Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

Utilização de algoritmos de mineração de dados para estimativa de intervalos de referência de biomarcadores laboratoriais – relato de experiência

Use of data mining algorithms to estimate reference intervals for laboratory biomarkers – experience report

Gustavo Oliveira Gonçalves[1,2], Alan Carvalho Dias[3], Leonardo de Souza Vasconcellos[4,5]

1Doutor pelo Programa de Pós-graduação em Ciências Aplicadas à Saúde do Adulto da Faculdade de Medicina da Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Brasil.
2Docente dos cursos de Biomedicina, Enfermagem, Farmácia, Medicina, Nutrição e Odontologia da Faculdade de Minas – FAMINAS BH. Belo Horizonte, Minas Gerais, Brasil.
3Sabin Medicina Diagnóstica. Brasília, Distrito Federal, Brasil.
4Docente do Departamento de Propedêutica Complementar, Faculdade de Medicina da Universidade Federal de Minas Gerais Belo Horizonte, Minas Gerais, Brasil.
5Orientador do Programa de Pós-graduação em Ciências Aplicadas à Saúde do Adulto da Faculdade de Medicina da Universidade Federal de Minas Gerais Belo Horizonte, Minas Gerais, Brasil.

Resumo

O laboratório clínico é protagonista na medicina personalizada e na segurança assistencial. A partir das análises laboratoriais é possível tomar decisões clínicas importantes. A utilização de intervalos de referência (IR) adequados à população atendida passa a ser uma prioridade da medicina laboratorial para atendimento às normas regulamentadoras e à crescente demanda por maior precisão na interpretação dos resultados das análises laboratoriais. O laboratório precisa documentar os critérios, metodologia e as etapas utilizadas para a determinação dos IRs. O objetivo geral deste trabalho é descrever a experiência na utilização dos principais algoritmos de mineração de dados derivados do sistema de informação laboratorial para estimativa de intervalos de referência dos biomarcadores laboratoriais por meio da abordagem do método indireto proposto nas ferramentas computacionais que aplicam o método Bhattacharya, Kosmic, refineR e LabRI. Trata-se de um estudo descritivo, do tipo relato da experiência, que descreve o trabalho desenvolvido pelo Grupo de Pesquisa em Patologia Clínica/Medicina Laboratorial da Universidade Federal de Minas Gerais (GPPCML/CNPq). As abordagens mais recentes aplicam algoritmos de mineração de dados utilizando ferramentas computacionais (softwares) desenvolvidos em linguagem de programação. Considerando as vantagens, desvantagens, facilidade de utilização das ferramentas computacionais, aplicação de critérios rigorosos para seleção e algoritmos cada vez mais robustos é possível considerar a utilização das diferentes ferramentas computacionais que aplicam a abordagem indireta propostos pelos métodos Bhattacharya, Kosmic, refineR e LabRI.

Palavras-chave: Valores de Referência; Biomarcadores; Testes Laboratoriais; Testes de Química Clínica; Segurança do Paciente.

Abstract

The clinical laboratory is a protagonist in personalized medicine and healthcare safety. From laboratory analyzes it is possible to make important clinical decisions. The use of reference intervals (RI) appropriate to the population served becomes a priority in laboratory medicine to comply with regulatory standards and the growing demand for greater precision in the interpretation of laboratory analysis results. The laboratory needs to document the criteria, methodology and steps used to determine the IRs. The general objective of this work is to describe the experience in using the main data mining algorithms derived from the laboratory information system to estimate reference intervals of laboratory biomarkers through the indirect method approach proposed in computational tools that apply the Bhattacharya method, Kosmic, refineR and LabRI. This is a descriptive study, of the experience report type, which describes the work developed by the Clinical Pathology/Laboratory Medicine Research Group at the Federal University of Minas Gerais (GPPCML/CNPq). The most recent approaches apply data mining algorithms using computational tools (software) developed in a programming language. Considering the advantages, disadvantages, ease of use of computational tools, application of rigorous criteria for selection and increasingly robust algorithms, it is possible to consider the use of different computational tools that apply the indirect approach proposed by the Bhattacharya, Kosmic, refineR and LabRI methods.

Keywords: Reference values; Biomarkers; Laboratory Tests; Clinical Chemistry Tests; Patient safety.

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

1 INTRODUÇÃO

Os profissionais que atuam na assistência ao paciente, dentro do ecossistema de saúde, público ou privado, dependem cada vez mais do apoio diagnóstico e o grande desafio é fazer com que o laboratório clínico seja protagonista na medicina personalizada e na segurança assistencial. A partir das análises laboratoriais é possível tomar decisões clínicas importantes que terão como resultado a confirmação de um diagnóstico, a exclusão de uma hipótese, o monitoramento terapêutico e o prognóstico de uma patologia (HALLWORTH, 2011; NAYUPE; MBULAJE; MUNHARO; PATEL *et al.*, 2023; PLEBANI, 2004; ROHR; BINDER; DIETERLE; GIUSTI *et al.*, 2016).

O laudo laboratorial precisa estar em consonância com os aspectos clínico-epidemiológicos do paciente. A utilização de intervalos de referência (IR) adequados à população atendida passa a ser uma prioridade da medicina laboratorial para atendimento às normas regulamentadoras e à crescente demanda por maior precisão na interpretação dos resultados das análises laboratoriais(OZARDA; HIGGINS; ADELI, 2018; VASARHELYI; DEBRECZENI, 2017).

Intervalo de referência pode ser definido como a faixa de normalidade esperada para um determinado biomarcador em uma dada população (BURTIS; BURNS, 2016). A estimativa de IR específicos é uma tarefa difícil, trabalhosa e onerosa para os laboratórios clínicos, por isto, na maioria das vezes são utilizados intervalos que foram definidos pelo fabricante do kit reagente, baseado em uma população diferente daquela que é atendida na instituição (OZARDA, 2016).

Segundo Melillo (1993), a comparação de resultados laboratoriais da população pediátrica ou da pessoa idosa utilizando intervalos estimados para adultos pode levar a interpretação inadequada e, muitas vezes, tratamentos desnecessários devido ao diagnóstico incorreto. O apoio à interpretação de testes laboratoriais com base nos IR devidamente validados para a população de acordo com a faixa etária é uma das principais preocupações dos laboratórios médicos na atualidade (ABEBE; MELKU; ENAWGAW; BIRHAN *et al.*, 2018; HUBER; MOSTAFAIE; STANGL; WOROFKA *et al.*, 2006; MELILLO, 1993; NILSSON; EVRIN; TRYDING; BERG *et al.*, 2003).

O laboratório precisa documentar os critérios, metodologia e as etapas utilizadas para a determinação dos IRs. Haeckel *et al.* (2023), Coskun *et al.* (2022), Yang; Su; Zhao (2022) e muitos outros autores descreveram o percurso metodológico com detalhamento das atividades que podem ser realizadas no laboratório clínico para a estimativa dos IRs (COSKUN; SANDBERG; UNSAL; SERTESER *et al.*, 2022; HAECKEL; ADELI; JONES; SIKARIS *et al.*, 2023; YANG; SU; ZHAO, 2022). O *Clinical and Laboratory Standards Institute* (CLSI) aprovou a Diretriz EP28-A3c, que descreve as etapas para definir, estabelecer e verificar os IRs no laboratório clínico pelo método direto ou método indireto (CLSI, 2010; OZARDA; ICHIHARA; JONES; STREICHERT *et al.*, 2021).

O método direto é aquele em que há uma seleção de população referência aplicando-se critérios de inclusão e exclusão previamente definidos para eleger indivíduos saudáveis. Este método já foi publicado em diferentes trabalhos de autores como Ozarda (2016); Henny *et al.* (2016) e Adeli *et al.* (2017) que descreveram em seus estudos a dificuldade em aplicar todos os critérios definidos na diretriz, especialmente no caso de estudos envolvendo pacientes pediátricos e pessoa idosa (ADELI; HIGGINS; TRAJCEVSKI; WHITE-AL HABEEB, 2017; CLSI, 2010; HENNY; VASSAULT; BOURSIER; VUKASOVIC *et al.*, 2016; OZARDA, 2016).

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

Jones *et al.* (2018) e Ozarda *et al.* (2021) descreveram em seus trabalhos que a aplicação prática do método indireto consiste em utilizar os resultados de dosagens de biomarcadores laboratoriais de pacientes que foram atendidos na instituição e estão disponíveis no Sistema de Informação Laboratorial (SIL). O emprego do método indireto possui como vantagens: (a) maior agilidade na geração do IR; (b) redução do custo e (c) não envolvem inconveniência, desconforto ou riscos associados à coleta de novas amostras biológicas e informações de saúde do paciente (JONES; HAECKEL; LOH; SIKARIS *et al.*, 2018; MA; ZOU; HOU; YIN *et al.*, 2022; OZARDA; ICHIHARA; JONES; STREICHERT *et al.*, 2021; VELEV; LEBIEN; ROCHE-LIMA, 2023; ZHONG; MA; HOU; YIN *et al.*, 2023).

A legislação brasileira vigente não traz a obrigatoriedade em estabelecer intervalos de referência, porém os laboratórios clínicos que participam de algum programa de acreditação brasileiro precisam estar atentos aos requisitos das normas que direcionam para a importância de interpretar adequadamente os laudos laboratoriais. O Manual do Sistema Nacional de Acreditação, vinculado à Sociedade Brasileira de Análises Clínicas define conceitualmente que os intervalos de referências devem ser obtidos a partir de uma população de referência biológica, considerando o intervalo central de 95%. Já a norma do Programa de Acreditação do Laboratório Clínico, vinculado à Sociedade Brasileira de Patologia Clínica e Medicina Laboratorial determina que os laboratórios deverão adequar os intervalos de referência para correta interpretação do laudo laboratorial. O estabelecimento de IR devem ser baseados em "população de tamanho e características adequados, seguindo procedimentos estatísticos válidos (incluindo porções por sexo, idade e outras, quando apropriadas)" (BRASIL, 2023; SBPC/ML, 2021; SNA-DICQ, 2023).

O objetivo geral deste trabalho é descrever a experiência na utilização dos principais algoritmos de mineração de dados derivados do sistema de informação laboratorial para estimativa de intervalos de referência dos biomarcadores laboratoriais por meio da abordagem indireta proposta nas ferramentas computacionais que aplicam os métodos Bhattacharya, Kosmic, refineR e LabRI.

2 MATERIAL E MÉTODOS

Trata-se de um estudo descritivo, do tipo relato da experiência, que descreve as principais ferramentas computacionais que aplicam os métodos Bhattacharya, Kosmic, refineR e LabRI utilizadas no trabalho desenvolvido pelo Grupo de Pesquisa em Patologia Clínica/Medicina Laboratorial da Universidade Federal de Minas Gerais (GPPCML/CNPq). Este artigo pretende apresentar características dos algoritmos utilizados na construção do projeto que culminou na tese intitulada "DETERMINAÇÃO DE INTERVALOS DE REFERÊNCIA DE EXAMES LABORATORIAIS BIOQUÍMICOS NA PESSOA IDOSA POR MÉTODOS DIRETO E INDIRETO VIA ALGORITMOS DE MINERAÇÃO DE DADOS" foi apresentada ao Programa de Pós-Graduação em Ciências Aplicadas à Saúde do Adulto da Faculdade de Medicina da Universidade Federal de Minas Gerais. Belo Horizonte, Minas Gerais, Brasil.

3 RELATO DE EXPERIÊNCIA

Haeckel *et al.* (2023), Coskun *et al.* (2022), Yang; Su; Zhao (2022) e muitos outros autores descreveram os diferentes métodos que podem ser empregados no laboratório clínico para determinação dos intervalos de referência (COSKUN; SANDBERG; UNSAL; SERTESER *et al.*, 2022; HAECKEL; ADELI; JONES; SIKARIS *et al.*, 2023;

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

YANG; SU; ZHAO, 2022). As abordagens mais recentes aplicam algoritmos de mineração de dados utilizando ferramentas computacionais (*softwares*) desenvolvidos em linguagem de programação: R e o RStudio (*R Foundation for Statistical Computing, Vienna, Austria*), JAVA (*Oracle Corporation, Califórnia, EUA*), Python (*Python Software Foundation, Delaware, EUA*) e C (*Bell Laboratories, Wisconsin, EUA*). Entende-se por linguagem de programação um método padronizado, formado por um conjunto de regras e instruções precisas para implementação de um código fonte que levam à solução de um problema por meio da elaboração de modelos estatísticos aplicados em um banco de dados. Este é o fundamento da abordagem por métodos indiretos para determinação de IR (LIU; WILSON; BEDNY, 2024; STAPLES, 2023; WINKEL, 1990; ZHONG; MA; HOU; YIN *et al.*, 2023).

Sabe-se que a definição de IR apropriados à população atendida pelo laboratório é um desafio, mas aumenta a qualidade assistencial e, consequentemente, auxilia os profissionais da saúde no diagnóstico e monitoramento de doenças. Ceriotti; Hinzmann; Panteghini (2009) discutiram os aspectos teóricos e principais desafios para responder "Qual caminho deve ser seguido para determinação dos intervalos de referência?", reforçando a possibilidade de estabelecimento de IR por meio de método direto ou indireto (CERIOTTI; HINZMANN; PANTEGHINI, 2009).

O método direto com seleção *a priori* ou *a posteriori* de membros da população de referência, tem sido amplamente utilizada e padronizada especialmente em função da recomendação da diretriz internacional emitida pelo *Clinical and Laboratory Standards Institute (CLSI)*, porém é trabalhosa e cara, conforme descrito por Martínez-Sánchez *et al.* (2021) (MARTÍNEZ-SÁNCHEZ; MARQUES-GARCÍA; OZARDA; BLANCO *et al.*, 2021; OZARDA, 2016; OZARDA; ICHIHARA; JONES; STREICHERT *et al.*, 2021). A **Figura 1** apresenta um quadro com a síntese com as principais características de cada método utilizado na determinação do intervalo de referência.

Figura 1 – Principais características do método direto e método indireto para determinação de intervalos de referência

	MÉTODO DIRETO	**MÉTODO INDIRETO**
VANTAGENS	Por se tratar de uma avaliação clínico-laboratorial é possível investigar o estado de saúde dos indivíduos e informações relevantes, como histórico familiar de doenças, uso de medicamentos, tabagismo, etilismo e outros interferentes analíticos.	Economia de tempo, despesas, trabalho e redução da necessidade de pessoal técnico envolvido no recrutamento de indivíduos de referência saudáveis, levantamento epidemiológico, coleta de amostras e outras etapas do processo.
	O grupo dos indivíduos de referência saudáveis é bem caracterizado, controlado e representativo.	Método adequado para pediatria, geriatria e outros grupos especiais, como líquido cefalorraquidiano. Pode ser repetida a qualquer momento, sem a necessidade de recrutar novos indivíduos.
	A definição dos valores de referência e o protocolo são padronizados, incluindo todos os indivíduos de referência na análise e cálculo.	Tem um alto potencial de estratificação, bem como a possibilidade de derivar limites de referência contínuos (dinâmicos), de modo que a verificação da transferência não seja necessária por padrão.
	Os métodos estatísticos simples podem ser realizados para calcular os intervalos de referência direta (ou seja, método não paramétrico).	Eliminação da possibilidade de iatrogenia e desconfortos ao paciente gerados pela coleta das amostras biológicas.
	A qualidade dos dados é controlada durante todo o processo, garantindo precisão dos resultados.	Processamento dos dados garantindo a confidencialidade, pois a identidade do paciente é tratada de forma sigilosa, o que evita questões éticas.
DE	Definir claramente quais são os critérios de inclusão e exclusão especialmente pela subjetividade e controvérsias do conceito de saúde.	O possível efeito de subpopulações doentes nos intervalos de referência derivados.
	Obter número suficiente de indivíduos que concordem com a coleta da amostra biológica ou	Ausência de padronização validada para verificar se os intervalos de referência obtidos são corretos e válidos.

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

SVANTAGENS	recrutamento de indivíduos em idades extremas (pediátricas e geriátricas).	
	Pode ocorrer viés de seleção, devido à complexidade da seleção e ao tamanho relativamente pequeno da amostra.	Alterações ou inconsistências no processo laboratorial que podem levar a possíveis erros.
	Não é viável determinar intervalos de referência dependentes de idade/sexo para testes que altera com a idade e difere entre homens e mulheres.	O método de exclusão de outliers pode afetar o resultado do IR, pois esses valores extremos têm potencial de serem afetados por doença ou pelo estado subclínico.
	O processo de obtenção de indivíduos de referência na fase inicial é pesado e complexo. Composto por muitas etapas, consumindo recursos (tempo, despesas e trabalho).	Vários métodos estatísticos foram propostos, mas ainda não há consenso ou recomendações oficiais sobre "qual método usar quando".

Fonte: Martínez-Sánchez *et al.* (2020); Ozarda *et al.* (2021); Yang; Su; Zhao (2022).

Os primeiros trabalhos publicados abordando a aplicação de ferramentas computacionais para estimar IR por métodos indiretos são da década de 1960. Desde então, houve grande avanço na inovação tecnológica e da aplicação prática da ciência de dados. Apesar de não ser alvo desta pesquisa, o método Hoffman foi desenvolvido em 1963 e, de acordo com a literatura científica consultada trata-se da primeira ferramenta computacional com algoritmo gráfico. Os dados de distribuições mistas são apresentados em diferentes regiões lineares para identificar os "indivíduos doentes", utilizando assim conjuntos de dados não doentes para estabelecer intervalos de referência (MA; YU; QIU, 2023). Para aplicar o método Hoffmann, os dados devem ser normalmente distribuídos (YANG; LANG; WANG; FANG *et al.*, 2023).

Yang; Su; Zhao (2022) concluíram em seu estudo que o método indireto é "simples, fácil e promissor para a pesquisa de intervalos de referência" e que os riscos são compensados pelos benefícios envolvidos com a agilidade e economia do uso dos bancos de dados para o estabelecimento de IR. O processo para estimar IR pelo método indireto em cinco etapas, a saber: (1) seleção das amostras no Sistema de Informação Laboratorial; (2) limpeza e pré-processamento dos dados; (3) transformação de dados; (4) tratamento de *outliers* e (5) o estabelecimento do IR (YANG; SU; ZHAO, 2022).

A seguir estão descritas as principais características de cada um dos algoritmos de mineração de dados derivados do sistema de informação laboratorial utilizados para estimativa de intervalos de referência dos biomarcadores laboratoriais por meio da abordagem do método indireto proposto nas ferramentas computacionais que aplicam os métodos Bhattacharya, Kosmic, refineR e LabRI.

3.1 MÉTODO DE BHATTACHARYA

O método Bhattacharya foi desenvolvido em 1965. De acordo com Ozarda *et al.* (2021) trata-se de "um método gráfico é usado para identificar esta distribuição central e assume que a distribuição dos dados de origem do SIL consiste em pelo menos uma distribuição gaussiana, com a predominante representando indivíduos saudáveis" (MA; ZOU; HOU; YIN *et al.*, 2022; OZARDA; ICHIHARA; JONES; STREICHERT *et al.*, 2021).

Utiliza o *software* Bellview versão 2.0.1, exigindo um ambiente Java (1.8 ou posterior). Os códigos R e pacotes para implementar a análise Bhattacharya estão disponíveis em www.Lab-R-torian.com. Os processos de transformação logarítmica do conjunto de dados são realizados manualmente pelos autores, gerando muitas vezes, insegurança no momento da interpretação dos resultados apresentados (HOLMES; BUHR, 2019; OZARDA; ICHIHARA; JONES; STREICHERT *et al.*, 2021).

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

3.2 Método Kosmic

O método Kosmic é a versão atualizada do *Truncated Maximum Likelihood (TML)* e foi desenvolvido em 2020 por Zierk *et al.* (2020). Este algoritmo usa modelagem de distribuição normal de potência. Primeiro aplica a transformação Box-Cox aos dados e depois ajusta uma distribuição gaussiana aos dados truncados. A distância Kolmogorov-Smirnov entre a distribuição truncada observada e a distribuição gaussiana foi calculada selecionando a menor distância como o IR de indivíduos saudáveis (CHEN; FAN; YANG; YANG, 2024; ZIERK; ARZIDEH; KAPSNER; PROKOSCH *et al.*, 2020).

Agaravatt *et al.* (2023) descreveram que a abordagem pelo algoritmo Kosmic utiliza uma combinação de resultados de testes patológicos e fisiológicos para estimar a distribuição dos resultados dos testes fisiológicos. Isto é conseguido por uma transformação Box-Cox e, em seguida, por uma distribuição gaussiana ajustada a uma parte dos dados que foi truncada (AGARAVATT; KANSARA; KHUBCHANDANI; SANGHANI *et al.*, 2023).

O algoritmo Kosmic está disponível como *software* de código aberto linguagem de programação Python em https://gitlab.miracum.org/KOSMIC. Uma ferramenta baseada na *web* acessível em https://KOSMIC.diz.uk-erlangen.de permite o uso do aplicativo Kosmic sem a necessidade de instalação local (AMMER; SCHUTZENMEISTER; PROKOSCH; ZIERK *et al.*, 2022; ZIERK; ARZIDEH; KAPSNER; PROKOSCH *et al.*, 2020).

3.3 MÉTODO refineR

O algoritmo refineR foi desenvolvido em 2021 por Ammer *et al.* e utiliza o ambiente de linguagem de programação R para estimar os intervalos de referência de um conjunto de dados laboratoriais utilizando a abordagem pelo método indireto. Trata-se de um *software* livre que realiza um conjunto integrado de técnicas e recursos estatísticos robustos usando o pacote de código aberto do 'refineR', disponível em https://CRAN.R-project.org/package=refineR (AGARAVATT; KANSARA; KHUBCHANDANI; SANGHANI *et al.*, 2023; AMMER; SCHUTZENMEISTER; PROKOSCH; RAUH *et al.*, 2021).

Ammer *et al.* (2021) publicaram o primeiro artigo com a descrição detalhada da versão 1.0 do algoritmo (AMMER; SCHUTZENMEISTER; PROKOSCH; RAUH *et al.*, 2021):

> O algoritmo refineR para a estimativa de intervalos de referência é baseado na suposição de que a maioria dos dados laboratoriais de rotina é composta de resultados de testes não patológicos. Além disso, assume-se que a distribuição destas amostras não patológicas pode ser modelada com uma distribuição normal transformada por Box-Cox, significando uma distribuição pode acomodar distribuições normais e distorcidas. Além disso, o algoritmo presume que existe um intervalo de resultados de testes onde a proporção de resultados de testes patológicos é insignificante (AMMER; SCHUTZENMEISTER; PROKOSCH; RAUH *et al.*, 2021).

Este método emprega um processo de modelagem inversa e imparcial de três etapas que começa determinando a área e o pico de pesquisa do parâmetro, seguido por uma pesquisa de grade multinível para encontrar os melhores parâmetros do modelo e, finalmente, extrai os intervalos de referência do modelo ideal (AGARAVATT; KANSARA; KHUBCHANDANI; SANGHANI *et al.*, 2023; AMMER;

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

SCHUTZENMEISTER; PROKOSCH; RAUH *et al.*, 2021). Meyer *et al.* (2023) afirmaram que este algoritmo é semelhante ao método TML, já que uma distribuição normal transformada por Box-Cox é gerada para o cálculo dos intervalos de referência. O refineR estende o intervalo de valores lambda para ajustar também distribuições distorcidas à esquerda (MEYER; MÜLLER; HOFFMANN; SKADBERG *et al.*, 2023).

Ammer *et al.* (2023) afirmaram que para implementação do algoritmo refineR são necessárias medições de rotina de testes diagnósticos, contendo amostras patológicas e não patológicas como entrada. A ferramenta "utiliza métodos estatísticos robustos para derivar um modelo que descreve a distribuição das amostras não patológicas. Esta distribuição pode então ser usada para derivar intervalos de referência" (AMMER; SCHUTZENMEISTER; RANK; DOYLE, 2023).

O algoritmo pode estimar IRs a partir de dados do mundo real que consistem em uma distribuição mista de resultados de testes não patológicos e patológicos. Supõe-se que a maioria dos resultados dos testes no conjunto de dados de entrada não são patológicos e que a sua distribuição pode ser descrita por uma distribuição normal transformada por Box-Cox. Além disso, presume-se que existe uma região de concentrações de resultados de testes, onde a fração de resultados de testes patológicos é insignificante. A forma da distribuição dos resultados dos testes patológicos pode ser arbitrária (AMMER; SCHUTZENMEISTER; RANK; DOYLE, 2023).

3.4 MÉTODO laBRI

O método LabRI emprega uma abordagem multicritério automatizada, incorporando o algoritmo Expectation-Maximization (EM) e integrando vários algoritmos em diferentes estágios para identificar a distribuição de referência ótima relacionada à subpopulação de referência e calcular os limites de referência (RLs) juntamente com seus respectivos intervalos de confiança de 90%. O método automatiza o processo de limpeza de dados e, se necessário, realiza transformação de dados, desconvolução de misturas e truncamento de distribuição. O algoritmo LabRI utiliza o ambiente de linguagem R para estimar os intervalos de referência de um conjunto de dados laboratoriais utilizando a abordagem pelo método indireto pelo método LabRI. Trata-se de um *software* livre que realiza um conjunto integrado de técnicas e recursos estatísticos robustos.

Após a aplicação dos critérios de exclusão mencionados acima, procedeu-se à estimação indireta dos IRs, empregando uma abordagem computacional que integra uma seleção de pacotes R para otimização e precisão. O pacote MixR foi utilizado para desconvolução de mistura, ehile modeest e multimode auxiliando na identificação e análise de modos dentro de distribuições. O pacote univOutl desempenhou um papel crucial na detecção e tratamento de outliers, garantindo a qualidade dos dados. Por último, o pacote de momentos facilitou os cálculos das principais medidas estatísticas, como assimetria e curtoses instruções para acesso e utilização do pacote compactado, que inclui o arquivo *Rmarkdown* para instalação dos pacotes R necessários e a ferramenta LabRI, estão disponíveis em https://grupolabr.com/LabRI_Packed.html.

4 CONSIDERAÇÕES FINAIS

A utilização de algoritmos de mineração de dados é uma realidade que tem evoluído a partir do desenvolvimento tecnológico, revolução da automação e o aumento da complexidade dos dados. Estas situações tem gerado a necessidade de aplicação de metodologias estatísticas mais sofisticadas para identificar a distribuição não patológica

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

na determinação de intervalos de referência. Sabe-se que as técnicas tradicionais de análise de dados apresentam algumas limitações especialmente em grandes conjuntos de dados, portanto a implementação dos algoritmos de mineração de dados pode reduzir a interferência de resultados discrepantes.

Diretrizes internacionais já descrevem há alguns anos, o processo para estabelecer, de forma direta, os intervalos de referência em uma população saudável e esta abordagem ainda é a que muitos autores sugerem. Contudo, é uma tarefa difícil, trabalhosa e onerosa para os laboratórios clínicos, por isto, na maioria das vezes são utilizados IR que não foram definidos na instituição.

Considerando as vantagens, desvantagens, facilidade de utilização das ferramentas computacionais, aplicação de critérios rigorosos para seleção e algoritmos cada vez mais robustos é possível considerar a utilização das diferentes ferramentas computacionais que aplicam a abordagem indireta propostos pelos métodos Bhattacharya, Kosmic, refineR e LabRI.

Ainda é preciso elucidar alguns pontos e estratégias para redução do risco de selecionar e estratificar os resultados disponíveis no SIL, no entanto, os resultados possibilitam concluir que a abordagem pelo método indireto usando diferentes ferramentas computacionais e algoritmos de mineração de dados é replicável e que o método indireto mostrou ser uma importante estratégia para que os laboratórios atualizem o IR de forma sistemática.

REFERÊNCIAS BIBLIOGRÁFICAS

ABEBE, M.; MELKU, M.; ENAWGAW, B.; BIRHAN, W., *et al.* Reference intervals of routine clinical chemistry parameters among apparently healthy young adults in Amhara National Regional State, Ethiopia. **PLoS One**, v. 13, n. 8, p. e0201782, 2018.

ADELI, K.; HIGGINS, V.; TRAJCEVSKI, K.; WHITE-AL HABEEB, N. The Canadian laboratory initiative on pediatric reference intervals: A CALIPER white paper. **Crit Rev Clin Lab Sci**, v. 54, n. 6, p. 358-413, set. 2017.

AGARAVATT, A.; KANSARA, G.; KHUBCHANDANI, A.; SANGHANI, H., *et al.* Verification of Reference Interval of Thyroid Hormones With Manual and Automated Indirect Approaches: Comparison of Hoffman, KOSMIC and refineR Methods. **Cureus**, v. 15, n. 5, p. e39066, maio 2023.

AMMER, T.; SCHUTZENMEISTER, A.; PROKOSCH, H. U.; RAUH, M., *et al.* refineR: A Novel Algorithm for Reference Interval Estimation from Real-World Data. **Sci Rep**, v. 11, n. 1, p. 16023, 6 ago. 2021.

AMMER, T.; SCHUTZENMEISTER, A.; PROKOSCH, H. U.; ZIERK, J., *et al.* RIbench: A Proposed Benchmark for the Standardized Evaluation of Indirect Methods for Reference Interval Estimation. **Clin Chem**, v. 68, n. 11, p. 1410-1424, 3 nov. 2022.

AMMER, T.; SCHUTZENMEISTER, A.; RANK, C. M.; DOYLE, K. Estimation of Reference Intervals from Routine Data Using the refineR Algorithm-A Practical Guide. **J Appl Lab Med**, v. 8, n. 1, p. 84-91, 4 jan. 2023.

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

BRASIL. Resolução de Diretoria Colegiada nº 786, de 5 de maio de 2023. Dispõe sobre os requisitos técnico-sanitários para o funcionamento de Laboratórios Clínicos, de Laboratórios de Anatomia Patológica e de outros Serviços que executam as atividades relacionadas aos Exames de Análises Clínicas (EAC) e dá outras providências. **Diário Oficial da União**: Seção 1, Brasília, DF, p. 161, 5 maio 2023.

BURTIS, C. A.; BURNS, D. E. Tietz Fundamentos de Química Clínica e Diagnóstico Molecular. Tradução: RODRIGUES, F. S. M. Rio de Janeiro: Elsevier, 2016.
CERIOTTI, F.; HINZMANN, R.; PANTEGHINI, M. Reference intervals: the way forward. **Ann Clin Biochem**, v. 46, n. Pt 1, p. 8-17, jan. 2009.

CHEN, J.; FAN, L.; YANG, Z.; YANG, D. Comparison of results and age-related changes in establishing reference intervals for CEA, AFP, CA125, and CA199 using four indirect methods. **Pract Lab Med**, v. 38, p. e00353, jan. 2024.

CLSI. EP28-A3c: Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guidelines - Third Edition. Wayne: **Clinical and Laboratory Standards Institute**, 2010.

COSKUN, A.; SANDBERG, S.; UNSAL, I.; SERTESER, M., *et al.* Personalized reference intervals: from theory to practice. **Crit Rev Clin Lab Sci**, p. 1-16, 17 maio 2022.

HAECKEL, R.; ADELI, K.; JONES, G.; SIKARIS, K., *et al.* Definitions and major prerequisites of direct and indirect approaches for estimating reference limits. **Clin Chem Lab Med**, v. 61, n. 3, p. 402-406, 23 fev. 2023.
HALLWORTH, M. J. The '70% claim': what is the evidence base? **Ann Clin Biochem**, v. 48, n. Pt 6, p. 487-488, nov. 2011.

HENNY, J.; VASSAULT, A.; BOURSIER, G.; VUKASOVIC, I., *et al.* Recommendation for the review of biological reference intervals in medical laboratories. **Clin Chem Lab Med**, v. 54, n. 12, p. 1893-1900, 1 dez. 2016.

HOLMES, D. T.; BUHR, K. A. Widespread Incorrect Implementation of the Hoffmann Method, the Correct Approach, and Modern Alternatives. **Am J Clin Pathol**, v. 151, n. 3, p. 328-336, 4 fev. 2019.

HUBER, K. R.; MOSTAFAIE, N.; STANGL, G.; WOROFKA, B., *et al.* Clinical chemistry reference values for 75-year-old apparently healthy persons. **Clin Chem Lab Med**, v. 44, n. 11, p. 1355-1360, 2006.

JONES, G. R. D.; HAECKEL, R.; LOH, T. P.; SIKARIS, K., *et al.* Indirect methods for reference interval determination - review and recommendations. **Clin Chem Lab Med**, v. 57, n. 1, p. 20-29, 19 dez. 2018.

LIU, Y. F.; WILSON, C.; BEDNY, M. Contribution of the language network to the comprehension of Python programming code. **Brain Lang**, v. 251, p. 105392, abr. 2024.

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024
MA, C.; YU, Z.; QIU, L. Development of next-generation reference interval models to establish reference intervals based on medical data: current status, algorithms and future consideration. **Crit Rev Clin Lab Sci**, p. 1-19, 26 dez. 2023.

MA, C.; ZOU, Y.; HOU, L.; YIN, Y., *et al.* Validation and comparison of five data mining algorithms using big data from clinical laboratories to establish reference intervals of thyroid hormones for older adults. **Clin Biochem**, v. 107, p. 40-49, set. 2022.

MARTÍNEZ-SÁNCHEZ, L.; MARQUES-GARCÍA, F.; OZARDA, Y.; BLANCO, A., *et al.* Big data e intervalos de referencia: motivación, prácticas actuales, prerrequisitos de armonización y estandarización y futuras perspectivas en el cálculo de intervalos de referencia mediante métodos indirectos. **Advances in Laboratory Medicine / Avances en Medicina de Laboratorio**, v. 2, n. 1, p. 17-25, 2021.

MELILLO, K. D. Interpretation of laboratory values in older adults. **Nurse Pract**, v. 18, n. 7, p. 59-67, jul. 1993.

MEYER, A.; MÜLLER, R.; HOFFMANN, M.; SKADBERG, Ø., *et al.* Comparison of three indirect methods for verification and validation of reference intervals at eight medical laboratories: a European multicenter study. **Clin Chem Lab Med**, v. 47, n. 4, p. 155-163, 2023.

NAYUPE, S. F.; MBULAJE, P.; MUNHARO, S.; PATEL, P., *et al.* Medical laboratory practice in Malawi - Current status. **Afr J Lab Med**, v. 12, n. 1, p. 1921, 2023.

NILSSON, S. E.; EVRIN, P. E.; TRYDING, N.; BERG, S., *et al.* Biochemical values in persons older than 82 years of age: report from a population-based study of twins. **Scand J Clin Lab Invest**, v. 63, n. 1, p. 1-13, 2003.
OZARDA, Y. Reference intervals: current status, recent developments and future considerations. **Biochem Med** (Zagreb), v. 26, n. 1, p. 5-16, 2016.

OZARDA, Y.; HIGGINS, V.; ADELI, K. Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations. **Clin Chem Lab Med**, v. 57, n. 1, p. 30-37, 19 dez. 2018.

OZARDA, Y.; ICHIHARA, K.; JONES, G.; STREICHERT, T., *et al.* Comparison of reference intervals derived by direct and indirect methods based on compatible datasets obtained in Turkey. **Clin Chim Acta**, v. 520, p. 186-195, set. 2021.,

PLEBANI, M. Towards quality specifications in extra-analytical phases of laboratory activity. **Clin Chem Lab Med**, v. 42, n. 6, p. 576-577, 2004.

ROHR, U. P.; BINDER, C.; DIETERLE, T.; GIUSTI, F., *et al.* The Value of In Vitro Diagnostic Testing in Medical Practice: A Status Report. **PLoS One**, v. 11, n. 3, p. e0149856, 2016.

Artigo recebido: 26/08/2024
Artigo aceito para publicação: 28/11/2024

SOCIEDADE BRASILEIRA DE PATOLOGIA CLÍNICA / MEDICINA LABORATORIAL (SBPC/ML). Norma PALC: Programa de Acreditação de Laboratórios Clínicos. São Paulo: **SBPC/ML**, 2021.

SOCIEDADE BRASILEIRA DE ANÁLISES CLÍNICAS (SBAC). Manual para Acreditação: Sistema de Gestão da Qualidade de Laboratórios Clínicos. 8. ed. Rio de Janeiro: SBAC, 2023.

STAPLES, T. L. Expansion and evolution of the R programming language. **R Soc Open Sci**, v. 10, n. 4, p. 221550, abr. 2023.

VASARHELYI, B.; DEBRECZENI, L. A. Lab Test Findings in the Elderly. **EJIFCC**, v. 28, n. 4, p. 328-332, dez. 2017.

VELEV, J.; LEBIEN, J.; ROCHE-LIMA, A. Unsupervised machine learning method for indirect estimation of reference intervals for chronic kidney disease in the Puerto Rican population. **Sci Rep**, v. 13, n. 1, p. 17198, 11 out. 2023.

WINKEL, P. A programming language and a system for automated time- and laboratory test level dependent decision-making during patient monitoring. **Comput Biomed Res**, v. 23, n. 5, p. 426-446, out. 1990.

YANG, C.; LANG, L.; WANG, S.; FANG, H., *et al.* Application of the Hoffmann, Bhattacharya, nonparametric test, and Q-Q plot methods for establishing reference intervals from laboratory databases. **Clin Biochem**, v. 113, p. 9-16, mar. 2023.

YANG, D.; SU, Z.; ZHAO, M. Big data and reference intervals. **Clin Chim Acta**, v. 527, p. 23-32, 15 fev. 2022.

ZHONG, J.; MA, C.; HOU, L.; YIN, Y., *et al.* Utilization of five data mining algorithms combined with simplified preprocessing to establish reference intervals of thyroid-related hormones for non-elderly adults. **BMC Med Res Methodol**, v. 23, n. 1, p. 108, 2 maio 2023.

ZIERK, J.; ARZIDEH, F.; KAPSNER, L. A.; PROKOSCH, H. U., *et al.* Reference Interval Estimation from Mixed Distributions using Truncation Points and the Kolmogorov-Smirnov Distance (kosmic). **Sci Rep**, v. 10, n. 1, p. 1704, 3 fev. 2020.

Autor correspondente:
Gustavo Oliveira Gonçalves (gustavo.ineti@gmail.com)
Faculdade de Minas – FAMINAS BH. Belo Horizonte, Minas Gerais, Brasil.

Conflitos de interesses: Esta pesquisa não foi financiada ou possui qualquer relação com qualquer tipo de instituição. Os autores não possuem conflitos de interesse.