SEPAROVIC; SAVIANO; LOURENÇO, 2018), um estudo verificou a importância de realizar a avaliação da incerteza de medicamentos realizando a análise em lotes de Amoxicilina. O estudo estabeleceu que o critério de aceitação de variabilidade do teor da de amoxicilina nesse medicamento poderia variar entre 95% e 105%, avaliando 45 medicamentos. A Figura 3 mostra que todos os resultados que são obtidos diretamente do sinal do equipamento estão dentro do intervalo de confiança de 95% a 105%, ou seja, todos os resultados estão aprovados dentro da uma faixa de aceitação pré-estabelecido

Figura 3 - Conteúdo dos comprimidos de amoxicilina em 45 lotes com base nos resultados dos ensaios de HPLC e suas incertezas.



Fonte TRAPLE et al. Using measurement uncertainty to assess the fitness for purpose of an HPLC analytical method in the pharmaceutical industry, **Measurement**. p.44. 2014

Porém, esses são os resultados obtidos de um equipamento, eles, na verdade, não são uma estimativa do valor real baseado na capacidade do método de ensaio e seus equipamentos de forma conjunta. Desta forma, esses resultados são estatisticamente prováveis de estarem dentro de determinada faixa de valores, chamada de incerteza, apesar de neste caso, considerar apenas a dispersão de valores em relação ao equipamento e não a todo o processo de medição.

A partir do gráfico (Figura 3), é possível observar que alguns pontos não estão dentro dessa faixa de variação de incerteza estabelecidos, ou seja, existe uma probabilidade estatística de que podem estar fora do intervalo de confiança estabelecido (95%-105%). E esta é a grande importância de incerteza de medição ser avaliada. No caso destes estudos, por exemplo, a partir dos limites de incerteza estabelecidos, foi possível determinar se existe o risco de que este medicamento se encontrar fora da faixa e aceitação especificada. Ou seja, nesse estudo a incerteza se mostrou importante na avaliação da conformidade de produtos farmacêuticos. Além disso, apesar de utilizar métodos oficiais validados, neste caso específico, os resultados de incerteza associados demonstraram a probabilidade estatística de se obter um resultado fora de uma faixa de especificação.

Apresentar as conclusões de uma validação de métodos com respectiva a incerteza de medição associada ao valor do mensurando na qual o procedimento

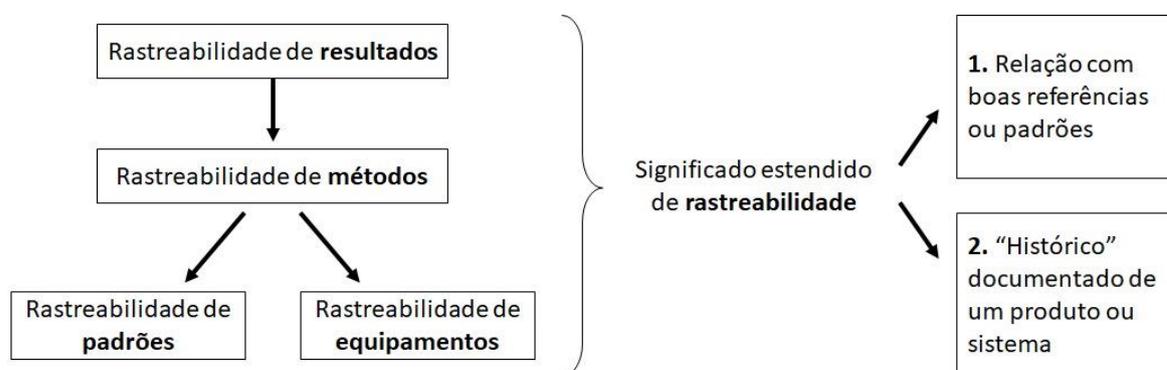
analítico tem como objetivo determinar, garantir e confirmar a validade e qualidade (ou não) dos resultados e valores que podem ser determinados utilizando tal método desenvolvido.

### 1.5.2.10 Relatórios de Validação

Ao final do desenvolvimento de um método analítico, enfim se fornecem as evidências objetivas de que o método é adequado para o uso pretendido. A forma mais comum de se evidenciar esta atividade é feita mediante apresentação da validação de métodos em um relatório. Um dos objetivos mais importantes dos relatórios de validação é fornecer rastreabilidade adequada pois, para produzir resultados confiáveis, a rastreabilidade é essencial, especialmente em química analítica (OLIVARES; LOPES, 2012).

A palavra “rastreabilidade” tem origem na metrologia. A rastreabilidade metrológica é uma propriedade de um resultado de medição, em que o resultado pode ser relacionado a uma referência por meio de uma cadeia ininterrupta de calibrações documentada, cada uma contribuindo para a incerteza da medição (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). A rastreabilidade dos resultados está relacionada à rastreabilidade de um método, que por sua vez está ligada à rastreabilidade dos padrões e à rastreabilidade do equipamento utilizado no procedimento analítico (OLIVARES; LOPES, 2012). A Figura 4 mostra em resumo um esquema que representa os diferentes conceitos e significados de rastreabilidade

Figura 4 - Diferentes conceitos e o significado estendido de rastreabilidade



Fonte: OLIVARES; LOPES, 2012. *Essential steps to providing reliable results using the Analytical Quality Assurance Cycle*. Trends in Analytical Chemistry. Vol. 35. p 112, 2012

Desta forma, para apresentação de um relatório de validação, o Guia de Validação de Métodos Bioanalíticos da Agência Europeia de Medicina (EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012) traz as seguintes recomendações:

- Em um relatório de validação, dependendo do nível de detalhe das informações fornecidas, apenas a referência aos procedimentos utilizados para etapas específicas de análise pode ser suficiente. Caso contrário, esses procedimentos devem ser anexados ao relatório. Todos os dados de origem devem estar disponíveis em seu formato original e disponíveis mediante solicitação;
- Qualquer desvio do protocolo de validação deve ser registrado;
- O relatório de validação deve incluir pelo menos as seguintes informações:
  - resumo dos desempenhos de validação;
  - detalhes do método analítico aplicado e, quando apropriado, a fonte do método analítico (referências da literatura e / ou modificações no procedimento);
  - detalhes do procedimento de ensaio (analito, padrão interno, pré-tratamento da amostra, extração e análise);
  - padrões de referência (origem, número do lote, certificado de análise, estabilidade e condições de armazenamento);
  - padrões de calibração e amostras de Controle de Qualidade (matriz, preparação, datas de preparação e condições de armazenamento);
  - atendimento aos critérios de aceitação;
  - atendimento aos critérios de análise;
  - tabela de todas as execuções analíticas com datas de análise, se foram aprovadas ou reprovadas e o motivo da falha;
  - tabela de resultados de calibração de todas as execuções analíticas aceitas, incluindo faixa de calibração, função de resposta, concentrações calculadas de volta e precisão;
  - tabela de resultados de Controle de Qualidade de todas as corridas analíticas aceitas (precisão e exatidão dentro e entre as corridas); valores fora dos critérios de aceitação devem ser claramente marcados;

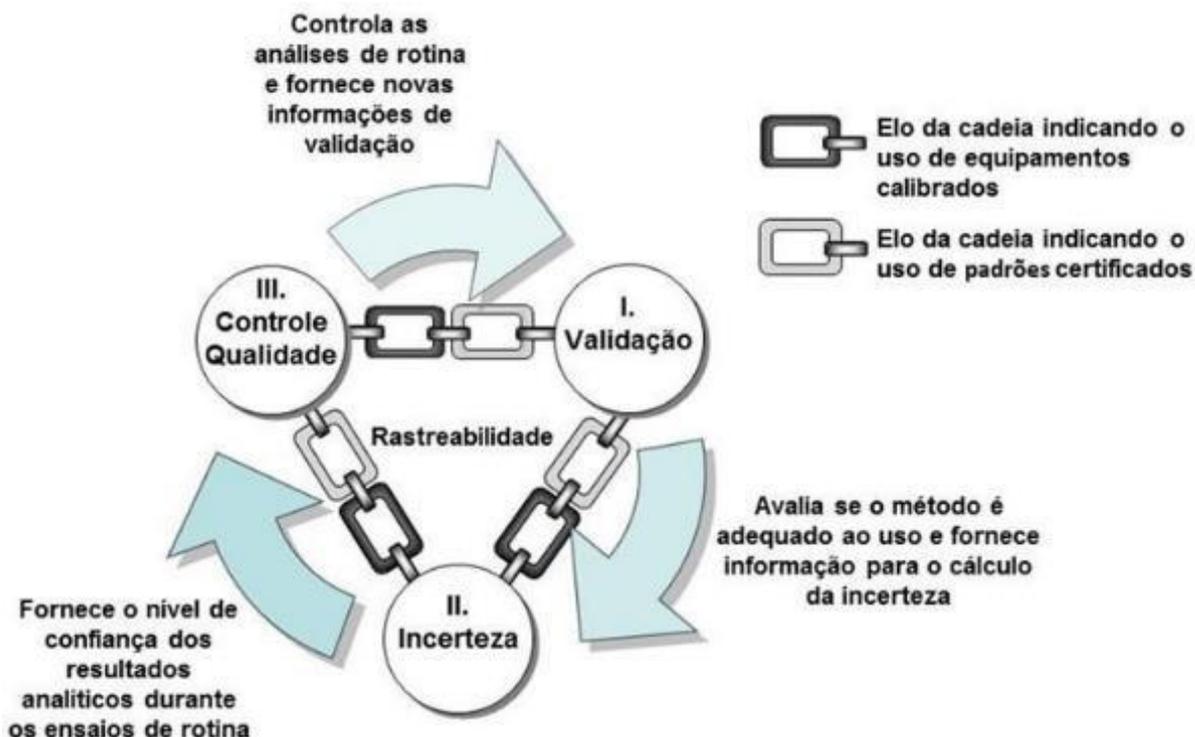
- dados de estabilidade da solução de estoque, solução de trabalho, QC, cobrindo as condições de armazenamento aplicadas;
  - dados sobre seletividade, LQ, efeito residual, efeito de matriz se aplicável, integridade de diluição;
  - resultados inesperados obtidos durante a validação com justificativa completa da ação tomada;
  - desvios do método e/ou procedimento oficial ou operacional padrão (descrição dos desvios, impacto no estudo, dados de suporte);
- Todas as medições com as concentrações individuais calculadas devem ser apresentadas no relatório de validação.

### **1.5.3 Ciclo de Garantia de Qualidade Analítica (*Analytical Quality Assurance Cycle (AQAC)*)**

Dentro da garantia da qualidade analítica podem-se diferenciar vários tipos de operações necessárias para garantir o controle de qualidade das medições analíticas, que estão inseridos na estrutura mais ampla da metrologia química e rastreabilidade (OLIVARES; LOPES, 2012).

Fundamentado então nestes conceitos, para garantir a validade dos resultados deste trabalho, a ferramenta de qualidade analítica que orientou todo o desenvolvimento do trabalho foi o Ciclo De Garantia de Qualidade Analítica (Ciclo AQAC) publicado em 2012 no artigo “Essential steps to providing reliable results using the Analytical Quality Assurance Cycle” (OLIVARES; LOPES, 2012). Assim como o Ciclo PDCA (Plan, Do, Check, Act), uma clássica ferramenta para gestão de qualidade de processos, o Ciclo AQAC consiste em um diagrama, que tem como objetivo facilitar a interpretação e aplicação da relação entre validação de métodos, incerteza de medição e controle de qualidade analítica (Figura 5).

Figura 5 - Ciclo De Garantia de Qualidade Analítica (Ciclo AQAC)



Fonte: OLIVARES; LOPES, 2012. *Essential steps to providing reliable results using the Analytical Quality Assurance Cycle*. Trends in Analytical Chemistry. Vol. 35. p 112, 2012.

As relações dispostas por intermédio do diagrama, que fecham um ciclo, simplificam a rastreabilidade adequada requerida para o controle e garantia da validade de resultados, porém os autores chamam a atenção para o fato de que conhecimento da relação entre eles é extremamente necessário e que, caso ocorra uma ação inadequada em algum dos três processos, todos os outros podem estar prejudicados.

A principal motivação do desenvolvimento do ciclo, segundo os autores, foi a importância que validação de métodos, estimativa da incerteza de medição e garantia de qualidade dos resultados de ensaio de fato tem, com base na forma que são abordadas pela ISO/IEC 17025:2017, como parte dos requisitos para a acreditação de laboratórios, mas que apesar da notória reputação desses três itens, foram citados como os mais difíceis de serem aplicados, quando comparados aos demais itens desta mesma norma, segundo um estudo envolvendo mais de 240 laboratórios químicos, todos credenciados na ISO/IEC 17025 (OLIVARES; LOPES, 2012).

O ciclo é utilizado como ferramenta central neste trabalho uma vez que a crise

de reprodutibilidade potencialmente repousa sobre os pilares da validação e da incerteza. Assim, o software utilizado (PACCES; OLIVARES, 2020) para a avaliação da correta aplicação dos cálculos envolvidos nas validações constantes no documento por este trabalho avaliadas, baseia-se nos princípios do Ciclo AQAC.

#### **1.5.4 Softwares ConfLab**

Os Softwares ConfLab são validados com o objetivo de assegurar que o programa cumpra com suas especificações, atendendo as exigências dos usuários (PACCES; OLIVARES, 2020b, 2020c). Baseiam-se no conceito do Ciclo de Qualidade Analítica (AQAC) (OLIVARES; LOPES, 2012), que relaciona validação, incerteza e controles de qualidade, apresentando três módulos: um que contempla todos os testes estatísticos capazes de atender aos protocolos de validação utilizados; o segundo, que faz uma adequada estimativa da incerteza da medição utilizando-se informações provenientes da validação (linearidade, precisão, exatidão,...) e de outras fontes, como equipamentos, massa, volume, amostragem, carta de controle, aplicados na execução do método; e um terceiro, o qual controla a qualidade do resultado analítico durante a rotina do laboratório a partir do uso de cartas de controle de média e amplitude. A Figura 6 apresenta a tela inicial dos softwares ConfLab, apresentando o acesso para cada software na forma do diagrama do Ciclo AQAC.

Figura 6 - Tela inicial dos Softwares ConfLab, com disposição baseada no Ciclo AQAC. Cada software possui acesso independente.

O que deseja fazer?

control. qualidade

validação/verificação

incerteza

CONF-036 C [DMA] - Determinação de Arsênio em arroz

EDITAR PROJETO

Empresa

ConfLab  
Avenida Trabalhador Sankarlenso, 400  
Parque Arnold Schmidt 13566-590 Universidade de São Paulo  
São Carlos SP  
BRASIL  
(16) 3373-6670

Pessoal

Responsável  
Vitor Hugo Polissé Paccos

Analistas  
Analista 1  
Analista 2  
Analista 3

Identificação

Método  
Determinação de hormônios sintéticos em urina

Analito  
Hormônios sintéticos

Matriz  
Urina animal

Documentos  
EMA - Guideline on bioanalytical method validation

Parâmetros

Precisão  
CV  $\leq$  15 %

Precisão intermediária  
CV  $\leq$  15 %

Exatidão  
CV  $\leq$  15 %  
Exatidão  $\geq$  85 % e  $\leq$  115 %  
Desvio  $\leq$  5 %  
Erro Normalizado

Linearidade  
 $r^2$   
 $r$   
Desvio  $\leq$  15 %  
CV  $\leq$  15 %  
CV (1º Ponto) = 20 %  
Resíduo Relativo  $\leq$  15 %

Rastreabilidade

Grupos de Equipamentos para Execução de Métodos:

Configuração 1

- UPLC Thermo Accela
- Espectrômetro de massa Thermo Q Exactive Heated Electrospray Ionization Source II (HESI II)

Insumos/Reagentes

Padrões de estanozolol, taleranol (8-zearalanol), zearalanona, zeranol (a-zearalanol), hexestrol, dienesterol, 17 $\alpha$ -etionlestradiol, dietilestilbestrol e trenbolona foram obtidos da Sigma-Aldrich (St. placa de poços (12  $\times$  8) (Waters Conn. Milford MA.

Fonte: (PACCES; OLIVARES, 2020a)

#### 1.5.4.1 ConfLab Validação (Versão 1.0.1.0)

Com foco na ISO 17025:2017, mediante a inserção dos valores experimentais, os cálculos de diferentes parâmetros estatísticos são realizados de forma instantânea, apresentando os resultados que podem ser impressos em formato de um relatório de validação personalizável, com os dados de rastreabilidade exigidos pelos sistemas de gestão da qualidade. Além da ISO 17025:2017, é também compatível e atende as exigências das normas neste trabalho. Os parâmetros de validação que este módulo calcula são:

- Precisão (Repetibilidade)
- Precisão Intermediária
- Exatidão
  - Exatidão com Amostras Fortificadas
  - Exatidão com Uso de Material de Referência
- Linearidade
  - Avaliação da Homocedasticidade
  - Regressão dos Mínimos Quadrados Ordinários (MMQO) e Regressão dos Mínimos Quadrados Ponderada (MMQP)
  - Testes Gerais
  - Normalidade

- Independência
- Anova
- Significância do Intercepto
- Resíduos Absolutos
- Resíduos Relativos
- Outliers (Grubbs por Nível)
- Outliers (Grubbs total)
- Exatidão
- Coeficiente de Variação
- Regressão dos Mínimos Quadrados Ordinários (MMQO) Versus Regressão dos Mínimos Quadrados Ponderados (MMQP)
- LOD e LOQ
  - Cálculos Teóricos - Desvios padrão do Branco da Amostra (ou menor nível quantificável) - IUPAC/INMETRO
  - Cálculos Teóricos - Desvios padrão do Branco da Amostra (ou menor nível quantificável) - FDA/ANVISA
  - Cálculos Teóricos - Desvio padrão de Intercepto do Sinal - FDA/ANVISA
  - Cálculos Experimentais
- Efeito Matriz
  - Diferenças Entre As Curvas De Calibração
  - Desvios Relativos aos FMNs
- Robustez
  - Teste de Youden

#### 1.5.4.2 ConfLab Incerteza (*Versão 1.0.1.0*)

O Software de Incerteza ConfLab estima as variações decorrentes dos processos analíticos na forma de um resultado aplicado ao mensurando. É baseado nos conceitos abordados pelo Guia Eurachem/CITAC de estimativa de incerteza para química analítica (EURACHEM; CITAC, 2012).

Cada fonte de incerteza, pode ser inserida diretamente em um diagrama de Ishikawa disposto em tela (estas entradas do diagrama podem ser orientadas pela descrição da equação do mensurando que pode ser inserida no próprio software), ou também, é possível buscar os dados do módulo de validação para o de estimativa de incerteza. Os tipos de fonte de incerteza disponíveis estão citados abaixo, e podem ser escolhidas conforme a necessidade de cada aplicação de ensaio. Ao final da inserção das funções, todos os resultados são apresentados em um diagrama de Pareto que permite avaliar as maiores contribuições seguido da incerteza final. Os tipos de funções disponíveis são:

- Incerteza da Linearidade
  - Modelo e regressão ordinária;
  - Modelo de regressão ponderada;
- Incerteza do estudo de Precisão (Repetibilidade)
- Incerteza do estudo de Precisão (Intermediária)
- Incerteza da Exatidão (Amostras Fortificadas)
- Incerteza da Exatidão (Uso de Material de Referência)
- Incerteza da Massa
- Incerteza do Volume
- Incerteza na Padronização de Soluções
- Incerteza do Material de Referência
- Incerteza de Outras Fontes ou Equipamentos
- Incerteza da Amostragem
  - Anova;
  - Range Double Split;
- Incerteza da Carta de Controle

Desta forma, considerando:

- A necessidade de uma melhor compreensão das limitações do uso da estatística como um critério de aceitação de resultados, visto que foi um dos fatores que podem afetar a reprodutibilidade científica, segundo o artigo de Monya Baker (BAKER, 2016);
- E que dentro da química analítica, esses fatores podem ser mensurados, dentre os outros citados (BAKER, 2016);
- Com base em (GOODMAN; FANELLI; IOANNIDIS, 2018), pode existir uma necessidade, de um melhor entendimento da relação entre reprodutibilidade e evidências verdadeiras das afirmações científicas e;
- Podem existir dúvidas e controversas em relação as definições de quais dados precisam ser investigados para garantir a reprodutibilidade em química analítica;

Foram definidos os objetivos desse trabalho descritos no próximo tópico.

## 2 Objetivos

Valendo-se de resultados de métodos analíticos publicados e com base nos conceitos da Ciclo AQAC (OLIVARES; LOPES, 2012), os objetivos do trabalho foram:

- I. Avaliar se a percepção da existência de uma crise de reprodutibilidade em relação a validação de métodos analíticos da área de química analítica (e afins) é passível de ser confirmada e quantificada, e para isto:
  - 1.1 Avaliar se a aplicação da validação de métodos, foi adequada e suficiente e se foram utilizados protocolos de validação apropriados considerando os métodos desenvolvidos e publicados nos documentos científicos analisados;
  - 1.2 Avaliar com o apoio de softwares previamente validados (ConfLab Validação e ConfLab Incerteza) (PACCES; OLIVARES, 2020) se os cálculos matemáticos e testes estatísticos dentro do estudo da validação dos métodos desenvolvidos e publicados nos documentos científicos analisados são adequados e suficientes;
  - 1.3 Investigar os equívocos mais comuns cometidos durante a validação de métodos;
  - 1.4 Investigar a existência de relação entre os itens 1.1 a 1.3;
- II. Investigar a existência de relação entre as conclusões encontradas no item I com estimativa de incerteza de medição, com base na gestão da qualidade analítica utilizando o Ciclo AQAC (OLIVARES; LOPES, 2012);
- III. Investigar a existência de relação com item II com a crise de reprodutibilidade.

## 3 Desenvolvimento

### 3.1 Informações gerais

A busca por documentos científicos foi realizada em plataformas de busca de literatura acadêmico-científica *online*, bancos de dados e bibliotecas virtuais, com base em alguns critérios de seleção apresentados no item 3.2.1 para tornar possível as avaliações pretendidas.

Com o auxílio dos Softwares ConfLab (CONFLAB, 2020), foram recalculados os resultados dos parâmetros estatísticos de validação e incerteza de medição, utilizando os dados brutos disponíveis tal e qual se apresentavam nos documentos. Com base nos critérios de aceitação de resultados das normas, os valores retornados pelos softwares foram submetidos à avaliação da conformidade da validação do método em estudo.

Os resultados foram apresentados em forma resumida nos gráficos nos subitens no corpo deste trabalho, e as discussões mais detalhadas e relatórios, estarão disponíveis no material suplementar em formato .pdf e independentes (FERREIRA, 2020). No material suplementar 1, foram realizadas discussões mais detalhadas dos primeiros 18 documentos em relação as questões dos testes estatísticos de cada parâmetro de validação, de forma a demonstrar o grau de profundidade desta pesquisa. Nos Materiais suplementares 2 e 3 estão, respectivamente, os relatórios de validação e incerteza dos métodos estudados gerados pelos softwares ConfLab.

Ao todo, foram analisadas 91 validações de métodos, presentes em 42 documentos científicos, que geraram 139 relatórios de validação e 139 relatórios de incerteza. A quantidade de validações é diferente da quantidade de documentos científicos, pois, dentro de um mesmo documento, mais de um método com condições analíticas, matrizes e/ou analitos diferentes foram estudados. Em relação aos números de relatórios, estes foram maiores do que o número de validações, pois em alguns casos:

- além de mais de um método estar presente, dentro de cada método, em alguns casos foram verificados as melhores e os piores intervalos de abrangência;
- quando a regressão linear deveria ser realizada de forma ponderada, mas o

autor realizou de forma ordinária, foram gerados relatórios de incerteza comparando a diferença entre os valores de incerteza nos dois casos;

- dentro de cada método, várias curvas de calibração com intervalos de concentração diferentes foram estudadas.

Nos primeiros 18 documentos foram realizadas análises mais detalhadas para se compreender o desenvolvimento e aplicação das ferramentas em cada um destes projetos. Estes documentos abordam uma discussão mais focada no estudo de cada teste estatístico para cada parâmetro de validação separadamente.

## **3.2 Objetos do estudo**

Para manter a confidencialidade da pesquisa, preservando os autores, instituições e revistas, foi adotada uma nomenclatura codificada para padronizar o tratamento de dados, combinando letras e números de acordo com o projeto criado no software. A nomenclatura possui as letras iniciais COD, seguida de uma sequência de três números em algarismos crescente por ordem de documento a ser avaliado. Caso um mesmo documento apresentasse mais de uma validação em seu conteúdo (com diferentes métodos, ou diferentes analitos, matrizes etc.) é adicionado uma letra em ordem alfabética para caracterizar que aquela validação ou incerteza pertence a um mesmo documento.

### **3.2.1 Seleção dos documentos avaliados**

Para o objetivo pretendido, os documentos elegíveis foram selecionados de acordo com os seguintes critérios:

- se os métodos estavam presentes entre artigos, teses, dissertações e trabalhos de conclusão de curso (TCC).
- tratavam sobre desenvolvimento de método analítico com objetivo de detectar, quantificar, identificar, caracterizar algum analito em alguma matriz;
- continham os dados brutos obtidos durante o processo de validação (e não apenas as respectivas médias) em pelo menos dois parâmetros, de preferência, sendo um deles o da linearidade.

### 3.2.2 Seleção dos parâmetros de avaliação

Principais aspectos de avaliação foram:

- Cálculos matemáticos e estatísticos;
- Condições de realização dos ensaios dos parâmetros de validação;
- Critérios de aceitação de resultados dos parâmetros de validação;
- Apresentação dos resultados e rastreabilidade adequada.

Os parâmetros de validação selecionados para avaliação foram:

- Linearidade;
- Precisão repetibilidade;
- Precisão intermediária;
- Exatidão;
- Limite de quantificação;
- Limite de detecção;
- Estimativa de incerteza de medição expandida.

Os motivos desta escolha foram:

- Estão presentes em todos os protocolos de validação utilizados;
- Seria possível avaliar junto ao ConfLab e utilizá-lo para obter os resultados de referência;
- Existiam mais dados disponíveis para o estudo.

### 3.2.3 Ciclo de Garantia de Qualidade Analítica (Analytical Quality Assurance Cycle (AQAC))

De posse das informações até aqui relatadas, o desenvolvimento deste trabalho teve como base o conceito de garantia de qualidade analítica, que atende à necessidade de fornecer resultados confiáveis que atendam aos requisitos de qualidade, dentro da gestão da qualidade (OLIVARES, 2019; OLIVARES; LOPES, 2012; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020).

### 3.2.4 Aplicação dos Softwares ConfLab

Mediante os relatórios de validação e incerteza fornecidos pelos softwares, foi possível fornecer evidências da conformidade, ou não, dos métodos aos requisitos especificados pelas legislações vigentes, comparando os resultados obtidos pelo

documento e pelo software. Além disso, também foi realizado a avaliação da conformidade da validação do método em relação aos critérios de aplicação do respectivo protocolo, comparando a forma na qual foi exigida a aplicação do parâmetro na norma e a forma que foi realizada no documento científico. Ao relacionar os resultados encontrados (material suplementar 1, 2 e 3) (FERREIRA, 2020), foi possível fornecer discussões e dados que permitiram mostrar se os métodos atenderam ou não aos requisitos especificados.

Os relatórios de validação e incerteza dos métodos estudados foram gerados automaticamente pelos Softwares ConfLab Versão 1.0.1.0 (PACCES; OLIVARES, 2020a), com o auxílio das licenças concedidas de forma gratuita para uso específico neste trabalho.

Nas análises dos documentos neste trabalho avaliados, foram utilizadas como fonte de incerteza os estudos de linearidade, precisão (repetibilidade) e (intermediária) e exatidão, visto que nenhum documento apresentou outros tipos de dados que pudessem entrar como outras fontes confiáveis para se introduzir o cálculo de incerteza. Portanto, a incerteza aqui apresentada é um resultado mínimo deste valor, sem considerar outras fontes de contribuições como equipamentos, amostragem, material de referência, preparo da solução padrão ou outros. Porém, é notório que as maiores contribuições para incerteza provêm dos parâmetros da validação (OLIVARES; LOPES, 2012).

#### **3.2.4.1 Relatórios de Validação e Incerteza**

Após a inserção dos dados experimentais nos softwares, todos os projetos são armazenados e podem ser acessados novamente a qualquer momento. Quando o projeto é concluído, por parte do usuário, é possível gerar o relatório de cada projeto.

Os relatórios são gerados em formato PDF, com todas as informações de entrada de dados, resultados encontrados e observações realizadas pelo usuário a fim de se garantir a rastreabilidade adequada relacionada ao método estudado.

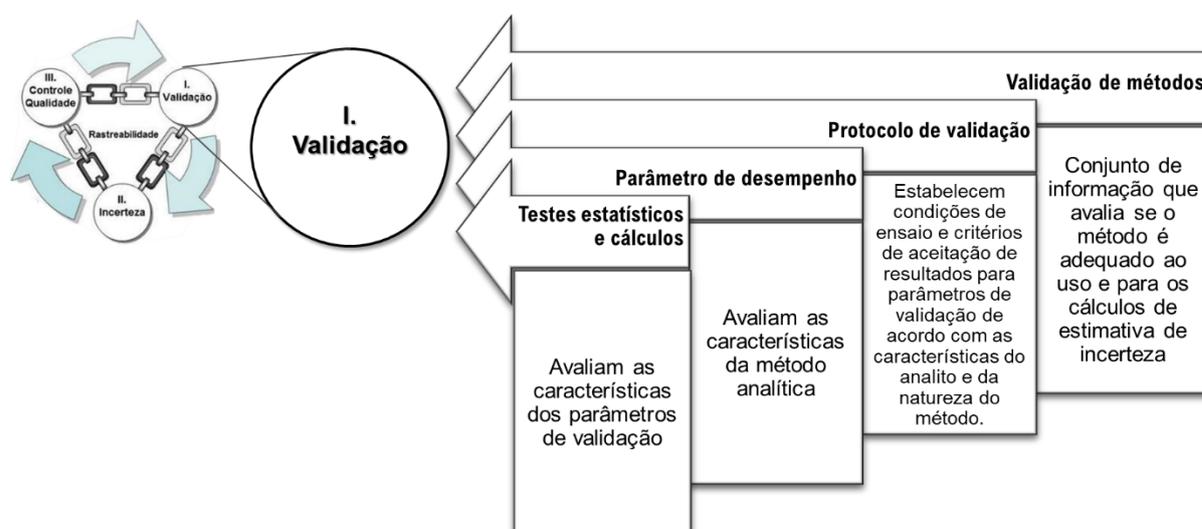
Os relatórios de validação e incerteza estão disponíveis respectivamente nos Materiais suplementares 2 e 3 (FERREIRA, 2020), nos quais para cada documento avaliado, foi gerado um relatório de validação e três relatórios de incerteza. Esta

proporção se deu pois para os estudos de estimativa de incerteza foram avaliados níveis de concentração diferentes (baixa, média e alta), de acordo com cada método, com o propósito de verificar como a concentração impacta as fontes no resultado da incerteza padrão expandida.

### 3.3 Avaliação da percepção da existência de uma crise de reprodutibilidade em relação a validação de métodos analíticos

Com base no Ciclo AQAC e nos itens listados de como realizar uma adequada e eficaz validação de procedimentos analíticos, o foco das avaliações e investigações foram os parâmetros de validação e suas características. De modo mais amplo, foram avaliados dois aspectos diferentes, para depois investigar suas relações: um deles foi o uso da validação de forma adequada e suficiente, com base nas condições de ensaios e critérios de aceitação de resultados, ambos utilizando protocolos de validação como referência. O outro aspecto, foi se os cálculos e testes estatísticos eram adequados e suficientes com base na comparação de resultados encontrados nos documentos com os encontrados pelo Software ConfLab. A Figura 7 traz a relação entre os referenciais teóricos e as ferramentas utilizadas.

Figura 7 - Esquema relacionando referenciais teóricos



Fonte: (OLIVARES; LOPES, 2012 adaptado)

### **3.3.1 Caracterização e avaliação da aplicação da validação de métodos nos documentos analisados**

O método utilizado para avaliar a conformidade das validações, foi buscar evidências se houve uso equivocado ou insuficiente dos parâmetros de validação e/ou testes estatísticos, bem como a apresentação dos seus resultados.

Neste trabalho, um dos pontos primordiais é a avaliação das validações de métodos descritas em documentos científicos. Assim, avaliar a conformidade consiste na investigação e verificação de evidências objetivas da capacidade do método e do desenvolvedor deste em demonstrar que o método desenvolvido é adequado ao uso pretendido, por meio da busca e conferência dos resultados de ensaio e análise crítica realizada pelos autores.

Para isto, foram estabelecidos alguns critérios para se compreender a correta aplicação dos conceitos de validação e estudo da incerteza de medições, mesmo que esta última tenha sido ignorada pelos autores, pois ela pode fornecer importantes informações a respeito do comportamento do método. Esta etapa envolve o estabelecimento de qual o protocolo (item 1.5.2.1, Tabela 1), e assim avaliar o documento científico no processo de determinação de critérios para a condução dos ensaios de validação, bem como de critérios de aceitação de resultados obtidos mediante a análise crítica do autor, avaliando as escolhas do pesquisador no uso correto de protocolos e ferramentas estatísticas. No decorrer deste item os procedimentos e critérios de avaliação seguirão com mais detalhes, mas resumidamente o processo de avaliação consiste em:

- I. Verificação do atendimento dos resultados dos estudos dos parâmetros de validação e dos testes estatísticos aplicados. Se estes atenderam aos critérios de aceitação de resultados do protocolo.
  - a) Inserir os dados obtidos dos ensaios (dados brutos) no ConfLab.
  - b) Gerar os respectivos relatórios
  - c) Comparar e avaliar os resultados dos estudos dos parâmetros de validação do software com os do documento científico

## II. Verificação do atendimento da validação de métodos aos critérios de condições de ensaio sugerido no protocolo

- a) Comparar as condições de ensaio sugerida ou exigida nos protocolos citados pelo documento ou corretos para o escopo e objetivo do método;
- b) Comparar e avaliar o estabelecimento das condições de ensaio apresentadas no documento e no protocolo.

### **3.3.1.1 Avaliação do uso dos protocolos de validação**

Para avaliar as condições definidas para o processo de validação foram verificadas informações em relação as interpretações e forma que o autor conduziu a validação do método. O objetivo foi identificar informações sobre o uso do protocolo correto no contexto do método, critérios e condições dos ensaios exigidos para cada parâmetro de validação, ou seja, o objetivo de uso do método proposto no documento em estudo deveria ser o mesmo do protocolo escolhido pelo autor.

#### **3.3.1.1.1 Verificação da compreensão do contexto e objetivo de aplicação do método pelo autor**

Para caracterizar a abrangência de entendimento dos autores em relação ao contexto de aplicação do seu método, com base na adequabilidade do uso dos protocolos, foram exploradas informações sobre a semelhança dos objetivos do método com o protocolo escolhido pelo autor. Foram estabelecidos alguns pontos para avaliar se a escolha e o uso do protocolo estavam adequados ao propósito, de forma que o protocolo escolhido deveria atender ao objetivo do método desenvolvido, bem como, se as exigências e critérios eram realizadas da forma que se apresentavam conforme as recomendações ou exigências do protocolo. Os seguintes pontos foram explorados:

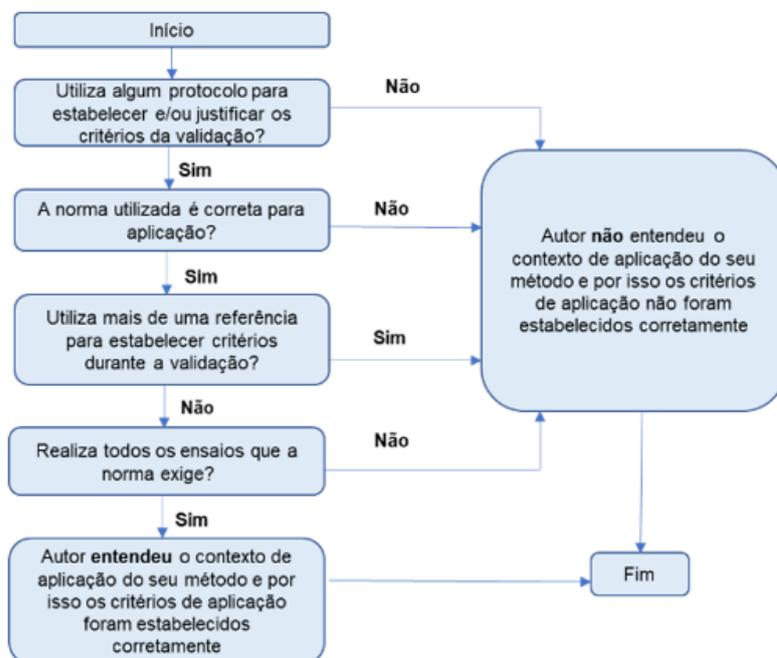
- O uso de algum protocolo para estabelecer e/ou justificar os critérios da validação;

- **Critério:** foi considerado que o autor utilizou algum protocolo quando o texto do documento menciona parâmetros relacionados a um determinado protocolo para estabelecer e/ou justificar a especificação dos requisitos para a determinação das características de desempenho dos métodos e resultados obtidos, detalhando sua adequação ao uso pretendido.
- Exemplo 1: para o ensaio de linearidade um determinado texto cita: “foram realizadas três curvas em três dias diferentes, com sete níveis de concentração e o branco da amostra, com sete replicatas cada uma junto com o uso do padrão do analito adicionadas de padrão interno, que atende a RDC nº27/2012 da ANVISA (ANVISA, 2012)”.
- Exemplo 2: para os resultados de precisão intermediária um determinado texto cita: “os valores dos coeficientes de variação (CV) encontrados a partir do estudo da precisão intermediária variaram entre 2,3% e 12,8%. Esses resultados atendem aos requisitos da RDC nº27/2012 da ANVISA (ANVISA, 2012), que exige um valor máximo para coeficiente de variação de 20%”.
- Se o protocolo escolhido para estabelecer os critérios de aceitação de resultado e aplicação foi o correto para aplicação;
  - **Critério:** foi considerado que o protocolo escolhido pelo autor era correto quando o protocolo escolhido apresentava o mesmo objetivo de aplicação que o método proposto pelo documento.
  - Exemplo: o método aborda estudo de fármacos em uma matriz biológica. Os protocolos corretos a serem utilizados são protocolos farmacêuticos que dispõe sobre os requisitos para validação de métodos bioanalíticos, como a RDC nº27/2012 (ANVISA, 2012) ou o Guia para validação de método bioanalíticos da Agência Europeia de Medicina (EMA, 2012).

- **Critério:** caso tenha sido utilizado mais de um protocolo em uma mesma validação, ou seja, quando mais de uma referência foi citada no texto do documento para estabelecer e/ou justificar a especificação dos requisitos para a determinação das características de dos métodos e para a declaração sobre a validade do método, detalhando sua adequação ao uso pretendido, foi considerado como uso incorreto dos protocolos de validação para estabelecer os critérios.
- Exemplo 1: para o ensaio de linearidade o texto cita: “foram realizadas três curvas em três dias diferentes, com sete níveis de concentração e o branco da amostra, com sete replicatas cada uma junto com o uso do padrão do analito adicionadas de padrão interno, que atende a RDC nº27/2012 da ANVISA (ANVISA, 2012)”. Porém, na discussão dos resultados cita: “os resultados de linearidade se mostraram suficientes com o valor do coeficiente de determinação  $r^2= 0,9224$ , segundo RIBANI, 2004 (INMETRO, 2003).
- Exemplo 2: para o ensaio de exatidão um determinado texto cita: “foram realizados nove níveis de concentrações, com cinco replicatas cada uma junto com o uso do padrão do analito adicionadas de padrão interno, que atende a (ICH, 2005)”. Porém, para o ensaio de linearidade o referido texto cita: “foram realizadas três curvas em três dias diferentes, com sete níveis de concentração e o branco da amostra, com sete replicatas cada uma junto com o uso do padrão do analito adicionadas de padrão interno, que atende a RDC nº27/2012 da ANVISA (ANVISA, 2012).”

A Figura 8 - Representação esquemática do procedimento de avaliação da conformidade da validação do método com os de critérios de aplicação para caracterização do estudo segundo os protocolos utilizados dispõe o esquema que representa quais os critérios e como foram coletadas essas informações.

Figura 8 - Representação esquemática do procedimento de avaliação da conformidade da validação do método com os de critérios de aplicação para caracterização do estudo segundo os protocolos utilizados



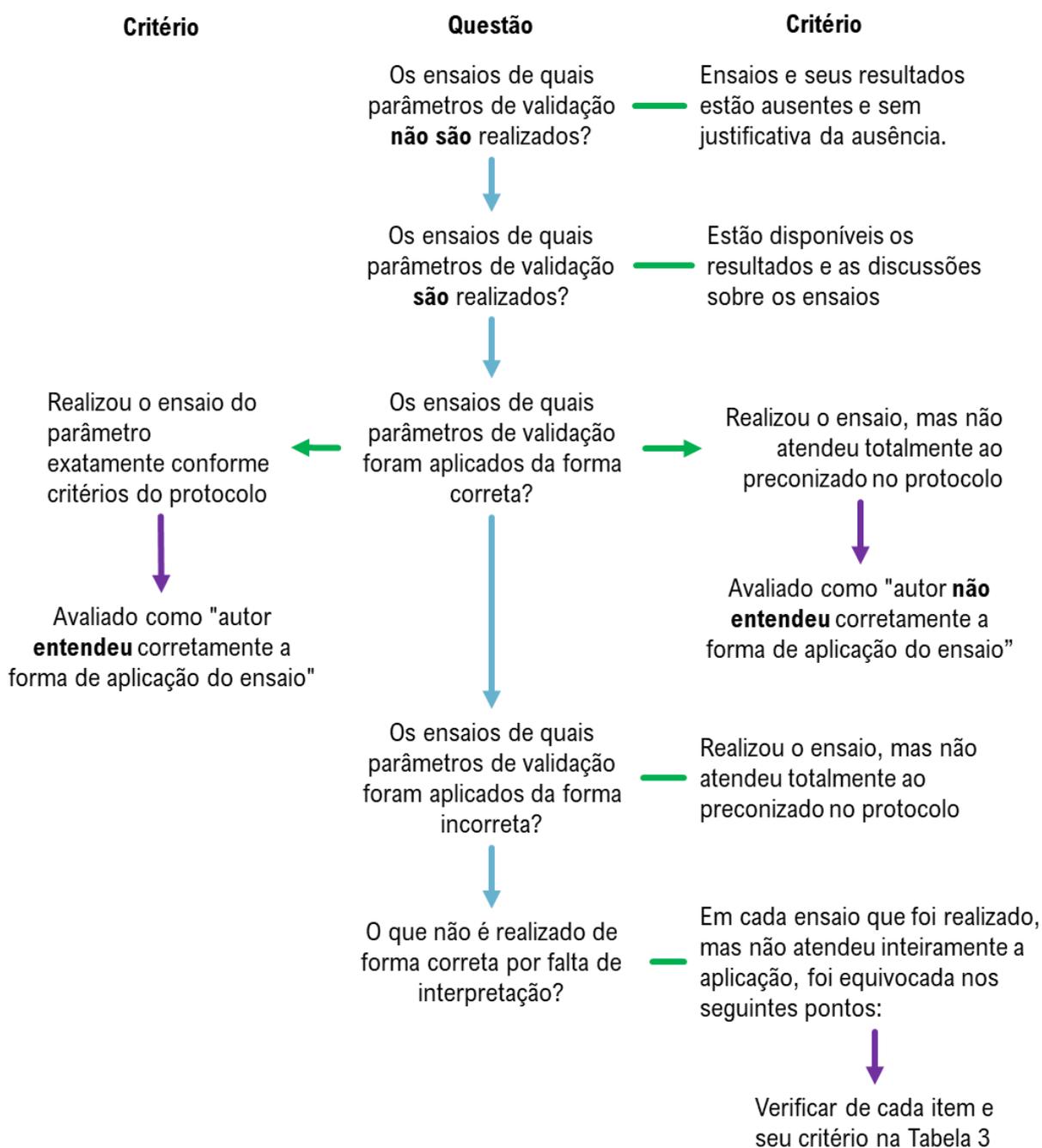
Fonte: Própria do autor

### 3.3.1.1.2 Verificação dos critérios e condições de parâmetro de validação e obtenção de dados

As condições de ensaio orientam e descrevem os requisitos do método a fim de se obter os dados necessários para a análise crítica dos resultados, dos testes estatísticos e evidenciar se o método é adequado ao uso pretendido. Por exemplo: como e quais os parâmetros de validação devem ser estudados; qual o intervalo de concentrações do estudo; quantos níveis de concentrações; quantas replicatas; como devem ser apresentados os resultados; etc.

Para confirmar se o autor entendeu e realizou corretamente a obtenção de dados da validação, com base no protocolo correto, foi investigado e conferido no documento se existiam evidências de que as condições exigidas para realização dos ensaios dos parâmetros de validação foram realizadas de forma correta e suficiente. As seguintes informações foram coletadas, seguindo os critérios propostos como mostra a Figura 9.

Figura 9 - Representação do procedimento de avaliação da conformidade da validação do método para caracterização do estudo segundo a aplicação dos ensaios de desempenho



Fonte: Própria do autor

Em relação a questão 5 da Figura 9 as justificativas dos motivos nos quais foram avaliados como incorretos estão na Tabela 3.

Tabela 3 - Critérios para avaliar e caracterizar os equívocos ocorridos em cada parâmetro de validação

Equívoco	Critério	Exemplo
Níveis de concentrações e/ou replicatas	Quando os níveis de concentrações e/ou replicatas exigidas para o ensaio do parâmetro de validação foi menor do que a preconizada no protocolo	Protocolo exigiu: 3 (três) concentrações: baixa, média e alta, com 3 (três) réplicas em cada nível e o método realizou menos do que de 3 concentrações ou 3 replicatas
Sem definição, ou definição incorreta de faixa de quantificação ou Intervalo contemplado pela curva	Ensaio do parâmetro foi realizado fora do intervalo em estudo	A curva para estudo da linearidade estuda uma faixa de concentrações entre 2 mg/L e 10 mg/L. Porém, precisão e exatidão são realizadas em 15, 20, e 25 mg/L, Também, LQ é considerado como sendo o primeiro ponto da curva, porém nos relatórios do ConfLab, esse ponto não possui precisão e adequados, segundo o protocolo.
Inconclusivo por falta de dados	Calcula o parâmetro, mas não disponibiliza os dados	Os dados para verificar os cálculos não estão disponíveis
Não está descrito e justificado a determinação por razão sinal/ruído	Protocolo exigiu confirmação da estimativa de LD por sinal/ruído e esta informação não está disponível	Estima valor de LQ ou LD com base do DP e <i>slope</i> de três curvas, ou algum outro tipo de estimativa sugerida pelo protocolo, mas não confirmada pela proporção sinal/ruído

Equívoco	Critério	Exemplo
Equívocos de cálculos e/ou estudos de linearidade	<p>Equação do ConfLab diferente da do documento</p> <p>Regressão linear citada no documento diferente do ConfLab.</p> <p>Critérios de aceitação de resultados para <math>r^2</math> não disponíveis ou não atenderam aos critérios de aceitação de resultados do protocolo e seus resultados estão diferentes do ConfLab</p>	O documento citou ou mostrou de alguma forma (por exemplo, disponibilizou uma planilha do Excel, ou resultados de testes estatísticos) que o modelo utilizado para a regressão foi o dos mínimos quadrados ordinários, porém no resultado ConfLab, o teste de Cochran mostra que o modelo deveria utilizar ponderação ou vice-versa
Uma vez estimado o valor teórico do LQ, não é confirmado experimentalmente por meio da análise de amostras independentes no mesmo nível de concentração/propriedade do LQ.	Valor de LQ teórico diferente do teórico ConfLab, e não existe evidência de confirmação da análise experimental.	Utiliza valores estimados para determinar LQ, mas não realiza confirmação experimental de precisão e exatidão

Fonte: Produção própria do autor

### 3.3.1.2 Avaliação dos cálculos e testes estatísticos quanto a adequação e suficiência

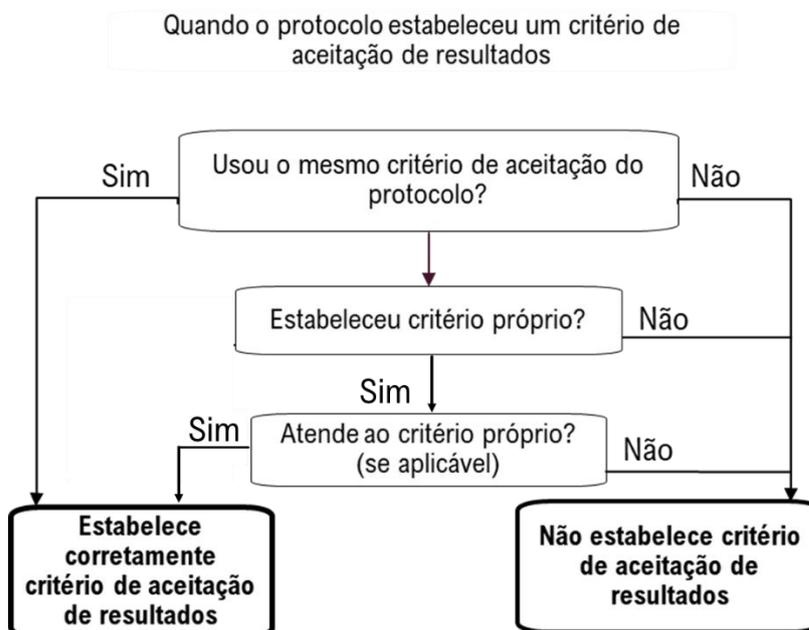
Considerando a Figura 7, os cálculos e testes estatísticos são os elementos de base para o estudo da validação de método, pois são as ferramentas nas quais entregam as características específicas dos parâmetros de validação em estudo. Desta forma, para avaliação destes pontos, foram os principais focos de investigação aspectos ligados aos critérios de aceitação de resultados e rastreabilidade adequada, também com base no Ciclo AQAC (OLIVARES; LOPES, 2012), que toma a rastreabilidade como item central na relação entre a validação, incerteza e controle de qualidade.

A primeira etapa foi a inserção dos dados brutos (tal e qual mencionados nos documentos); submetendo-os aos testes estatísticos dos softwares ConfLab, obtendo os resultados e gerando os respectivos relatórios de validação e incerteza. E, de posse dos resultados, foi avaliado o atendimento dos valores encontrados contra os critérios de aceitação e a realização correta dos cálculos.

### 3.3.2 Avaliação dos critérios de aceitação de resultado

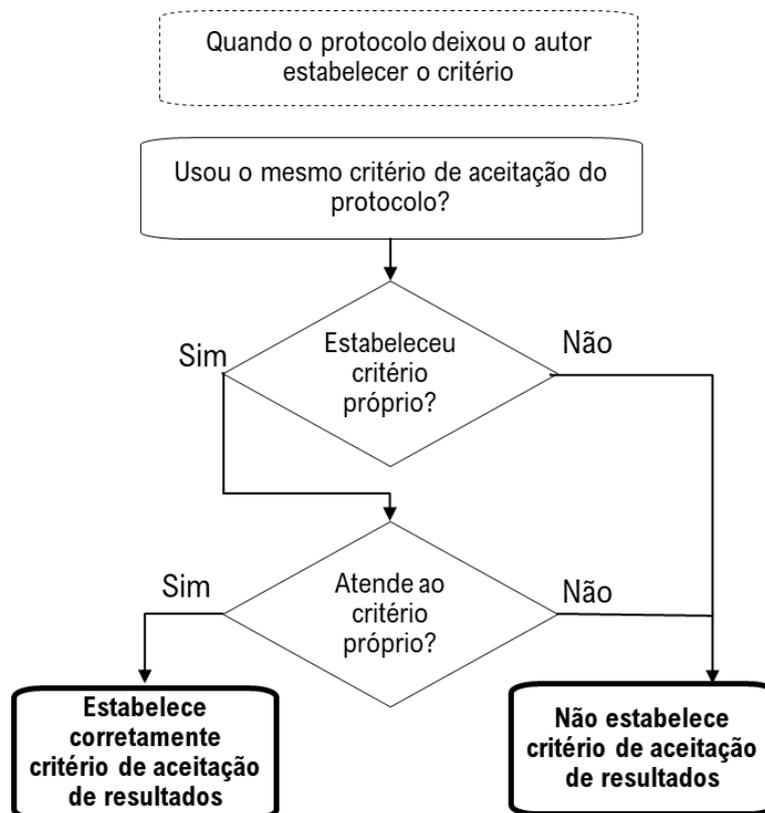
Para verificar as características em relação à capacidade dos autores em estabelecer os critérios de aceitação de resultados, a partir da comparação dos apresentados no documento científico e aqueles obtidos a partir do Software ConfLab. Algumas condições foram estabelecidas, como nas Figura 10 e Figura 11, e foram realizadas conforme exemplos da Tabela 4

Figura 10 - Esquema de avaliação dos critérios de aceitação de resultados quando o protocolo estabeleceu critério de aceitação de resultado para o parâmetro de validação



Fonte: Própria do autor

Figura 11 - Esquema de avaliação dos critérios de aceitação de resultados quando o protocolo deixou o autor estabelecer um critério de aceitação para o parâmetro de validação



Fonte: Própria do autor

Para a compilação das respostas foi criada uma planilha onde cada validação possuía uma sequência de todos os parâmetros disposta em uma tabela. Na tabela foram estabelecidas células para a entrada dos resultados de três níveis de concentração obtidos do documento em avaliação e dos relatórios do ConfLab, e na sequência as colunas com as verificações supracitadas, como no exemplo da Tabela 4:

Tabela 4 - Exemplo da distribuição de dados para a comparação dos valores obtidos pelo software e pelo documento científico em estudo, submetendo-os aos critérios de aceitação do respectivo protocolo.

Coeficiente de determinação ( $r^2$ )								
Código	Resultado documento	Resultado ConfLab	Critério Protocolo	Usou mesmo critério do protocolo?	Estabeleceu próprio critério ou outro critério diferente do protocolo?	Atende ao critério do protocolo?	Atendeu ao critério estabelecido?	Cálculo Correto?
<b>COD-030</b>	0,9992	0,9766	Critério do autor	Sim	Sim	Sim	Não	Não
<b>COD-031</b>	0,9988	0,9533	Critério do autor	Não	Não	Não	Não	Não
<b>COD-032</b>	0,9999	0,9241	0,99	Sim	Não	Não	Não	Não
<b>COD-033</b>	0,9997	0,9025	0,99	Não	Não	Não	N.A.*	Não
<b>COD-034</b>	0,9998	0,9998	0,99	Sim	Não	Sim	Sim	Sim
<b>COD-035</b>	0,9991	0,9991	Critério do autor	Sim	Sim	Sim	Sim	Sim

Fonte: Própria do autor

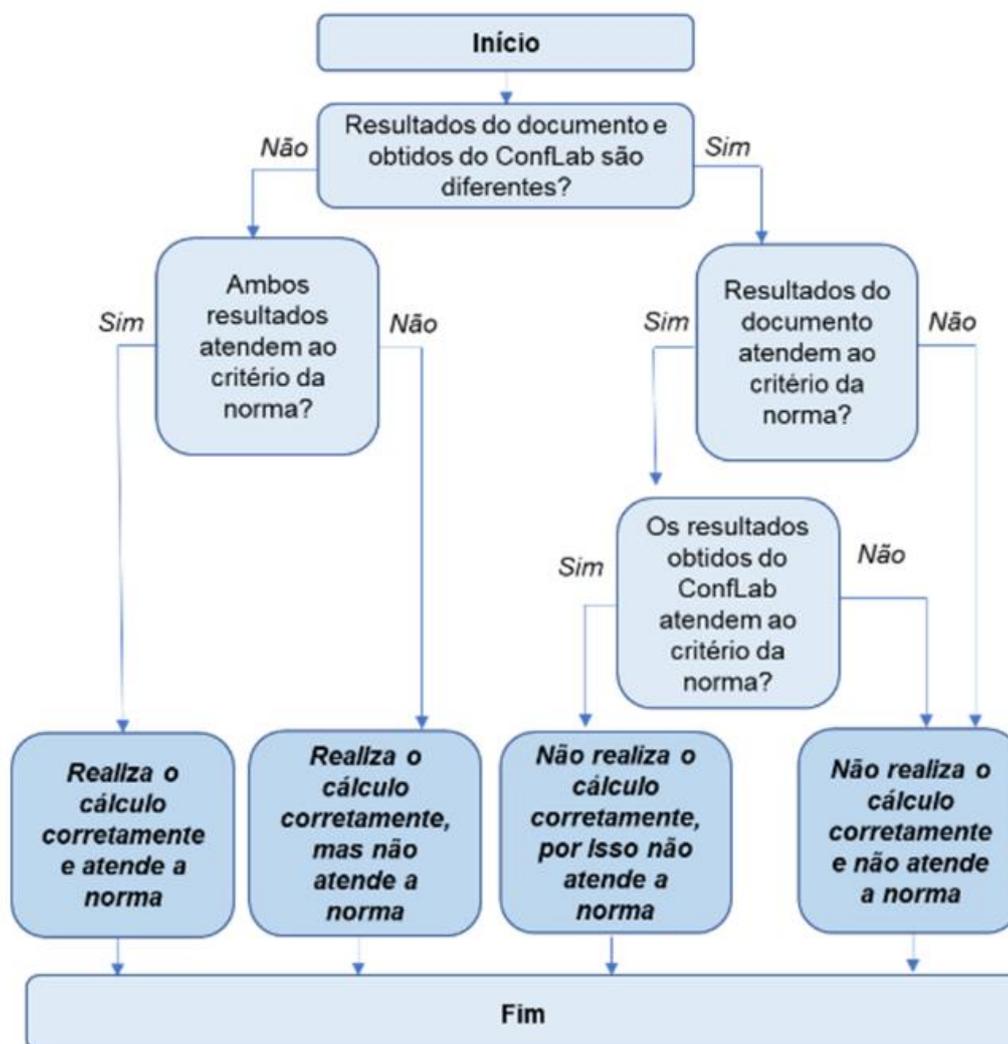
\* Não aplicável

### **3.3.2.1 Verificação dos Cálculos**

Para avaliar se os resultados dos parâmetros de validação do método atendiam aos critérios de aceitação de resultados dos protocolos ou estabelecidos pelo próprio autor, além de confirmar se os resultados estavam dentro dos limites estabelecidos pelo protocolo, foram também procuradas evidências de que os cálculos foram realizados corretamente no documento científico, bem como verificadas as características que levam a confiabilidade no resultado.

Uma comparação dos resultados disponíveis no documento com os obtidos pelo ConfLab foi realizada. Esta comparação pode ser realizada uma vez que os cálculos do software ConfLab são validados e baseiam-se em equações reconhecidas na bibliografia mundial, sendo amplamente aplicadas e aceitas nas áreas de conhecimento do estudo. Assim, os resultados dos cálculos foram utilizados como referência para a avaliação dos parâmetros estabelecidos. Para cada parâmetro de validação estudado, foi aplicada a sequência de perguntas representadas no esquema da Figura 12 para que fosse possível realizar caracterizações individuais de cada parâmetro. Porém, para considerar se o documento avaliado é adequado ao propósito, todos os resultados de todos os parâmetros deveriam estar dentro dos critérios de aceitação dos protocolos.

Figura 12 - Representação do esquema de avaliação da conformidade dos cálculos descrito no documento científico aos critérios de aceitação de resultados do protocolo e dos cálculos



Fonte: Produção própria do autor

Quando não havia registros sobre os cálculos realizados, os critérios sobre os cálculos tiveram como base o item 7.5. Registros técnicos da ABNT NBR ISO/IEC 17025 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017):

#### Registros técnicos

O laboratório deve assegurar que os registros técnicos para cada atividade de laboratório contenham os resultados, o relatório e as informações suficientes para facilitar, se possível, a identificação de fatores que afetem o resultado de medição e sua incerteza de medição associada, bem como para possibilitar que a atividade de laboratório seja repetida em condições o mais próximo possível das condições originais. Os registros técnicos devem incluir a data e a identificação do pessoal responsável por cada atividade de laboratório e pela conferência dos dados e resultados. Observações: dados

e cálculos originais devem ser registrados quando são realizados e devem ser identificáveis à tarefa específica a que se referem (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017).

Desta forma, na falta de registros originais dos cálculos, quando os resultados do documento não eram iguais aos do software, mesmo que o valor nos documentos estivessem dentro dos limites, foi considerado que o critério de aceitação de resultados não foi atendido.

Também foram verificados quais os equívocos que ocorreram durante a condução dos cálculos e análise de dados. A

Tabela 5 traz algumas características e critérios, com base nos protocolos de validação corretos para a aplicação

Tabela 5 - Critérios para a avaliação dos equívocos de cálculos dos parâmetros de validação

<b>Equívocos na condução dos cálculos e análise de dados</b>							
	<b>LD</b>	<b>LQ</b>	<b>Precisão</b>	<b>Precisão Intermediária</b>	<b>Exatidão</b>	<b>Regressão</b>	<b>r<sup>2</sup></b>
<b>Estimativa não confirmada e/ou diferente de pela razão sinal/ruído</b>	Estima o valor de LD, mas não confirma por meio da razão sinal/ruído conforme proporção indicada no protocolo	Estima o valor de LQ, mas não confirma por meio da razão sinal/ruído conforme proporção indicada no protocolo	N.A.	N.A.	N.A.	N.A.	N.A.
<b>O erro do cálculo é, ou pode ser, consequência de aplicação incorreta</b>	Quando os equívocos nas condições de estudos (Tabela 3) consequentemente provocaram equívocos de cálculos						
<b>O estudo de linearidade possui erros e/ou insuficiências, o que prejudica os valores dos desvios e coeficientes</b>	Quando os equívocos nas condições de estudos de linearidade (Tabela 3) consequentemente provocaram equívocos de cálculos em outros parâmetros						

<b>Equívocos na condução dos cálculos e análise de dados</b>							
	<b>LD</b>	<b>LQ</b>	<b>Precisão</b>	<b>Precisão Intermediária</b>	<b>Exatidão</b>	<b>Regressão</b>	<b>r<sup>2</sup></b>
<b>Precisão e exatidão da estimativa não confirmada e/ou diferente da razão sinal/ruído</b>		Depois de estimado o valor de LQ, não confirma precisão e exatidão, conforme critérios do protocolo	N.A.	N.A.	N.A.		
<b>Usa unicata ou a média das replicatas para fazer a curva</b>	Para estimativa de LD usou os valores obtidos de uma curva que utilizou unicata ou a média das replicatas	Para estimativa de LQ usou os valores obtidos de uma curva que utilizou unicata ou a média das replicatas	N.A.	N.A.	N.A.	Usou unicata ou a média das replicatas no estudo de linearidade, desta forma, pode ser que a regressão linear fora realizada de forma incorreta	Usou unicata ou a média das replicatas no estudo de linearidade, desta forma, pode ser que a regressão linear fora realizada de forma incorreta
<b>Não utiliza DP de, pelo menos, 3 curvas</b>	Citou que DP é obtido a partir de 3 curvas e que realiza estimativa pela fórmula $LD = \frac{3.3 \sigma}{S}$ mas na verdade utiliza residual de apenas 1 linha de regressão	Citou que DP é obtido a partir de 3 curvas e que realiza estimativa pela fórmula $LQ = \frac{10 \sigma}{S}$ mas na verdade utiliza residual de apenas 1 linha de regressão	N.A.	N.A.	N.A.	N.A.	N.A.

<b>Equívocos na condução dos cálculos e análise de dados</b>							
	<b>LD</b>	<b>LQ</b>	<b>Precisão</b>	<b>Precisão Intermediária</b>	<b>Exatidão</b>	<b>Regressão</b>	<b>r<sup>2</sup></b>
<b>Calcula recuperação ao invés de CV, ou vice-versa</b>	N.A.	N.A.	Calcula a recuperação do método e discutiu com base neste resultado que o método é preciso	N.A.	Calcula o desvio padrão do método e discutiu com base neste resultado que o método é exato	N.A.	N.A.
<b>Calculou com base no sinal do equipamento, e não na curva construída</b>	N.A.	N.A.	Realiza o cálculo de CV das replicatas dos sinais obtido direto do equipamento, e não dos resultados retornado por meio da curva	N.A.	Realiza o cálculo de recuperação das replicatas dos sinais obtido direto do equipamento, e não dos resultados retornado por meio da curva	N.A.	N.A.
<b>Inconclusivo por falta de dados</b>	A indisponibilidade dos dados impediu alguma conclusão sobre o equívoco do cálculo						
<b>Não reporta/calcula /avalia precisão intermediária como CV todas as replicatas de todos os dias naquela concentração, mas sim como CV de repetibilidade individual de cada dia (ou outra diferença)</b>	N.A.	N.A.	N.A.	Ao invés de calcular o CV de todos os valores juntos, ou reporta apenas os valores de CV individuais de cada dia (ou condição diferente) e não discute, ou calcula as médias dos valores de cada condição diferente	N.A.	N.A.	N.A.

<b>Equívocos na condução dos cálculos e análise de dados</b>							
	<b>LD</b>	<b>LQ</b>	<b>Precisão</b>	<b>Precisão Intermediária</b>	<b>Exatidão</b>	<b>Regressão</b>	<b>r<sup>2</sup></b>
<b>Utiliza MMQO quando deveria utilizar MMQP, ou vice versa</b>	N.A.	N.A.	N.A.	N.A.	N.A.	Quando as equações dos documentos e as encontradas pelo ConfLab eram diferentes	Quando as equações dos documentos e as encontradas pelo ConfLab eram diferentes

Fonte: Própria do autor

Especificamente para os estudos de linearidade, com base na

Tabela 5, foram adotados critérios para que este parâmetro fosse considerado como correto, pelo menos se as seguintes condições fossem atendidas:

- cálculos dos coeficientes de determinação ( $r^2$ ) ou correlação ( $r$ ), precisão (CV) da curva, exatidão da curva (erro relativo ou desvio), deveriam estar corretos;
- teste para avaliar se a regressão linear utilizou o modelo correto (ponderado ou ordinário de acordo com teste estatístico para verificar a homogeneidade das variâncias);

### 3.3.2.2 Verificação de características de rastreabilidade

Neste trabalho foram avaliadas algumas características que conduzem a confiabilidade do resultado. Desta forma, com relação aos documentos em estudo, para que na avaliação fosse identificado se seus resultados atenderam ao protocolo, além dos resultados apresentados pelo autor dentro dos limites de aceitação do protocolo, os cálculos para a obtenção do resultado também deveriam estar corretos.

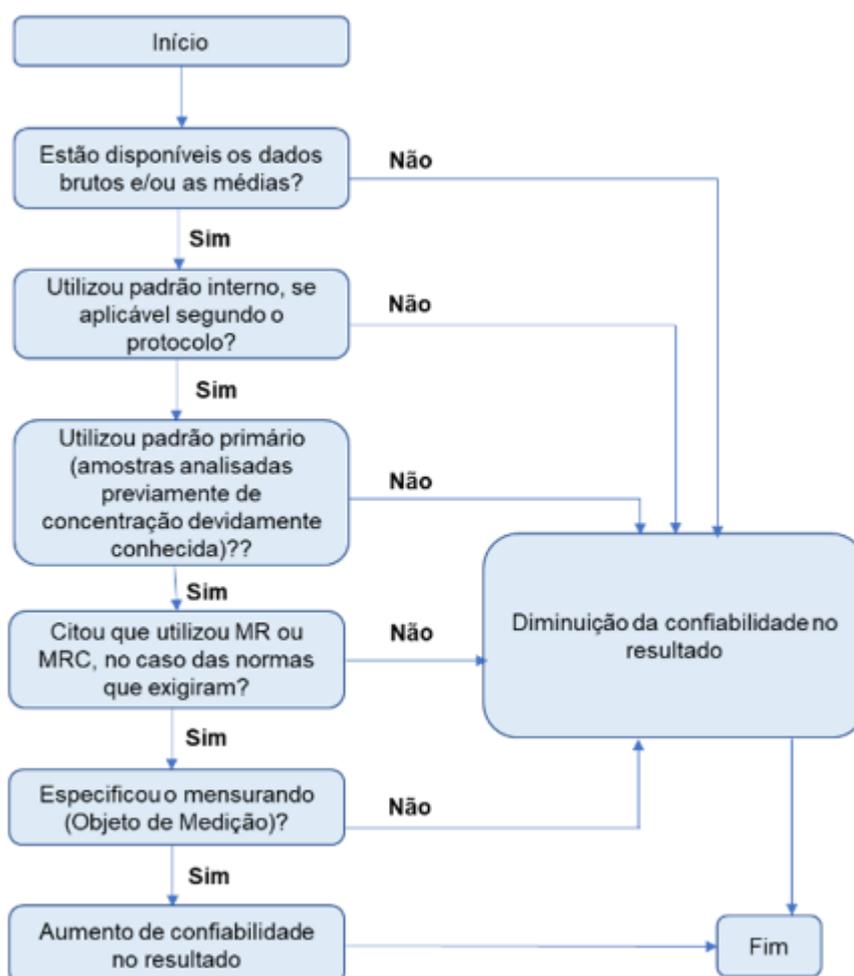
Para obter informações dos tipos de características que possibilitassem associações dos resultados a um valor de referência no documento, foram explorados os seguintes pontos: cálculos corretos, uso de instrumentos de medição calibrados, materiais de referência (MR) e materiais de referência certificados (MRC), padrão interno, padrão primário e se especificação do mesurando estava presente de forma clara (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017). Ou seja, as características que podem atribuir confiabilidade ao resultado com relação a validação. Quando o documento não demonstrou, não especificou ou não citou, de alguma forma estas informações, o método foi avaliado com características que levam à diminuição da confiabilidade no resultado. Foi considerado o uso de:

- **Instrumentos de medição calibrados** - o autor citou que o equipamento tinha passado por calibração ou verificação
- **Materiais de referência (MR)** - o autor citou e especificou como MR, ou em caso de fármacos, utilizou um medicamento comercial produzindo por uma indústria farmacêutica.
- **Materiais de referência certificados (MRC)** - citou que o material era certificado de fato, com marca e lote especificados.

- **Padrão interno** - citou o uso, o nome e a justificativa científica de ser escolhido para padrão interno. (Somente quando exigido pelo protocolo)
- **Padrão primário** - parte do preparo de uma solução mãe.
- **Especificação do mesurando** - as unidades de concentração, dos sinais dos equipamentos e das conversões dos sinais em concentração estavam claramente especificadas nos textos.

Na Figura 13 - Representação do esquema utilizado na busca informações dos tipos de características que possibilitassem associações dos resultados a um valor de referência no documento estão os critérios para cada a busca dessas informações

:Figura 13 - Representação do esquema utilizado na busca informações dos tipos de características que possibilitassem associações dos resultados a um valor de referência no documento



### **3.3.2.3 Verificação da apresentação e qualidade dos resultados**

Os protocolos estabelecem limites mínimos, máximos ou intervalos aceitáveis dentro do escopo de sua aplicação. Por meio da validação de métodos, os estudos descritos nos documentos científicos devem dispor de meios objetivos para comprovar que os métodos desenvolvidos conduzem a resultados confiáveis e adequados à qualidade pretendida.

Neste contexto, o documento científico foi considerado como um relatório de validação e que, portanto, deveria conter evidências de que o método é adequado ao uso pretendido. Desta forma, as evidências deveriam apresentar as seguintes características:

- uso adequado do protocolo, mediante o atendimento das condições de ensaio e critérios de aceitação de resultados,
- evidências do uso cálculos corretos
- um conjunto de evidências que levam ao aumento da confiabilidade no resultado com base na rastreabilidade adequada
- uso de cálculos de estimativa de incerteza, bem como o resultado reportado com o valor desta associada ao mensurando.

Os equívocos mais comuns nos estudos dos parâmetros de validação, foram identificados para posterior apresentação e discussão para comunidade científica com a finalidade de se investigar a existência de equívocos frequentes possíveis de serem corrigidos ou a avaliação da necessidade de outra investigação mais direcionada.

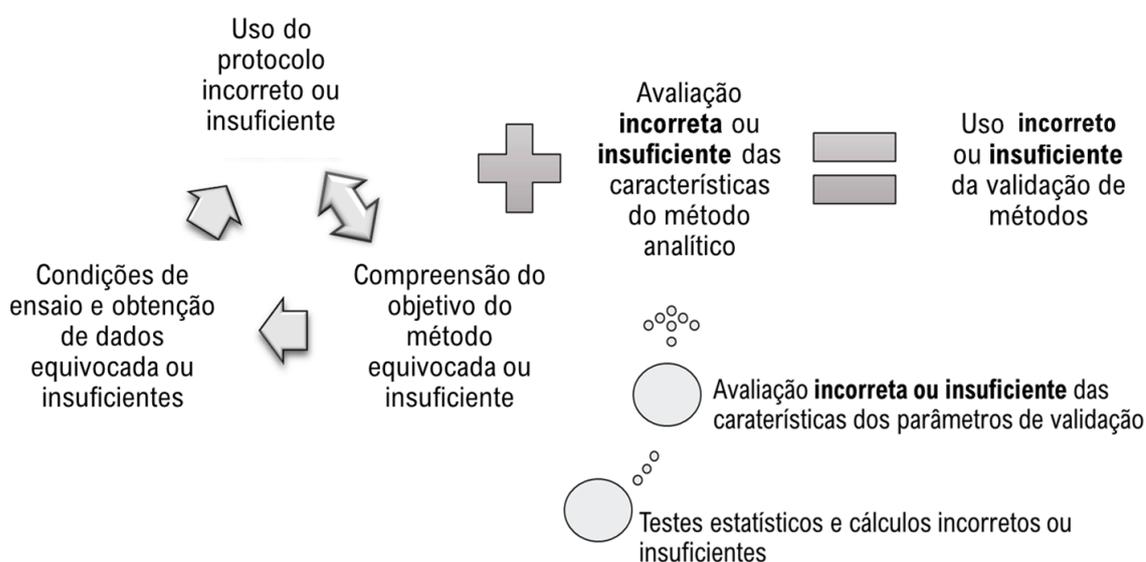
### **3.3.3 Investigação da existência de relação entre os itens de uso dos protocolos de avaliação, cálculos e testes estatísticos e os erros mais comuns cometidos durante a validação;**

A Figura 5 (Ciclo AQAC) utiliza como instrumento de avaliação aspectos dentro da validação de métodos para identificar a percepção da crise de reprodutibilidade com relação ao uso equivocado ou insuficiente das ferramentas estatísticas. Desta forma foram investigadas as relações dos aspectos encontrados e que partiram de uma possível relação entre o uso equivocado ou insuficiente da validação de métodos

com uso incorreto de protocolos de validação somados às análises estatísticas equivocadas e insuficientes e avaliações incompletas das características e qualidade do método é apresentada na Figura 14:

Para entender melhor a possível relação proposta na Figura 14, é preciso retomar a Figura 7, que, resumidamente, demonstra que para a validação de métodos ocorrer de uma forma eficaz, testes estatísticos suficientes e adequados devem avaliar as características dos parâmetros, os quais irão fornecer as características do método analítico em desenvolvimento e, desta forma, apresentar os dados de estudos da validação.

Figura 14 - Possível relação para investigar o uso equivocado e insuficiente da validação de métodos



Fonte: Própria do autor

Então, explicando melhor a possível proposição sobre o uso equivocado e insuficiente da validação de métodos aqui apresentada, caso ocorra um ciclo de insuficiências e equívocos em relação a compreensão do objetivo do método, uso de protocolo, condições de ensaio e obtenção de dados, somados à uma avaliação incorreta e suficiente das características do método analítico, que podem ser causadas pela avaliação incorreta ou insuficiente das características dos parâmetros de validação, pois os testes estatísticos para avaliar esses itens também podem ter sido realizados de forma equivocada ou insuficiente. Assim, a essa proposição

apresentada (Figura 14) permite evidenciar o fato de que o uso equivocado e insuficiente de validação de métodos é um aspecto que poderia confirmar a percepção da crise de reprodutibilidade em relação aos métodos analíticos.

### **3.4 Investigar a existência de relação entre as conclusões sobre o uso de validação com a estimativa de incerteza de medição;**

Com base em gestão de qualidade analítica e utilizando o Ciclo AQAC, é conhecido que a validação de métodos fornece insumos para o cálculo da estimativa de incerteza do método. Desta forma, se identificado o uso equivocado ou insuficiente da validação de métodos, seria provável que os resultados dos cálculos de estimativa da incerteza dos resultados obtidos da avaliação dos parâmetros de validação dos método avaliados poderiam se apresentar inconsistentes.

Utilizando os dados dos relatórios de validação do ConfLab, foram também emitidos os relatórios de incerteza, que forneceram os valores de incerteza expandida das avaliações. Considerando que a incerteza de medição do método fornece o intervalo e o nível de confiança associada a um mensurando, para apresentação de resultados deste trabalho foram estabelecidas as seguintes condições:

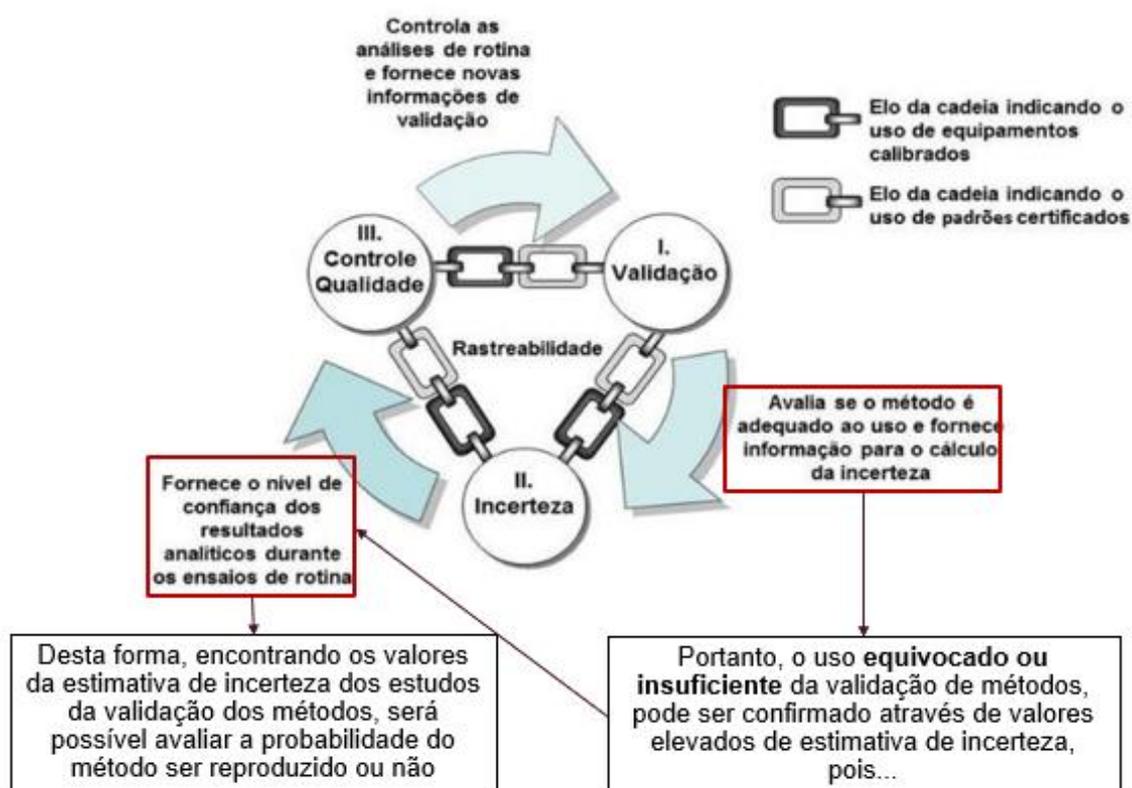
- Porcentagem de todas as validações avaliadas, quantas apresentaram os cálculos ou os valores da incerteza de medição associada ao mensurando do respectivo método;
- Das que não apresentaram a incerteza, quantas delas tiveram uma incerteza de valor questionável, neste trabalho considerados entre 40% e 100%;
- Quantos estudos obtiveram valores de incerteza calculados no ConfLab considerados “aberrantes”, ou seja, maiores do que 100% de incerteza.

### **3.5 Investigar a existência de relação da incerteza de medição com a crise de reprodutibilidade**

A probabilidade é uma linguagem matemática para quantificar a incerteza (WASSENNAN, 2004). Então, com base nos valores de incerteza encontrados, as probabilidades de resultados serem encontrados utilizando os métodos avaliados,

poderiam enfim confirmar a relação entre a crise de reprodutibilidade (uma vez confirmada nos itens anteriores), causada pela “Replicação insuficiente no laboratório” ou “Baixo poder estatístico” (BAKER, 2016), com os fatores que ajudariam a solucionar a crise como “Projeto experimental mais robusto”, “melhores estatísticas” e “Validação em laboratórios externos”. Essa relação está esquematicamente descrita na Figura 15.

Figura 15 - Relação entre a crise de reprodutibilidade e incerteza e medições, com base no ciclo AQAC



Fonte: (OLIVARES; LOPES, 2012 adaptado)

A partir dos dados obtidos conforme a relação da Figura 14, sobre o uso equivocado e insuficiente da validação de método, os valores de incerteza obtidos do item 3.4 e no Ciclo AQAC (OLIVARES; LOPES, 2012) (item 1.5.3), outra possível conclusão sobre a crise de reprodutibilidade foi levantada e segue apresentada na Figura 15.

Diante dos resultados obtidos da verificação dos valores de incerteza, citados no item 3.4, tornou-se possível avaliar a relação da crise de reprodutibilidade com o uso

insuficiente e equivocado da estatística pelos desenvolvedores dos métodos analíticos, bem como relacionar se o uso de ferramentas estatísticas, como validação de métodos e cálculos de incerteza de medição, seriam, de fato, uma solução para o problema da crise.

## 4 Resultados e discussão

A apresentação dos resultados está na forma de gráficos que resumem as observações provenientes do estudo de cada documento. Para o detalhamento dos resultados e discussões, foram disponibilizados materiais suplementares ([https://1drv.ms/u/s!BOi9HR4EVIszgsdRH-I\\_a9kZZ440w?e=kHCH7p](https://1drv.ms/u/s!BOi9HR4EVIszgsdRH-I_a9kZZ440w?e=kHCH7p)) organizados e apresentados de acordo com a Tabela 6.

Tabela 6 - Representação da disposição do material suplementar onde estão disponíveis os dados e discussões em materiais suplementares independentes

Material suplementar	Título
1	Discussões Detalhadas sobre Validações de Métodos Analíticos em parte dos Documentos Científicos
2	Relatórios de Validação do Método - ConfLab Validação
3	Relatórios da Incerteza de Medição - ConfLab Incerteza

Fonte: Própria do autor

Ao todo, foram analisadas 91 validações de métodos, presentes em 42 documentos científicos, 139 relatórios de validações e 139 relatórios de incerteza. Destes, nas primeiras 18 validações foram realizadas análises mais detalhadas e descritivas para entender e explorar em profundidade a abordagem de cada um destes projetos, resultando em uma discussão mais focada no estudo de cada teste estatístico em cada parâmetro de validação separadamente. Nos demais documentos o foco foi avaliar, de forma amostral, um maior número de projetos não discutindo em detalhes cada um dos desvios, em cada teste estatístico de cada parâmetro de validação, mas sim realizando uma avaliação direta dos resultados a partir dos softwares ConfLab. A Figura 16 traz a quantidade de materiais produzidos pelo trabalho.

Figura 16 - Quantidade de materiais estudados e gerados pelo trabalho

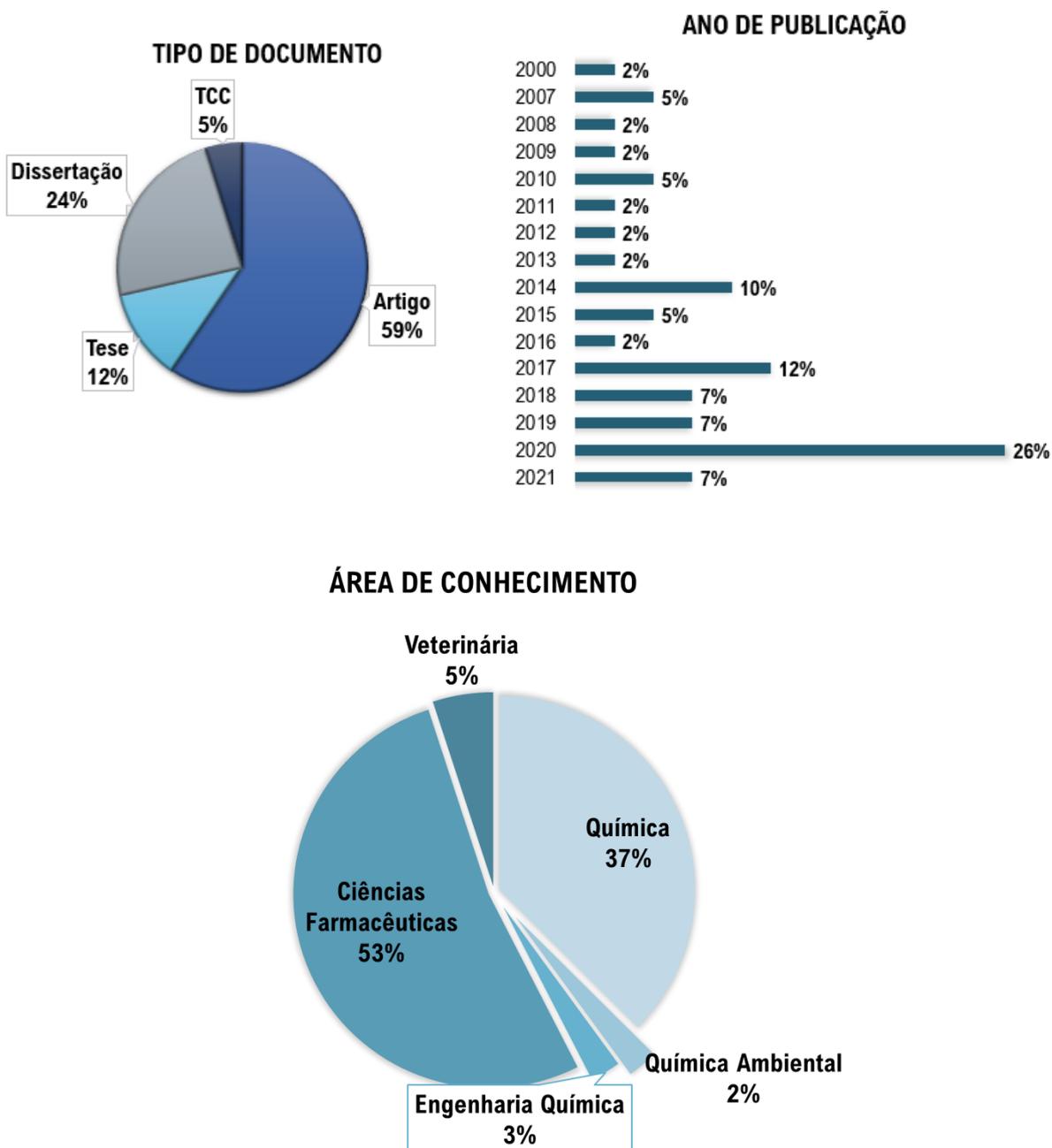


Fonte: Própria do autor

## 4.1 Caracterização do Estudo

A Figura 17 demonstra a distribuição e caracterização dos documentos analisados quanto ao tipo, área e ano da publicação.

Figura 17 - Caracterização do estudo em relação ao tipo, ano e área de conhecimento



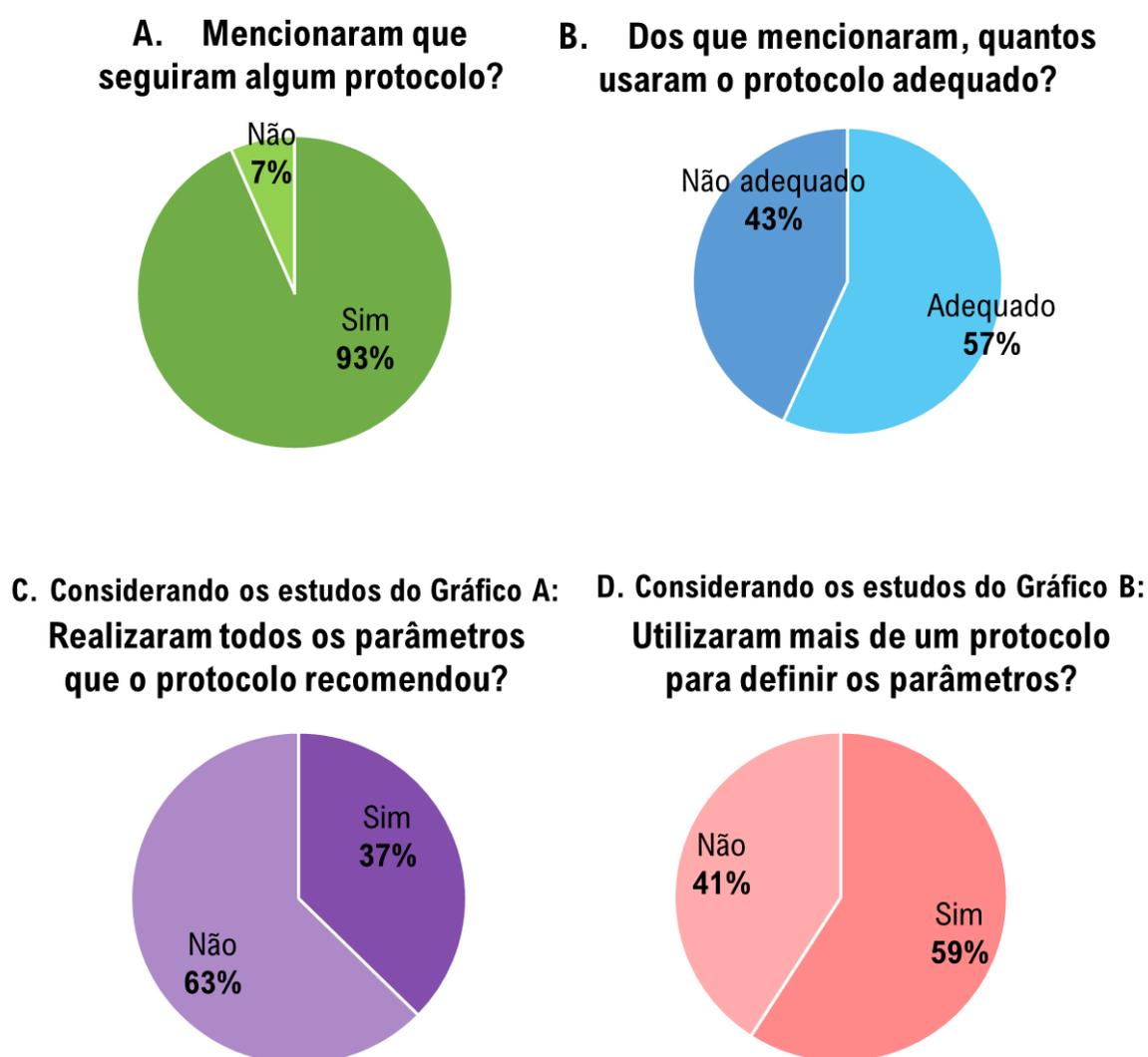
Fonte: Própria do autor

## 4.2 Resultados gerais

### 4.2.1 Uso dos protocolos de validação

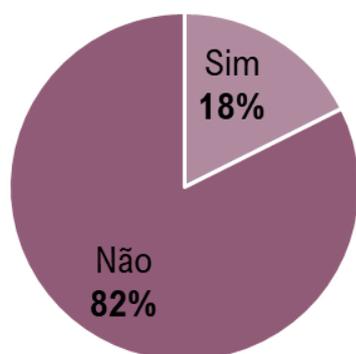
Os resultados em relação ao uso do protocolo, estão na Figura 18 e discutidas posteriormente.

Figura 18 - Características do uso dos protocolos de validação



Fonte: Própria do autor

Figura 18 - Características do uso dos protocolos de validação (continuação)

**E. Entenderam o contexto de aplicação do método?****F. Entenderam os critérios dos parâmetros de validação?**

Fonte: Própria do autor

Os dados da Figura 18 mostraram que 93% dos autores mencionaram que seguiram algum protocolo (Figura 18 (A)), desses, 57% utilizaram o protocolo adequado para a aplicação (Figura 18 (B)). Quando não foram mencionados uso de protocolo, foram atribuídos quais os protocolos mais adequados para realizar a avaliação (Tabela 1), e dentre todas as validações, que mencionaram ou não o uso de algum protocolo da Figura 18 (B), 63% dos autores não realizam todos os parâmetros recomendados pelo protocolo como apresentado na Figura 18 (C). Dos que mencionaram uso de algum protocolo (Figura 18 (B)), 59% utilizaram mais de um protocolo para definir os critérios de aceitação e condições de ensaio dos parâmetros de validação como apresentado na Figura 18 (D). Dentre todas as validações (mesmo espaço amostral considerado na Figura 18 (A)), 82% dos autores não entenderam o contexto do objetivo de seu método (Figura 18 (E)) e 81% não entenderam as condições e critérios dos parâmetros de validação que foram sugeridos pelo protocolo (Figura 18 (F))

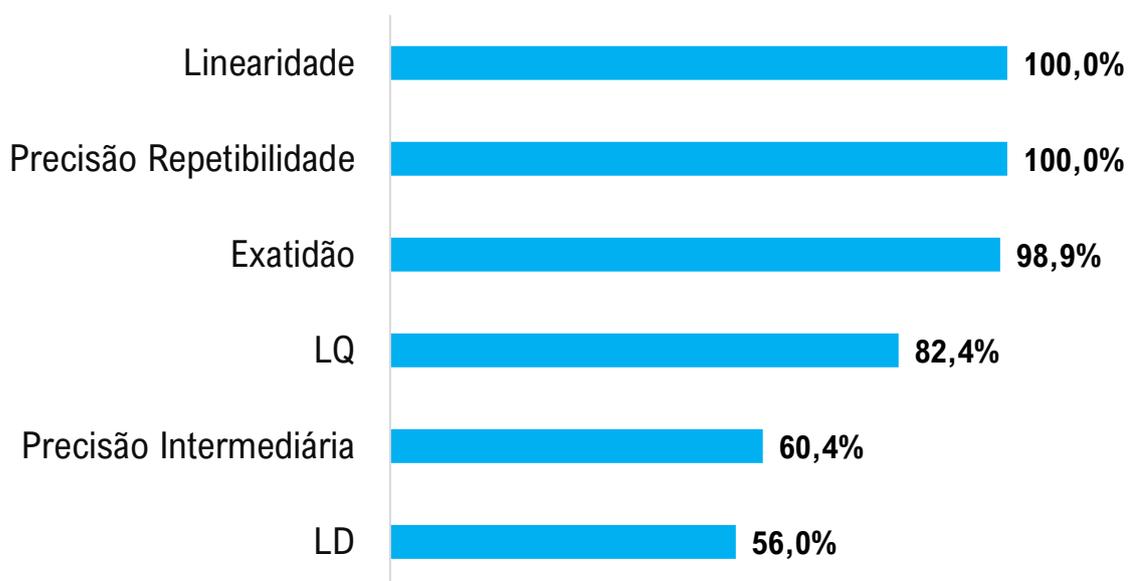
Os resultados (Figura 18) caracterizam que, apesar da maioria dos autores citarem o uso de protocolos para realizar a validação do método, existem inconsistências em relação a capacidade dos autores em conduzir o processo de validação de métodos em sua totalidade. Desta forma, mesmo a validação de métodos baseando-se em algum protocolo de validação, quando esse protocolo não é o correto para o propósito, ou seja, objetivo do método não é o mesmo do protocolo, bem como

não é o único a ser utilizado para estabelecer condições e critérios de parâmetros de validação e aceitação de resultados, resulta em uma cadeia de equívocos em relação ao processo de validação, (RAPOSO; IBELLI-BIANCO, 2020). Vale ressaltar, também, que, ao utilizar mais de um protocolo para a validação de um método, não é possível determinar se tal escolha foi feita em razão de algum parâmetro não ser atendido pelo primeiro protocolo, sendo necessário o uso de um segundo para que a aprovação ocorra.

Em relação aos critérios e condições dos parâmetros de validação, na Figura 19, dentre todas as validações (Figura 18 (A)), os parâmetros que menos foram realizados, quando exigidos foram respectivamente limite de detecção (LD) (56,0%), precisão intermediária (60,4%) e limite de quantificação (LQ) (82,4%). Este tópico considerou apenas se o parâmetro foi estudado ou não dentro do processo de validação, independente da forma de aplicação ou cálculos e resultados corretos e apropriados.

Figura 19 - Frequência que os parâmetros de validação foram estudados dentro das validações de métodos

### Parâmetros de validação mais presentes quando exigidos



Fonte: Própria do autor

Os dados apresentados na Figura 19, mostram características em relação a frequência que os parâmetros foram estudados no processo de validação, quando recomendado pelo protocolo. Os dados da Figura 20 mostram as características das vezes que esses parâmetros foram estudados, nos quais o limite de detecção (LD) foi o parâmetro que menos atendeu as condições e critérios dos parâmetros de validação recomendados nos protocolos, seguido da linearidade e do limite de quantificação (LQ). Os ensaios de exatidão foram os que mais atenderam as condições de ensaio.

Os equívocos que ocorreram na condução de ensaio e obtenção de dados em relação ao uso do protocolo, estão descritos abaixo: nas Figura 21, Figura 22, Figura 23, Figura 24, Figura 25 e Figura 26 separados por parâmetro de desempenho.

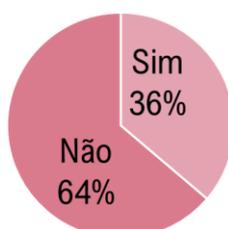
Figura 20 - Comparação entre as frequências que os parâmetros de validação foram estudados dentro de condições adequadas segundo o protocolo mais apropriado para o objetivo do método

Foram avaliados dentro das condições recomendadas pelo protocolo?

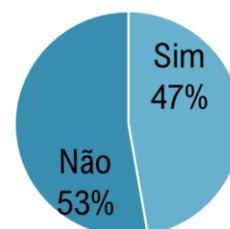
**A. Limite de Detecção**



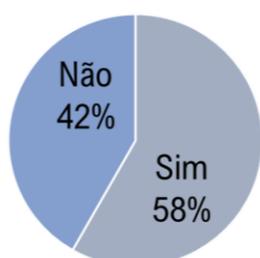
**B. Linearidade**



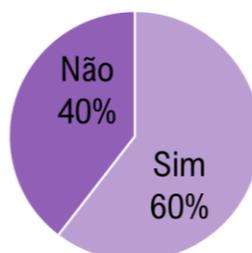
**C. Limite de Quantificação**



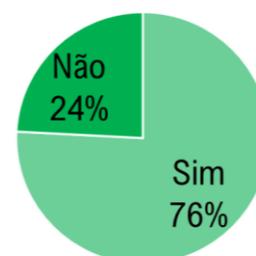
**D. Precisão Intermediária**



**E. Precisão**

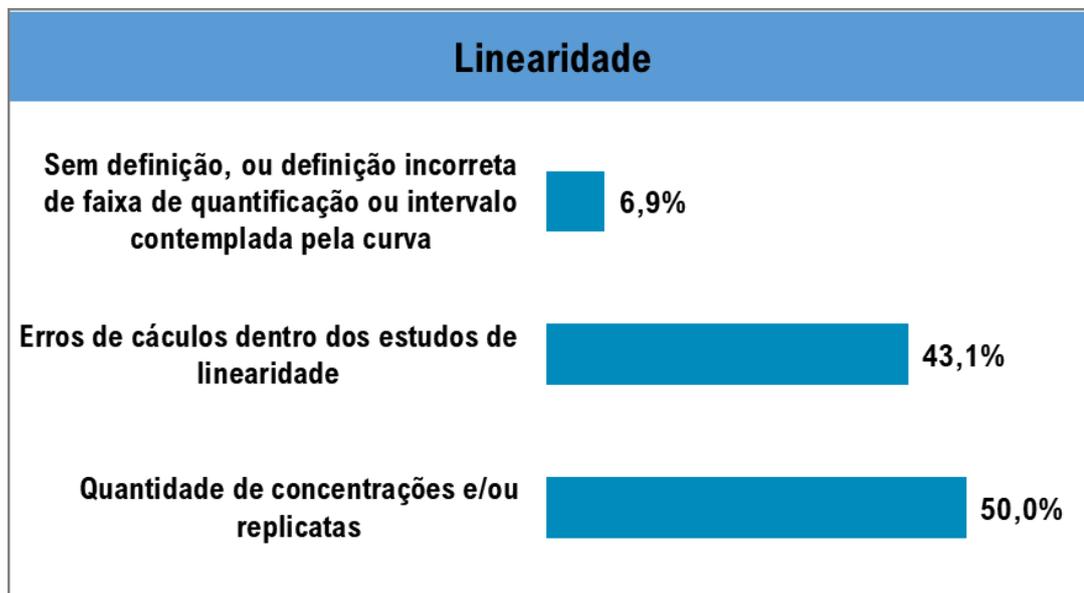


**F. Exatidão**



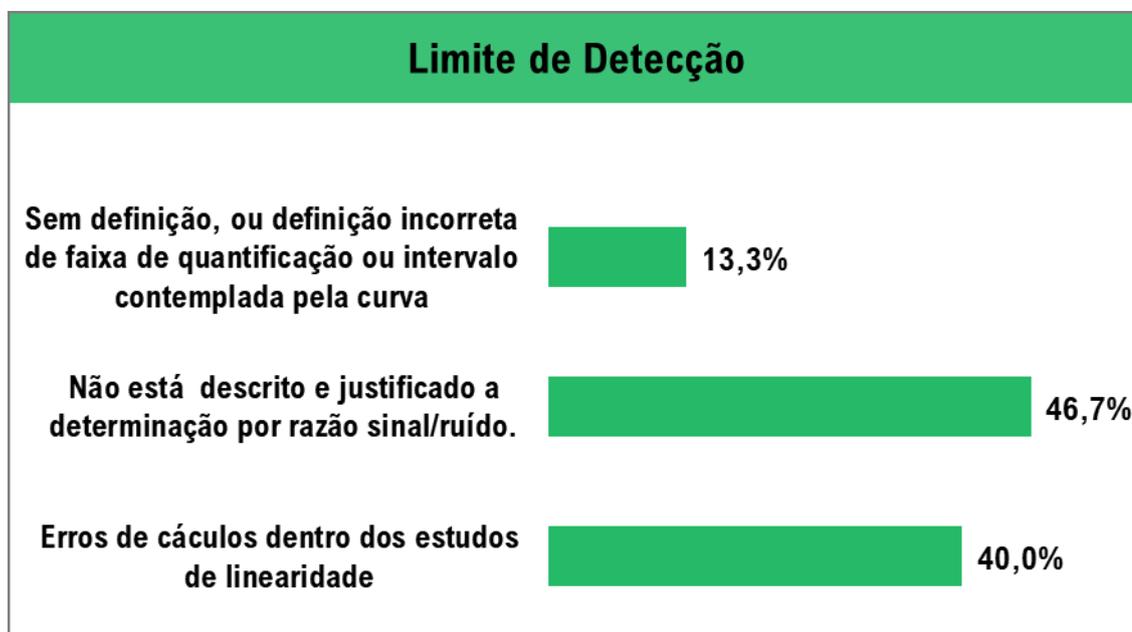
Fonte: Própria do autor

Figura 21 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos de linearidade



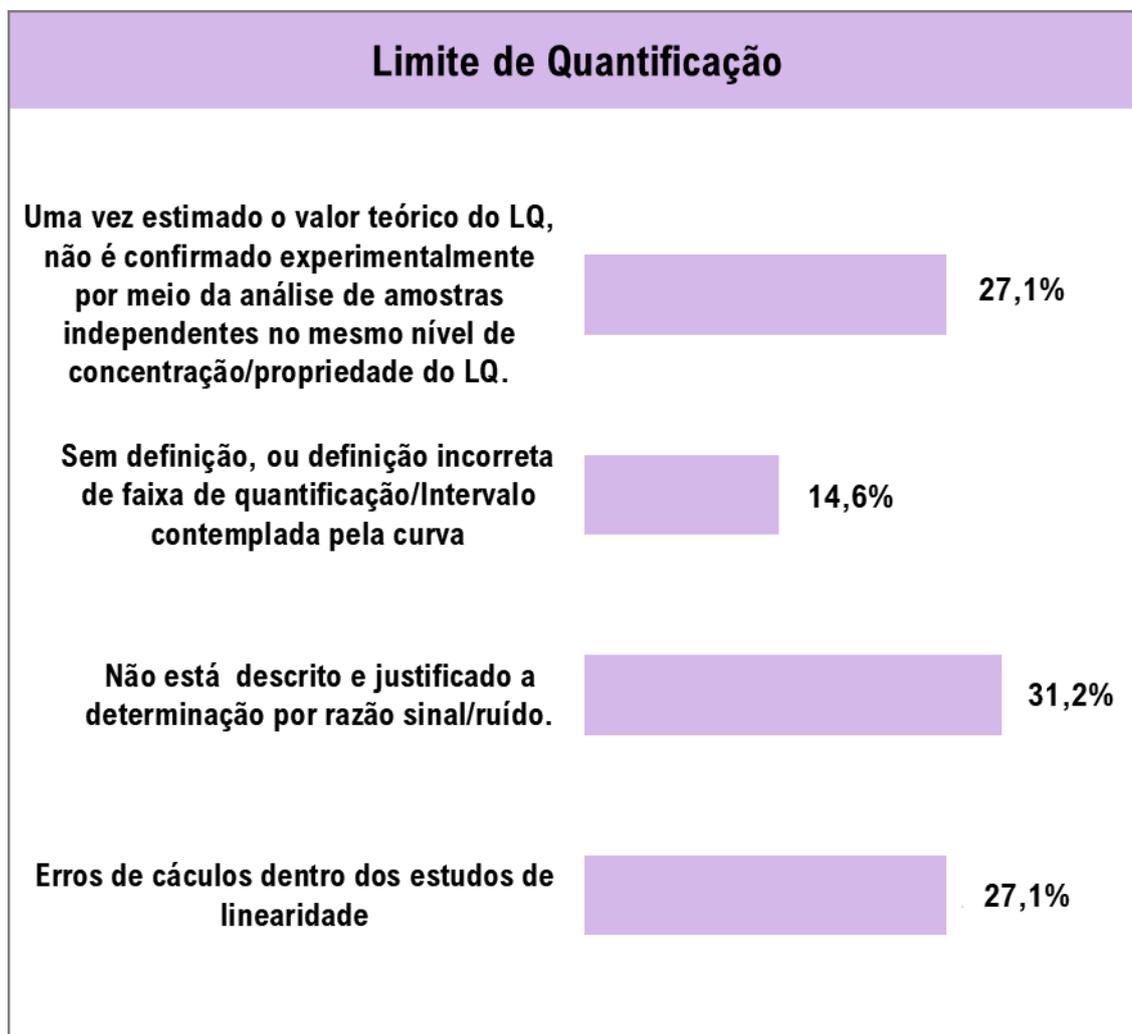
Fonte: Própria do autor

Figura 22 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos do limite de detecção (LD)



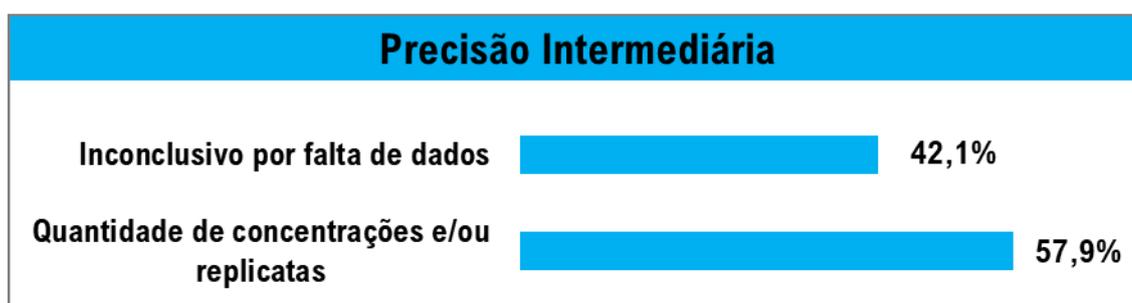
Fonte: Própria do autor

Figura 23 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos do limite de quantificação (LQ)



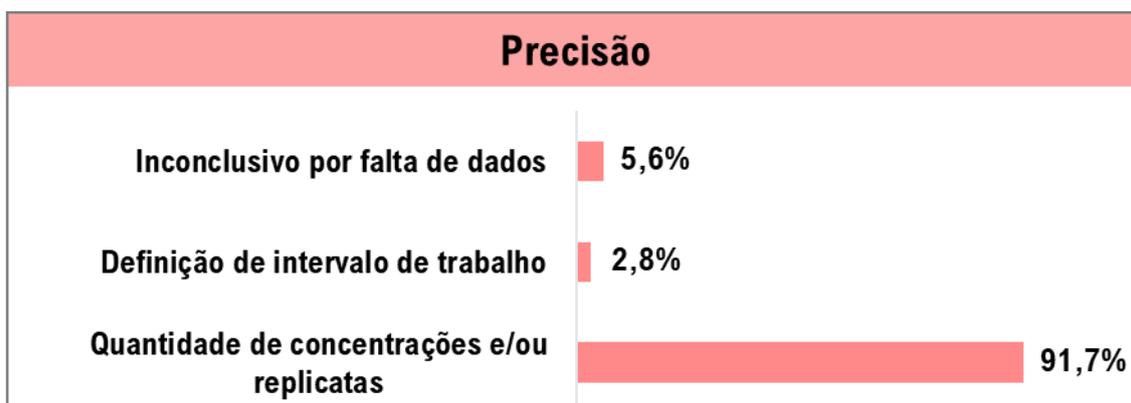
Fonte: Própria do autor

Figura 24 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos da precisão intermediária



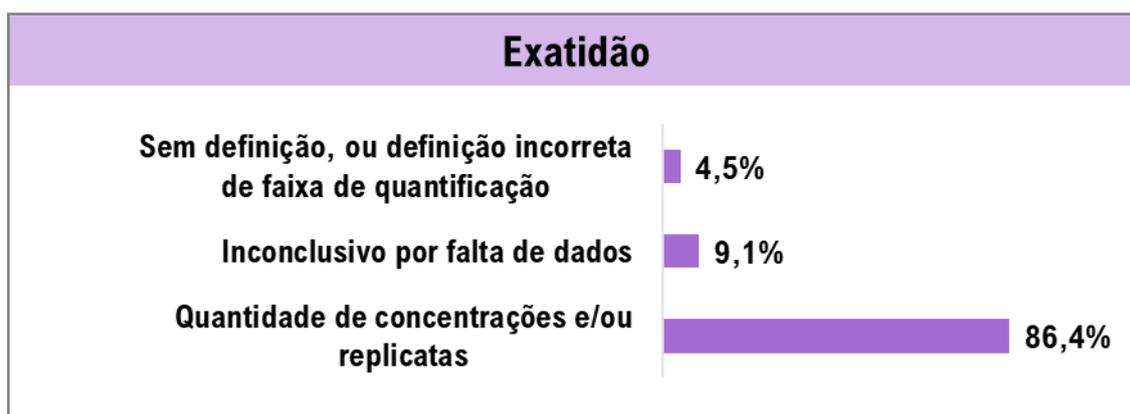
Fonte: Própria do autor

Figura 25 - Equívocos em relação à condução do ensaio e obtenção de dados nos da precisão



Fonte: Própria do autor

Figura 26 - Equívocos em relação à condução do ensaio e obtenção de dados nos estudos da exatidão

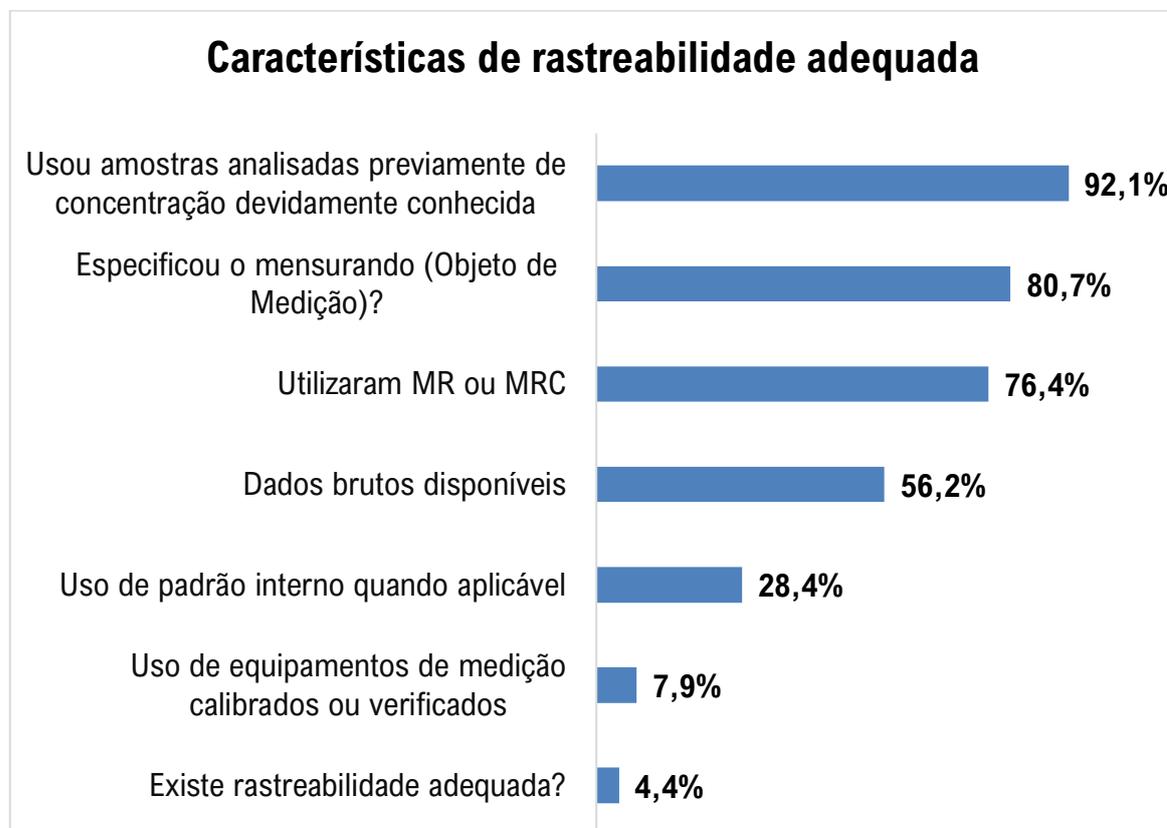


Fonte: Própria do autor

#### 4.2.2 Características de rastreabilidade adequada

A Figura 27, traz informações que permite caracterizar aspectos relacionados a confiabilidade de resultados, que se mostraram insuficientes com base na dificuldade de encontrar documentos científicos com dados de validação disponíveis, bem como a tendência da utilização de recursos e referências de forma isolada para estabelecer rastreabilidade adequada, e não um conjunto de recursos e referências que possibilitem estabelecer uma cadeia ininterrupta e documentada de calibrações (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2017; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

Figura 27 - Resultados sobre a avaliação das características que levam a confiabilidade no resultado no documento científico



Fonte: Própria do autor

Durante a busca, foram raros os documentos com os dados dos resultados com replicatas disponíveis de todos os parâmetros de validação. Quando estavam disponíveis os dados normalmente foram apresentados em tabelas com os valores das médias dos sinais de resposta do equipamento, e mesmo que planilhas fossem oferecidas, as fórmulas não estavam disponíveis.

Dos artigos analisados, foi possível observar que a partir de 2020, a disponibilização de materiais suplementares, segundo critério de algumas revistas e jornais científicos tenderam a ser mais frequentes. Mesmo assim, as buscas foram realizadas com o uso de filtros avançados para que fosse possível encontrar esses dados, os quais nem sempre se apresentam de forma a conseguir refazer os cálculos de validação com dados obtidos pelo documento, talvez devido as condições de ensaio insuficientes, como mostraram os resultados do item 4.2.1.

Em relação as características de rastreabilidade adequada, as quais são consideradas características que levam a confiabilidade nos resultados, todos os documentos continham dados brutos referentes a pelo menos dois parâmetros de validação, sendo 56,2% (Figura 27) deles disponibilizados em documentos suplementares como planilhas, cromatogramas etc., ou pelo menos parte dos dados (como os sinais das replicadas). Mesmo com o critério estabelecido para a busca dos documentos (item 3.2.1), ainda assim os dados brutos não estavam disponíveis para todos os parâmetros que o documento citou estudar. Apenas 7,9% (Figura 27) dos documentos avaliados mencionaram ou evidenciaram de alguma forma o uso de equipamentos calibrados ou verificados, porém 92,1% realizaram análises prévias de amostras com concentração conhecida (Figura 27). Também, apenas 28,4% utilizaram padrão interno quando exigido (Figura 27), 76,4% utilizaram material de referência ou material de referência certificado (Figura 27). As condições mínimas para que o método demonstrasse características que levam a confiabilidade do resultado, deveriam atender a pelo menos esses requisitos. Desta forma, apenas 4,4% dos métodos demonstraram características de rastreabilidade (Figura 27), que levam a confiabilidade nos resultados.

#### **4.2.3 Cálculos e testes estatísticos**

##### **4.2.3.1 Cálculos**

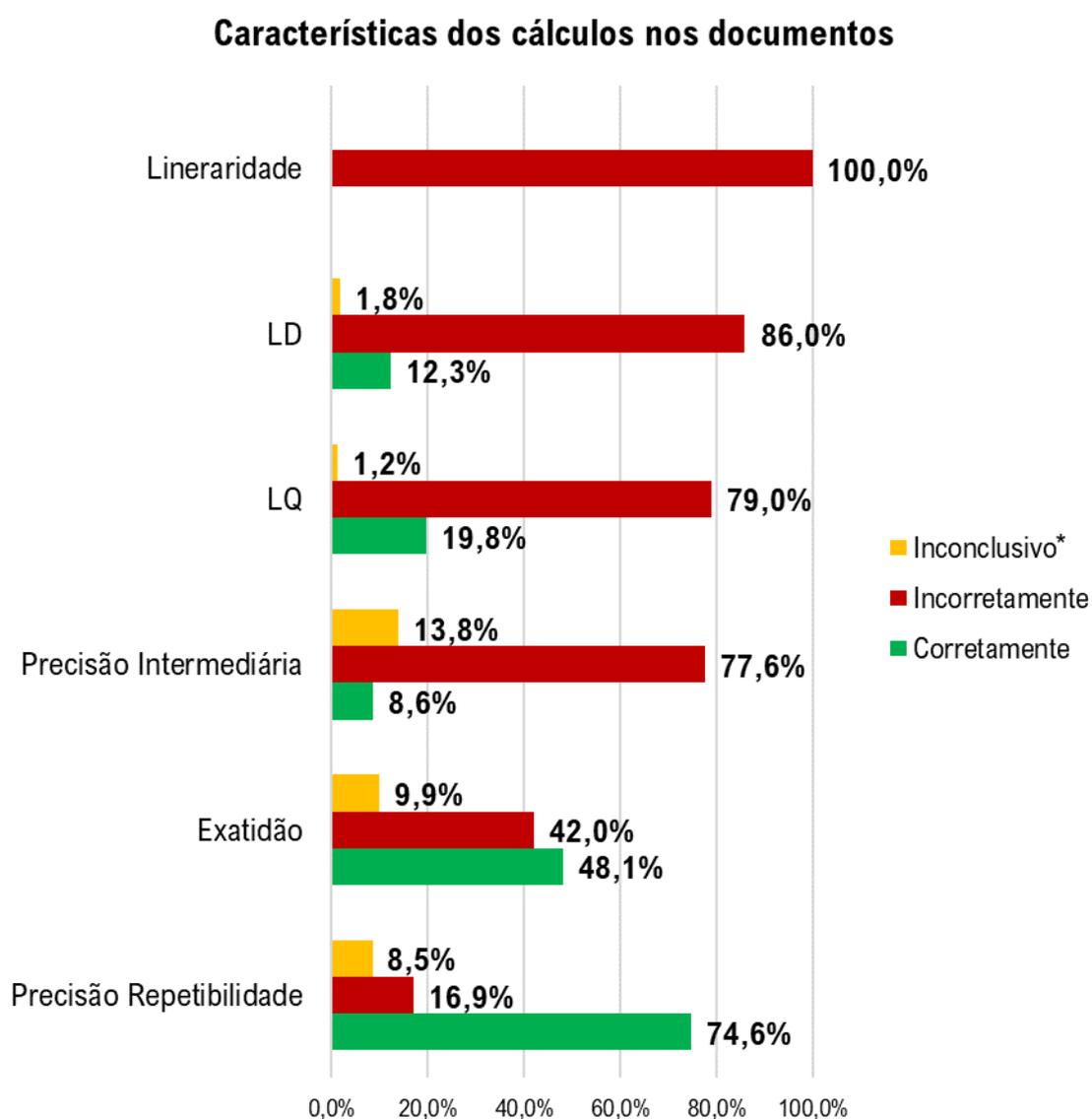
Para os cálculos no estudo de linearidade serem avaliados corretos, pelo menos os seguintes itens deveriam ser atendidos:

- cálculos dos coeficientes de determinação ( $r^2$ ) ou correlação ( $r$ ) corretos
- cálculos de precisão (CV) dos pontos da curva;
- exatidão dos pontos da curva corretos;
- testes para avaliar e aplicar o modelo correto de regressão linear de acordo com a homogeneidade das variâncias;

Os resultados obtidos em relação aos cálculos e testes estatísticos, quando realizados, apenas na exatidão e na repetibilidade número de cálculos corretos foi maior do que incorretos (Figura 28). Os outros parâmetros de validação, mostraram

mais resultados de cálculos incorretos do que corretos, sendo a linearidade o parâmetro de validação que apresentou 100% dos cálculos realizados incorretamente.

Figura 28 - Frequência que os cálculos dos parâmetros de validação foram realizados corretamente quando foram estudados.



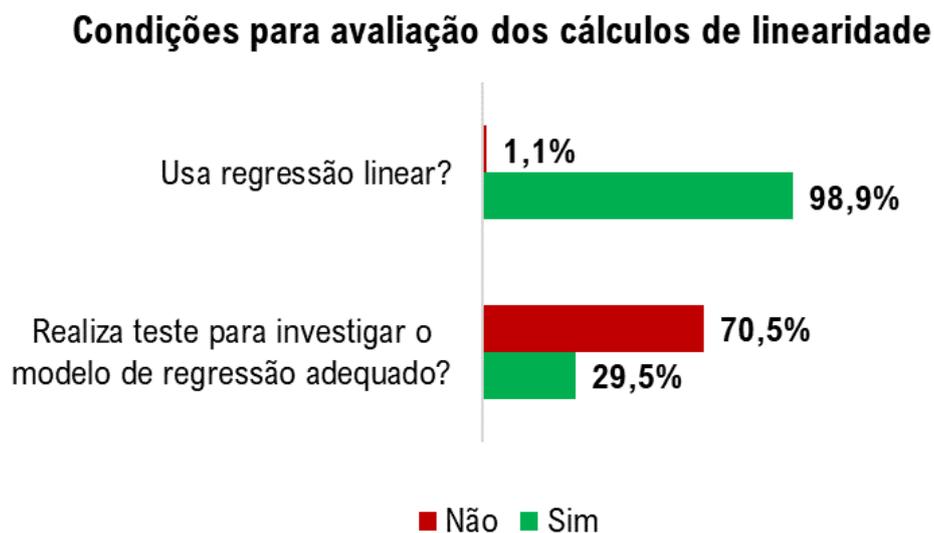
Fonte: Própria do autor

\* Quando no documento original os dados sobre o parâmetro não estavam disponíveis para inserir no software (PACCES; OLIVARES, 2020), foi considerado como inconclusivo a avaliação em relação aos cálculos do parâmetro, sendo a precisão intermediária com a maior frequência neste aspecto.

De acordo com os resultados obtidos e considerando o tratamento incorreto nos cálculos do parâmetro linearidade Figura 28 alguns aspectos mais específicos em relação a este parâmetro de validação, seus cálculos e testes estatísticos foram avaliados, e de acordo com a Figura 29, a maioria dos documentos utilizaram regressão linear, porém a minoria realizou o teste para investigar qual modelo de regressão adequado.

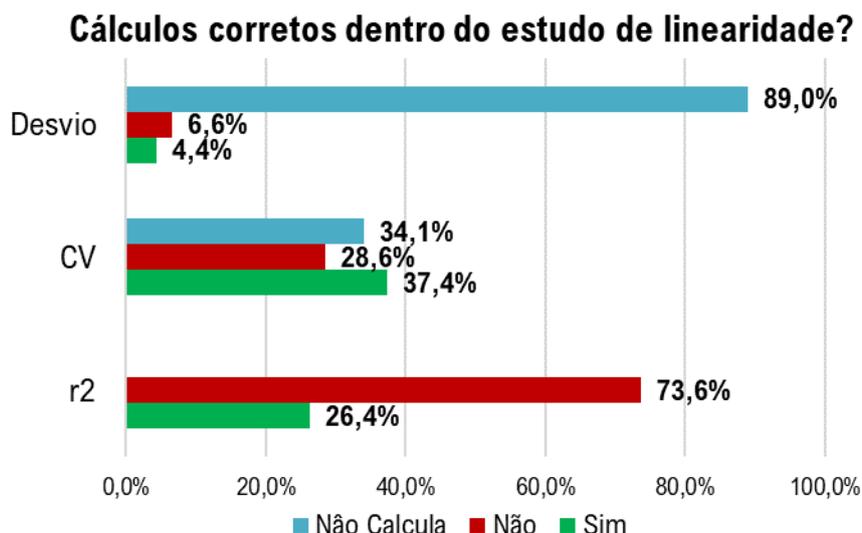
Em relação as características dos cálculos relacionados ao estudo de linearidade, é possível observar na Figura 30 que o coeficiente de variação ( $r^2$ ) foi calculado de forma incorreta, quando calculado; que o desvio (erro relativo) dos pontos da curva não é frequentemente realizado; e que a precisão dos pontos da curva, se apresenta corretamente calculada em apenas 37,4% das vezes, quando calculado.

Figura 29 - Características das condições em relação a regressão linear nos estudos da linearidade



Fonte: Própria do autor

Figura 30 - Características cálculos relacionados estudos de linearidade



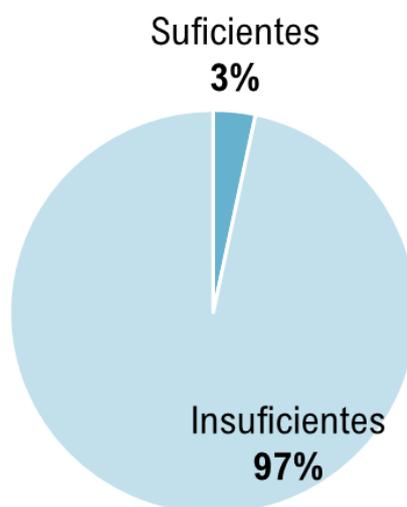
Fonte: Própria do autor

Na Figura 30, também é possível observar que o coeficiente de determinação ( $r^2$ ) é calculado em todas as validações realizadas e aqui estudadas, mas, apesar disso, como discutido por Raposo (RAPOSO, 2016), apenas este aspecto não sustenta sozinho as conclusões dentro do estudo de linearidade, mesmo que utilizados para fins comparativos. Bem como não sustenta evidências de uma análise crítica dos resultados analíticos nos ensaios de linearidade, pois o coeficiente de determinação ( $r^2$ ) não está completamente relacionado a vários outros fenômenos que podem ocorrer durante o estudo de linearidade (KOMSTA, 2012). Assim, se a avaliação da linearidade é reportada apenas com base na equação da reta obtida da regressão linear e pelo coeficiente de determinação ( $r^2$ ), a avaliação da linearidade é realizada de forma insuficiente.

Dentro desse contexto, para o estudo de linearidade, não apenas os cálculos precisavam estarem corretos, mas também, como mencionado no item 1.5.2.3 e na Tabela 2 para que o estudo de linearidade seja suficiente, testes estatísticos para avaliar a qualidade dos resultados obtidos da relação entre as variações das concentrações com os sinais obtidos dos instrumentos de medição utilizados. A Figura 31 mostra que a frequência que a linearidade foi avaliada com auxílio de testes estatísticos suficientes em apenas 3% das validações avaliadas.

Figura 31 - Características sobre a avaliação dos dados da curva mediante uso de testes estatísticos

## Testes estatísticos para estudo de linearidade



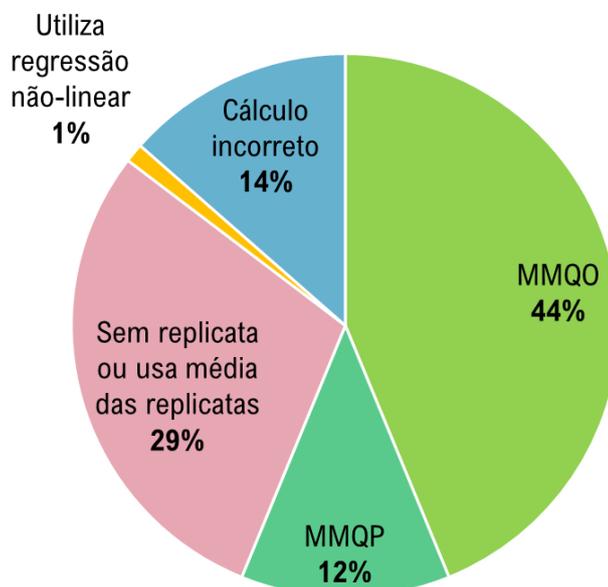
Fonte: Própria do autor

Os resultados da Figura 31, demonstram que, mesmo que, neste trabalho, apenas alguns dos aspectos que devem ser avaliados no estudo da linearidade (Tabela 2 **Erro! Fonte de referência não encontrada.**), ainda assim a avaliação de itens básicos essenciais, frequentemente não são avaliados, tais como precisão e exatidão na curva de calibração.

Algumas justificativas para os resultados da Figura 31, ainda podem ter como fundamento alguns resultados em relação as características em relação aos modelos de regressão linear. A Figura 32 demonstra que, quando foi possível identificar o modelo de regressão utilizado no documento, a maioria das regressões lineares presentes nos documentos utilizaram os mínimos quadrados ordinários (MMQO), 29% não utilizam replicatas para a curva ou utilizaram a média das replicatas para encontrar a equação da reta, sem que fossem realizados o estudo dos resíduos, em 14% das avaliações apresentaram-se com o cálculo incorreto, pois não foi possível identificar o modelo de regressão, uma vez que as equações citadas no documento, não estavam de acordo com nenhuma das fornecidas pelo software ou não existiam dados para identificar o modelo utilizado.

Figura 32 - Características das regressões lineares nos documentos avaliados

### Características das regressões lineares presentes nos documentos



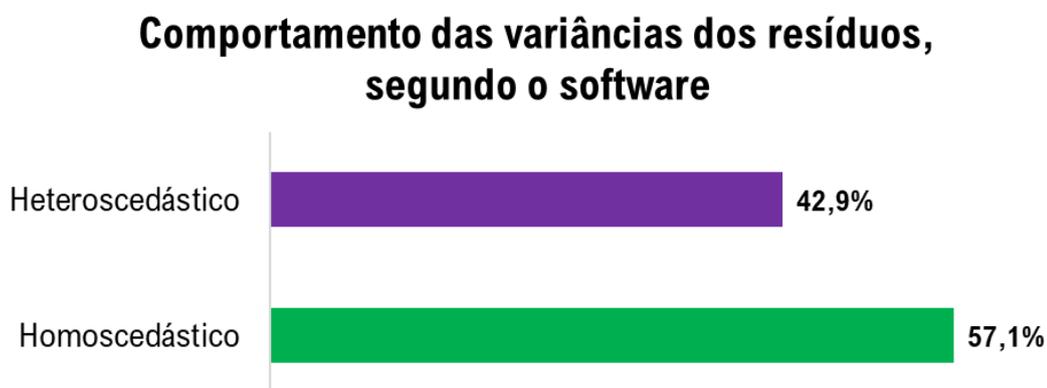
MMQO - Regressão linear realizada através dos mínimos quadrados ordinários  
MMQP - Regressão linear realizada através dos mínimos quadrados ponderados

Fonte: Própria do autor

Em relação aos 29% que não utilizam replicatas para a curva ou utilizaram a média das replicatas para encontrar a equação da reta, sem que fossem realizados o estudo dos resíduos, é importante citar que, então, o modelo de regressão utilizado, automaticamente foi dos mínimos quadrados ordinários (MMQO), mas que na Figura 32 não foi contabilizado como MMQO.

As características em relação ao comportamento da variância dos resíduos obtidas dos relatórios de validação do ConfLab Validação (FERREIRA, 2020, vol. 2; PACCES; OLIVARES, 2020), demonstram que a maioria dos testes de Cochran realizados (MILLER; MILLER, 2010) resultaram em um comportamento homoscedástico, como mostra a Figura 33.

Figura 33 - Comportamento das variâncias dos resíduos, segundo os testes de Cochran realizados pelo software ConfLab Validação



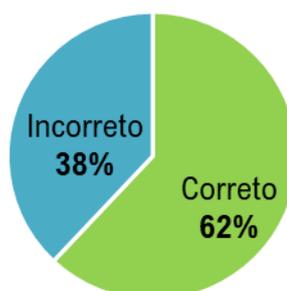
Fonte: Própria do autor

Dentre os documentos nos quais foi possível identificar qual o modelo de regressão utilizado, 62% deles demonstraram realizar a regressão de forma correta, que, segundo os critérios deste trabalho, tinham como base os resultados oferecidos pelos relatórios de validação do ConfLab Validação (PACCES; OLIVARES, 2020), como mostra a Figura 34.

Figura 34 - Porcentagem de validações que, dos documentos nos quais foi possível identificar o modelo de regressão empregado, foram corretas e incorretas

### Características das aplicações dos modelos de regressões lineares (ordinários ou ponderados)

- Resultados do documento iguais aos valores de referência\* (corretos)
- Resultados do documento diferentes aos valores de referência\* (incorretos)



\*Os valores de referência são os obtidos através do software ConfLab (PACCES; OLIVARES, 2020)

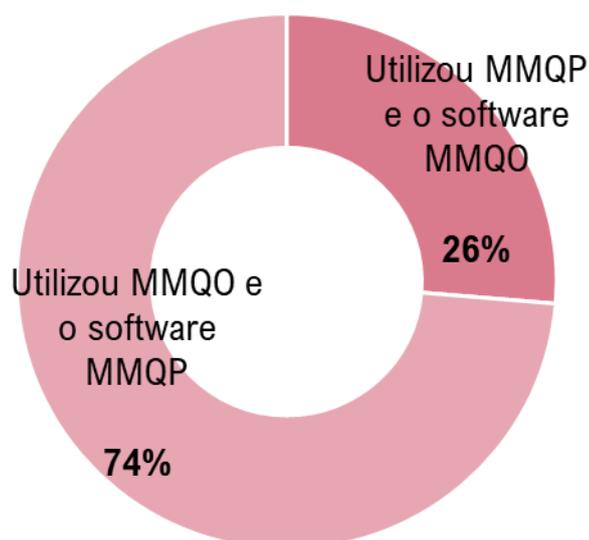
Fonte: Própria do autor

Das validações nas quais o modelo de regressão realizado no documento foi incorreto, ou seja, não estavam iguais ao do software (PACCES; OLIVARES, 2020), em 74% o motivo foi devido ao modelo utilizado no documento recorrer aos mínimos quadrados ordinários (MMQO) e o valor referência (software) indicou que o modelo mais adequado era fazendo uso dos mínimos quadrados ponderados (MMQP), como mostra a

Figura 35.

Figura 35 - Frequência dos equívocos em relação ao modelo de regressão linear utilizado no documento, comparados ao valor referência (software), dentre os modelos de regressão linear avaliados como incorretos.

**Modelo de regressão linear utilizada no documento  
*diferente* do valor referência**



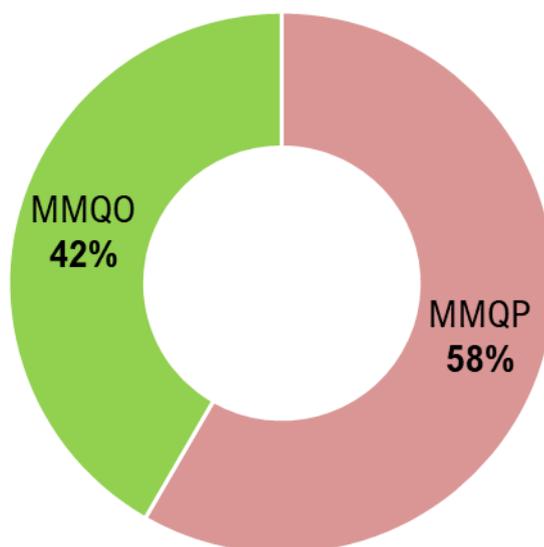
Fonte: Própria do autor

Dos 14% (Figura 32) dos documentos nos quais os cálculos dos modelos de regressão foram incorretos no documento, segundo os resultados do software, o modelo de regressão mais adequado, na maioria deles, seria o dos mínimos quadrados ponderados (MMQP).

Dentre as regressões lineares que os relatórios de validação demonstraram ser mais adequado realizar por meio dos mínimos quadrados ordinários (MMQO) (Figura 33), 47, 1% das validações, os respectivos documentos não citaram testes para verificar o melhor modelo de regressão, e para os resultados dos relatórios que indicaram os mínimos quadrados ponderados (MMQP) como modelo mais adequado (Figura 33), em 74,1% dos documentos não foram realizados testes estatísticos para o melhor modelo de regressão (Figura 37).

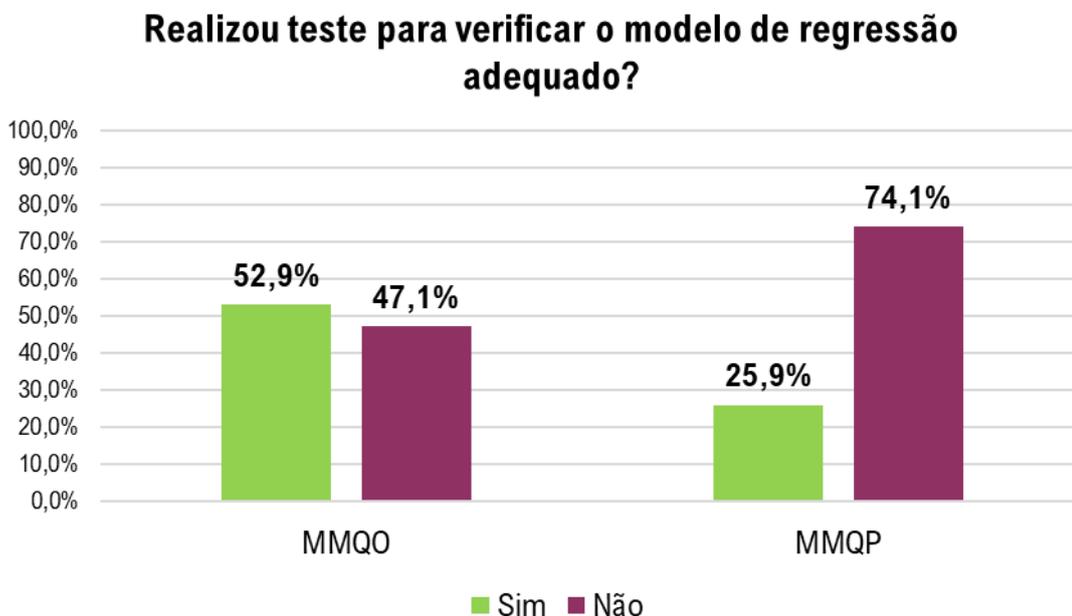
Figura 36 - Dentre as validações que não foi possível identificar qual modelo de regressão linear empregado, de acordo com os resultados do software, quais os modelos mais adequados para realizar a regressão.

### **Modelos de regressão linear com cálculos incorretos: Quais seriam os adequados?**



Fonte: Própria do autor

Figura 37 - Das validações nas quais foi possível avaliar o modelo de regressão linear utilizado, e dentre essas, as avaliações foram corretas, ou seja, as que os modelos de regressão do software estavam iguais ao do documento, quantas das validações realizaram os testes para verificar o modelo de regressão linear adequado.



Fonte: Própria do autor

Mesmo quando não é verificado o modelo de regressão linear mais adequado, ao realizar regressão linear pelos mínimos quadrados, esta é automaticamente realizada na forma ordinária. Considerando este aspecto, e relacionando os resultados da Figura 37 aos resultados da Figura 34, como na maioria dos testes para verificar a homoscedasticidade os comportamentos das variâncias dos resíduos foram homoscedásticos (Figura 33), e conseqüentemente, a maioria das regressões lineares realizadas pelo software foram mediante os mínimos quadrados ordinários (MMQO), mesmo que a maioria das validações tenham utilizado os modelos de regressão lineares da forma correta, pois obtiveram os mesmo resultados segundo o ConfLab, 47,1% (Figura 37), não realizaram os testes para verificar o melhor modelo de regressão, desta forma, na verdade, caso o comportamento das variâncias nesses estudos, não fossem homogêneas, talvez os resultados não estivessem corretos. Um dos impactos de uma avaliação mais específica em relação as características dentro dos estudos da linearidade é que, muitas vezes, os estudos de limite de detecção (LD) e limite de quantificação (LQ) podem ser estimados conforme os desvios padrão e

coeficientes com base na curva analítica estudada. Neste caso, se a avaliação da linearidade tenha sido realizada de forma insuficiente, incorreta ou insatisfatória, a probabilidade das estimativas e estudos dos valores de LD e LQ, também podem ser insuficientes e incorretos. E, se esses valores estimados com base na curva não forem confirmados com auxílio de outros meios, a probabilidade de obter um resultado inadequado e incorreto é maior ainda.

Por exemplo, nos estudos de limite de quantificação (LQ) e limite de detecção (LD) o texto do documento diz que é realizado o estudo de uma forma teórica por meio do desvio padrão de 3 curvas, mas na verdade o estudo de linearidade não foi realizado em 3 curvas analíticas, mas sim em apenas uma e em unicata. Então, na discussão sobre LQ e LD, é possível perceber que as estimativas dos valores desses parâmetros ocorrem com base no desvio padrão dos valores obtidos dos sinais do equipamento (respostas) em cada nível diferente de concentração estudado na curva. Ao final, para reportar os valores de LQ e LD, não existe confirmação desses valores de alguma forma, por exemplo, replicação do valor estimado, verificação de valores próximos aos valores estimados, relação sinal-ruído etc.

Este exemplo traz aspectos de alguns resultados aqui mencionados, como a forma de estabelecer as condições analíticas e critérios de parâmetros de validação. Isso é pode ser realizado de uma forma confusa, não existindo evidências de cálculos ou dados brutos disponíveis para confirmação do que possa estar ocorrendo. Não existem referências sólidas e consolidadas sobre critérios e condições, uma vez que não foram estabelecidos de forma adequada pelos autores ou foi utilizado mais de um protocolo na validação. Assim, se não existe confirmação dos valores estimados de LQ e LD, não é possível saber se os valores encontrados estão dentro de critérios de aceitação adequados, nem se os valores são reais.

Essas características se relacionam com o fato de que os limites de quantificação (LQ) e limites de detecção (LD) foram, depois da linearidade, os parâmetros que apresentaram mais equívocos nos cálculos Figura 28 sendo considerados entre os três parâmetros de validação que menos foram realizados dentro das condições adequadas de ensaio Figura 28.

A partir da observação das características nos estudos de linearidade foi possível observar que o coeficiente de determinação ( $r^2$ ) foi o parâmetro mais frequente para avaliação da linearidade porém, por ser inconclusivo e quando utilizado

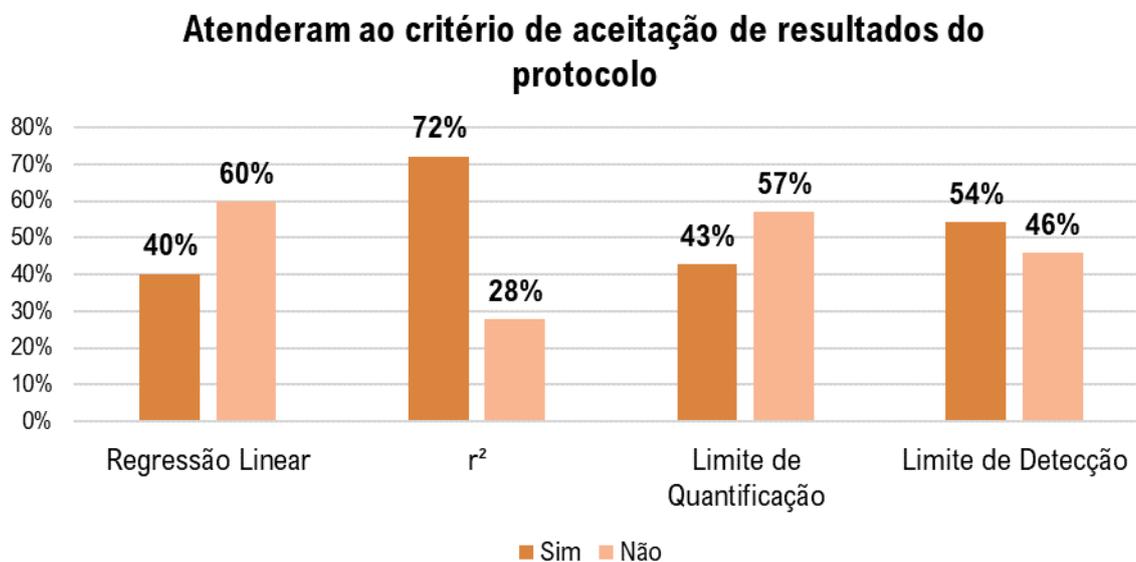
sozinho, segundo a literatura (RAPOSO; IBELLI-BIANCO, 2020), não sustenta os estudos de linearidade das validações realizadas, além de ser diretamente dependente de um teste de homogeneidade das variâncias dos resíduos consistente (DESHARNAIS et al., 2017; HARRIS, 2010; NASCIMENTO et al., 2010; RAPOSO, 2016; SAYAGO; ASUERO, 2004; SKOOG, 2014; SONAWANE et al., 2019), ou seja, testes para investigar o melhor modelo de regressão (ponderado ou ordinário), quando utilizado mínimos quadrados. Testes estes que não foram realizados em 70,5% das vezes que as regressões lineares foram realizadas recorrendo-se aos mínimos quadrados Figura 29 e que mesmo que maioria dos modelos de regressão linear tenham apresentado uma aplicação correta quando comparado aos resultados do software, como a maioria da regressões adequadas seriam pelos mínimos quadrados ordinários Figura 33, em aproximadamente metade dessas validações “corretas” Figura 34, como não foram realizados os testes de homogeneidade da variância dos resíduos, esses modelos de regressão ainda poderiam estar incorretos, caso o comportamento das variâncias dos resíduos fossem heteroscedásticas, o que confirma os testes estatísticos significativamente insuficientes em relação aos estudos de linearidade .

#### **4.2.3.2 Critérios de aceitação de resultado**

A verificação das características dos resultados de validação de métodos em relação aos critérios de aceitação de resultado recomendados pelos protocolos foi realizada em relação os critérios de regressão linear, coeficiente de determinação e limite de quantificação (LQ), que foram os estudos que obtiveram mais insuficiências e equívocos nas condições de aplicação, cálculos e testes estatísticos.

A Figura 38 mostra que o coeficiente de determinação ( $r^2$ ) esteve dentro dos critérios de aceitação em 72% dos estudos de linearidade; os modelos de regressão atenderam ao critério de aceitação do protocolo; 40% dos estudos e os limites de quantificação não atenderam aos critérios de aceitação de resultados em 57% das vezes que foi avaliado dentro do estudo de validação.

Figura 38 - Porcentagem que os resultados atenderam aos critérios de aceitação do protocolo

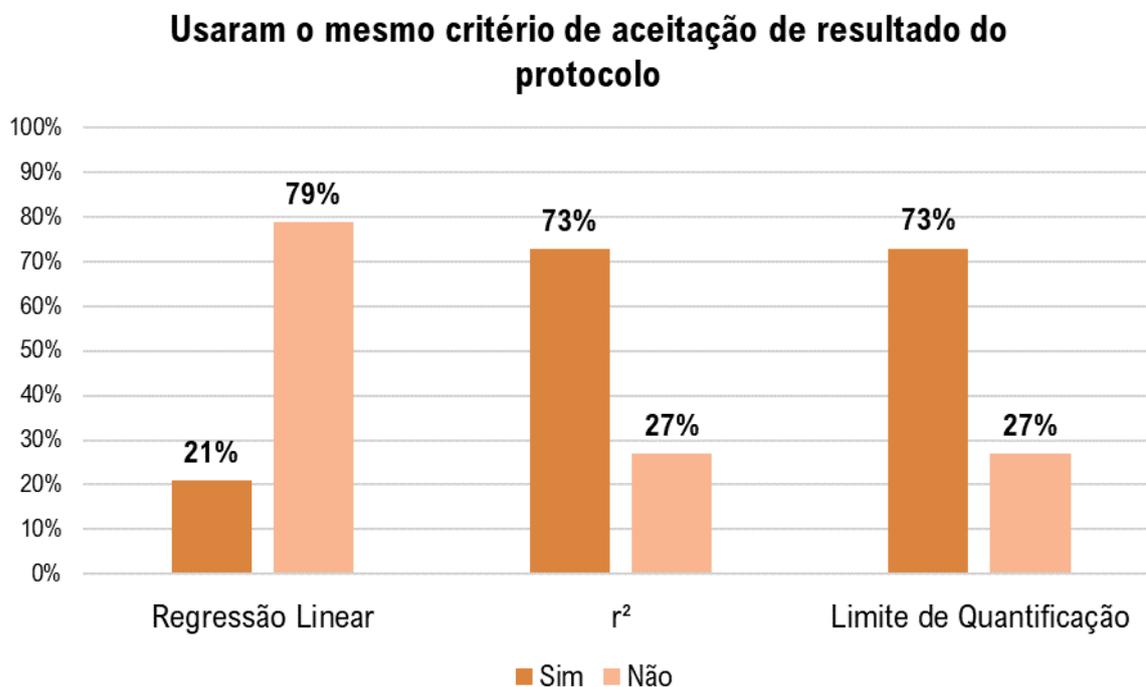


Fonte: Própria do autor

\* Os critérios de aceitação de resultados para LQ se diz respeito à precisão (repetibilidade) e exatidão do respectivo valor de concentração de LQ estimados.

Com relação ao estabelecimento de critérios de aceitação de resultados, em 79% das validações, considerando os critérios para a escolha do modelo de regressão, os autores não utilizaram o mesmo critérios de aceitação do protocolo (Figura 39). Para o  $r^2$  e limite de quantificação (LQ) em 73% das validações os autores utilizaram o mesmo critério dos protocolos (Figura 39). Quando o protocolo deixou a critério do autor estabelecer os critérios de aceitação resultado dos parâmetros de validação, foi considerado que o autor atendeu o critério de aceitação de resultados, mas que usou o próprio critério de aceitação de resultados.

Figura 39 - Frequência que utilizaram o mesmo critério de aceitação de resultados do protocolo

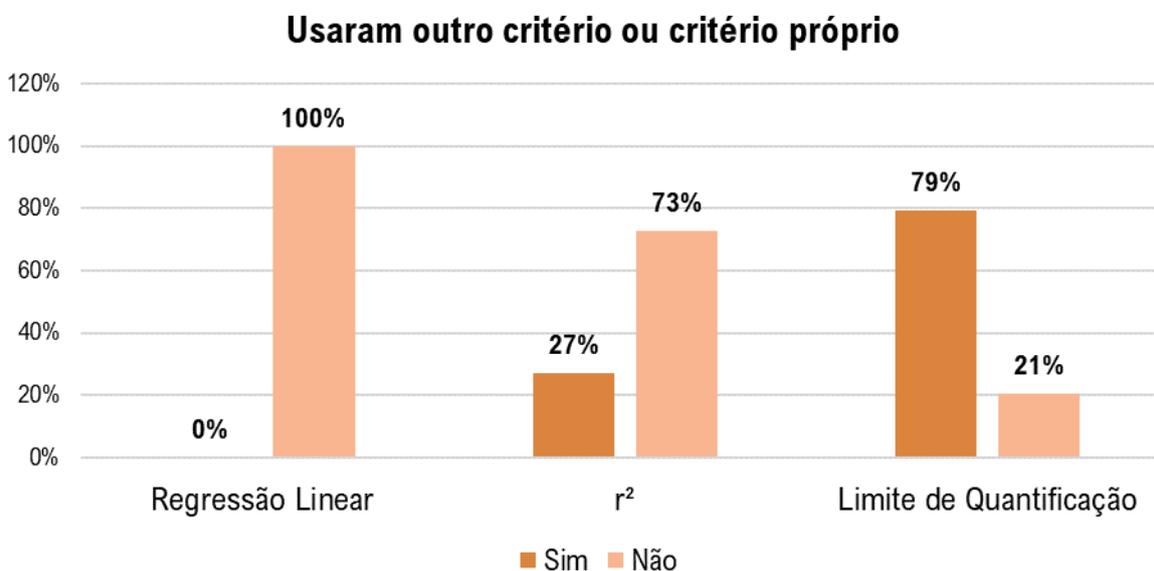


Fonte: Própria do autor

\*Os critérios de aceitação de resultados para LQ se diz respeito à precisão (repetibilidade) e exatidão do respectivo valor de concentração de LQ estimados.

Quando não utilizou o mesmo critério de aceitação do protocolo, ou o protocolo deixou a critério do pesquisador o estabelecimento do mesmo, foi considerada que a validação foi realizada sem nenhum critério de aceitação estabelecido (Figura 40). Desta forma, é possível que estes resultados demonstrem que a avaliação dos dados obtidos e análise das características do método, foi então realizada sem fundamentação crítica para aceitação ou rejeição de seus resultados, não sendo possível informar e evidenciar se o método é adequado ou não ao uso pretendido.

Figura 40 - Quando não utilizou o mesmo critério de aceitação do protocolo, ou o protocolo deixou a critério do pesquisador estabelecer o critério, estão dispostas as frequências que foram utilizados outros critérios



Fonte: Própria do autor

\* Os critérios de aceitação de resultados para LQ se diz respeito à precisão (repetibilidade) e exatidão do respectivo valor de concentração de LQ estimados.

O propósito de avaliar as características em relação ao estabelecimento de critérios de aceitação de resultados, não foi apenas verificar se os autores realizavam cálculos e testes estatísticos corretamente ou não, nem apenas se atendiam ou não o critério do protocolo em relação aos parâmetros de validação, mas também a capacidade que os desenvolvedores dos métodos tinham em avaliar esses resultados obtidos com fundamentação adequadamente científica, estatística e lógica., e então poder informar e evidenciar as características do método em questão com fundamentação crítica. Ao avaliar os resultados dos cálculos e testes estatísticos, para dizer se os parâmetros estão adequados ou não, critérios para aceitação ou rejeição desses resultados precisam estar pré-estabelecidos para que seja realizada a análise crítica dos mesmos.

Nos resultados obtidos da Figura 38, Figura 39 e Figura 40 , é possível observar em relação ao coeficiente de determinação ( $r^2$ ) que a maioria dos documentos atende ao critério de aceitação do protocolo. No entanto, dos que não utilizaram o critério de aceitação do protocolo, 73% não estabeleceram critério algum.

Para a regressão linear, 60% não atenderam ao critério de aceitação de resultados do protocolo. Apenas 21% do total utilizou os mesmos critérios de aceitação do protocolo. E, finalmente, dos que não utilizaram os critérios do protocolo, nenhum estabeleceu um critério próprio para regressão linear.

Para o limite de quantificação (LQ), 73% utilizaram os mesmos critérios de aceitação do protocolo e 79% estabeleceram um critério próprio. Apesar disso apenas em 43% os resultados estavam dentro dos critérios de aceitação estabelecidos.

Juntando essas características com as obtidas em relação aos cálculos (item 4.2.3) com o uso do protocolo (item 4.2.1) e das características sobre rastreabilidade (item 4.2.2) têm-se um conjunto de informações que caracterizam os documentos avaliados neste trabalho, que serão discutidos nos próximos itens.

### 4.3 Valores de incerteza de medição

Os valores de incerteza neste item reportados, se referem as incertezas expandidas, que são o produto das incertezas-padrão combinadas de cada parâmetro de desempenho estudado pelo documento e possível de ser avaliado pelo software ConfLab, por um fator de abrangência igual a 2 ( $k=2$ ).

Para analisar os valores encontrados, estes foram separados em grupo de acordo com as concentrações em cada curva de calibração:

- **Primeiro ponto da curva:** menor concentração que foi contemplada pela curva, podendo ou não ser o limite quantificação (LQ);
- **Ponto de média concentração:** o valor de concentração contemplado pela curva que mais se aproximava do valor médio da amplitude da curva;
- **Último ponto de concentração:** maior concentração que foi contemplada pela curva.

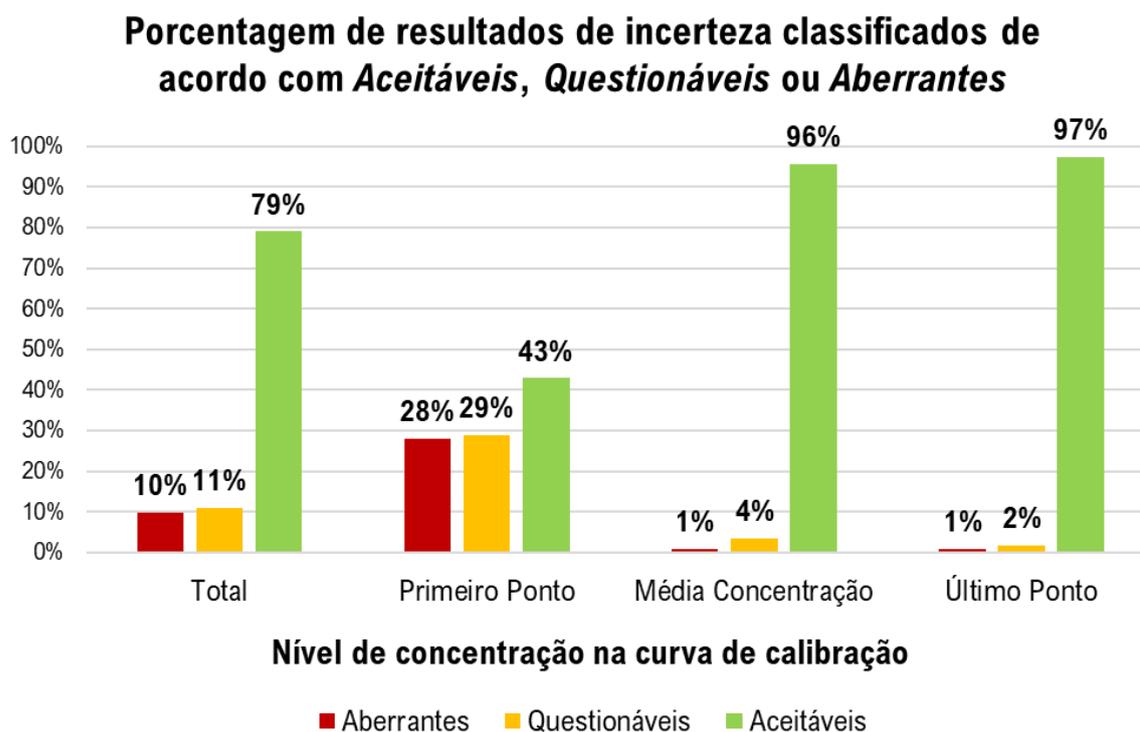
Além disso, os valores encontrados foram classificados como:

- **Aceitáveis:** valores de incerteza menores do 40%
- **Questionáveis:** Valores de incerteza entre 40% e 100%;
- **Aberrantes:** valores de incerteza maiores do que 100%.

Desta forma, os resultados obtidos pela Figura 41 demonstram que, no total (todas os pontos de concentração juntos), a maioria dos valores foram aceitáveis, porém no primeiro ponto a maioria (57%) obtiveram valores questionáveis (29%) ou aberrantes (28%). Nos últimos pontos e nas médias concentrações, quase todos os resultados estiveram dentro dos valores aceitáveis.

Os resultados Figura 41 que demonstraram que a maioria dos valores totais estavam dentro do aceitável (menor que 40%), estão relacionados ao fato de quase 100% dos resultados das concentrações médias e nos últimos pontos também estarem dentro do mencionado aceitável.

Figura 41 - Valores de incerteza expandida obtidas dos resultados dos parâmetros de validação, separados no gráfico pelos por três pontos diferentes de concentrações de cada curva de calibração estudada. São eles: primeiros pontos das curvas de calibração, os pontos de média concentração e últimos pontos.



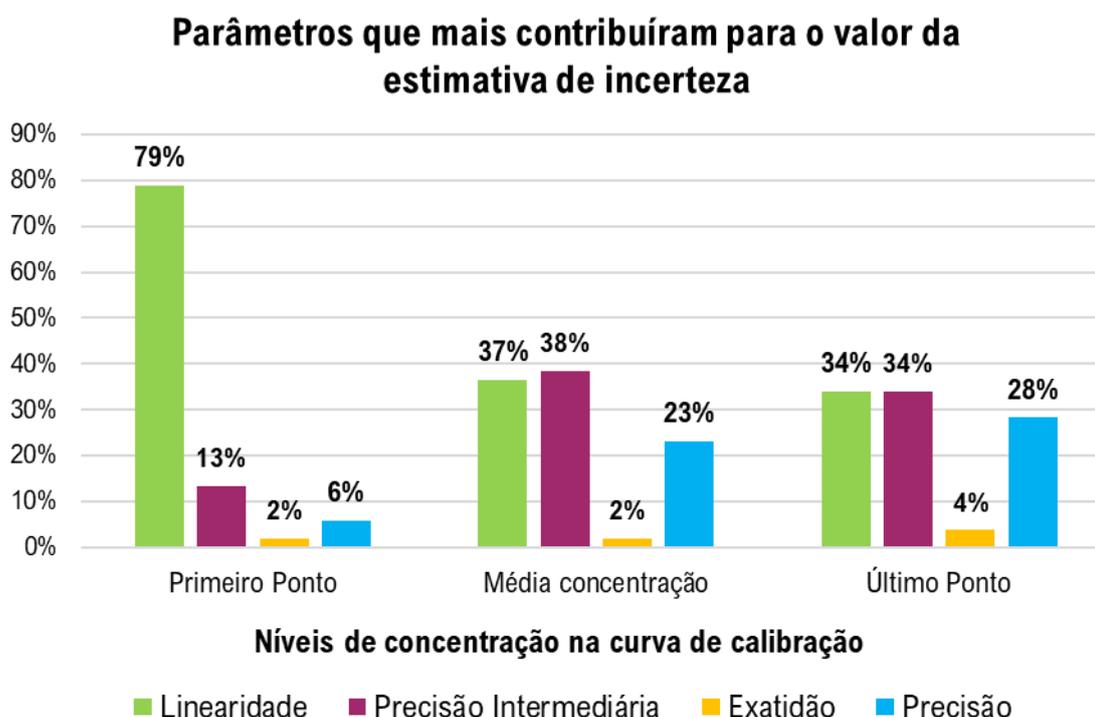
Fonte: Própria do autor

Os valores de incerteza questionáveis e aberrantes, apesar de somarem uma minoria de 22% do total Figura 41, em relação aos primeiros ponto da curva, somam 57% (Figura 41), e este resultado demonstra que a probabilidade de se encontrar os valores das concentrações nos primeiros pontos, utilizando esses métodos, em 29% deles, está entre 0% e 60%, e em 28%, não existe probabilidade encontrar algum resultado, visto que a incerteza é maior do que 100%.

Dentre os parâmetros que mais contribuíram para as estimativas de incerteza (Figura 34), a linearidade foi o parâmetro mais frequente no primeiro (81%) e no último ponto (41%) de concentração da curva, e nas médias concentrações, a linearidade praticamente se manteve no mesmo nível de contribuição (38%) com a precisão intermediária (40%).

Os resultados apresentados mostraram que valores das estimativas de incerteza foram elevados, o que indica que a probabilidade estatística de ser obter um resultado utilizando os métodos aqui avaliados, em alguns casos foram, de certa forma, nulas apresentando resultados maiores do que 100%.

Figura 42 - Parâmetros que mais contribuíram para a estimativa a incerteza expandida



Também, em relação aos valores de incerteza e quais os parâmetros que mais contribuíram para seus valores no primeiro ponto, com base no Ciclo AQAC, demonstram que existe a possibilidade de relação entre os equívocos insuficiências em relação aos estudos de linearidade e sua relação com as estimativas equivocadas dos valores dos limites de detecção e quantificação, visto que os valores de incerteza, no primeiro ponto, na maioria das vezes se apresentaram maiores de 40%, e dentre estes, quase a metade apresentaram valores maiores do que 100%.

#### **4.4 Discussões dos resultados**

Muitos conceitos sobre qualidade analítica ainda parecem estar distantes da pesquisa científica, porém o propósito da publicação de uma validação de método analítico não é diferente do que o de um “POP”, em resumo, o de “minimizar a ocorrência de desvios na execução de tarefas fundamentais para a qualidade do ensaio, independente de quem as faça” (PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020). A validação de métodos precisa ser realizada de forma que seja possível “aumentar a previsibilidade de seus resultados, minimizando as variações causadas por imperícia e adaptações aleatórias dos métodos” (PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020), com base no uso correto de testes e ferramentas estatísticas, sustentado por uma compreensão adequada e suficiente a fim de diminuir a probabilidade de erros aleatórios chegarem aos resultados, na análise crítica de dados e suas conclusões.

O contexto mundial durante a pandemia da COVID-19 deixou claro que cálculos estatísticos, a competência e conhecimento dos cientistas sobre seu uso suficiente e correto são imprescindíveis e necessários para fornecer evidências e sustentar decisões pautadas na qualidade de dados e resultados para a resolução de grandes problemas. Neste contexto, o uso correto e suficiente de ferramentas e testes estatísticos precisam ser utilizados pelos desenvolvedores de métodos analíticos de forma satisfatória, completa, eficaz, adequada, suficiente e segura para que ofereçam resultados não equivocados sobre as características do método.

Diferente de um laboratório de rotina, na pesquisa científica, o mais importante ao desenvolver e publicar um novo método analítico, não é que este seja somente “sensível”, “seletivo”, “preciso”, “exato”, que quantifica ou detecta baixíssimas

concentrações ou possui uma grande faixa linear para quantificação, mas sim, deve fornecer dados das características daquelas condições analíticas, pois este é o caráter investigativo da ciência: fornecer evidência sobre a verdade (ou não) dos fenômenos, processos, etc. Ou seja, não se busca ou se força um parâmetro ideal; apenas o quantifica, determinando o seu significado com relação ao futuro uso do método.

Para a pesquisa científica, se um procedimento analítico não for adequado ao uso pretendido, ou seja, se, de acordo com os critérios adequados ao objetivo do método, este não for seletivo, sensível, exato ou preciso em determinado intervalo de concentração, esses dados também trazem informações úteis e importantes, uma vez que evitam vários tipos de desperdícios e fornecem informações para que os mesmos ou outros pesquisadores otimizem esses métodos. Os resultados reportados com a sua respectiva incerteza, também permitem informar a todos os interessados o real contexto daquele estudo e, assim, o leitor daquela pesquisa pode avaliar se o método é adequado para uso.

Desta forma, talvez o mais importante, ao desenvolver e publicar um método analítico, não seja necessariamente que este apresente valores “impecáveis”, mas sim, que sejam baseados na veracidade e adequação dos dados obtidos, refletindo as suas reais características.

## 5 Conclusão

Com base na *survey* de Mônia Baker (BAKER, 2016), as investigações sobre soluções pautadas em melhorias nas definições de condições e critérios de parâmetros de validação, esforços em estudos colaborativos de reprodução em laboratórios externos, análise adequada, suficiente e eficaz de dados e, conseqüentemente, uma adequada apresentação de resultados, apresentaram uma probabilidade estatística a ser considerada em relação a existência de uma crise de reprodutibilidade causada pelo uso equivocado e insuficiente da validação de métodos como ferramenta estatística, sustentada pelos elevados valores de incerteza expandida de medição associadas as validações de métodos publicadas em documentos científicos.

A Figura 43 traz um resumo dos resultados dos aspectos que estiveram adequados, conforme estabelecido neste trabalho.

Figura 43 - Resumo dos resultados nos quais os autores realizaram os estudos conforme as recomendações do protocolo, os resultados dos parâmetros de validação atenderam aos critérios de aceitação de resultados dos protocolos, no primeiro ponto e que os valores de incerteza associadas aos resultados estavam dentro dos valores aceitáveis, ou seja, menores do que 40%, também no primeiro ponto.



Fonte: Própria do autor

Esses resultados podem passar a impressão ao leitor de que os valores de incerteza em sua maioria estiveram dentro do aceitável, o que iria contra os resultados encontrados em relação à validação de métodos e os estudos dos parâmetros realizados pelos autores. Mas, essa diferença está relacionada aos seguintes pontos:

- 1- O critério para avaliar os valores de incerteza foi maleável considerando incertezas abaixo de 40% como aceitáveis:
  - 40% é o dobro das dispersões (CV) aceitas pelos protocolos utilizados, nos quais admitem 20% de precisão, recuperação e desvio da curva;
  - Se fossem considerados como aceitáveis valores menores do que 20%, essa porcentagem cairia de 43% para 34%, valor que se aproxima da soma dos valores sobre as validações e compreensão dos autores (Figura 43).
- 2- Foram utilizados apenas dados da validação para o cálculo de incerteza;
- 3- A validação foi avaliada com valores pré-estabelecidos pelos protocolos, ou seja, valores “cravados”, e a incerteza foi estabelecida como critério próprio, pois o objetivo foi encontrar de fato valores com baixa probabilidade de reprodução.
- 4- A maioria utilizou cromatografia e espectrometria de massas, que são equipamentos que possuem boa precisão, que proporciona uma menor variabilidade.
- 5- É também importante ressaltar que a incerteza foi ignorada, e não calculada pelos autores.

## **5.1 Avaliação da percepção da existência da crise de reprodutibilidade e o uso equivocado de validação de métodos**

A partir das considerações levantadas nas Figura 14 e Figura 15 (itens 3.3.3 e 3.5), e os resultados obtidos no item 4.2, foi possível perceber uma tendência de características nos documentos científicos publicados e estudados neste trabalho que aumentam a probabilidade de que o conhecimento sobre validações de métodos seja de fato insuficiente, como sugeriram os pesquisadores entrevistados por Monya Baker (BAKER, 2016), bem como a existência de uma probabilidade significativa de que

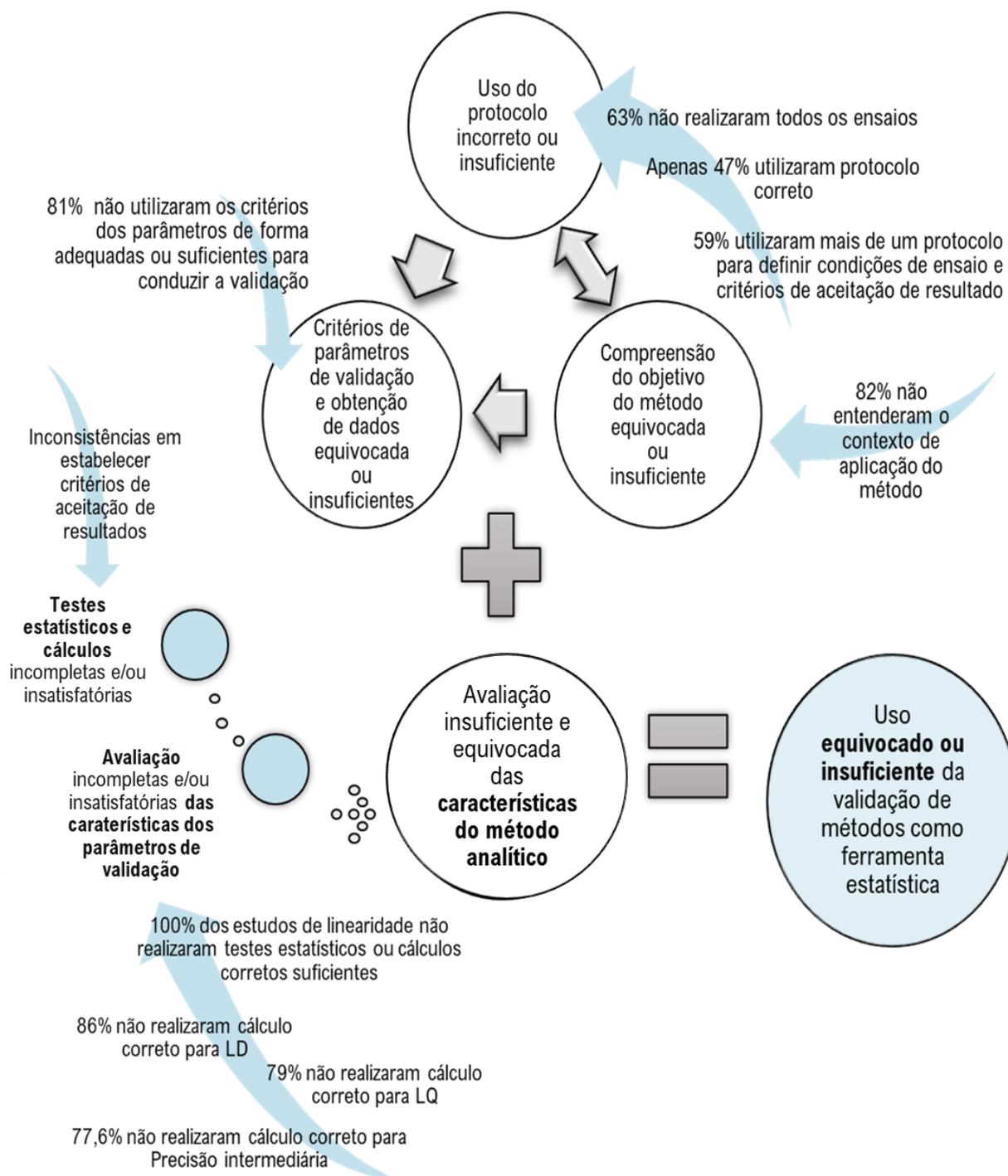
sejam estes a causa da dificuldade de reprodução desses métodos em outros laboratórios. Além disso, com base nas Figuras Figura 17 e Figura 28 também foi possível caracterizar que as reproduções nos próprios laboratórios é baixa, como citado também pela *survey* (BAKER, 2016), visto que a precisão intermediária foi o parâmetro de validação que aparece com menor frequência, e quando avaliados seus cálculos realizados.

A partir desses pontos, a avaliação das considerações apresentadas, de que um ciclo de insuficiências e equívocos na compreensão do objetivo do método, uso do protocolo, condições de ensaio e obtenção dos dados de validação, somadas a uma avaliação insatisfatória ou insuficiente das características do método analítico, causadas pelo uso incorreto ou incompleto de testes e cálculos estatísticos, seriam a causa do uso inadequado e insuficiente da validação de métodos, considerando:

- ✓ Quando o desenvolvedor do método não obtinha conhecimento suficiente sobre o objetivo e escopo do método em desenvolvimento (Figura 18) não houve uso do protocolo correto, ou não foi utilizado de forma suficientemente (completa) e, conseqüentemente, as condições e critérios de estudo de parâmetros de validação e obtenção de dados, também não;
- ✓ As possíveis avaliações incompletas e/ou insatisfatórias dos testes estatísticos, provocaram uma avaliação incorreta dos parâmetros de validação, que podem ter como consequência uma possível avaliação insuficiente e equivocada das características do método analítico em desenvolvimento;
- ✓ Os resultados podem aumentar a ocorrência de dificuldade na reprodução de métodos analíticos devido ao uso insuficiente de validações de métodos como ferramenta estatística na pesquisa científica na área de química analítica e afins.

A Figura 44 traz um esquema relacionando os resultados encontrados com o uso insuficiente e inadequado da validação de métodos.

Figura 44 - Relação entre os resultados encontrados e o uso de validação de métodos



Fonte: Própria do autor

O infográfico apresentado na Figura 44 estabelece conexões entre os resultados para verificar a coerência da possível conclusão a respeito da probabilidade da existência da relação entre as características encontradas com o uso da validação de métodos, que mostrou-se razoável com base nos resultados de Monya Baker (BAKER, 2016) e com algumas afirmações de Goodman, et.al. mencionadas. (GOODMAN; FANELLI; IOANNIDIS, 2018)

## **5.2 Características da apresentação e da qualidade dos resultados**

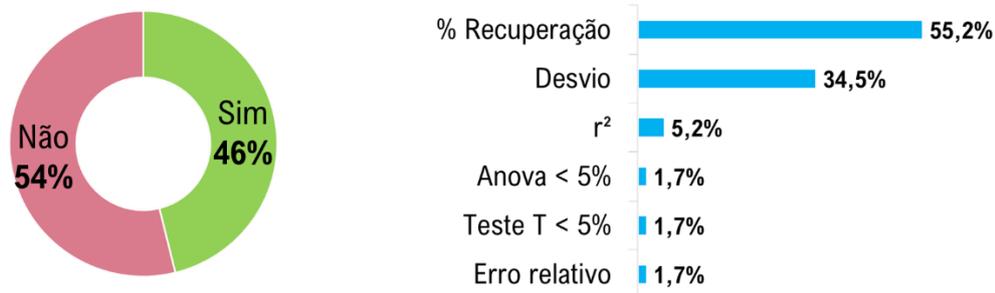
Retomando os critérios estabelecidos para identificar uma apresentação e análise apropriada e suficiente dos dados de validação:

- uso adequado do protocolo, mediante o atendimento das condições de ensaio e critérios de aceitação de resultados;
- evidências do uso correto dos cálculos;
- conjunto de evidências que levam ao aumento da confiabilidade no resultado com base na rastreabilidade adequada;
- uso de cálculos de estimativa de incerteza, bem como, o seu resultado reportado com o valor associado ao mensurando.

A Figura 45 permite evidenciar que 54% das validações não relatam o intervalo de abrangência dos resultados, e dos que apresentaram, a maioria reportou como porcentagem de recuperação seguido de coeficiente de variação (CV).

Figura 45 - Forma de apresentação e reporte de resultados das validações de métodos estudadas

**a) Apresentou intervalo de aplicação?    b) Como apresentou o intervalo de aplicação?**



Fonte: Própria do autor

A maioria dos documentos não forneceram o intervalo de abrangência do método (53,8%; Figura 45 (A)), e dessa maioria 55,2% dos autores (Figura 45 (B)), ao utilizarem os valores de recuperação, demonstraram entender que estes são os parâmetros que fornecem informações suficientes para evidenciar o intervalo de abrangência do método, ou seja, informações sobre o intervalo que contém o conjunto de valores verdadeiros do mensurando que se pretende obter, aplicando o determinado método, com probabilidade determinada.

Como mencionado, na abordagem de erro para descrever as medições, o valor verdadeiro de uma grandeza é considerado único e, na prática, impossível de ser conhecido (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a). Também retomando os aspectos da reprodutibilidade de métodos e de resultados com base na abordagem de incerteza (GOODMAN; FANELLI; IOANNIDIS, 2018; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a), e relacionando-os ao Ciclo AQAC (OLIVARES; LOPES, 2012) no que diz respeito a qualidade de resultados, é possível verificar que estes demonstram características que definem uma baixa frequência de preocupação com a forma de apresentação e, se a informação que está sendo fornecida é satisfatória para o fim que se pretende, o método é adequado ao uso.

Essas características também aproximam os conceitos sobre uso insuficiente e inadequado da estatística (BAKER, 2016), pois um intervalo de abrangência pode ser estimado por meio de uma incerteza de medição expandida, na qual os resultados obtidos dos estudos da validação está entre as informações que contribuem para seus

valores, de acordo com o Ciclo AQAC (OLIVARES; LOPES, 2012).

De todas as validações de métodos estudadas, nenhuma delas reportou a incerteza total do método e 43% obtiveram os valores de estimativa de incerteza expandida menores do que 40% no primeiro ponto da curva de calibração, como mostra a Figura 46. Desses, 28% obtiveram valores maiores do que 100% de incerteza, também na Figura 46.

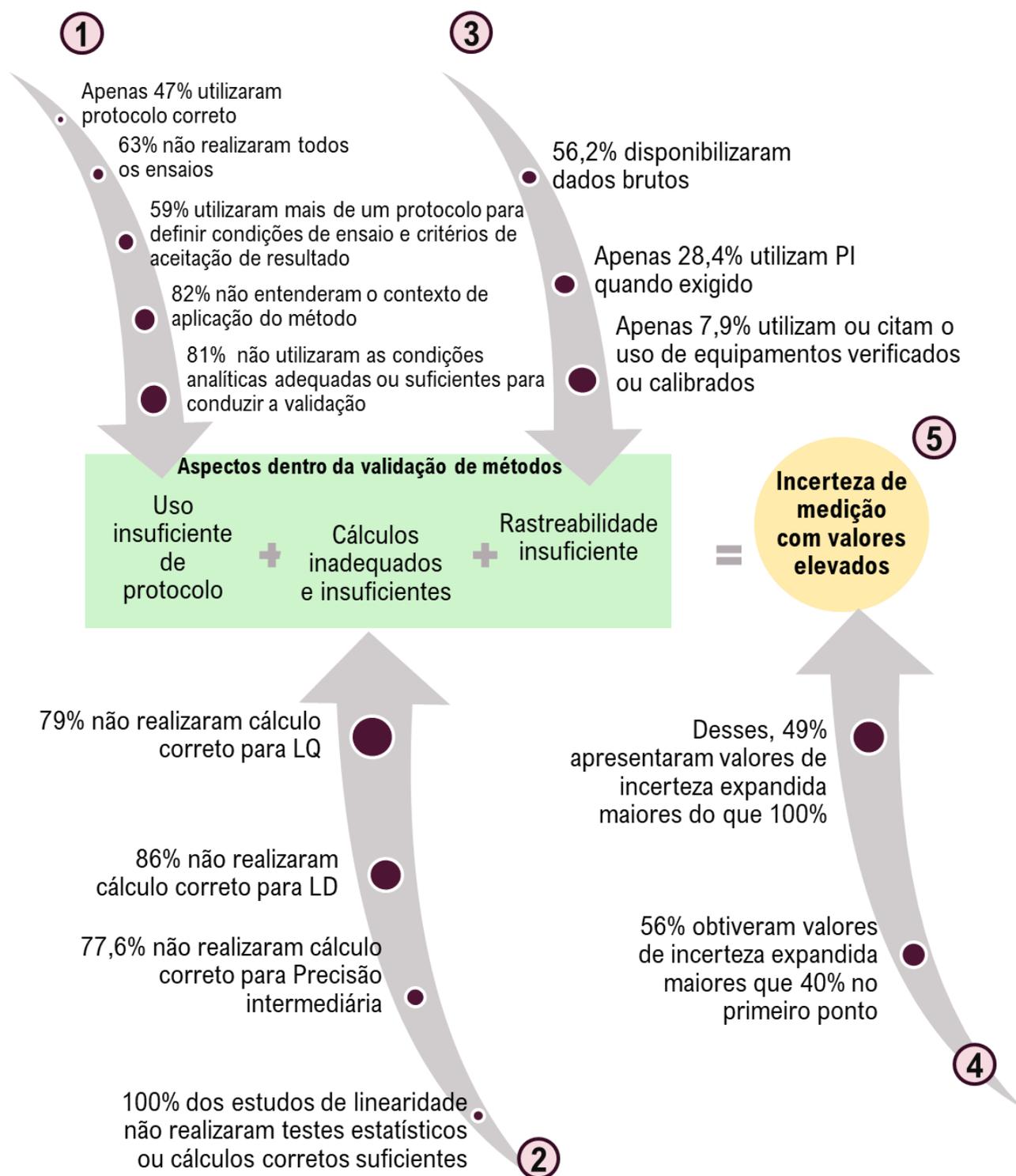
A Figura 47 traz um esquema, que fazendo uso de conexões dos resultados obtidos, estabelece a possibilidade da relação entre uso equivocado insuficiente das validações de métodos com os valores de estimativa de incerteza encontrados.

Figura 46 - Valores das estimativas de incerteza expandida associadas as validações de métodos estudadas no primeiro ponto da curva de calibração

<b>Incertezas dentro de um limite aceitável para métodos</b>	<b>Questionável</b>	<b>Aberrantes</b>
Validações com incerteza, no 1º ponto, menor que 40%	Validações com incerteza, no 1º ponto, entre 40% e 100%	Validações com incerteza, no 1º ponto, maiores que 100%
<b>43%</b>	<b>29%</b>	<b>28%</b>

Fonte: Própria do autor

Figura 47 - Relação entre uso equivocado de validação de métodos e incerteza de medição



Fonte: Própria do autor

Com o auxílio do Ciclo AQAC (OLIVARES; LOPES, 2012), foi possível relacionar as características obtidas em relação ao uso de validação de métodos a incerteza de medição, de forma que, como permite revelar a partir do esquema da Figura 46, há coerência na existência da probabilidade de que os altos valores de estimativas de incerteza sejam consequência das insuficiências e equívocos cometidos durante a realização dos processos de validação de métodos.

### **5.3 Relação entre incerteza de medição e crise de reprodutibilidade**

Com base na conclusão sobre proposta na Figura 15, no Ciclo AQAC e nos resultados até aqui obtidos, é possível relacionar a incerteza de medição com a crise de reprodutibilidade, atendendo a própria definição de incerteza de medição, pois os valores desta fornecem informações que caracterizam a dispersão dos valores que poderiam ser razoavelmente atribuídos ao mensurando (EURACHEM, 2012; INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a)). Desta forma, quanto maior o valor da incerteza relacionado ao resultado analítico obtido para o método validado, maior a dispersão dos valores em relação a um valor do mensurando que for encontrado utilizando o respectivo método.

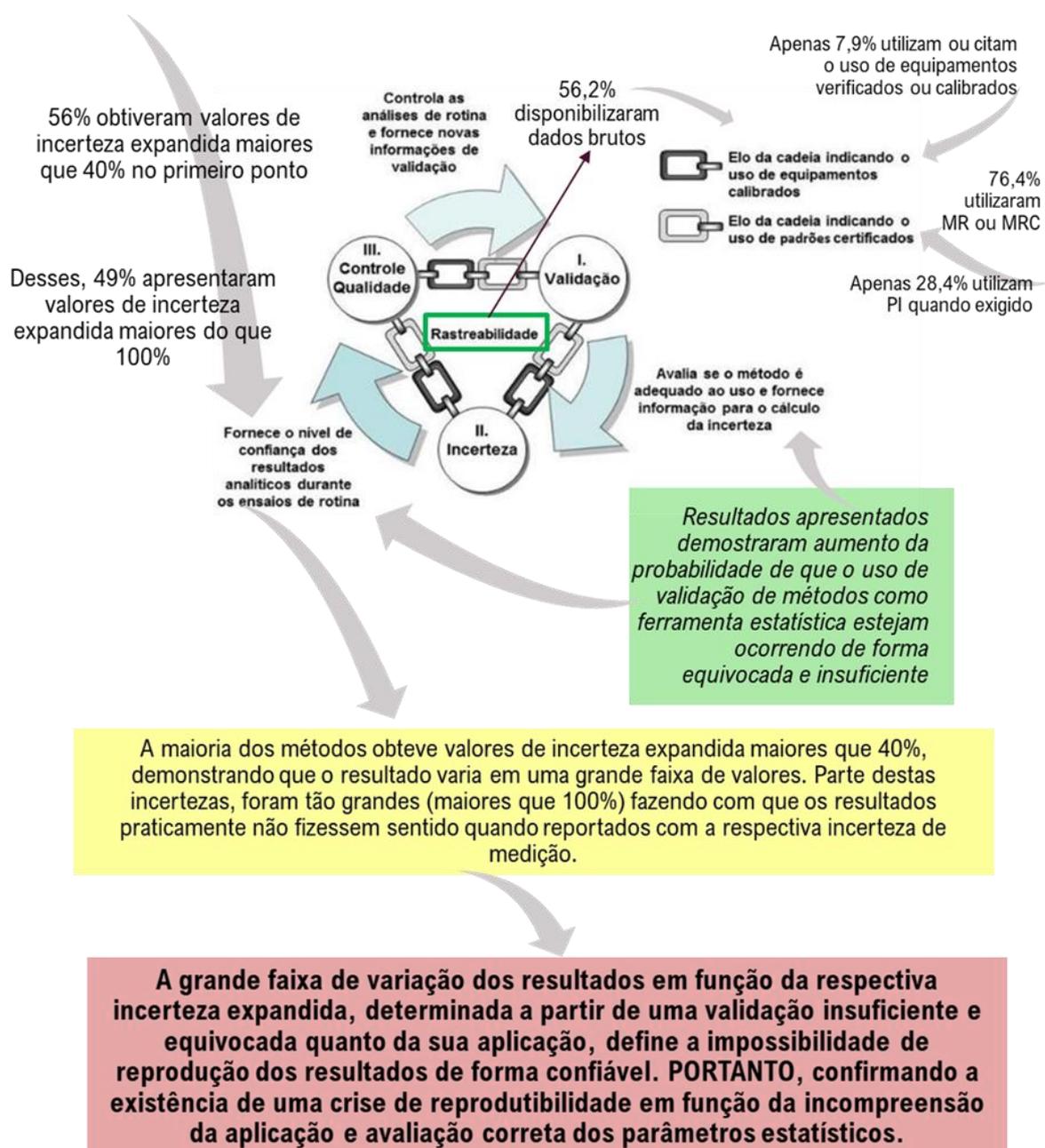
Com base nos resultados, é possível verificar que a maioria dos métodos aqui estudados possuem valores de incerteza expandida maiores que 40%. Destes, 49% possuem valores de incerteza expandida maiores que 100%. Estes valores resultam do fato que estes métodos desenvolvidos pretendem quantificar ou detectar um intervalo de aplicação (faixa de trabalho) muito extenso. Além disso, ao atribuir uma incerteza de 100%, pode-se afirmar que existe uma probabilidade estatística inexistente de se encontrar o valor do objeto que se pretende mensurar.

Com os valores elevados das incertezas encontradas, é possível a relação entre as conclusões sobre o uso de validação de métodos não ser realizada de forma estatisticamente adequada e suficiente e a dificuldade de se reproduzir estes métodos. Uma vez que, esses valores elevados indicam uma baixa probabilidade estatística de se obter um resultado consistente e confiável ao se utilizar os métodos aqui avaliados, dentro dos intervalos mencionados pelos documentos. É possível que este seja o sentido que os entrevistados pelo artigo (BAKER, 2016) disseram em

relação a uma “crise de reprodutibilidade causado por uso insuficiente da estatística”.

A Figura 48 demonstra um infográfico que relaciona o Ciclo AQAC, os resultados encontrados e a reprodutibilidade.

Figura 48 - Relação entre incerteza de medição e reprodutibilidade de métodos analíticos



Fonte: (OLIVARES; LOPES, 2012 adaptado)

## **5.4 Considerações a respeito de medidas na tentativa de amenizar a crise**

As pesquisas precisam incorporar, na impossibilidade de uma implantação formal de um sistema de gestão da qualidade, pelo menos os conceitos que estes abordam para se garantir a competência técnica destes laboratórios. Uma alternativa à implantação formal de um sistema de gestão da qualidade, seria o emprego de forma contundente do Ciclo AQAC (OLIVARES; LOPES, 2012) e seus princípios, com controles rígidos do processo de manutenção, calibração e, quando aplicável, qualificação dos equipamentos envolvidos, considerando um período razoável para repetição destas atividades ao longo das pesquisas desenvolvidas nestes laboratórios.

Uso de softwares ou planilhas validadas baseado em normas, guias ou manuais de acordo com a natureza do método empregado, observando a forma prática de execução dos ensaios.

De forma geral, ainda é necessária uma mudança de paradigma na pesquisa científica sobre o foco de se desenvolver e publicar novos métodos analíticos, com o objetivo de quebrar a cobrança pela alta produtividade de artigos, o que pode acarretar uma menor atenção na aplicação das ferramentas estatísticas de forma suficiente e adequada. Para isso, talvez os dados apresentados possam refletir a necessidade de inclusão de disciplinas relativas à gestão da qualidade, validação de métodos e estatística inferencial nos currículos de Química e, principalmente, na pós-graduação das áreas afins também.

Com base nas referenciais (EUROPEAN MEDICINE AGENCY (EMA); U. S. FOOD AND DRUG ADMINISTRATION (FDA), 2012; NATIONAL INSTITUTES OF HEALTH (EUA), 2021; OLIVARES; LOPES, 2012; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020; WILSON; BOTHAM, 2021) e uma vez atendidas as formas de operação, devem ser definidos critérios mínimos para aceitação de trabalhos para publicação:

- a. resultados dos parâmetros de validação adequados e de acordo com protocolos apropriados para aplicação em função do método estudado e, na ausência destes protocolos, as devidas justificativas para o emprego das

ferramentas estatísticas assumidas.

- b. cálculo da incerteza das medições envolvidas com avaliação dos resultados obtidos de acordo com critérios adequados ao uso pretendido do método
- c. caso os resultados não tenham atendido aos critérios de aceitação de resultado pré-estabelecidos ou a incerteza de medição associada ao resultado seja alta, ainda assim, o intervalo de abrangência deve estar especificado e podem ser apresentadas sugestões de melhorias;
- d. Apresentação dos dados brutos, recomendações ou exigências de materiais suplementares que evidenciem os dados do planejamento, condução e cálculos envolvidos no estudo;

## REFERÊNCIAS

- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR ISO/IEC 17025: requisitos gerais para a competência de laboratórios de ensaio e calibração. Rio de Janeiro: ABNT, 2017. Disponível em: [www.abnt.org.br](http://www.abnt.org.br). Acesso em: 30 ago. 2021.
- BAKER, M. Over half of psychology studies fail reproducibility test. *Nature News & Comment*, Berlin, 2015. DOI: 10.1038/nature.2015.18248. Disponível em: <https://www.nature.com/articles/nature.2015.18248>. Acesso em: 30 ago. 2021.
- BAKER, M. 1,500 scientists lift the lid on reproducibility. *Nature*, Berlin, v. 533, n. 7604, p. 452–454, 2016. DOI: 10.1038/533452a. Disponível em: <http://www.nature.com/articles/533452a>. Acesso em: 2 set. 2021.
- BARWICK, V. Planning and reporting method validation studies: supplement to Eurachem guide on the fitness for purpose of analytical methods. London: EURACHEM, 2019. Disponível em: <https://www.eurachem.org/index.php/publications/guides/planning-validation-studies>. Acesso em: 29 ago. 2021.
- BEGLEY, C. G.; BUCHAN, A. M.; DIRNAGL, U. Robust research: institutions must do their part for reproducibility. *Nature*, Berlin, v. 525, n. 7567, p. 25–27, 2015. DOI: 10.1038/525025a. Disponível em: <https://www.nature.com/articles/525025a>. Acesso em: 28 ago. 2021.
- BETTENCOURT DA SILVA, R.; WILLIAMS, A. Setting and using target uncertainty in chemical measurement. London: EURACHEM, 2015. p. 30. Disponível em: [www.eurachem.org](http://www.eurachem.org). Acesso em: 25 ago. 2021.
- BIPM. International Bureau of Weights and Measures. Joint Committee for Guides in Metrology. JCGM 106:2012 - Evaluation of measurement data: the role of measurement uncertainty in conformity assessment. [S. l.]: JCGM, 2012. DOI: 10.1515/ci.2013.35.2.22.
- BRASIL. Ministério da Agricultura Pecuária e Abastecimento. Manual da garantia da qualidade analítica - resíduo e contaminantes em alimentos, Brasília, DF: Ministério da Agricultura Pecuária e Abastecimento, 2011. p. 125. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/laboratorios/credenciamento-e-laboratorios-credenciados/legislacao-metodos-credenciados/residuos-e-contaminantes-em-alimentos>.
- BRASIL. Ministério da Saúde. Diretoria Colegiada da Agência Nacional de Vigilância Sanitária. ANVISA. Resolução - RDC No 27 , de 17 de maio de 2012. Diário Oficial da União: seção 1, Brasília, DF, 22 maio. 2012. p. 98. Disponível em: [bvsms.saude.gov.br/bvs/saudelegis/anvisa/2012/rdc0027\\_17\\_05\\_2012.html](http://bvsms.saude.gov.br/bvs/saudelegis/anvisa/2012/rdc0027_17_05_2012.html). Acesso em: 30 ago. 2021.
- BRASIL. Ministério da Saúde. Diretoria Colegiada da Agência Nacional de Vigilância Sanitária. Resolução - RDC no 166, de 24 de julho de 2017. Diário Oficial da União, Ed 141, Brasília, DF, 25 jul. 2017. p. 87. Disponível em: [http://portal.anvisa.gov.br/documents/10181/2721567/RDC\\_166\\_2017\\_COMP.pdf/d5fb92b3-6c6b-4130-8670-4e3263763401](http://portal.anvisa.gov.br/documents/10181/2721567/RDC_166_2017_COMP.pdf/d5fb92b3-6c6b-4130-8670-4e3263763401). Acesso em: 30 ago. 2021.
- CAMARGO, R. S.; OLIVARES, I. R. B. Quality management system for application of the analytical quality assurance cycle in a research project. *Journal of Physics: conference series*, Bristol, v. 733, p. 012046, 2016. DOI: 10.1088/1742-6596/733/1/012046. Disponível em: [https://www.researchgate.net/publication/305800770\\_Quality\\_management\\_system\\_for\\_application\\_of\\_the\\_analytical\\_quality\\_assurance\\_cycle\\_in\\_a\\_research\\_project](https://www.researchgate.net/publication/305800770_Quality_management_system_for_application_of_the_analytical_quality_assurance_cycle_in_a_research_project). Acesso em: 2 set. 2021.
- COLLINS, F. S.; TABAK, L. A. NIH plans to enhance reproducibility. *Nature*, Berlin, v. 505, n. 7485, p. 612–613, 2014. DOI: 10.1038/505612a. Disponível em: <https://www.nature.com/articles/505612a>. Acesso em: 2 set. 2021.

CONFLAB. Softwares Conflab: validação, incerteza e controle de qualidadevalidação, incerteza e controle de qualidade. [S. l.], 2020. Disponível em: <https://www.conflab.com.br/home>. Acesso em: 2 set. 2021.

CYRANOSKI, D. China introduces sweeping reforms against misconduct. *Nature*, Berlin, v. 558, p. 171, 2018. DOI: <https://doi.org/10.1038/d41586-018-05359-8>. Disponível em: <https://www.nature.com/articles/d41586-018-05359-8>. Acesso em: 30 ago. 2021.

DESHARNAIS, B.; CAMIRAND-LEMYRE, F.; MIREAULT, P.; SKINNER, C. D. Procedure for the selection and validation of a calibration model II-theoretical basis. *Journal of Analytical Toxicology*, Cary, v. 41, n. 4, p. 269–276, 2017. DOI: 10.1093/jat/bkx002. Disponível em: <https://academic.oup.com/jat/article/41/4/269/2965649>. Acesso em: 28 ago. 2021.

EURACHEM. CITAC Working Group. Guide CG 4 - Quantifying uncertainty in analytical measurement. Sussex, 2012. p. 132. Disponível em: [https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012\\_P1.pdf](https://www.eurachem.org/images/stories/Guides/pdf/QUAM2012_P1.pdf). Acesso em: 2 set. 2021.

EUROPEAN MEDICINE AGENCY. Science Medicines Health. Guideline on bioanalytical method validation. London, 2012. Seção 8, p. 865–868. DOI: 10.4155/bio.12.44. Disponível em: [https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation\\_en.pdf](https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-bioanalytical-method-validation_en.pdf). Acesso em: 2 set. 2021.

GOODMAN, S. N.; FANELLI, D.; IOANNIDIS, J. P. A. What does research reproducibility mean? *Science Translational Medicine*, Washington, v. 8, n. 341, p. 96–102, 2018. DOI: 10.1126/scitranslmed.aaf5027. Disponível em: <https://www.science.org/doi/10.1126/scitranslmed.aaf5027>. Acesso em: 3 set. 2021.

HARRIS, D. C. *Quantitative chemical analysis*. 8. ed. New York: W. H. Freeman and Company, 2010.

INTERNATIONAL COUNCIL FOR HARMONISATION OF TECHNICAL REQUIREMENTS FOR PHARMACEUTICALS FOR HUMAN USE. Quality guidelines. Q2(R1) Validation of analytical procedures: text and methodology. [S. l.], 2005. Disponível em: <https://www.ich.org/page/quality-guidelines>. Acesso em: 3 set. 2021.

INSTITUTO NACIONAL DE METROLOGIA (BRASIL). Grupo de trabalho luso-brasileiro. Vocabulário internacional de termos de metrologia legal - VIM, Rio de Janeiro: INMETRO, 2012a.

INSTITUTO NACIONAL DE METROLOGIA (BRASIL). Grupo de trabalho para tradução do GUM 2008. Avaliação de dados de medição - guia para a expressão de incerteza de medição, Duque de Caxias: INMETRO, 2012b. Disponível em: [https://www.gov.br/inmetro/pt-br/centrais-de-conteudo/publicacoes/documentos-tecnicos-em-metrologia/gum\\_final.pdf/view](https://www.gov.br/inmetro/pt-br/centrais-de-conteudo/publicacoes/documentos-tecnicos-em-metrologia/gum_final.pdf/view). Acesso em: 1 set. 2021.

INSTITUTO NACIONAL DE METROLOGIA (BRASIL). Coordenação Geral de Acreditação. DOQ-CGCRE-008 - orientação sobre validação de métodos analíticos. Rio de Janeiro: INMETRO, 2020. p. 1-30. Disponível em: [http://www.inmetro.gov.br/credenciamento/organismos/doc\\_organismos.asp?torganismo=calibensaio](http://www.inmetro.gov.br/credenciamento/organismos/doc_organismos.asp?torganismo=calibensaio). Acesso em: 2 set. 2021.

JURADO, J. M.; ALCÁZAR, A.; MUÑIZ-VALENCIA, R.; CEBALLOS-MAGAÑA, S. G.; RAPOSO, F. Some practical considerations for linearity assessment of calibration curves as function of concentration levels according to the fitness-for-purpose approach. *Talanta*, Amsterdam, v. 172, p. 221–229, 2017. DOI: 10.1016/j.talanta.2017.05.049. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0039914017305647>. Acesso em: 3 set. 2021.

KOMSTA, L. Chemometric and statistical evaluation of calibration curves in pharmaceutical analysis-A short review on trends and recommendations. *Journal of AOAC International*, Cary, v. 95, n. 3, p. 669–672, 2012. DOI: 10.5740/jaoacint.SGE\_Komsta. Disponível em: <https://academic.oup.com/jaoac/article/95/3/669/5655250>. Acesso em: 3 set. 2021.

LEEKA, J. T.; PENG, R. D. Opinion: Reproducible research can still be wrong: Adopting a prevention approach. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, DC, v. 112, n. 6, p. 1645–1646, 2015. DOI: 10.1073/pnas.1421412111. Disponível em: <https://www.pnas.org/content/112/6/1645>.

LEITE, F. Validação em análise química. 5. ed. Campinas: Átomo, 2008.

LUND-JOHANSEN, F.; TRAN, T.; MEHTA, A. Towards reproducibility in large-scale analysis of protein–protein interactions. *Nature News & Views*, Berlin, v. 18, n. 7, p. 720–721, 2021. DOI: 10.1038/s41592-021-01202-7. Disponível em: <http://dx.doi.org/10.1038/s41592-021-01202-7>. Acesso em: 3 set. 2021.

MILLER, J. N.; MILLER, J. C. *Statistics and chemometrics for analytical chemistry*. 6. ed. Gosport, UK: Pearson Education, 2010.

NASCIMENTO, R. S.; FROES, R. E. S.; SILVA, N. O. C.; NAVEIRA, R. L. P.; MENDES, D. B. C.; NETO, W. B.; SILVA, J. B. B. Comparison between ordinary least squares regression and weighted least squares regression in the calibration of metals present in human milk determined by ICP-OES. *Talanta*, Washington, DC, v. 80, n. 3, p. 1102–1109, 2010. DOI: 10.1016/j.talanta.2009.08.043. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0039914009006808?via%3Dihub>. Acesso em: 3 set. 2021.

OLIVARES, I. R. B. *Gestão de qualidade em laboratórios*. 4. ed. Campinas: Átomo e Alínea, 2019.

OLIVARES, I. R. B.; LOPES, F. A. Essential steps to providing reliable results using the Analytical Quality Assurance Cycle. *TrAC - Trends in Analytical Chemistry*, Kidlington, v. 35, p. 109–121, 2012. DOI: 10.1016/j.trac.2012.01.004. Disponível em: <http://dx.doi.org/10.1016/j.trac.2012.01.004>. Acesso em: 3 set. 2021.

PACCES, V. H. P. Como redigir um procedimento. In: SOUZA, G. B.; OLIVARES, I. R. B.; SOBRINHO, M. R.; PACCES, V. H. P.; BOZA, Y. (org.). *Aplicação prática dos conceitos de garantia da validade dos resultados com foco na ISO/IEC 17025:2017*. São Paulo: Rede Metrológica do Estado de São Paulo, 2020. p. 77–86. E-book. Disponível em: <https://ctqaremesp.wixsite.com/ctqa>.

PETER C. M.; RICHARD, E. Z. *Statistical methods in analytical chemistry*. 2. ed. New York: Wiley Intersciences Publications, 2000.

PRINZ, F.; SCHLANGE, T.; ASADULLAH, K. Believe it or not: how much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, Berlin, v. 10, n. 9, p. 712–713, 2011. DOI: 10.1038/nrd3439-c1. Disponível em: <https://www.nature.com/articles/nrd3439-c1>. Acesso em: 3 set. 2021.

RAPOSO, F. Evaluation of analytical calibration based on least-squares linear regression for instrumental techniques: a tutorial review. *TrAC - Trends in Analytical Chemistry*, Kidlington, v. 77, p. 167–185, 2016. DOI: 10.1016/j.trac.2015.12.006. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0165993615301242>. Acesso em: 3 set. 2021.

RAPOSO, F.; IBELLI-BIANCO, C. Performance parameters for analytical method validation: controversies and discrepancies among numerous guidelines. *TrAC Trends in Analytical Chemistry*, Kidlington, v. 129, n. 115913, p. 1–12, 2020. DOI: <https://doi.org/10.1016/j.trac.2020.115913>. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0165993620301424?via%3Dihub>. Acesso em: 3 set. 2021.

SANAGI, M.; NASIR, Z.; LING, S. L.; HERMAWAN, D.; WAN IBRAHIM, W. A.; NAIM, A. A. A practical approach for linearity assessment of calibration curves under the international union of pure and applied chemistry (IUPAC) guidelines for an in-house validation of method of analysis. *Journal of AOAC International*, Washington, DC, v. 93, n. 4, p. 1322–1330, 2010. DOI: 10.1093/jaoac/93.4.1322. Disponível em: <https://academic.oup.com/jaoac/article/93/4/1322/5655698>. Acesso em: 3 set. 2021.

- SANCHEZ, J. M. Estimating detection limits in chromatography from calibration data: ordinary least squares regression vs. weighted least squares. *Separations*, Basel, v. 5, n. 4, 2018. DOI: 10.3390/separations5040049. Disponível em: <https://www.mdpi.com/2297-8739/5/4/49>. Acesso em: 3 set. 2021.
- SAYAGO, A.; ASUERO, A. G. Fitting straight lines with replicated observations by linear regression: Part II. Testing for homogeneity of variances. *Critical Reviews in Analytical Chemistry*, Basel, v. 34, n. 3–4, p. 133–146, 2004. DOI: 10.1080/10408340490888599. Disponível em: <https://www.mdpi.com/2297-8739/5/4/49>. Acesso em: 3 set. 2021.
- SEPAROVIC, L.; SAVIANO, A. M.; LOURENÇO, F. R. Using measurement uncertainty to assess the fitness for purpose of an HPLC analytical method in the pharmaceutical industry. *Measurement: Journal of the International Measurement Confederation*, Kidlington, v. 119, p. 41–45, 2018. DOI: 10.1016/j.measurement.2018.01.048. Disponível em: <https://doi.org/10.1016/j.measurement.2018.01.048>. Acesso em: 3 set. 2021.
- SHAPIRO, S.; WILK, M. B. An Analysis of Variance Test for Normality ( Complete Samples ) Published. *Biometrika*, Oxford, v. 52, n. 3/4, p. 591–611, 1965. Disponível em: <http://www.jstor.org/stable/2333709>. Acesso em: 3 set. 2021.
- SKOOG, Douglas. *Fundamentos de química analítica*. 9. ed. São Paulo: Editora Thompson, 2014.
- SONAWANE, S. S.; CHHAJED, S. S.; ATTAR, S. S.; KSHIRSAGAR, S. J. An approach to select linear regression model in bioanalytical method validation. *Journal of Analytical Science and Technology*, Daejeon, v. 10, n. 1, 2019. DOI: 10.1186/s40543-018-0160-2. Disponível em: <https://jast-journal.springeropen.com/articles/10.1186/s40543-018-0160-2>. Acesso em: 5 set. 2021.
- STEPHENS, M. False discovery rates: a new deal. *Biostatistics*, Oxford, v. 18, n. 2, p. 275–294, 2016. DOI: 10.1093/biostatistics/kxw041. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5379932/>. Acesso em: 5 set. 2021.
- TENNY, S.; ABDELGAWAD, I. *Statistical significance*. Treasure Island: StatPearls Publishing; 2021. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK459346/>. Acesso em: 5 set. 2021.
- THOMPSON, M.; ELLISON, S. L. R.; WOOD, R. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). *Pure and Applied Chemistry*, Berlim, v. 74, n. 5, p. 835–855, 2002. DOI: 10.1351/pac200274050835. Disponível em: <https://www.degruyter.com/document/doi/10.1351/pac200274050835/html>. Acesso em: 5 set. 2021.
- TRAPLE, M. A. L.; SAVIANO, A. M.; FRANCISCO, F. L.; LOURENÇO, F. R. Measurement uncertainty in pharmaceutical analysis and its application. *Journal of Pharmaceutical Analysis*, Amsterdam, v. 4, n. 1, p. 1–5, 2014. DOI: 10.1016/j.jpha.2013.11.001. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2095177913001317>. Acesso em: 8 set. 2021.
- UNITED STATES. Department of Health & Human Services. National Institutes of Health. ACD Working Group on enhancing rigor, transparency, and translatability in animal research. Bethesda, 2020. Disponível em: <https://acd.od.nih.gov/working-groups/epar.html>. Acesso em: 30 ago. 2021.
- WASSENAN, L. *All of statistics: a concise course in statistical inference*. New York: Springer Science, 2004. DOI: 10.1007/978-0-387-21736-9. (Springer texts in statistics).
- WILSON, J. L.; BOTHAM, C. M. Three questions to address rigour in your proposal. *Nature Career Column*, Berlim, v. 596, p. 609–610, 2021. DOI: <https://doi.org/10.1038/d41586-021-02286-z>. Disponível em: <https://www.nature.com/articles/d41586-021-02286-z>. Acesso em: 6 set. 2021.
- YEH, R. W.; VALSDOTTIR, L. R.; YEH, M. W.; SHEN, C.; KRAMER, D. B.; STROM, J. B.; SECESKY, E. A.; HEALY, J. L.; DOMEIER, R. M.; KAZI, D. S.; NALLAMOTHU, B. K. Parachute use to prevent death and major trauma when jumping from aircraft: randomized controlled trial. *BMJ*, London, v. 363, p. k5094, 2018. DOI: 10.1136/bmj.k5094. Disponível em: <https://www.bmj.com/content/363/bmj.k5094>. Acesso em: 6 set. 2021.

ZHAO, Y.; SAMPSON, M. G.; WEN, X. Quantify and control reproducibility in high-throughput experiments. *Nature Methods*, Berlim, v. 17, n. 12, p. 1207–1213, 2020. DOI: 10.1038/s41592-020-00978-4. Disponível em: <http://dx.doi.org/10.1038/s41592-020-00978-4>. Acesso em: 6 set. 2021.

b

## GLOSSÁRIO

**Abordagem de Erro:** O objetivo da medição na Abordagem de Erro é determinar uma estimativa do valor verdadeiro que esteja tão próxima quanto possível deste valor verdadeiro único. O desvio do valor verdadeiro é composto de erros aleatórios e sistemáticos. Os dois tipos de erros, supostos como sendo sempre distinguíveis, têm que ser tratados diferentemente. Nenhuma regra pode ser estabelecida quanto à combinação deles para se chegar ao erro total caracterizando um determinado resultado de medição, tido geralmente como a estimativa (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Abordagem de Incerteza:** A Abordagem de Incerteza consiste no reconhecimento de que, devido à quantidade intrinsecamente incompleta de detalhes na definição de uma grandeza, não existe um valor verdadeiro único, mas sim um conjunto de valores verdadeiros consistentes com a definição. Entretanto, este conjunto de valores é, em princípio e na prática, impossível de ser conhecido. Outras abordagens evitam completamente o conceito de valor verdadeiro de uma grandeza e avaliam a validade dos resultados de medição com auxílio do conceito de compatibilidade metrológica (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Balanço de incerteza:** Formulação e apresentação de uma incerteza de medição e de suas componentes, assim como de seu cálculo e combinação. Convém que, em um balanço de incerteza, sejam incluídos o modelo de medição, as estimativas e incertezas de medição associadas às grandezas consideradas no modelo de medição, as covariâncias, os tipos de funções de densidade de probabilidade utilizadas, os graus de liberdade, os tipos de avaliação da incerteza de medição e qualquer fator de abrangência (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Erro de medição:** Diferença entre o valor medido de uma grandeza e um valor de referência (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 21). O conceito de “erro de medição” pode ser utilizado:

- a) quando existe um único valor de referência, o que ocorre se uma calibração for realizada por meio de um padrão de medição com um valor medido cuja incerteza de medição é desprezável, ou se um valor convencional for fornecido; nestes casos, o erro de medição é conhecido (INSTITUTO NACIONAL DE METROLOGIA

(BRASIL), 2012a, p. 21);

b) caso se suponha que um mensurando é representado por um único valor verdadeiro ou um conjunto de valores verdadeiros de amplitude desprezável; neste caso, o erro de medição é desconhecido (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 21).

Não se deve confundir erro de medição com erro de produção ou erro humano (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 21).

**Fator de abrangência (Brasil); fator de expansão (Portugal):** número maior do que um pelo qual uma incerteza-padrão combinada é multiplicada para se obter uma incerteza de medição expandida (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Grandeza:** propriedade de um fenômeno, corpo ou substância, que pode ser expressa quantitativamente sob a forma de um número e uma referência (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Incerteza de medição:** parâmetro não negativo que caracteriza a dispersão dos valores atribuídos a um mensurando, com base nas informações utilizadas. A incerteza de medição inclui componentes provenientes de efeitos sistemáticos, tais como componentes associadas a correções e a valores atribuídos a padrões, assim como a incerteza definicional. Algumas vezes, não são corrigidos efeitos sistemáticos estimados, mas, em vez disso, são incorporadas componentes de incerteza de medição associadas. A incerteza de medição geralmente engloba muitas componentes. Algumas delas podem ser estimadas por uma avaliação do Tipo A da incerteza de medição, a partir da distribuição estatística dos valores provenientes de séries de medições e podem ser caracterizadas por desvios-padrão. As outras componentes, as quais podem ser estimadas por uma avaliação do Tipo B da incerteza de medição, podem também ser caracterizadas por desvios-padrão estimados a partir de funções de densidade de probabilidade baseadas na experiência ou em outras informações. Geralmente para um dado conjunto de informações, subentende-se que a incerteza de medição está associada a um determinado valor atribuído ao mensurando.

**Incerteza de medição expandida; incerteza expandida:** Produto de uma incerteza-padrão combinada por um fator maior do que o número um. O fator depende do tipo de distribuição de probabilidade da grandeza de saída e da probabilidade de abrangência escolhida. O termo “fator” nesta definição se refere ao fator de abrangência. A incerteza de medição expandida é chamada de “incerteza global” no parágrafo 5 da Recomendação INC-1 (1980) (ver o GUM) e simplesmente “incerteza” nos documentos IEC.

**Incerteza definicional:** Componente da incerteza de medição que resulta da quantidade finita de detalhes na definição de um mensurando. Incerteza mínima que se pode obter, na prática, em qualquer medição de um dado mensurando. Qualquer modificação nos detalhes descritivos conduz a uma outra incerteza definicional (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Incerteza-alvo; incerteza de medição pretendida:** Incerteza de medição especificada como um limite superior e escolhida de acordo com o uso pretendido dos resultados de medição (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Incerteza-padrão combinada:** Incerteza-padrão obtida ao se utilizarem incertezas-padrão individuais associadas às grandezas de entrada em um modelo de medição. Em caso de correlações entre grandezas de entrada em um modelo de medição, as covariâncias também devem ser levadas em consideração no cálculo da incerteza-padrão combinada.

**Incerteza-padrão:** Incerteza de medição expressa na forma de um desvio-padrão.

**Intervalo de abrangência:** Intervalo, baseado na informação disponível, que contém o conjunto de valores verdadeiros de um mensurando, com uma probabilidade determinada. Um intervalo de abrangência não está necessariamente centrado no valor. Não é recomendável que um intervalo de abrangência seja denominado "intervalo de confiança" para evitar confusão com o conceito estatístico. Um intervalo de abrangência pode ser deduzido de uma incerteza de medição expandida (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Mensurando:** Grandeza que se pretende medir. A especificação de um mensurando requer o conhecimento da natureza da grandeza e a descrição do estado do fenômeno, do corpo ou da substância da qual a grandeza é uma propriedade, incluindo qualquer constituinte relevante e as entidades químicas envolvidas (VIM, 2012).

**Método de medição:** Descrição genérica de uma organização lógica de operações utilizadas na realização de uma medição.

**Metrologia:** Ciência da medição e suas aplicações. Engloba todos os aspectos teóricos e práticos da medição, qualquer que seja a incerteza de medição e o campo de aplicação (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Probabilidade de abrangência; probabilidade de expansão:** probabilidade de que o conjunto de valores verdadeiros de um mensurando esteja contido em um intervalo de abrangência especificado. Esta definição se refere à Abordagem de Incerteza como apresentado no GUM. Não é recomendável que este conceito seja confundido com o conceito estatístico de nível de confiança, embora o termo “*confidence level*” seja utilizado em inglês no GUM.

**Procedimento de medição:** Descrição detalhada de uma medição de acordo com um ou mais princípios de medição e com um dado método de medição, baseada em um modelo de medição e incluindo todo cálculo destinado à obtenção de um resultado de medição. É geralmente documentado em detalhes suficientes para permitir que um operador realize uma medição. Pode incluir uma declaração referente à incerteza alvo. Algumas vezes é chamado de procedimento operacional padrão (POP) (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a; PACCES; COMITÊ TÉCNICO DE QUÍMICA DE ALIMENTOS, 2020).

**Protocolo:** Documentos de referências que definem as condições e critérios de parâmetros de validação e de aceitação de resultados normalmente denominados de resoluções, diretrizes, protocolos, guias, manuais etc. A Tabela 1 traz exemplos desses tipos de documentos que foram utilizados.

**Rastreabilidade metrológica:** propriedade de um resultado de medição pela qual tal resultado pode ser relacionado a uma referência por intermédio de uma cadeia ininterrupta e documentada de calibrações, cada uma contribuindo para a incerteza de medição (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a).

**Reprodutibilidade de métodos:** capacidade de implementar, da forma mais exata possível, os procedimentos experimentais e computacionais, com os mesmos dados e ferramentas, para obter os mesmos resultados (GOODMAN; FANELLI; IOANNIDIS, 2016, p. 2)

**Reprodutibilidade de resultados:** refere-se à obtenção dos mesmos resultados da aplicação de um estudo independente cujos procedimentos sejam o mais próximos possível do experimento original. (GOODMAN; FANELLI; IOANNIDIS, 2018)

**Reprodutibilidade na pesquisa científica:** capacidade de um indivíduo em duplicar os resultados de um estudo anterior usando os mesmos materiais que foram usados pelo investigador original. Condição mínima necessária para que uma descoberta seja, realizável e aceitavelmente informativa (GOODMAN; FANELLI; IOANNIDIS, 2018).

**Resultado de medição:** conjunto de valores atribuídos a um mensurando, juntamente com toda outra informação pertinente disponível. Um resultado de medição geralmente contém “informação pertinente” sobre o conjunto de valores, alguns dos quais podem ser mais representativos do mensurando do que outros. Isto pode ser expresso na forma de uma função de densidade de probabilidade (FDP). Um resultado de medição é geralmente expresso por um único valor medido e uma incerteza de medição. Caso a incerteza de medição seja considerada desprezável para alguma finalidade, o resultado de medição pode ser expresso como um único valor medido. Em muitos domínios, esta é a maneira mais comum de expressar um resultado de medição. Na literatura tradicional e na edição anterior do VIM, o resultado de medição era definido como um valor atribuído a um mensurando obtido por medição, que podia se referir a uma indicação, ou um resultado não corrigido, ou um resultado corrigido, de acordo com o contexto (VIM, 2012).

**Repetibilidade:** Condição de medição em um conjunto de condições, as quais incluem o mesmo procedimento de medição, os mesmos operadores, o mesmo sistema de medição, as mesmas condições de operação e o mesmo local, assim como

medições repetidas no mesmo objeto ou em objetos similares durante um curto período. Uma condição de medição é uma condição de repetibilidade apenas com respeito a um conjunto especificado de condições de repetibilidade. Em química, o termo “condição de precisão intra-série” é algumas vezes utilizado para designar este conceito (INSTITUTO NACIONAL DE METROLOGIA (BRASIL), 2012a, p. 22).

**Valor verdadeiro de uma grandeza:** valor de uma grandeza compatível com a definição da grandeza. Na Abordagem de erro para descrever as medições, o valor verdadeiro de uma grandeza é considerado único e, na prática, impossível de ser conhecido (VIM, 2012).

**Validação de métodos:**

- ✓ *ABNT NBR ISO/IEC 17025:2017:* “Validação é a confirmação por exame e fornecimento de evidência objetiva de que os requisitos específicos para um determinado uso pretendido são atendidos”;
- ✓ *ANVISA RDC nº 27/2012:* “Validação é a confirmação por ensaio e fornecimento de evidência objetiva de que os requisitos específicos para um determinado uso pretendido são atendidos”;
- ✓ *VIM 2012 Vocabulário Internacional de Metrologia:* “Verificação na qual os requisitos especificados são adequados para um uso pretendido.”
- ✓ *ANVISA RDC nº 166/2017:* “Validação analítica é a avaliação sistemática de um método por meio de ensaios experimentais de modo a confirmar e fornecer evidências objetivas de que os requisitos específicos para seu uso pretendido são atendidos”;
- ✓ *EMA 2012 Guideline on bioanalytical method validation:* “O principal objetivo da validação de método é demonstrar a confiabilidade de um método específico para a determinação de uma concentração de analito em uma matriz”;
- ✓ *ICH Q2(R1) Validation of analytical procedures:* “O objetivo da validação de um procedimento analítico é demonstrar que ele é adequado para a finalidade pretendida. Um somatório tabular das características aplicáveis à identificação, controle de impurezas e procedimentos de ensaio está incluído”.

## APÊNDICE A - DISTRIBUIÇÃO DOS EQUÍVOCOS DE CÁLCULOS MAIS COMUNS

Figura 49 - Distribuição dos equívocos de cálculos mais comuns



Fonte: Própria do autor