

## Metodi indiretti per la determinazione e la verifica degli intervalli di riferimento

Matteo Vidali<sup>1</sup>, Ruggero Dittadi<sup>2</sup>, Ferruccio Ceriotti<sup>1</sup>, Duilio Brugnoli<sup>3</sup>, Davide Giavarina<sup>4</sup>

<sup>1</sup>SC Patologia Clinica, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano

<sup>2</sup>UOC Medicina di Laboratorio, AULSS 3 Serenissima, Ospedale Dell'Angelo, Mestre

<sup>3</sup>SC Laboratorio Centrale di Analisi Chimico-Cliniche, ASST Spedali Civili, Brescia

<sup>4</sup>UOC Medicina di Laboratorio, AULSS 8 Berica, Ospedale san Bortolo, Vicenza

### ABSTRACT

#### Indirect methods to establish and verify reference intervals

Indirect methods are data mining techniques for establishing or verifying reference intervals. They present many advantages over the traditional direct approach, including reduced costs, closeness of preanalytical and analytical conditions to routine conditions and usefulness in specific age groups or with uncommon sample types. After discussing advantages and limitations of the indirect approach, in this paper the main statistical algorithms, and their implementation in open-source software as well, are briefly introduced by a practical example. Moreover, key aspects of data selection and preparation are discussed. The indirect methods represent a powerful and valuable approach for the definition or validation of reference intervals, whose knowledge and correct use should be part of the laboratorian's toolbox.

**Parole chiave:** *metodi indiretti, metodi diretti, intervalli di riferimento*

### INTRODUZIONE: DEFINIZIONI E CONTESTO

Perché una misura effettuata in un laboratorio clinico diventi un'informazione, vengono identificate tre condizioni necessarie:

- la domanda, ossia la condizione di uno stato di conoscenza del richiedente, che possa essere modificato dalla comunicazione di un dato;
- una misura affidabile, che rispetti specifiche di qualità sufficienti per discriminare informazioni diverse;
- un sistema di valutazione, o di riferimento, che attribuisca alla misura un significato o una qualità. Si tratta in sintesi di limiti o valori decisionali che separano o classificano stati di salute e/o di malattia o condizioni fisiopatologiche comunque rilevanti.

È opportuno precisare la differenza tra limiti di

riferimento, o intervalli di riferimento (IR), e limiti decisionali. Gli IR sono calcolati da misure effettuate su una popolazione considerata "sana", utilizzando metodi statistici; la loro applicazione informa se un soggetto appartiene o meno, con una data probabilità (in genere 95%), a questa popolazione di riferimento. I limiti decisionali dipendono da scelte o quesiti clinici, sono derivati da studi clinici o definiti attraverso un processo di consenso di esperti, ed il loro utilizzo consente di stimare un rischio finalizzato alla gestione clinica di un paziente (1). Se i limiti decisionali sono uno strumento molto più affilato ed efficace nella definizione delle informazioni e nella successiva gestione dei pazienti, essi non sempre sono definiti o definibili, per carenza di studi, sovrapposizione di più condizioni alternative, incertezza o contraddittorietà tra le raccomandazioni.

Corrispondenza a: Matteo Vidali, Patologia Clinica, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milano, Italia  
Via F. Sforza 35, 20122, Milano, Italia, E-mail: matteo.vidali@policlinico.mi.it

Ricevuto: 02.04.2023

Revisionato: 14.04.2023

Accettato: 08.05.2023

Publicato on-line: 29.05.2023

DOI: 10.19186/BC\_2023.037

Riconoscere una condizione di “normalità” e quindi definire un intervallo di valori “normali” o di riferimento, parrebbe invece apparentemente più semplice; tuttavia, tale attività presenta difficoltà e criticità teoriche e pratiche. Infatti, le stesse definizioni di “normalità”, di “benessere” o di “salute” non sono di semplice risoluzione. L’approfondimento e la riflessione su questi concetti sono di una certa complessità (2-3) ed esulano dagli scopi di questa rassegna, che è, invece, focalizzata sugli aspetti tecnico-pratici del calcolo degli IR, in particolare mediante l’approccio dei metodi indiretti. A tal fine, dopo aver introdotto le differenze tra metodi diretti e indiretti, con i vantaggi e gli svantaggi, saranno discussi gli elementi critici dell’approccio indiretto, e presentati alcuni algoritmi relativi ai metodi indiretti nonché una breve descrizione dei software attualmente disponibili per la loro esecuzione.

### Approccio diretto e indiretto al calcolo degli intervalli di riferimento: caratteristiche

La definizione degli IR rappresenta un’attività fondamentale e una responsabilità del laboratorio. La pratica diffusa di utilizzare, senza opportuna verifica, i limiti suggeriti dal produttore di una strumentazione o di un kit commerciale, non risulta molto sostenibile scientificamente e può portare a decisioni cliniche errate. Errori del produttore (misurazione o calcolo), differenze tra metodi o tra popolazioni a cui l’esame è destinato, possono essere alla base di tali problematiche. La necessità di revisionare periodicamente gli IR utilizzati è, inoltre, ribadita dalla norma ISO 15189 per l’accreditamento dei laboratori medici (4). È necessario, tuttavia, sottolineare che la definizione *ex-novo* degli IR non è l’unica strategia a disposizione del laboratorista; altri approcci possibili includono il trasferimento e la verifica di IR determinati nello stesso laboratorio o da terzi e l’adozione di IR comuni (1,5). Nell’ambito della definizione e verifica degli IR, il documento CLSI EP28-A3c del 2010 e la recente guida CLSI EP28IG del 2022 rappresentano la letteratura di riferimento (6-7). In presenza di intervalli di riferimento definiti da terzi (produttori, altri laboratori, studi multicentrici), qualora la procedura di definizione dell’IR risulti ben documentata (inclusi la popolazione, gli aspetti pre-analitici ed analitici), il laboratorio può decidere di trasferire l’IR (nel senso di utilizzare lo stesso IR) determinato altrove dopo opportuna verifica. Questa include la selezione di un ridotto numero (in genere 20) di soggetti di riferimento presi dalla popolazione del laboratorio ricevente e l’esecuzione di un calcolo statistico (eliminazione di outliers e test binomiale) (5-6). Quando un laboratorio, con un precedente IR validato, cambia il metodo analitico, potrebbe non essere necessario ridefinire nuovi IR, quanto piuttosto ricalcolare i limiti dell’IR tramite i parametri di regressione, ottenuti da uno studio di comparazione tra metodi. In questo specifico scenario, il procedimento richiede un numero maggiore di campioni (in genere almeno 40) con concentrazioni che coprano l’intero IR, ma può essere condotto anche con individui non di riferimento.

L’alternativa dell’uso di “intervalli comuni” rappresenta

un approccio recente al problema degli IR. In presenza di metodi analitici correttamente standardizzati e popolazioni con caratteristiche simili, l’adozione di IR comuni dovrebbe essere la strategia più logica e non dovrebbe presentare particolari difficoltà. A dispetto della semplicità del concetto, tuttavia, la reale applicazione degli IR comuni risulta difficilmente realizzabile, in quanto la loro adozione richiede una serie di importanti requisiti analitici e clinici non sempre possibili, inclusi l’esistenza di metodi commerciali riferibili al sistema di riferimento, la loro corretta implementazione e verifica da parte dei laboratori, la definizione accurata degli IR e la disponibilità di informazioni accurate relative all’influenza di fattori biologici e ambientali (1).

In assenza di queste condizioni, quando non sia possibile trasferire un IR determinato da terzi, o per analiti nuovi, si rende necessaria la determinazione *ex-novo* dell’IR. A tal fine può essere utilizzato l’approccio diretto o indiretto. L’approccio diretto, o tradizionale, alla definizione degli IR prevede le seguenti fasi:

- selezione degli individui di riferimento, mediante criteri di inclusione o esclusione (ad esempio allattamento, obesità, alcolismo, tossicodipendenza, tabagismo, terapie farmacologiche, comorbidità); la selezione può essere effettuata a priori o a posteriori se gli individui sono selezionati, rispettivamente, prima o dopo il campionamento;
- suddivisione degli individui di riferimento in categorie tramite criteri di partizione (ad esempio etnia, età, sesso, fase del ciclo mestruale, settimana di gravidanza, fumo, esercizio fisico);
- preparazione dei soggetti;
- raccolta e analisi dei campioni;
- elaborazione statistica dei dati (identificazione degli outliers, calcolo dei limiti di riferimento e dei loro intervalli di confidenza) (6).

Relativamente a questo ultimo aspetto, per il calcolo dei limiti dell’IR, a  $\text{media} \pm 2 \text{ DS}$  (metodo parametrico), applicabile solo quando la distribuzione dei valori di riferimento è perfettamente gaussiana, il CLSI preferisce e raccomanda da sempre l’utilizzo dei percentili (generalmente 2,5 e 97,5) (metodo non parametrico), applicabile anche in presenza di distribuzioni asimmetriche o non gaussiane (6). Il limite di questo approccio non parametrico dei percentili è rappresentato dall’elevato numero di osservazioni richiesto: infatti, per determinare i limiti dell’IR (come percentili 2,5 e 97,5) sono necessarie 40 osservazioni, mentre il calcolo dei rispettivi intervalli di confidenza richiede 120 osservazioni (6). Poiché tale numerosità è da intendersi per ogni partizione, in presenza di sottogruppi (ad esempio per sesso ed età) il numero di individui da reclutare aumenta significativamente e l’esperimento di determinazione dell’IR risulta difficilmente gestibile e fuori dalla portata della maggior parte dei laboratori, a causa delle risorse necessarie. Per tentare di superare tale criticità, il CLSI, pur continuando a raccomandare il metodo non parametrico e in aggiunta alle alternative riportate sopra (trasferimento e verifica o applicazione di intervalli comuni), nelle ultime versioni del documento di riferimento ha proposto l’utilizzo di un metodo robusto,

un processo iterativo basato su mediana e deviazione mediana assoluta (MAD) (6). Questo metodo richiede, invero, un numero di osservazioni minore rispetto al metodo non parametrico dei percentili; tuttavia, il CLSI non raccomanda il suo utilizzo con meno di 60-80 osservazioni (6). Inoltre, questo metodo risulta ancora poco implementato nella maggior parte dei software statistici.

La definizione del soggetto di riferimento, il riconoscimento dei soggetti che aderiscono alla definizione, la numerosità campionaria necessaria (ridotta solo in parte se si utilizza il metodo robusto), le difficoltà pratiche ed etiche nella raccolta di campioni per alcune classi di ripartizione (come i valori di riferimento pediatrici o della popolazione anziana) e di esami non frequenti e molto ripartiti, sono tutti fattori che concorrono a determinare un rilevante costo nella costruzione di appropriati IR, sia in termini monetari che economici in senso lato. L'effetto comune, come è già stato osservato in precedenza, è che si utilizzino IR forniti dai produttori dei metodi analitici o presi dalla letteratura scientifica. Le difficoltà sono quindi evidenti e si complicano quando si cercano "limiti" nell'ambito di situazioni bi- o multi-modali, come ad esempio nel caso di un marcatore tumorale, oppure ove esistano gradi diversi dello stesso stato di salute (ad esempio, eccellente, molto buono, buono, modesto, scarso). Queste considerazioni hanno suggerito l'idea di poter utilizzare dati anche di soggetti "non di riferimento" con un qualche sistema statistico robusto in grado di discriminare e/o pesare diversamente sottopopolazioni di risultati. L'approccio costituisce, evidentemente, un rischio rilevante per una stima corretta di un limite discriminante. Quando però si possono utilizzare campioni molto numerosi, il riconoscimento della sottopopolazione di appartenenza, normale o

patologica, del singolo risultato è più agevole. È questo il caso delle banche dati dei laboratori clinici, dove sono registrati centinaia di migliaia di dati per ogni singolo costituente (8). Se, infine, si considera che, soprattutto per la popolazione non ricoverata, la maggior parte di questi dati deriva da accertamenti di screening o si riferisce ad individui senza patologia correlata, il riconoscimento della sottopopolazione maggioritaria "normale", su cui stimare un IR, può essere un obiettivo raggiungibile. Questa è essenzialmente l'idea alla base dell'approccio indiretto, o dei cosiddetti metodi indiretti, alla definizione degli IR: utilizzare tutti i dati relativi ad un certo analita, raccolti in un certo intervallo di tempo e già disponibili all'interno del sistema gestionale del laboratorio, sotto l'ipotesi che la maggior parte di questi risultati appartengano a individui sani o comunque a individui senza una patologia correlata all'analita di interesse. Le differenze principali tra i metodi diretti e indiretti sono evidenziate nella Tabella 1.

A differenza dei metodi diretti, dove gli individui di riferimento sono attentamente selezionati in base a specifici criteri, la popolazione di individui utilizzata per i metodi indiretti è estremamente eterogenea perché i campioni raccolti e i dati ottenuti non erano originariamente destinati a definire un IR, ma a scopo di screening, diagnosi o monitoraggio. Una delle principali critiche mosse ai fautori dei metodi diretti è che spesso la preparazione del paziente e le strette condizioni di raccolta dei campioni degli individui di riferimento sono molto diverse da quanto poi avviene nella routine di un ambulatorio o di un centro prelievi, con il rischio di definire IR scarsamente generalizzabili. Altre differenze significative tra i due approcci sono rappresentate dalla numerosità del campione e dalla complessità dell'analisi statistica.

**Tabella 1**

*Principali differenze tra metodi diretti e indiretti nella definizione dell'intervallo di riferimento*

	Metodi diretti	Metodi indiretti
Popolazione	Individui di riferimento scelti in base a criteri di inclusione/esclusione	Individui sottoposti a quel determinato esame per motivi diversi dalla definizione dell'intervallo di riferimento. Sono inclusi individui con differenti patologie
Selezione delle osservazioni	A priori (se prima della raccolta con criteri di esclusione) o a posteriori (dopo la misurazione)	Osservazioni patologiche escluse in base ad altre informazioni presenti nel database di laboratorio o ad altri database (big data)
Variabili pre-analitiche e preparazione del paziente	Le principali variabili pre-analitiche devono essere note e controllate per minimizzare la variabilità pre-analitica	La variabilità pre-analitica e le condizioni di raccolta del campione sono le stesse o molto simili a quelle della routine del laboratorio
Dimensione del campione	Per il metodo dei percentili si raccomandano 120 individui per partizione. Con il metodo robusto 60-80 individui per partizione	Sono necessarie migliaia di osservazioni
Analisi statistica	Percentili, metodo robusto	Complessa: diversi algoritmi disponibili con differenti prestazioni

È necessario sottolineare che i metodi indiretti possono essere applicati solo in presenza di un campione molto numeroso. È difficile fissare un limite inferiore accettabile per la dimensione campionaria, perché essa dipende dal livello di contaminazione della popolazione normale da parte di osservazioni patologiche (cioè di individui con una patologia correlata all'analita in studio). Non vi sono dubbi, tuttavia, che l'analisi statistica per i metodi indiretti è particolarmente complessa, richiede competenze specifiche e particolari software. La Tabella 2 presenta vantaggi e svantaggi dei metodi diretti e indiretti in termini

di aspetti teorici, organizzativi, economici, analitici e post-analitici. Come si osserva dalla tabella, in aggiunta a quanto già riportato sopra, gli aspetti che maggiormente differenziano i metodi diretti da quelli indiretti sono quelli economici, statistici, etici e di privacy. Mentre, infatti, per i metodi indiretti i dati sono già disponibili nel gestionale di laboratorio e quindi non vi sono eventuali costi da includere (se si escludono quelli necessari a coprire eventuali spese di consulenza a causa delle competenze richieste), per i metodi diretti essi possono essere molto onerosi e includere spese per la raccolta di tutte le informazioni

**Tabella 2**

*Vantaggi e svantaggi dei metodi diretti e indiretti*

Aspetti		Metodi diretti	Metodi indiretti
Teorici	Vantaggi	- minima o assente contaminazione di risultati patologici - variabilità pre-analitica minimizzata	- condizioni identiche o simili alle operazioni di routine
	Svantaggi	- necessaria la conoscenza di tutte le variabili con influenza sull'analita e di complete informazioni cliniche dei soggetti	- possibile contaminazione di risultati patologici (con allargamento dell'IR) - difficoltà a trasferire l'IR ad altri laboratori per non conoscenza completa della popolazione e del processo
Organizzativi	Vantaggi	- privacy: consenso all'utilizzo dei dati clinici ottenuto direttamente dai partecipanti	- i risultati sono già all'interno del gestionale, devono solo essere estratti - etici: utilizzabili anche in popolazioni pediatriche o anziane
	Svantaggi	- condurre le operazioni di raccolta del campione in condizioni controllate - etici: difficilmente realizzabile in pazienti pediatriche o anziani	- privacy (se per l'eliminazione dei risultati contaminanti si utilizzano altri database)
Economici	Vantaggi		- costi assenti o minimi (dati già disponibili)
	Svantaggi	- costi per la selezione degli individui (questionari, esami aggiuntivi, eventuali esami strumentali) - costi per l'esecuzione dell'esame (l'esame è effettuato con lo scopo preciso di definire l'IR)	
Analitici	Vantaggi	- analisi statistica semplice (percentili) e disponibile in qualunque software	
	Svantaggi	- analisi statistica più complessa per il metodo robusto	- elevata numerosità (migliaia di dati) - analisi statistica complessa e scarsamente implementata a livello di software
Post-analitici	Vantaggi	- facilmente comunicabile	
	Svantaggi	- tutte le fasi devono essere attentamente e accuratamente descritte	- procedimento difficilmente comunicabile - possibile allargamento dei limiti a causa della presenza di risultati patologici

*IR, intervallo(i) di riferimento.*

cliniche necessarie per la selezione (conduzione di questionari o interviste, visite ambulatoriali, esecuzione di esami ematici o strumentali), i costi necessari per l'esecuzione dell'esame per cui si vuole definire l'IR e i costi di tutto il personale coinvolto nella procedura. Relativamente all'analisi statistica, in alternativa ai percentili e al metodo robusto dell'approccio diretto, per i metodi indiretti possono essere utilizzati numerosi algoritmi caratterizzati da diversa complessità teorica e computazionale. Da un punto di vista etico, il prelievo di campioni da numerosi pazienti pediatrici o anziani, al solo fine della definizione di un IR, rappresenta un importante limite per l'approccio diretto, ma non per i metodi indiretti, dove i risultati utilizzati per la definizione dell'IR sono ottenuti grazie a prelievi già eseguiti per necessarie precedenti indagini cliniche. Viceversa, mentre la privacy costituisce una criticità per i metodi indiretti, in particolare per accedere a database contenenti informazioni cliniche al fine di escludere eventuali risultati contaminanti, per i metodi diretti le informazioni cliniche sono ottenute da soggetti che hanno già fornito in precedenza il proprio consenso informato.

**METODI INDIRETTI**

Prima di descrivere i differenti algoritmi, storici e più recenti, utilizzati nell'approccio indiretto, è essenziale discutere alcuni aspetti critici di questi metodi relativi alla selezione dei dati e al loro trattamento prima dell'analisi vera e propria con l'algoritmo specifico.

**Selezione e pretrattamento dei dati per i metodi indiretti**

*Origini dei dati e selezione*

Poiché, come si è visto, l'ipotesi sottostante ai metodi indiretti è che la maggior parte dei dati presenti nel gestionale di laboratorio, e analizzati con l'algoritmo specifico, appartengano a individui normali o senza patologia correlata all'analita di interesse, la selezione dei dati è fondamentale per ottenere risultati accurati con

questo approccio (9). Lo scopo, infatti, della selezione è quella di limitare il più possibile la contaminazione con risultati patologici.

A tal fine sono possibili due scenari non mutualmente esclusivi (10):

*Selezione delle fonti di dati in ingresso.* Le fonti di dati utilizzabili si differenziano in base al grado di contaminazione con risultati patologici. Tra le fonti con minima o assente contaminazione possono essere annoverati, a titolo di esempio, database con risultati di donatori, individui sottoposti a screening periodico, individui sottoposti a screening genetico o che si sono sottoposti a piccoli interventi presso ambulatori pubblici o privati (10). Queste fonti presentano, tuttavia, dei limiti importanti:

- i dati possono non essere disponibili per motivi di privacy;
- i database sono generalmente di dimensioni molto limitate;
- l'analita per cui si vuole definire l'IR non è compreso nel pannello di esami di screening o eseguito prima di una specifica procedura.

*Fonti miste: individuazione ed eliminazione dei risultati con alta probabilità di essere patologici.* Più spesso si avranno a disposizione fonti miste di dati, con la presenza di risultati sia normali che patologici. È questo il caso di risultati ottenuti da pazienti ricoverati o afferenti ad un centro prelievi. In questo frequente scenario la percentuale di contaminazione è estremamente variabile e può dipendere da diversi fattori, inclusi l'origine dei dati ed il tipo di analita. È verosimile, ad esempio, che la contaminazione con valori patologici sia maggiore per esami richiesti per indagini specialistiche (ad esempio ormoni) che per esami generalmente prescritti per controlli periodici, oppure tra i risultati ottenuti in un'unità operativa ospedaliera o in un ambulatorio rispetto a quelli eseguiti presso un centro prelievi. Una possibile eccezione può essere rappresentata dai prelievi ai soggetti pediatrici, frequentemente richiesti nel caso di specifici quesiti e approfondimenti clinici o per il monitoraggio di terapie, più che per la valutazione periodica dello stato di salute. Sebbene la maggior parte dei metodi indiretti,

**Tabella 3**

*Possibili strategie, e relative criticità, per l'individuazione e l'eliminazione di risultati probabilmente patologici di database misti*

Criterio	Criticità
- risultato chiaramente patologico	- scelta soggettiva o limite massimo non noto
- esame ripetuto in un determinato intervallo temporale	- l'individuo potrebbe aver eseguito l'esame in altri laboratori
- associazione con risultati patologici di esami correlati	- i risultati di altri esami potrebbero non essere disponibili (non eseguiti o eseguiti in altra sede) o essere normali (ad esempio soggetto in terapia)
- associazione con una determinata terapia o patologia o diagnosi di dimissione	- l'informazione potrebbe non essere disponibile - scelta soggettiva
- provenienza della richiesta	- scelta soggettiva

in particolare quelli più recenti, funzioni correttamente anche con contaminazioni fino al 20-30%, è indubbio che minore è la contaminazione e più accurata sarà la stima dell'IR. La Tabella 3 riporta diverse strategie o criteri, e relative criticità, che possono essere utilizzati al fine di individuare i risultati probabilmente patologici. Un primo criterio è quello di eliminare risultati con valori chiaramente patologici, molto bassi o elevati (come >2-3 volte il limite considerato normale; ad esempio, eliminare le osservazioni con valori di tirotropina (TSH) >10 mU/L. La scelta del limite, superato il quale l'osservazione deve essere eliminata, rimane tuttavia soggettiva ed è possibile solo quando l'analita è già noto e il suo IR già riportato in letteratura. Possono essere eliminate con una certa confidenza osservazioni di pazienti che hanno ripetuto tale esame più volte in un determinato intervallo temporale (11). Il rationale di questo criterio si basa sull'ipotesi che quando un individuo ripete lo stesso esame a breve distanza, o comunque più volte in un determinato intervallo temporale, probabilmente è affetto da qualche condizione patologica correlata a quel determinato analita. Questa ipotesi è molto probabile per analiti specialistici e non comuni, ma non necessariamente per analiti molto comuni. Se questo criterio viene applicato ad un determinato database, tutte le osservazioni, compresa la prima occorrenza, devono essere eliminate (si sottolinea comunque il fatto che, indipendentemente dal database e dal metodo considerato, per la determinazione di un IR, solo un risultato per individuo deve essere considerato). Un'altra strategia è quella di eliminare dal database un individuo se presenta risultati patologici per esami correlati all'analita di cui si vuole determinare l'IR (ad esempio, in caso di IR per TSH, eliminare individui con risultati patologici per FT3, FT4 o anticorpi tiroidei). Il rationale di questo criterio è facilmente intuibile; l'informazione però potrebbe non essere disponibile (esami eseguiti in altre sedi) oppure non facilmente reperibile (con alcuni gestionali di laboratorio può non risultare semplice recuperare e confrontare queste informazioni; per evitare lunghe ricerche manuali, quando le osservazioni sono molto numerose come avviene generalmente nei metodi indiretti, è obbligatorio estrarre i risultati degli analiti correlati e programmare uno script in qualche linguaggio di programmazione che automatizzi la procedura di ricerca dei risultati per ogni paziente e valuti la presenza di risultati patologici). Inoltre, se per alcuni analiti la scelta degli esami correlati è semplice (vedi l'esempio precedente relativo alla funzionalità tiroidea), per altri è più complesso e rischia di essere soggettiva. La selezione di eventuali esami associati a quello di interesse, può essere basata su opinioni di esperti o metodi statistici, come algoritmi multivariati (12). In alternativa a eventuali risultati di altri esami, è possibile considerare altre informazioni come concomitanti terapie, la presenza di patologie associate o la diagnosi di dimissione. I limiti di questi approcci sono simili a quelli descritti in precedenza: informazione non disponibile o non facilmente reperibile (si pensi al tempo necessario per consultare eventuali cartelle cliniche anche se informatizzate) o scelta soggettiva. È interessante

notare come i big data e le tecniche di machine learning potrebbero impattare positivamente e significativamente proprio in questo ambito (10,13-15). Quando si consideri i big data nella loro accezione più ampia, non quindi solo superficialmente o unicamente come grande mole di dati ma come database interconnessi recanti informazioni di varia natura, l'utilità di questa tecnologia, nell'individuare ed eliminare gli eventuali risultati contaminanti, è facilmente comprensibile. Se alla mole di informazioni dei big data, per quantità e qualità si aggiunge la capacità dei metodi di machine learning di individuare condizioni mediche, o terapie, o altri esami potenzialmente associati a risultati patologici o contaminanti (talora l'associazione può non essere immediatamente evidente al ricercatore da un punto di vista fisiopatologico o clinico), si intuisce come i metodi indiretti possano rappresentare, almeno in teoria, un'alternativa valida e accurata per la determinazione degli IR. Ad esempio, nello studio LIMIT (Laboratory Mining for Individualized Threshold) gli autori, utilizzando un metodo di machine learning non supervisionato, hanno identificato i codici diagnostici della classificazione ICD9 significativamente associati a risultati estremi per l'analita di interesse, per escludere dal calcolo dell'IR, con un procedimento iterativo, i soggetti che presentavano tali codici (13). Infine, un elemento di selezione facilmente utilizzabile può essere rappresentato dall'informazione relativa al richiedente (tipo di ambulatorio, o di unità operativa, o centro prelievi) da cui la richiesta proviene. Sebbene anche questo criterio presenti, come i precedenti, una certa soggettività, vi è il vantaggio che l'informazione della provenienza della richiesta è quasi sempre disponibile e facilmente recuperabile nel gestionale di laboratorio (una possibile eccezione è quando le richieste provengano da spoke periferici e l'hub registri nel gestionale solo il nome dello spoke e non dell'unità operativa dello spoke che ha richiesto l'esame).

#### *Pretrattamento dei dati*

Successivamente alla selezione dei dati in ingresso e prima della vera e propria analisi statistica mediante l'algoritmo specifico selezionato, è necessario pretrattare i dati.

Questa fase include:

- verifica dell'assenza di deriva. Poiché generalmente i risultati considerati per l'approccio indiretto sono raccolti in un periodo di più anni, è necessario verificare l'assenza di significative differenze tra i diversi periodi considerati, che possono verificarsi sia per fattori analitici (bias, cambio strumenti o reagenti) oppure per modifiche della popolazione afferente al laboratorio. Tali differenze possono essere verificate confrontando la distribuzione dei risultati di diversi periodi (ad esempio comparando le mediane o i percentili relativi a diversi periodi) e/o verificando i dati del CQI o della VEQ. Se poi il database include risultati provenienti da laboratori diversi (studi multicentrici), che utilizzano metodi analitici diversi, può rendersi necessaria, in aggiunta ad una verifica, un'eventuale armonizzazione dei risultati (ad esempio mediante trasformazione matematica

basata sui parametri di regressione di uno studio di comparazione di metodi);

- eliminazione di outliers. Come per i metodi diretti, anche per quelli indiretti è fondamentale ripulire il database dai dati aberranti. Diversi metodi possono essere utilizzati, sia univariati (ad esempio, Dixon, Hampel, Horn (16) che multivariati;
- normalizzazione. Raramente le osservazioni di un database misto seguono una distribuzione gaussiana, che invece è richiesta da alcuni metodi di eliminazione di outliers o metodi indiretti. A tal fine è possibile applicare manualmente alcune trasformazioni matematiche basandosi, ad esempio, sull'entità dell'asimmetria (come radice quadrata, inversa, logaritmica), oppure applicando la trasformazione di Box-Cox, che individua il parametro ottimale della trasformazione massimizzando la funzione di verosimiglianza (17);
- eventuale partizione in categorie. La necessità di una partizione in categorie può essere dedotta da conoscenze pregresse oppure mediante l'applicazione di metodi statistici opportuni, al pari di quanto avviene per i metodi diretti. Per i metodi indiretti questa valutazione è più semplice e praticabile a causa della maggiore dimensione del campione in studio.

L'insieme di queste operazioni preliminari (selezione, pulizia, trasformazione, partizione) vanno a costituire una vera e propria "pipeline" di dati, che accetta dati non elaborati (i dati importati) in ingresso del dataset misto, li seleziona mediante opportuni criteri, li ripulisce eliminando i dati aberranti, li trasforma ed eventualmente li partiziona, prima dell'analisi vera e propria (Figura 1) (18).

**Metodi disponibili**

Numerosi sono i metodi indiretti che è possibile utilizzare per la definizione di un IR (19). Non è possibile fornire in questa sede una descrizione esaustiva, perché troppo tecnica e perché richiederebbe uno spazio eccessivo. Di seguito, invece, vengono presentati i metodi indiretti più utilizzati in letteratura, dai primi a

quelli più recenti, riportando gli elementi essenziali ed il loro razionale. Inoltre, i metodi indiretti presentati sono confrontati utilizzando un dataset misto simulato di due popolazioni.

**Metodo di Pryce**

Uno dei primi ad essere utilizzato su popolazioni miste, questo metodo (20) assume che la componente centrale della distribuzione dei dati osservati rappresenti le osservazioni di individui sani e che segua la distribuzione gaussiana. Per la stima di media e DS della popolazione sana, è necessario conoscere se gli eventuali valori contaminanti patologici siano presenti ad entrambe le estremità (come avviene ad esempio per gli elettroliti) o ad una sola coda della distribuzione (come il caso delle transaminasi).

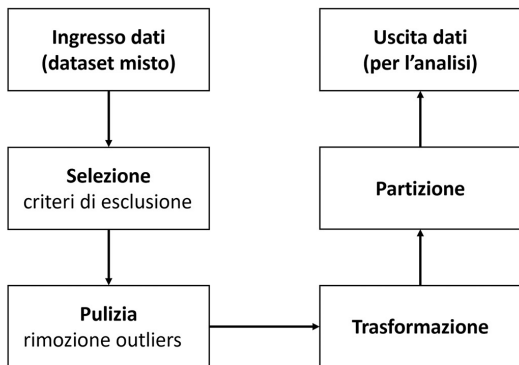
- Operativamente, nel caso di:
  - valori patologici ad entrambe le estremità, il metodo assume che i valori patologici non influenzino la distribuzione dei valori degli individui sani nell'intervallo  $\pm 1DS$  dalla media. Da ciò deriva che la media della popolazione sana è posta uguale alla media della distribuzione mista e che la DS della popolazione sana è calcolata come metà della distanza tra il percentile  $84^\circ$  e il percentile  $16^\circ$  (questo perché, nella distribuzione gaussiana, l'intervallo  $media \pm 1DS$  include il 68% delle osservazioni, che è appunto la differenza tra i percentili  $84^\circ$  e  $16^\circ$ );
  - valori patologici ad una sola estremità, la media della popolazione sana è posta uguale alla moda della distribuzione mista e la DS della popolazione sana è calcolata come differenza assoluta tra il percentile  $16^\circ$  e la moda (se le osservazioni patologiche cadono nella coda di destra) e tra il percentile  $84^\circ$  e la moda (se le osservazioni patologiche cadono nella coda di sinistra).

Il metodo ha il vantaggio della semplicità di calcolo, importante in passato quando non vi era disponibilità di macchine con elevata capacità computazionale. Tuttavia, il metodo di Pryce non risulta accurato in presenza di elevata contaminazione o quando la sovrapposizione delle distribuzioni (patologica e normale) è considerevole.

Approcci simili o successive modifiche sono state presentate da Becketl et al. (21) e Kairisto et al. (22-24).

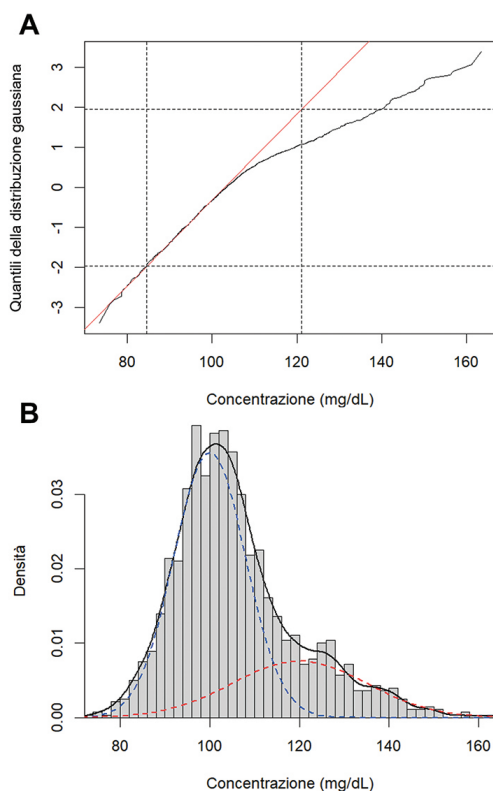
**Metodo di Hoffmann e sue modifiche**

Nel 1963 Hoffmann descrive un metodo grafico basato su alcune assunzioni: che la distribuzione dei dati sia gaussiana o quasi-gaussiana, che la distribuzione osservata sia un mix di due popolazioni (sana e patologica), che la maggior parte delle osservazioni appartengano a individui sani (25). L'approccio di Hoffmann è relativamente semplice e richiede di graficare la frequenza cumulativa delle osservazioni non su una scala lineare ma su una scala di probabilità gaussiana (asse y). Il grafico di frequenza cumulativa di una distribuzione gaussiana, qualora graficato su una scala lineare per entrambi gli assi, ha una forma sigmoidale, ma lineare quando graficato utilizzando una scala di probabilità gaussiana (si ottiene invero lo stesso grafico



**Figura 1**  
*Pipeline di dati indicanti le operazioni di pre-trattamento dei dati dall'ingresso, come database misto "contaminato", all'uscita verso il metodo indiretto di analisi.*

costruendo il cosiddetto "Normal probability plot", dove la distribuzione osservata è graficata verso i quantili teorici della distribuzione gaussiana). In questo secondo caso, dopo aver tracciato la miglior linea retta che passa per la parte centrale della distribuzione (e che quindi individua la componente gaussiana della distribuzione), si individuano per estrapolazione sui suoi prolungamenti i percentili 2,5° e 97,5° della distribuzione cumulativa, che corrispondono ai limiti dell'IR cercato. Il procedimento è rappresentato nella Figura 2. Si noti che i limiti teorici (impostati nello script di generazione della distribuzione) dell'IR per la popolazione sana simulata nella Figura 2 sono 84,3 (100-1,96x8) e 115,7 (100+1,96x8). Con il metodo di Hoffman si è trovato 84,6 e 121,1. Le stesse assunzioni del metodo di Hoffmann costituiscono in parte i suoi



**Figura 2**

*Descrizione del metodo di Hoffman.*

*A: dopo aver costruito il "Normal probability plot" una linea retta è aggiunta nella parte lineare del grafico (il grafico rappresentato contiene due sezioni lineari; la retta è stata aggiunta nella prima, dove è localizzata la distribuzione della popolazione sana). Sulla retta si estrapolano i percentili 2,5° e 97,5° (ai quantili -1,96 e +1,96) che corrispondono ai limiti dell'IR cercato.*

*B: simulazione di una distribuzione mista, costruita estraendo 1 000 osservazioni da una distribuzione teorica gaussiana con media=100 e DS=8 (popolazione sana) e 400 osservazioni da una distribuzione teorica gaussiana con media=120 e DS=15 (popolazione patologica). In blu, in rosso e in nero sono rappresentate rispettivamente le distribuzioni teoriche gaussiane delle 2 componenti sana e patologica e la distribuzione della popolazione mista. Script modificato da <https://labrorian.com>.*

limiti. Inoltre, la maggior limitazione del metodo originale consiste nella scelta arbitraria (definita "a occhio") della retta nella porzione lineare del grafico (26). Il metodo di Hoffmann è stato largamente utilizzato da molti autori per la determinazione dell'IR (27) nella sua forma originale o modificata, inclusa una versione automatizzata, o computerizzata, proposta da Katayev et al. (28-30). Quest'ultima modifica supera la soggettività della scelta della linea di regressione lungo la porzione lineare della distribuzione cumulativa, combinando il criterio di Chauvenet per l'eliminazione degli outliers, utilizzando l'errore residuo massimo (MRE) basato sulla variabilità biologica come criterio per la deviazione ammissibile per la porzione lineare della curva, e la trasformazione di Box-Cox in caso di distribuzione non gaussiana (29). In una recente pubblicazione del 2019, Holmes et al. hanno mostrato come in molti lavori di letteratura il metodo di Hoffmann sia stato utilizzato non correttamente, graficando la distribuzione cumulativa su scala lineare e non su una scala di probabilità gaussiana, con il rischio di aver riportato IR più stretti (l'errore dipende dal livello di contaminazione della popolazione sana) (27).

#### *Metodo di Bhattacharya e sue modifiche*

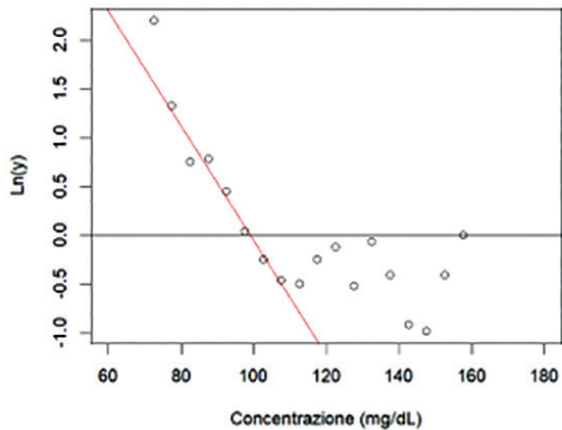
Nel 1967 Bhattacharya propone un metodo grafico per individuare in una popolazione mista due componenti gaussiane quando esse siano sufficientemente separate (31).

Il metodo di Bhattacharya prevede le seguenti fasi:

- la distribuzione di frequenza viene suddivisa in un numero arbitrario "i" di classi di uguale ampiezza "h";
- per ogni classe i-esima si calcola la frequenza della classe  $f_i$ ;
- per ogni coppia di classi vicine i e i+1, si calcola il logaritmo del rapporto delle due frequenze  $\ln(f_{i+1}/f_i)$ ;
- i logaritmi dei rapporti sono graficati verso il punto centrale della prima delle due classi vicine ( $x_i$ ), cioè  $y_i = \ln(f_{i+1}/f_i)$  vs  $x_i$ ;
- si individua la linea retta di regressione  $y_i = ax_i + b$ , che corrisponde alla componente gaussiana;
- media e varianza  $s^2$  della distribuzione della popolazione sana sono calcolate come  $m = -b/a + 0,5h$  e  $s^2 = -h/a - h^2/12$ , dove h è l'ampiezza delle classi;
- l'IR è quindi calcolato come  $m \pm 1,96s$ .

Nella Figura 3 il metodo di Bhattacharya è applicato alla stessa popolazione mista utilizzata con il metodo di Hoffmann della Figura 2. In questo caso i limiti dell'IR ottenuti sono pari a 83,6 e 119,3, simili a quelli individuati con Hoffmann e vicini a quelli teorici (84,3 e 115,7). Le principali criticità del metodo di Bhattacharya sono rappresentate dalla soggettività della scelta delle classi e della linea di regressione e dal fatto che questo metodo non fornisce stime accurate quando la distribuzione non è gaussiana o è molto contaminata.

Numerose modifiche sono state proposte per superare alcuni limiti del metodo originale di Bhattacharya. Naus et al. hanno proposto di applicare la procedura di lisciamento (smoothing) a 5 punti di Savitzky-Golay per ridurre l'impatto di basse frequenze sulla retta di regressione (32). Hemel et al. hanno suggerito di utilizzare un grafico di residui



**Figura 3**  
*Grafico di Bhattacharya: dopo suddivisione in un numero arbitrario di classi di ampiezza costante, i logaritmi dei rapporti delle frequenze di due classi adiacenti sono graficati verso i punti centrali della prima delle due classi. Una retta è individuata nella porzione lineare della prima parte del grafico. Media e DS, da cui sono calcolati i limiti dell'IR, sono derivati, come riportato nel testo a partire dai valori di intercetta e pendenza della retta di regressione. La popolazione simulata è la stessa della Figura 2. Script modificato da <https://labrtorian.com>.*

per evidenziare la deviazione dal modello gaussiano, di includere nel grafico di Bhattacharya solo punti con frequenza maggiore di una soglia prefissata (ad esempio 1/10 della frequenza modale) per limitare l'influenza di fluttuazioni casuali e, in presenza di deviazioni dal modello gaussiano, di utilizzare la funzione gamma (33). Infine, Baadenhuijsen et al. (34) e Oosterhuis et al. (35) hanno raccomandato una dimensione campionaria minima di 1 500 osservazioni e il requisito che la porzione lineare del grafico di Bhattacharya includa almeno il 40% della popolazione.

**Algoritmi di "expectation-maximization" (EM), metodi di Arzideh (TML), kosmic e TMC**

La crescente disponibilità di macchine ad elevata capacità computazionale ha consentito l'applicazione di nuove strategie analitiche al problema della definizione degli IR.

In particolare, alcuni autori hanno proposto il metodo della massima verosimiglianza ("maximum likelihood" o ML), mediante i cosiddetti algoritmi di "expectation-maximization" (EM) (36), per stimare i parametri di una distribuzione mista (37).

La teoria sottostante a questi algoritmi è estremamente complessa e non rientra negli scopi di questa rassegna. Tuttavia, la loro applicazione è favorita dallo sviluppo di alcuni software liberamente scaricabili (27,38-39). Un'applicazione di uno di questi software (pacchetto R mixtools) è riportata nella Figura 4. L'algoritmo utilizzato per la popolazione mista, già descritta negli esempi precedenti, identifica due componenti gaussiane e fornisce i seguenti parametri:

- componente 1 (popolazione sana):  $\lambda=0,739$ ;  $\mu=100,1$ ;  $\sigma=8,3$ ;
- componente 2 (popolazione patologica):  $\lambda=0,261$ ;  $\mu=121,9$ ;  $\sigma=13,6$ .

Utilizzando i dati della componente 1 si stimano i seguenti limiti dell'IR: 83,9 e 116,4. Come si può osservare sia i parametri delle due componenti che i limiti dell'IR sono molto vicini a quelli teorici.

La curva di densità di frequenza riportata in Figura 4, ottenuta dalla combinazione delle due componenti riportate sopra, si ottiene calcolando:

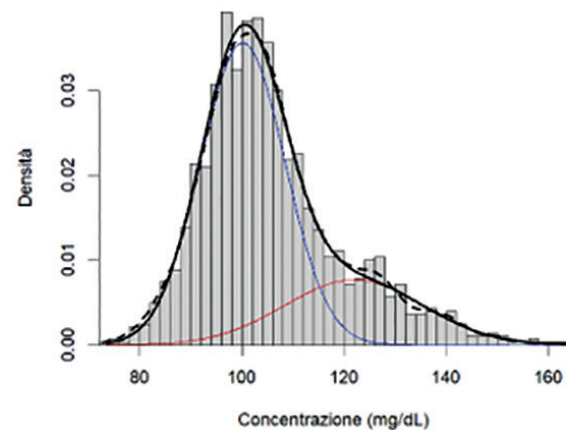
$$p(x) = \lambda_1 n(x; \mu_1, \sigma_1) + \lambda_2 n(x; \mu_2, \sigma_2),$$

dove  $n(x; \mu, \sigma)$  rappresenta la funzione di densità di probabilità gaussiana con media e DS rispettivamente delle due componenti.

Questi metodi sono molto versatili. I vari pacchetti di R (oltre al già citato mixtools, anche flexmix, mixmod, mixdist) offrono differenti metodi di deconvoluzione (scomposizione di una distribuzione mista nelle sue componenti), validi sia per distribuzioni parametriche che non-parametriche (40).

Una procedura più sofisticata rispetto alle precedenti è stata proposta da Arzideh et al. (41). Essa è nota anche come metodo TML (truncated maximum likelihood) e prevede i seguenti passaggi:

- stima di una funzione di densità "lisciata" della popolazione mista tramite metodo dei nuclei;
- modellazione della distribuzione della popolazione sana mediante trasformazione di Box-Cox e troncamento delle estremità (l'algoritmo assume che la parte centrale dei dati troncata alle due estremità, e contenente quasi unicamente dati non patologici, possa essere modellata secondo una distribuzione Box-Cox troncata). I parametri



**Figura 4**  
*Istogramma della popolazione mista e sovrapposizione delle due componenti gaussiane identificate dall'algoritmo iterativo di "expectation-maximization" del pacchetto mixtools. In blu, in rosso, in nero tratteggiato sono rappresentate rispettivamente le distribuzioni teoriche gaussiane delle 2 componenti sana e patologica e la distribuzione della popolazione mista. La linea nera continua indica la distribuzione ottenuta combinando le due componenti gaussiane (vedi testo). La popolazione simulata è la stessa della Figura 2. Script modificato da <https://labrtorian.com>.*

della distribuzione gaussiana, della trasformazione e dell'intervallo troncato, sono trovati mediante metodo della massima verosimiglianza;

- ottimizzazione mediante metodo di Kolmogorov-Smirnov;
- calcolo dei limiti dell'IR tramite percentili 2,5° e 97,5° (41).

L'algoritmo di Arzideh è stato implementato in un'applicazione per Microsoft Excel, basata sul linguaggio R, e distribuito dalla Società Tedesca di Chimica Clinica e Medicina di Laboratorio (DGKL) (per ulteriori informazioni su questa applicazione si veda la sezione "Software disponibili" di questa rassegna). Tale metodo è stato inoltre utilizzato in un certo numero di lavori in letteratura (42-45) e recensito favorevolmente in un recente opinion paper del Committee on Reference Intervals and Decision Limits dell'IFCC (46). Il metodo di Arzideh non presenta alcuni dei limiti propri dei metodi descritti in precedenza: non assume una distribuzione gaussiana per la popolazione sana e non richiede l'identificazione visiva soggettiva della porzione non contaminata o pura di tale componente. Tuttavia, vanno sottolineate alcune criticità: non è completamente automatizzato (il metodo è implementato in un ambiente misto Excel-Linguaggio R che richiede un intervento umano) e l'algoritmo definitivo non è stato descritto completamente.

Il più recente algoritmo kosmik (47) rappresenta un'evoluzione del metodo di Arzideh (che risulta peraltro tra gli stessi ideatori). Il metodo assume che la componente sana di una distribuzione mista possa essere modellata con una distribuzione gaussiana dopo trasformazione matematica di Box-Cox e che esista un certo intervallo troncato, o tronco, del dataset in cui la contaminazione con valori patologici è assente o praticamente trascurabile.

L'algoritmo minimizza la distanza di Kolmogorov-Smirnov (massima distanza tra due distribuzioni) tra una distribuzione normale  $F$  stimata (ottenuta variando  $\mu$  e  $\sigma$ ) e un intervallo tronco  $T$  della distribuzione osservata dopo trasformazione di Box-Cox. I parametri  $\mu$  e  $\sigma$  della distribuzione gaussiana stimata, il parametro  $\lambda$  per la trasformazione di Box-Cox e l'intervallo  $T$  sono tutti ottimizzati numericamente (47). La Figura 5 mostra l'elaborazione (istogramma della popolazione mista, curva gaussiana della componente sana e limiti dell'IR) ottenuta analizzando la popolazione mista simulata negli esempi precedenti con l'algoritmo di kosmik implementato nella pagina web <https://kosmik.diz.uk-erlangen.de/>. I parametri della componente gaussiana stimata e i limiti dell'IR sono risultati pari a: media=89,4, DS=7,7, limiti=83,6-118,0. Le due criticità principali dell'algoritmo kosmik sono: prestazioni peggiori con contaminazioni >20%, e tempo di computazione lungo, in particolare all'aumentare dei decimali.

Simile al metodo precedente, il metodo TMC (Truncated Minimum Chi-square) utilizza dati su scala intervallare e minimizza la distanza chi-quadro, entro un certo intervallo, tra la distribuzione stimata e quella delle frequenze osservate (48). Le assunzioni del metodo TMC sono simili a quelle degli altri metodi indiretti descritti in precedenza (48).

### Algoritmo refineR

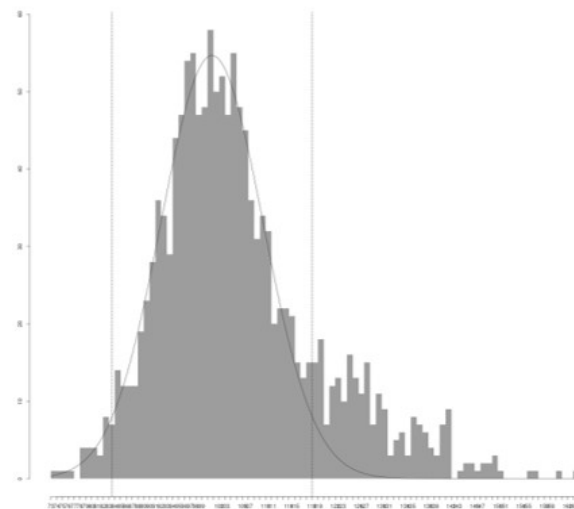
Metodo recentemente sviluppato per superare le citate criticità del metodo kosmik, il nuovo algoritmo refineR prevede una serie di fasi di descrizione e successiva modellazione dei dati:

- definizione del picco principale e selezione di una regione di dati utilizzata successivamente per la ricerca dei parametri ottimizzati di  $\mu$ ,  $\sigma$  e  $\lambda$  per la distribuzione gaussiana ottenuta mediante Box-Cox;
- le osservazioni entro la regione sono rappresentate con un istogramma;
- modellazione di una funzione parametrica, cioè di una distribuzione gaussiana ottenuta mediante Box-Cox facendo variare i parametri  $\mu$ ,  $\sigma$  e  $\lambda$ , per predire i valori attesi di ogni classe dell'istogramma della regione;
- calcolo di una funzione costo (discrepanza tra distribuzione modellata e osservata a livello della regione) e scelta dei parametri con i costi minimi (49).

La Figura 6 mostra l'analisi della stessa popolazione mista degli esempi precedenti ottenuta mediante algoritmo refineR. I limiti stimati dell'IR sono risultati pari a 84,1 e 119,0.

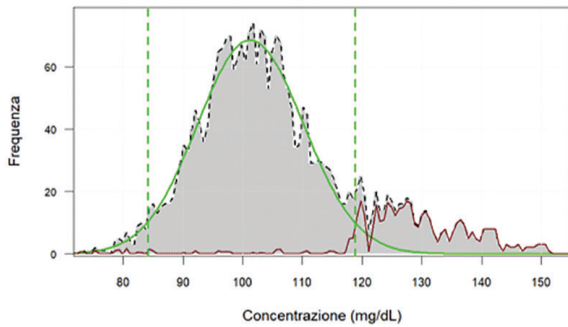
I vantaggi dell'algoritmo refineR, rispetto ai metodi precedenti, consistono in tempi di computazione relativamente contenuti, accurate prestazioni anche con popolazioni con contaminazioni >20% e disponibilità di un software open-source, relativamente facile da utilizzare. Si segnala, inoltre, la disponibilità di una guida pratica all'uso di tale algoritmo (50).

Recentemente è stato proposto uno strumento di benchmarking, implementato nel pacchetto R Rlbench,



**Figura 5**

Istogramma della popolazione mista e sovrapposizione della curva gaussiana relativa alla componente sana, calcolata tramite algoritmo kosmik implementato nell'applicazione web disponibile a <https://kosmik.diz.uk-erlangen.de/>. Le due linee verticali indicano i limiti dell'IR calcolati dall'algoritmo. La popolazione simulata è la stessa della Figura 2.



**Figura 6**  
*Analisi della popolazione mista tramite algoritmo refineR. In grigio ombreggiato la distribuzione di frequenza della popolazione mista. Le linee verde e rossa continue indicano, rispettivamente, la densità di frequenza della componente sana e patologica. Le linee verticali indicano i limiti dell'IR calcolato dai parametri della componente sana. La popolazione simulata è la stessa della Figura 2. Analisi ottenuta mediante pacchetto R refineR.*

per comparare le prestazioni dei principali metodi indiretti (51). Gli autori hanno evidenziato che i metodi indiretti moderni (Arzideh, TMC, cosmic e refineR), quando la frazione di osservazioni patologiche è <20% e le osservazioni sono >5 000, raggiungono prestazioni nella stima dell'IR comparabili o superiori a quelle dei metodi diretti (51).

La Tabella 4, per il parametro simulato in questa rassegna, riporta le stime dei limiti dell'IR ottenute con i differenti metodi indiretti presentati. Come si può osservare, con la popolazione mista utilizzata, in cui è stata simulata una contaminazione pari a circa il 28%, si sono ottenute delle stime vicine a quella teorica con tutti i metodi utilizzati. Gli ottimi risultati ottenuti, nonostante il livello di contaminazione non trascurabile, dipendono dal fatto che le due componenti simulate avevano una distribuzione gaussiana. In uno scenario reale, con distribuzioni non gaussiane, e un livello di contaminazione anche maggiore, è possibile che alcuni di questi algoritmi presentino prestazioni peggiori.

Complessivamente, alla luce della letteratura disponibile, i recenti metodi indiretti (EM, Arzideh, TMC,

**Tabella 4**  
*Stime dei limiti dell'IR del parametro simulato ottenute mediante i differenti metodi indiretti. L'unità di misura per i dati simulati è mg/dL.*

Metodo indiretto	Limite inferiore	Limite superiore
Teorico (simulato)	84,3	115,7
Hoffmann	84,6	121,1
Bhattacharya	83,6	119,3
EM	83,9	116,4
cosmic	83,6	118
refineR	84,1	119

IR, intervallo di riferimento; EM, Expectation-Maximization.

cosmic e refineR) presentano prestazioni migliori rispetto agli algoritmi proposti in precedenza (Pryce, Hoffmann, Bhattacharya). A conclusione di questa parte, un riassunto dei principali vantaggi e svantaggi dei metodi indiretti presentati in questa rassegna è riportato nella Tabella 5.

**Software disponibili**

Dai primi metodi indiretti con carta e penna e linee tracciate "a occhio", si hanno ora a disposizione metodi indiretti moderni in cui i parametri di ottimizzazione sono ottenuti per forza bruta (testando molte combinazioni di parametri). Questo procedimento necessita obbligatoriamente di macchine ad alta capacità computazionale e software specifici. Infatti, la disponibilità di software gratuiti e/o open-source è presupposto fondamentale sia per la verifica del metodo da parte della comunità scientifica, sia della diffusione dello stesso tra i professionisti. Ad oggi sono disponibili numerosi software per l'esecuzione dei differenti algoritmi. Essi differiscono per la complessità e per le conoscenze o abilità necessarie per la loro esecuzione. La Tabella 6 riporta un elenco non esaustivo dei principali software, con il relativo algoritmo implementato e una breve descrizione. Si può notare che nella maggior parte dei casi gli algoritmi sono stati implementati nel linguaggio R. In alcuni casi non serve alcuna conoscenza di R perché l'algoritmo accetta i dati in ingresso tramite interfaccia grafica e l'applicazione non necessita di installazione in locale; in altri vi è sempre un'interfaccia ma l'utente deve installare il linguaggio R e questo potrebbe rappresentare un ostacolo per l'utilizzo dell'applicativo; infine, per pacchetti come mixtools o refineR è necessario saper programmare in linguaggio R. I software riportati sono tutti gratuiti.

**Definizione *ex-novo* versus verifica dell'Intervallo di Riferimento**

Quando si contrappongono i metodi diretti a quelli indiretti ci si riferisce generalmente allo scenario della definizione *ex-novo* di IR. Questo specifico utilizzo dei metodi indiretti, pur presentando degli indubbi e non trascurabili vantaggi su quelli diretti (Tabelle 1-2), presenta tuttavia alcune importanti criticità che è bene sempre considerare. In particolare, se la numerosità non è molto elevata, se la contaminazione è significativa (>30%) e/o se la distribuzione della componente sana non è gaussiana, la stima dei limiti dell'IR può non essere accurata. Scenari associati a queste criticità includono, ad esempio, la definizione di IR di analiti misurati per esami specialistici (ad esempio ormoni) con dataset di dimensione ridotta e molto contaminati (e difficile da ripulire per mancanza di informazioni), oppure la definizione di IR di analiti recenti con dataset limitati, poche informazioni disponibili per selezionare eventuali contaminanti e pochi dati di letteratura con cui confrontare l'IR determinato.

In realtà, un'applicazione importante dei metodi indiretti in Medicina di Laboratorio potrebbe essere, non tanto la definizione *ex-novo*, quanto piuttosto la

**Tabella 5**

*Vantaggi e svantaggi dei principali metodi indiretti presentati in questa rassegna con i relativi riferimenti bibliografici*

Metodo indiretto (rif)	Vantaggi	Svantaggi
Pryce (20)	- semplicità di calcolo	- scarsa accuratezza in presenza di elevata contaminazione o quando la sovrapposizione delle distribuzioni patologica e normale è considerevole
Hoffmann (25)	- applicazione relativamente semplice	- richiede una distribuzione gaussiana o quasi-gaussiana dei dati (superato in alcune modifiche del metodo) - scelta arbitraria della retta lungo la porzione lineare del grafico (superato in alcune modifiche del metodo) - tollera un livello di contaminazione con risultati patologici inferiore a quello dei metodi più recenti
Bhattacharya (31)	- applicazione relativamente semplice	- soggettività della scelta delle classi e della retta di regressione - scarsa accuratezza in presenza di distribuzioni non gaussiane o molto contaminate
EM (36,37)	- versatilità (utilizzabile sia per distribuzioni parametriche che non-parametriche)	- richiede conoscenze teoriche e informatiche
Arzideh (41)	- non richiede una distribuzione gaussiana - popolazione sana identificata tramite criteri statistici non soggettivi	- procedura non completamente automatizzata - algoritmo definitivo non descritto completamente
kosmik (47)	- non richiede una distribuzione gaussiana - popolazione sana identificata tramite criteri statistici non soggettivi - disponibilità di un'applicazione web	- prestazioni peggiori in caso di contaminazione >20% - tempo di computazione lungo
refineR (49)	- non richiede una distribuzione gaussiana - popolazione sana identificata tramite criteri statistici non soggettivi - tempi di calcolo contenuti - prestazioni accurate anche per contaminazioni >20% - disponibilità di un software open-source	- necessità di conoscenze informatiche per l'utilizzo del software (comunque relativamente semplice rispetto ad altri applicativi)

verifica/validazione degli IR. Rispetto ai metodi diretti, un vantaggio non trascurabile sarebbe, ad esempio, la mancata necessità di utilizzare soggetti di riferimento (oltre ovviamente ai costi economici ridotti o assenti). Un secondo vantaggio importante sarebbe la possibilità di rivalutare periodicamente l'IR usato dal laboratorio. Non sono molti i lavori in letteratura relativi a possibili strategie per la verifica degli IR mediante metodi indiretti (5,15,46,52-53). Di seguito sono indicate alcune possibilità.

Le assunzioni sono le stesse utilizzate per la definizione *ex-novo* e, in particolare, che la maggior parte delle osservazioni siano normali, che la distribuzione della componente sana debba essere chiaramente identificata e che sia possibile applicare metodi di pre-selezione per eliminare la maggior parte delle osservazioni

patologiche. Ottenuti i limiti dell'IR (e i loro intervalli di confidenza), questo può essere utilizzato per verificare l'IR da trasferire, seguendo uno o più di questi approcci (alcuni sono più robusti di altri):

- confrontare i punti centrali (ad esempio le mediane) dei due IR (IR da trasferire e IR determinato con il metodo indiretto): questo semplice metodo consente di verificare eventuali bias dovuti a differenze di popolazioni o metodi analitici;
- comparare direttamente i due IR mediante: valutazione soggettiva, comparazione della differenza percentuale tra i limiti dei due intervalli con la differenza critica basata sulla variabilità analitica e biologica (54), verificare che i limiti dell'IR candidato da trasferire siano compresi negli intervalli di confidenza dei due limiti dell'IR ottenuto con il metodo indiretto.

**Tabella 6**

Elenco non esaustivo dei principali software disponibili per l'esecuzione dei metodi indiretti con i relativi riferimenti bibliografici

Algoritmo (rif)	Software	Descrizione
Bhattacharya (31)	Bellview	Applicazione Java ( <a href="https://sourceforge.net/projects/bellview/">https://sourceforge.net/projects/bellview/</a> )
	Bhattacharya Spreadsheet	Applicazione MS Excel ( <a href="http://www.syddpath.stvincents.com.au/">http://www.syddpath.stvincents.com.au/</a> )
EM (36,37)	Pacchetto R mixtools	Pacchetto aggiuntivo del Linguaggio R. Richiede R per l'utilizzo ( <a href="https://cran.r-project.org/web/packages/mixtools/">https://cran.r-project.org/web/packages/mixtools/</a> ) Nota: numerosi altri pacchetti sono disponibili per questa famiglia di algoritmi (flexmix, mixmod, mixdist) (19,51)
Arzideh (41)	Reference Limit Estimator	Script R con interfaccia in MS Excel. Distribuito dalla Società Tedesca di Chimica Clinica e Medicina di Laboratorio (DGKL). Richiede R per l'utilizzo. ( <a href="https://www.dgkl.de/en/activities/sections/entscheidungsgrenzenrichtwerte">https://www.dgkl.de/en/activities/sections/entscheidungsgrenzenrichtwerte</a> )
TMC (48)	TMC13	Script R con programma di installazione. Richiede R per l'utilizzo ( <a href="http://www.math.uni-bremen.de/~c05c/TMC13_public_2022-11-15/">http://www.math.uni-bremen.de/~c05c/TMC13_public_2022-11-15/</a> )
kosmic (47)	kosmicGUI	Applicazione web scritta in linguaggio R ( <a href="https://kosmic.diz.uk-erlangen.de/">https://kosmic.diz.uk-erlangen.de/</a> )
refineR (49)	Pacchetto R refineR	Pacchetto aggiuntivo del Linguaggio R. Richiede R per l'utilizzo ( <a href="https://cran.r-project.org/web/packages/refineR/">https://cran.r-project.org/web/packages/refineR/</a> )

## CONCLUSIONI

Schematicamente, i concetti essenziali relativi all'utilizzo dei metodi indiretti sono raggruppati di seguito.

- I metodi indiretti rappresentano uno strumento importante per la definizione *ex-novo* e, in particolare, per la verifica di IR candidati e/o da trasferire.
- Rispetto all'approccio diretto, i metodi indiretti presentano numerosi vantaggi, tra cui costi ridotti, condizioni pre-analitiche più vicine a quelle della routine di un laboratorio, possibilità di utilizzo in popolazioni pediatriche o anziane.
- È fondamentale comprendere, tuttavia, che non è possibile ottenere stime accurate per i limiti di un IR semplicemente utilizzando un algoritmo, fosse anche molto performante, e un dataset grezzo estratto dal gestionale del laboratorio. Come anche sottolineato nelle raccomandazioni del C-RIDL, i metodi indiretti non possono ridursi ad un mero "gioco statistico" (46). Come è stato evidenziato in questo lavoro, è fondamentale utilizzare numerose strategie per selezionare e ripulire le osservazioni probabilmente patologiche, al fine di ridurre il più possibile la componente non sana. In questa operazione svolgeranno un ruolo sempre più importante i big data e gli algoritmi di machine learning: i primi perché consentiranno di collegare dataset e quindi informazioni tra di loro fondamentali per aumentare la probabilità di individuare le osservazioni patologiche;

i secondi perché la ML può consentire di individuare, mediante associazioni statisticamente significative, quegli esami ematici o quelle condizioni associate a risultati anomali e quindi criteri di esclusione non noti o immediatamente non riconoscibili. Inoltre, è necessario sottolineare che gli algoritmi potrebbero essere meno accurati e performanti con analisi specialistiche, a causa della ridotta numerosità del dataset e della maggiore contaminazione con risultati patologici.

- Sono attualmente disponibili interfacce web e guide pratiche che consentono l'applicazione di alcuni metodi indiretti anche ai non esperti e senza dover obbligatoriamente conoscere i dettagli matematici degli algoritmi; tuttavia, la maggior parte dei pacchetti disponibili richiede conoscenze approfondite di statistica e di programmazione, in particolare del linguaggio R.
- Per la riproducibilità dei risultati è importante riportare i limiti dell'IR ottenuto con l'approccio indiretto unitamente ad una descrizione dettagliata di tutte le fasi seguite (selezione dataset, pretrattamento dati, gestione outliers, algoritmo applicato, dati componente gaussiana, calcolo limiti di confidenza, software utilizzato).

## CONFLITTO DI INTERESSE

Nessuno

## BIBLIOGRAFIA

1. Ceriotti F. Gli intervalli di riferimento nel nuovo millennio. *Bioch Clin* 2007;31:255-66.
2. Gräsbeck R, Saris NE. Establishment and use of normal values. *Scand J Clin Lab Invest* 1969;26 Suppl 110:62-3.
3. Galen RS, Gambino SR. Beyond normality: the predictive value and efficiency of medical diagnoses. New York: John Wiley and Sons, 1975.
4. ISO15189:2022 Medical laboratories: Requirements for quality and competence. 2022.
5. Ozarda Y, Higgins V, Adeli K. Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations. *Clin Chem Lab Med* 2019;57:30-7.
6. Clinical Laboratory Standards Institute (CLSI). Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline, 3rd ed. CLSI document EP28-A3c. Wayne, PA; 2010.
7. Clinical Laboratory Standards Institute (CLSI). Verification of reference intervals in the medical laboratory implementation guide, 1st ed. CLSI document EP28IG. Wayne, PA; 2022.
8. Giavarina D. Gli intervalli di riferimento. *RIMeL/IJLaM* 2006;2:50-6.
9. Sikaris KA. Separating disease and health for indirect reference intervals. *J Lab Med* 2021;45:55-68.
10. Farrell CL, Nguyen L. Indirect reference intervals: harnessing the power of stored laboratory data. *Clin Biochem Rev* 2019;40:99-111.
11. Arzideh F, Özcürümez M, Albers E, Haeckel R, Streichert T. Indirect estimation of reference intervals using first or last results and results from patients without repeated measurements. *J Lab Med* 2021;45:103-9.
12. Grossi E, Colombo R, Cavuto S, Franzini C. The REALAB project: a new method for the formulation of reference intervals based on current data. *Clin Chem* 2005;51:1232-40.
13. Poole S, Schroeder LF, Shah N. An unsupervised learning method to identify reference intervals from a clinical database. *J Biomed Inform* 2016;59:276-84.
14. Martinez-Sanchez L, Marques-Garcia F, Ozarda Y, Blanco A, Brouwer N, Canalias F, et al. Big data and reference intervals: rationale, current practices, harmonization and standardization prerequisites and future perspectives of indirect determination of reference intervals using routine data. *Adv Lab Med* 2021;2:9-16.
15. Yang D, Su Z, Zhao M. Big data and reference intervals. *Clin Chim Acta* 2022;527:23-32.
16. Solberg HE, Lahti A. Detection of outliers in reference distributions: Performance of Horns algorithm. *Clin Chem* 2005;51:2326-32.
17. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc* 1964;B26:211-52.
18. Bunch DR. Indirect reference intervals using an R pipeline. *J Mass Spectrom Adv Clin Lab* 2022;24:22-30.
19. Haeckel R, Wosniok W, Streichert T, Members of the Section Guide Limits of the DGKL. Review of potentials and limitations of indirect approaches for estimating reference limits/intervals of quantitative procedures in laboratory medicine. *J Lab Med* 2021;45:35-53.
20. Pryce JD. Level of haemoglobin in whole blood and red blood-cells, and proposed convention for defining normality. *Lancet* 1960;2:333-6.
21. Becktel JM. Simplified estimation of normal ranges from routine laboratory data. *Clin Chim Acta* 1970;28:119-25.
22. Kairisto V, Poola A. Software for illustrative presentation of basic clinical characteristics of laboratory tests – GraphROC for windows. *Scand J Clin Lab Invest* 1995;55:43-60.
23. Kouri T, Kairisto V, Virtanen A, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem* 1994;40:2209-15.
24. Kairisto V, Kouri T, Virtanen A, Uusipaikka E, Koivula T, Nantö. Estimation of reference change limits using patient data. *Scand J Clin Lab Invest Suppl* 1995;222:37-41.
25. Hoffmann RG. Statistics in the practice of medicine. *JAMA* 1963;185:864-73.
26. Zhang Y, Ma W, Wang G, Lv Y, Peng Y, Peng X. Limitations of the Hoffmann method for establishing reference intervals using clinical laboratory data. *Clin Biochem* 2019;63:79-84.
27. Holmes DT, Buhr KA. Widespread incorrect implementation of the Hoffmann method, the correct approach, and modern alternatives. *Am J Clin Pathol* 2019;151:328-36.
28. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol* 2010;133:180-6.
29. Katayev A, Fleming JK, Luo D, Fisher AH, Sharp TM. Reference intervals data mining: no longer a probability paper method. *Am J Clin Pathol* 2015;143:134-42.
30. Jones G, Horowitz G, Katayev A, Fleming JK, Luo D, Fisher AH, et al. Reference intervals data mining: getting the right paper. *Am J Clin Pathol* 2015;144:526-7.
31. Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. *J Biometric Soc* 1967;23:115-35.
32. Naus AJ, Borst A, Kuppens PS. The use of patient data for the calculation of reference values for some haematological parameters. *J Clin Chem Clin Biochem* 1980;18:621-5.
33. Hemel JB, Hindriks FR, van der Slik W. Critical discussion on a method for derivation of reference limits in clinical chemistry from a patient population. *J Automat Chem* 1985;7:20-30.
34. Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. *J Clin Chem Clin Biochem* 1985;23:829-39.
35. Oosterhuis WP, Modderman TA, Pronk C. Reference values: Bhattacharya or the method proposed by the IFCC?. *Ann Clin Biochem* 1990;27:359-65.
36. Redner RA, Walker HF. Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review* 1984;26:195-239.
37. Concordet D, Geffré A, Braun JP, Trumel C. A new approach for the determination of reference intervals from hospital-based data. *Clin Chim Acta* 2009;405:43-8.
38. Macdonald P, Juan DU. Mixdist: Finite Mixture Distribution Models. <https://cran.r-project.org/web/packages/mixdist/mixdist.pdf>. (ultimo accesso: marzo 2023).
39. Benaglia T, Chauveau D, Hunter DR, Young DS. mixtools: An R Package for Analyzing Mixture Models. *J Stat Soft* 2009;32:1-29.
40. Lee SX, McLachlan GJ. EMMIXskew: An R Package for the Fitting of a Mixture of Canonical Fundamental Skew t-Distributions. *J Stat Soft (Internet)* 2018;83:1-32.
41. Arzideh F, Wosniok W, Gurr E, Hinsch W, Schumann G, Weinstock N, et al. A plea for intra-laboratory reference limits. Part 2. A bimodal retrospective concept for determining reference limits from intra-laboratory databases demonstrated by catalytic activity concentrations of enzymes. *Clin Chem Lab Med* 2007;45:1043-57.
42. Arzideh F, Wosniok W, Haeckel R. Reference limits of plasma and serum creatinine concentrations from intra-laboratory data bases of several German and Italian

- medical centres: Comparison between direct and indirect procedures. *Clin Chim Acta* 2010;411:215-21.
43. Arzideh F, Brandhorst G, Gurr E, Hinsch W, Hoff T, Roggenbuck L, et al. An improved indirect approach for determining reference limits from intra-laboratory data bases exemplified by concentrations of electrolytes. *J Lab Med* 2009;33:52-66.
  44. Lo Sasso B, Vidali M, Scazzone C, Agnello L, Ciaccio M. Reference interval by the indirect approach of serum thyrotropin (TSH) in a Mediterranean adult population and the association with age and gender. *Clin Chem Lab Med* 2019;57:1587-94.
  45. Bohn MK, Adeli K. Application of the TML method to big data analytics and reference interval harmonization. *J Lab Med* 2021;45:79-85.
  46. Jones GR, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. IFCC Committee on Reference Intervals and Decision Limits. Indirect methods for reference interval determination – review and recommendations. *Clin Chem Lab Med* 2018;57:20-9.
  47. Zierk J, Arzideh F, Kapsner LA, Prokosch HU, Metzler M, Rauh M. Reference interval estimation from mixed distributions using Truncation Points and the Kolmogorov-Smirnov Distance (kosmic). *Sci Rep* 2020;10:1704.
  48. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. *Clin Chem Lab Med* 2019;57:1933-47.
  49. Ammer T, Schützenmeister A, Prokosch HU, Rauh M, Rank CM, Zierk J. refineR: a novel algorithm for reference interval estimation from real-world data. *Sci Rep* 2021;11:16023.
  50. Ammer T, Schützenmeister A, Rank CM, Doyle K. Estimation of reference intervals from routine data using the refiner algorithm—a practical guide. *J Appl Lab Med* 2023;8:84-91.
  51. Ammer T, Schützenmeister A, Prokosch HU, Zierk J, Rank CM, Rauh M. Rlbench: a proposed benchmark for the standardized evaluation of indirect methods for reference interval estimation. *Clin Chem* 2022;68:1410-24.
  52. Jones GR. Validating common reference intervals in routine laboratories. *Clin Chim Acta* 2014;432:119-21.
  53. Özcürümez M, Haeckel R, Gurr E, Streichert T, Sack U. Determination and verification of reference interval limits in clinical chemistry. Recommendations for laboratories on behalf of the working group guide limits of the dgkl with respect to ISO Standard 15189 and the guideline of the german medical association on quality assurance in medical laboratory examinations (Rili-BAEK). *J Lab Med* 2019;43:127-33.
  54. Yildiz Z, Dağdelen LK. Reference intervals for thyroid disorders calculated by indirect method and comparison with reference change values. *Biochem Med (Zagreb)*. 2023;33:010704.