© 2026 EDIZIONI MINERVA MEDICA
Versione online su https://www.minervamedica.it

La Rivista Italiana della Medicina di Laboratorio 2026 Sep 11
DOI: 10.23736/S1825-859X.26.00349-X

RACCOMANDAZIONI E LINEE GUIDA

Intervalli di riferimento e limiti decisionali nelle raccomandazioni SIPMeL per accreditamento ISO 15189 del laboratorio medico

Reference intervals and decision limits in the SIPMeL recommendations for ISO 15189 Accreditation of Medical Laboratories

Marco PRADELLA *

Società Italiana di Patologia Clinica e Medicina di Laboratorio, Commissione Nazionale Qualità e Accreditamento, Gruppo di Studio Informatica, Castelfranco Veneto, Treviso, Italia

*Autore di contatto: Marco Pradella, Piazza Duomo 18, 45026 Lendinara (RO), Italia. E-mail: labmedico@labmedico.it

RIASSUNTO

Il lavoro dei laboratori non si esaurisce con i risultati delle misurazioni o degli esami. Tuttavia, esistono differenze significative tra i laboratori medici e i laboratori di prova per quanto riguarda la gestione dei risultati. I laboratori di prova si sono sempre trovati a dover interpretare i risultati delle misurazioni per prendere decisioni, in relazione ai destinatari dei rapporti di prova, sulla base di valori soglia e dell'incertezza di misura. Nei laboratori medici, l'approccio è piuttosto diverso. Per i requisiti di accreditamento ISO 15189, ai destinatari dei risultati delle prove vengono forniti criteri interpretativi basati su intervalli di riferimento, limiti decisionali e differenze rispetto ai risultati precedenti. A tal fine sono disponibili linee guida in costante miglioramento. Tuttavia, emergono punti critici di cui i laboratori devono tenere conto, che riguardano sia aspetti formali che di contenuto. Alcune di queste questioni critiche sono state evidenziate nelle raccomandazioni ufficiali del SIPMeL. La ricerca scientifica ha fornito ulteriori contributi anche dopo la pubblicazione delle Raccomandazioni SIPMeL e questa nota ne raccoglie alcuni significativi, per esaminarli criticamente. La revisione dell'utilizzo degli intervalli di riferimento secondo le indicazioni attuali di standard e linee guida può modificare molto la pratica del laboratorio medico e favorire l'inquadramento più corretto delle prestazioni cliniche degli esami.

(*Per citare questo articolo:* Pradella M. Intervalli di riferimento e limiti decisionali nelle raccomandazioni SIPMeL per accreditamento ISO 15189 del laboratorio medico. Riv Ital Med Lab 2026 Sep 11. DOI: 10.23736/S1825-859X.26.00349-X)

ABSTRACT

The role of laboratories does not end with the results of measurements or tests. However, there are significant differences between medical laboratories and testing laboratories in how they handle results. Testing laboratories have always been confronted with the interpretation of measurement results to make decisions, in relation to the intended users of test reports, based on threshold values and measurement uncertainty. In medical laboratories, the approach is quite different. For ISO 15189 accreditation requirements recipients of test results are given interpretive criteria provided by reference intervals, decision limits and differences from previous results. Constantly improving guidelines are available for this. However, critical points emerge that laboratories must take into account, involving both formal and content aspects. Some of these critical issues have been highlighted in the official SIPMeL recommendations. Scientific research continued to contribute new insights even after the publication of the SIPMeL Recommendations, and this note compiles some of the most significant ones for critical analysis. Revising the use of reference intervals in accordance with current standards and guidelines can significantly change medical laboratory practice and promote a more accurate assessment of the clinical performance of examinations.

Key words: ISO 15189; ISO 17025; Reference intervals; Clinical decision limits; CLSI.

Introduzione: il Progetto M2 parte seconda, risultato dell'esame e regole decisionali

Grazie al lavoro congiunto dei Gruppi di Studio SIPMeL "Evidence Based Laboratory Medicine"[1], "Normativa sanitaria" (già "Management sanitario")[2], "Informatica"[3] e la Commissione Nazionale per la Qualità[4] si sta sviluppando il progetto "M2 Raccomandazioni per Accreditamento ISO 15189 del laboratorio medico: requisiti per le relazioni con pazienti e medici curanti". M2 fa seguito alle Raccomandazioni M1 ("requisiti per organizzazione e direzione"),[5-7] che partivano dai capitoli 5 (requisiti di organizzazione e direzione) e 8 (requisiti del sistema di gestione) dello standard ISO 15189.[8]

Lavorando alle Raccomandazioni M2, è emersa l'opportunità di dividere la materia in almeno tre parti: la prima, partendo dai requisiti ISO 15189 del capitolo 4 (imparzialità, riservatezza, garanzie dei pazienti), la seconda dalla clausola 7.3.5 (Intervalli biologici di riferimento e limiti decisionali clinici), la terza dalle clausole 7.3.6 (Documentazione delle procedure di esame) e 7.4.1 (Presentazione dei risultati degli esami). Quindi la parte seconda è stata completata e approvata dal Consiglio Nazionale il 5 ottobre 2025[1] ed è stata successivamente messa a disposizione dell'intera comunità scientifica, nazionale e internazionale.[2] Nella Tabella I sono estratte alcune Raccomandazioni del documento SIPMeL M2 parte seconda.

Decidere in base ai risultati nei laboratori di prova secondo ISO 17025

La versione vigente di ISO 15189 è stata confezionata partendo da ISO 17025, come prescritto dai vertici ISO.[3]

Sia i laboratori di prova accreditati secondo la norma ISO 17025 che i laboratori medici accreditati secondo la norma ISO 15189 devono stimare l'incertezza di misura e fornire criteri per l'interpretazione dei risultati delle misurazioni. Ma devono agire in modi diametralmente opposti.[4] ISO richiede ai laboratori di prova e di taratura di fornire, insieme al risultato, un'indicazione dell'incertezza — un dubbio (secondo la nuova definizione[5]) — sulla base della quale occorre comunque prendere una decisione.[6] Al contrario, ISO richiede ai laboratori medici di fornire a medici e pazienti una convinzione di sicurezza — una base sufficientemente solida su cui fondare le decisioni cliniche, procedere con gli accertamenti o le cure e andare avanti[7] (Tabella II). Non ci si deve stupire, perciò, se nella realtà si vedano comportamenti non sempre coerenti, perché le differenze tra i due tipi di laboratorio non sono sempre così nette. La materia non è così semplice come sembra, confrontare i due approcci aiuta a comprendere il contesto in cui si colloca il tema.

I laboratori di prova per ottenere l'accreditamento ISO 17025, devono definire regole decisionali per i risultati di ciascuna prova, comunicare tali regole ai clienti che richiedono una dichiarazione di conformità, documentare il livello di rischio considerato per ciascuna regola e applicare la regola anche per la taratura degli strumenti. La fonte normativa si trova nella Guida ISO 98, Parte 4,[8] dove compaiono le definizioni di regole decisionali, intervalli di copertura, intervalli di tolleranza, intervalli di accettazione e bande di sicurezza. Altri riferimenti sul tema sono il documento *International Laboratory Accreditation Cooperation* ILAC G8[9] e quello EURACHEM.[10] In estrema sintesi, secondo le guide la regola decisionale deriva dall'incertezza di misura, il concetto di «tolleranza» differisce da quello utilizzato in statistica (purtroppo) e il «margine di sicurezza» consiste in un intervallo aggiunto all'intervallo di accettazione quando il rischio di falsi rifiuti è considerato elevato (Figura 1).

EURACHEM utilizza il termine «specifica» come sinonimo di «tolleranza», distingue tra "accettazione semplice", accettazione con una "banda di salvaguardia", accettazione "condizionata" o "inconcludente" (che ILAC definisce "non binaria") e, infine, accettazione "in due fasi". La "fascia di sicurezza" è tipicamente un multiplo dell'incertezza, scelto in base alla probabilità del rischio

Tabella I.—*Estratto dalle Raccomandazioni SIPMeL M2 parte seconda.*[1]

5. intervalli di riferimento e limiti decisionali clinici devono restare distinti. Gli intervalli non sono collegabili a sensibilità e specificità
6. gli studi tradizionali e i metodi "verifica su larga scala" e "trasferimento mediante confronto", sono adatti per fabbricanti e grandi istituzioni.
7. il laboratorio medico può acquisire intervalli proposti da altri laboratori qualificati o da fabbricanti e verificarli con metodi riconosciuti.
8. ai laboratori con meno risorse si raccomanda il metodo binomiale su piccola scala.
9. ai laboratori medici che eseguono un volume elevato di esami e dispongono di strumenti informatici adeguati, si raccomandano i metodi "indiretti"
10. accompagnare i risultati degli esami preferenzialmente con limiti decisionali
11. curare l'armonizzazione dei risultati
12. La stima delle prestazioni cliniche degli esami richiede soglie o limiti decisionali
13. concordare con i medici curanti di riferimento le prestazioni diagnostiche (sensibilità, specificità, valori predittivi) dell'esame.

TABELLA II.—*Norme ISO relative alle decisioni basate sui risultati di misurazioni o analisi di laboratorio: un confronto tra la norma ISO 17025[6] e la norma ISO 15189.[7]*

ISO 17025	ISO 15189
3.7 Regola decisionale: ...descrive in che modo si tiene conto dell'incertezza di misura quando si dichiara la conformità a un requisito specificato	3.2 Intervallo di riferimento biologico: ... intervallo della distribuzione dei valori rilevati in una popolazione di riferimento biologica. ... comunemente definito come l'intervallo centrale al 95%. ... In alcuni casi, ... sarebbe inferiore o uguale a «x». Nota 4 alla voce: Termini quali «intervallo normale», «valori normali» e «intervallo clinico» sono ambigui e pertanto se ne sconsiglia l'uso.
7.1.3 ... una dichiarazione di conformità a una specifica o a una norma per la prova o la taratura (ad es. superato/non superato, entro tolleranza/fuori tolleranza); la specifica o la norma, ... la regola di decisione scelta deve essere comunicata al cliente e concordata con esso.	3.3 Limite decisionale clinico: risultato dell'esame (3.8) che indica un rischio maggiore di esiti clinici avversi o che è diagnostico per la presenza di una malattia specifica ...
7.8.6.1 Quando viene fornita una dichiarazione di conformità a una specifica o a una norma, il laboratorio deve documentare il criterio decisionale utilizzato, tenendo conto del livello di rischio (ad esempio, falsi positivi e falsi negativi e ipotesi statistiche)	7.3.5 ... I valori di riferimento biologici forniti dal produttore possono essere utilizzati dal laboratorio, purché sia verificata la popolazione su cui si basano tali valori ...
7.8.6.2 Il laboratorio deve indicare nella dichiarazione di conformità, ... c) il criterio decisionale applicato ...	7.4.1.6 Requisiti relativi alle relazioni: ... h) intervalli di riferimento biologici, limiti decisionali clinici, rapporti di verosimiglianza o diagrammi/nomogrammi a supporto della decisione clinica ...
A.2.3 ... — l'utilizzo di una regola decisionale adeguata per stabilire la conformità;	

La tabella B.2 della norma ISO 15189[7] (confronto tra la norma ISO/IEC 17025:2017 e il presente documento) può indurre in errore. Il punto 7.3.5 della norma ISO 15189 dovrebbe essere confrontato con i punti 3.7, 7.1.3, 7.8.6.1, 7.8.6.2 e A.2.3 della norma ISO 17025, che mancano nella tabella.

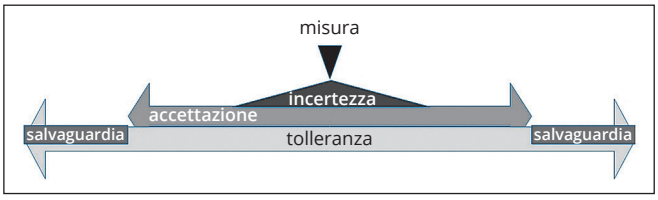


Figura 1.—Schema delle regole decisionali per i laboratori di prova per l'accreditamento ISO 17025, in conformità alla Guida ISO 98-4[8] a ILAC G8[9] e EURACHEM 2021.[10]

di accettare decisioni errate. Ad esempio, 1,64 per un rischio del 5%, 2,33 per un rischio dell'1%. Per valori di rischio pari a zero, si ha l'accettazione semplice o "rischio condiviso". Accettazione "condizionata" o "inconcludente" (che ILAC definisce "non binaria") si realizza quando l'incertezza si sovrappone ai limiti di accettazione e la decisione potrebbe essere formulata come "probabilmente fallita" o "probabilmente superata". L'accettazione "in due fasi" prevede una seconda misura se la prima rimane inconcludente.

ISO, EURACHEM e ILAC forniscono regole generali corrette dal punto di vista formale e probabilistico. Ma trascurano quasi completamente il peso di ciascun rischio e la sua rilevanza concreta — che sia di natura materiale, economica o etica, come nel caso degli esseri umani.

Intervalli di riferimento nei laboratori medici: requisiti e guide

Nei laboratori medici la situazione è diversa dai laboratori di prova ISO 17025.

Le implicazioni pratiche, etiche ed economiche del processo decisionale sono preoccupazioni quotidiane. La clausola 7.3.5 della norma ISO 15189:2022 è interamente dedicata agli intervalli di riferimento biologici (IR) e ai limiti di decisione clinica (LD). Tutti e due devono essere definiti e comunicati agli utenti e riflettere la popolazione di pazienti servita dal laboratorio, tenendo conto del rischio per i pazienti. Nella norma ISO 15189, gli intervalli non si limitano ai risultati quantitativi numerici, ma servono anche per i risultati qualitativi binari (come la presenza o l'assenza di un marcatore, ad esempio genetico). ISO prescrive che vengano verificati, prima del loro utilizzo, gli intervalli forniti dal produttore dei dispositivi, dei sistemi e dei reagenti, per confermare che la popolazione da cui derivano corrisponda a quella servita dal laboratorio. Intervalli e limiti devono essere rivisti periodicamente e, ogni volta che vengono apportate modifiche a un esame o a un metodo di raccolta dei campioni, gli utenti devono essere informati.

Coerentemente con le indicazioni ISO, lo standard ISO 15189 non dice come ottenere intervalli di riferimento. Ma i laboratori medici possono fare riferimento alle linee guida di *Clinical & Laboratory Standards Institute* (CLSI). Il documento CLSI EP44 (Stabilire intervalli di riferimento)[11] sostituisce il documento CLSI EP28-A3c del 2010.[9] EP44 distingue la definizione degli intervalli di riferimento dalla loro verifica, un processo trattato dal documento separato CLSI EP45.[12] La produzione di IR coinvolge ricercatori e sviluppatori (compresi sia i fabbricanti che i laboratori che sviluppano metodi in proprio), la verifica di IR, invece, è compito dei laboratori (Figura 2).

CLSI mantiene l'approccio tradizionale per la forma-

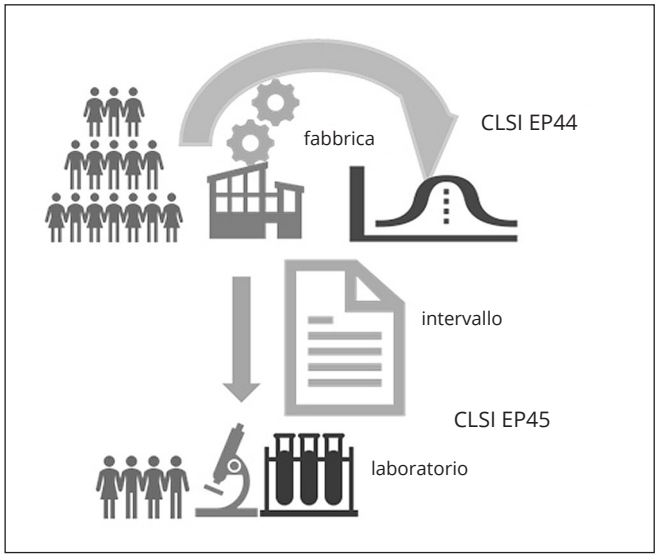


Figura 2.—Produzione e verifica di intervalli di riferimento secondo ISO 15189,[7] CLSI EP44[11] e CLSI EP45.[13]

zione degli intervalli: utilizza i risultati delle osservazioni, delle misurazioni o dei calcoli relativi a individui di riferimento, per poi calcolare l'intervallo centrale dei valori (ad esempio, l'intervallo centrale al 95%). Il CLSI descrive intervalli «associati alla salute», utilizzando metodi di campionamento «a priori» o «a posteriori». "A priori" vuol dire selezione degli individui prima della raccolta dei campioni, mentre "a posteriori" è la selezione dopo la raccolta dei campioni.

Secondo CLSI, gli intervalli possono anche essere continui, basati su covariate o univariati, con linee smussate — sia graficamente che mediante calcolo — per rappresentare la funzione continua, come rispetto all'età.

Un punto critico del processo è il numero di soggetti. Secondo il CLSI, uno studio con soli 120 soggetti (il metodo tradizionale) è piccolo. Sarebbe preferibile includere diverse centinaia di soggetti o anche di più. Tuttavia, per determinati esami o popolazioni specifiche (come quella pediatrica), il reclutamento può risultare molto difficile. Per CLSI anche gli studi di piccole dimensioni, basati ad esempio su 60 o 80 individui, possono comunque fornire indicazioni utili. Tuttavia, i percentili derivati da metodi non parametrici richiedono una quantità minima di dati, la rimozione dei valori anomali dai risultati richiede più dati, così come la suddivisione in sottogruppi, ad esempio per sesso o età. CLSI descrive partizionamenti che impiegano anche l'analisi di regressione multipla o l'analisi della varianza, ottenendo così fino a quindici fattori di suddivisione.

L'analisi statistica dei valori di riferimento interessa più i produttori che i laboratori, tuttavia possiamo menzionare in sintesi le indicazioni CLSI. CLSI distingue tra un metodo non parametrico (senza uso della distribuzione gaussiana), un metodo parametrico (basato sulla variabile normale o gaussiana) e un metodo robusto. I metodi parametrici quantile-quantile (PQQ) utilizzano la relazione tra i valori (o le loro trasformazioni) e i punteggi z della distribuzione gaussiana standard. Questi metodi non richiedono grandi competenze statistiche, ma solo semplici fogli di calcolo. Al contrario, per il metodo robusto è necessario il supporto di uno specialista biostatistico.

CLSI fornisce linee guida specifiche per la comunicazione degli intervalli dal fabbricante al laboratorio. In mancanza di ripartizioni può essere un unico intervallo, accompagnato da informazioni di base sullo studio condotto: strumento e metodo di analisi, numero e sede anatomica dei campioni, numero di soggetti, criteri di inclusione ed esclusione, preparazione dei soggetti (come lo stato di digiuno e la postura), ora del giorno in cui sono stati prelevati i campioni, proporzione dei soggetti per sesso, fascia d'età e razza e/o etnia. Si esprime l'intervallo centrale al 95%, così come l'intervallo al 99%, insieme agli intervalli di confidenza per i limiti dell'intervallo, mentre la mediana, il 25° percentile e il 75° percentile della distribuzione possono fornire informazioni sulla normalità o sull'asimmetria.

Le raccomandazioni SIPMeL M2-2 segnalano diversi limiti del metodo CLSI, sia per i laboratori che sviluppano metodi in casa (LDT), che per gli stessi produttori di strumenti e materiali di consumo. Fare ricerche originali e specifiche per ogni singolo esame richiede molto tempo, è incerto, laborioso e costoso. Richiede il reclutamento di un numero molto elevato di soggetti, che devono essere sottoposti a uno *screening* approfondito anche prima della raccolta dei campioni, nonché competenze statistiche altamente avanzate, oltre a competenze in materia di tecnologia di misurazione e medicina. La raccolta dei campioni richiede tempo e risorse umane dedicate e qualificate, con un costo significativo. Lo stesso vale per la conservazione, il trasporto e la manipolazione.

Il concetto di "buona salute" soffre di una ben nota incertezza di definizione. Il concetto di «razza» o «etnia» può essere inutile o addirittura dannoso.[14, 15] Distinguere i diversi gruppi in base alla razza/etnia, all'età, allo status sociale e così via è tutt'altro che facile. È difficile applicare criteri di selezione e classificazione, ma è anche difficile determinare a quale gruppo appartenga un singolo paziente sottoposto a esami di laboratorio (Figura 3).

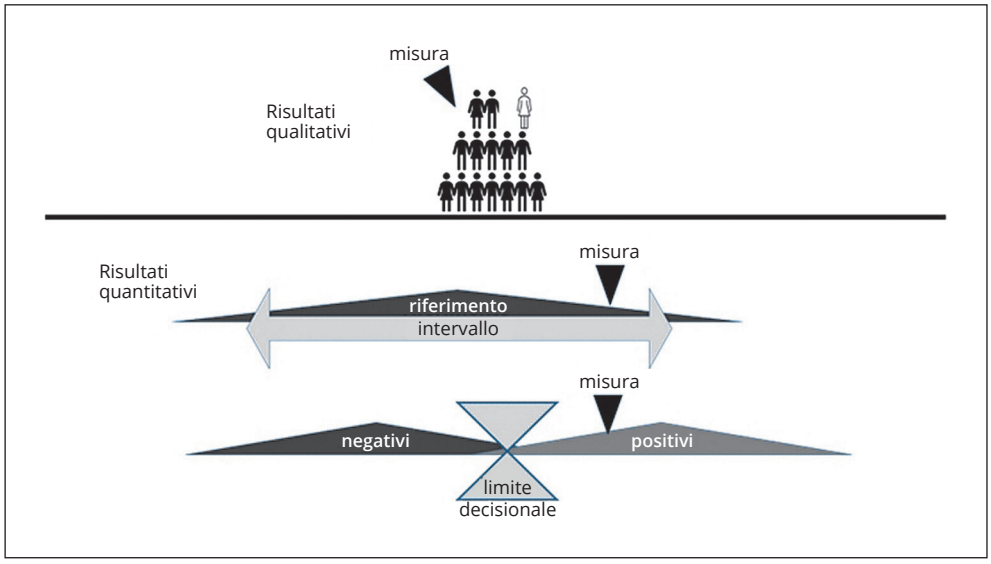


Figura 3.—Schema delle regole decisionali per risultati qualitativi e quantitativi nei laboratori medici con l'accreditamento ISO 15189.

Le statistiche di regressione richieste per il metodo "quantile-quantile" sono messe in discussione dallo stesso CLSI. L'approccio 'robusto' (intervallo predittivo "a doppia ponderazione") è raccomandato da alcune fonti,[16] ma criticato da altre,[17] che ne mettono in dubbio l'applicabilità nei laboratori medici tipici.

La Raccomandazione SIPMeL M2-2, sulla base della letteratura scientifica corrente, solleva dubbi sul fatto che la creazione di nuovi intervalli di riferimento rientri nelle capacità finanziarie e professionali dei laboratori e di molti produttori e, d'altra parte, sul fatto che tali intervalli siano realmente efficaci per classificare i risultati degli esami di laboratorio dei singoli pazienti. Gli intervalli, a prescindere dalla loro funzione di classificazione dei risultati in categorie, possono essere validi per descrivere il metodo utilizzato per l'esame, insieme al nome e alle unità di misura.

I risultati dello studio recente del gruppo di Colonia[18] sono particolarmente illuminanti. Mostrano che il profano che vuole interpretare il risultato di un esame nel 55% dei casi non considera gli intervalli di riferimento e nel 51% nemmeno le unità di misura. Il fenomeno è preoccupante soprattutto in relazione all'utilizzo crescente di strumenti interpretativi con intelligenza artificiale e alla diffusione di informazioni mediche in canali incontrollati, divulgativi, non scientifici e non professionali.

CLSI mira a distinguere chiaramente tra la definizione degli intervalli di riferimento e il loro utilizzo nel laboratorio clinico, suddividendoli in due linee guida distinte[11, 12] (Figura 4). La linea guida EP45[12] è destinata a sostituire la verifica degli intervalli di riferimento precedentemente inclusa nella linea guida EP28-A3c.[13] CLSI fornisce al laboratorio indicazioni sui vari studi sperimentali, sul numero di campioni da raccogliere e sulla valutazione dei risultati. Aggiunge la verifica degli intervalli mediante metodi indiretti. L'accettazione di un intervallo di riferimento proposto (ad esempio, dal produttore) richiede una comprensione della comparabilità dei metodi di analisi, della popolazione di pazienti e dei fattori pre-esame.

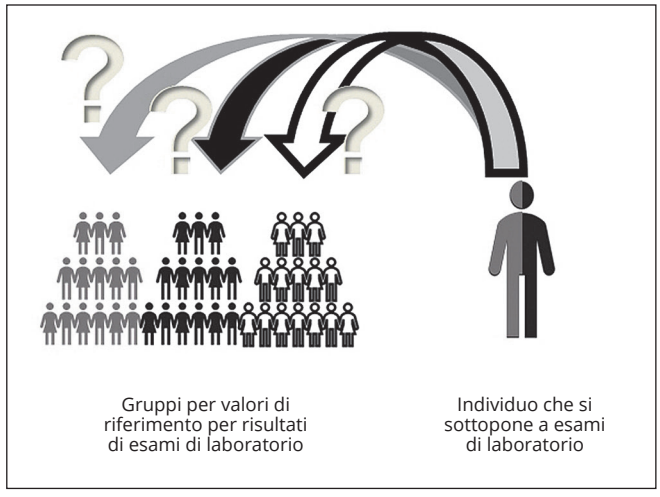


Figura 4.—Schema della difficoltà di assegnare un individuo a un gruppo specifico di riferimento.

CLSI propone tre tipi di protocolli sperimentali diretti: due sulla popolazione (Figura 5), di cui uno studio binomiale su piccola scala con pochi soggetti (solo 20) e uno su larga scala con un numero maggiore di soggetti (ad esem-

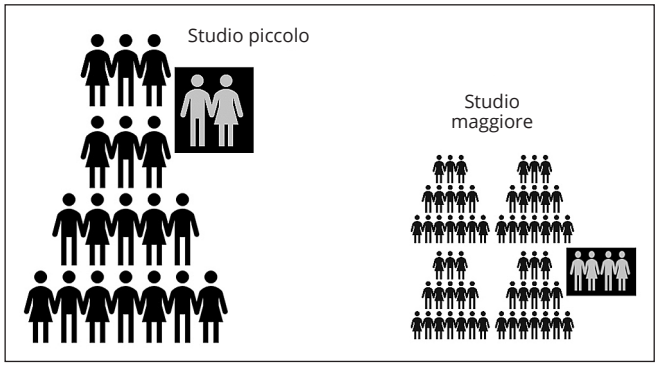


Figura 5.—Verifica di intervalli di riferimento sulla popolazione, approcci proposti da CLSI.[13]

pio, 60 o più, accanto a uno sul metodo di esame, tramite confronto e regressione. Tutti e tre presentano dei limiti. Con lo studio su piccola scala se la varianza nella popolazione di laboratorio fosse minore dell'originale avremmo l'accettazione di un intervallo eccessivamente ampio.

Lo studio su larga scala richiede più campioni e analisi statistica più raffinata. Con più dati si possono adottare diversi approcci per verificare un intervallo di riferimento. Ad esempio, una procedura euristica non parametrica può essere usata senza ricorrere ad analisi statistiche avanzate.

Per una dimensione del campione pari a 60, l'intervallo centrale al 95% può essere stimato utilizzando la 2ª e la 59ª osservazione in ordine di rango. In alternativa, l'analisi parametrica richiede metodologie statistiche più avanzate: si basa su dati distribuiti normalmente o su dati trasformati in modo da seguire una distribuzione normale o Gaussiana. È possibile utilizzare diverse metodologie per valutare la normalità di un insieme di dati alla distribuzione normale. Se i dati sono distribuiti normalmente, le stime di media, varianza e scarto tipo della popolazione possono essere utilizzate per calcolare gli intervalli di riferimento.

Infine, il confronto tra i metodi ignora le possibili differenze tra la popolazione di laboratorio e quella originale. La soluzione presente di EP28 chiamata da Westgard "giudizio divino", e fortemente criticata,[19] viene abbandonata dal nuovo EP45.

CLSI sostiene i metodi indiretti. Si tratta di metodi che utilizzano risultati di laboratorio ottenuti in precedenza e destinati a uno scopo diverso (come il controllo periodico, la diagnosi e il monitoraggio terapeutico).[20-26] Le difficoltà sono l'inclusione di popolazioni eterogenee, la mancanza di informazioni cliniche, la mancanza di dati storici, la complessità dell'analisi dei dati e la mancanza di dati "normali" per esami particolari come quelli ormonali.[27]

Conosciamo oggi due generazioni di metodi indiretti: quelli prima generazione (Hoffmann[28] e Bhattacharya[29]) e quelli di seconda generazione, tra cui il massimo di verosimiglianza troncato di Kosmic,[30] il chi-quadrato minimo troncato[31] e refineR.[32]

Nelle raccomandazioni SIPMeL si osserva che le linee guida CLSI sugli intervalli di riferimento sono disponibili da molto tempo e sono chiare, ma la pratica quotidiana dei laboratori sembra purtroppo essere andata in una direzione diversa. In molti casi, i laboratori clinici utilizzano gli intervalli di riferimento forniti dal produttore nelle istruzioni per l'uso, senza curarne bene la verifica. Sembra che gli intervalli forniti dal fabbricante siano preferiti dagli enti di accreditamento e di autorizzazione, ma è noto che ottenere prove della loro verifica in laboratorio è molto difficile. Per alcuni la verifica in laboratorio con studio su piccola scala presenta importanti limitazioni.[33]

Le linee guida CLSI presentano diverse debolezze. Manca una riflessione critica sulla fattibilità di studi onerosi e concettualmente complessi. Si attribuisce un ruolo forse eccessivo al confronto tra i metodi: non si considerano difficoltà e costi di tali studi,[34, 35] che dovrebbero vedere l'interazione diretta tra singolo laboratorio e un particolare fabbricante, di cui a oggi non si ha notizia.

I metodi indiretti sono probabilmente ancora sottoutilizzati e poco valorizzati anche da CLSI. Al contrario, esiste una ricca letteratura scientifica sull'argomento. Essi sono adatti, come minimo, ai laboratori di grandi dimensioni, per confermare o adeguare gli intervalli attualmente in uso.[36]

Infine, CLSI non affronta la questione della presentazione degli intervalli di riferimento con i risultati dei pazienti, in particolare nei messaggi digitali generati dai sistemi informatici. Il tema sarà oggetto di discussione nella terza parte del progetto di Raccomandazioni SIPMeL M2.

In alternativa agli intervalli del produttore o di grandi laboratori, è stato proposto l'uso di intervalli armonizzati su base regionale.[37-39] Si tratta di una soluzione interessante, ma non disponibile in molti paesi e per molti esami, per la necessità di rispettare stringenti prerequisiti.[40] SIPMeL M2-2 non propone ancora questa soluzione perché non risulta disponibile nell'area geografica italiana.

Realtà virtuale e mondo reale

Se ISO 15189:2022 (clausola 7.3.5) chiede ai laboratori di rivedere periodicamente gli IR in relazione alla popolazione servita dal laboratorio, l'approccio "indiretto" potrebbe essere più adeguato del tradizionale, a condizione che il pacchetto di dati analizzato sia ampio, come avviene nei laboratori centralizzati di grandi dimensioni.

Il tema dei "dati del mondo reale" va oltre la verifica degli intervalli di riferimento. ISO ha già pubblicato il documento 18727,[41] che descrive un caso di studio specifico relativo all'esperienza del Giappone nell'acquisizione di dati reali (RWD) attraverso l'attuazione di un sistema di gestione della qualità, nell'ambito del progetto MID-NET®. ISO 18727 definisce RWD come dati relativi allo stato di salute raccolti regolarmente da varie fonti. Questa esperienza ha permesso di ridurre sia i costi che i tempi e ha migliorato notevolmente l'affidabilità dei dati. Nella discussione sul progetto, è stato osservato che l'approccio sarebbe stato utile in vari campi, intervalli di riferimento compresi, ma i redattori hanno preferito limitarsi al progetto originale. Comunque, l'applicazione è riuscita a estrarre dalle cartelle elettroniche anche risultati di laboratorio, compresa la batteriologia, codificati con i sistemi giapponesi *Japan Laboratory Code 10* (JLAC10) e *Japan Nosocomial Infections Surveillance* (JANIS).

Non trascuriamo il fatto che le iniziative basate sui dati del mondo reale vengono prese sul serio da agenzie come *Food and Drug Administration* (FDA).[42-44]

Il laboratorio riceve proposte di intervalli di riferimento

Vediamo lo sforzo encomiabile CLSI per distinguere per i laboratori la tradizione dalla prassi corrente. Tuttavia, da tempo si ammette che stabilire i propri valori di riferimento in ogni laboratorio è impedito dall'elevato carico di lavoro dei laboratori stessi, obbligandoli a prendere gli intervalli da una fonte esterna.[45] Questo dato, a distanza di anni, non è ancora stato smentito. Anche le fasce d'età pediatrica e geriatrica continuano a presentare ulteriori difficoltà.

Le opinioni degli esperti convergono su alcuni punti fermi. Gli sforzi per produrre definizioni di "normalità", di "benessere" o di "salute" non hanno dato risultati robusti.[46]

Gli "intervalli comuni", su base regionale, non sembrano maturi per tutte le regioni, se si eccettuano singole esperienze. La recente rassegna prodotta in collaborazione tra Austalia, USA e Svezia[33] ha confermato la realtà esistente, ovvero che le principali fonti degli intervalli di riferimento sono nei foglietti illustrativi dei diagnostici in vitro. Le proposte per la verifica degli intervalli su piccola scala sono ritenute poco affidabili dagli autori di questa rassegna, specialmente se gli intervalli sono troppo ampi. Si raccomanda ai laboratori di consultare la letteratura scientifica per cercare studi più approfonditi sulla loro specifica strumentazione. Badrick e i suoi collaboratori hanno però poca fiducia sulla disponibilità di documentazione completa dai fabbricanti. Giungono a dubitare della utilità concreta degli intervalli per i pazienti e il pubblico generico. Ritengono da abbandonare l'idea di intervalli validi per tutti, mentre laddove sia prevalente la diagnostica su asintomatici (per "benessere") gli intervalli dovrebbero essere allargati per ridurre l'incidenza di falsi positivi. Lo studio riconosce che le alternative, come stime indirette su dati numerosi, limiti decisionali clinici, intervalli personalizzati e punteggi statistici, sembrano crescere progressivamente di rilevanza.

Le raccomandazioni SIPMeL M2-2 propendono sull'idea che gli intervalli possono essere interpretati come una descrizione della prevalenza dei risultati nella popolazione servita dal laboratorio, come conferma delle unità di misura utilizzate e della taratura del metodo, nonché come base per individuare i limiti decisionali clinici. Non come criterio di classificazione dicotomica (sano/malato). Le raccomandazioni SIPMeL M2-2 includono riferimenti a vari studi condotti in condizioni reali sugli intervalli di riferimento.

Produzione di intervalli di riferimento

La letteratura scientifica sugli intervalli di riferimento è sterminata e continua a proporre approcci vari, da valutare criticamente non solo sui risultati e sugli aspetti statistici, ma anche e soprattutto sulle modalità di esecuzione. Ad esempio, ancora oggi si utilizzano donatori di sangue.[47] Un gruppo internazionale si è esercitato sui marcatori ossei,[48] ma nonostante le risorse a disposizione, il numero di soggetti selezionati è rimasto ben al di sotto di quello richiesto dagli standard. Inoltre, nelle figure descrivono intervalli da 5% a 95%, ovvero al 90%, non al 95%.

Le stesse limitazioni sono visibili nel progetto *European Biological Variation Study* (EuBIVAS),[49, 50] che ha coinvolto 91 volontari presunti sani (53 femmine e 38 maschi; età compresa tra 21 e 69 anni) provenienti da 6 Paesi europei, con un questionario sullo stile di vita e presunto stato di salute, verificato da una serie di esami di laboratorio "di routine" (comuni). Si tratta di circa 40 maschi adulti, 40 donne fino ai 50 anni e 10 donne anziane. Nonostante gli sforzi profusi nell'analisi dei dati, anche con tecniche molto sofisticate, l'estrazione da EuBIVAS di percentili come intervalli di riferimento non sembra sostenuta dalla numerosità e dalle caratteristiche dei soggetti. L'esperienza EuBIVAS sembra quindi un esercizio speculativo utile per simulazioni, ma di scarsa utilità nel mondo reale.

Alti esempi riguardano HbA1c,[51] su 38 volontari presumibilmente sani (20 femmine e 18 maschi) di età compresa tra i 30 e i 60 anni, la maggior parte dei quali erano per-

sonale di laboratorio. Un altro studio[52] ha usato un gruppo di soggetti con 92 uomini e 148 donne raccolti in una sede, mentre 125 uomini e 125 donne sono stati raccolti altrove, per 28 esami di chimica clinica, da albumina ad acido urico. Dallo studio emergono difficoltà a fondere insieme dati di istituzioni diverse.

Si ammette da più parti che lo sviluppo di intervalli di riferimento è un processo al di là delle capacità di un singolo laboratorio perché complesso, costoso e richiede molto tempo.[53] Pertanto, un laboratorio può alternativamente trasferire e verificare intervalli stabilite da una fonte esterna (ad esempio, foglietti illustrativi dei fabbricanti, pubblicazioni).

É stata proposta la rappresentazione bivariata di IR.[54] L'intervallo bivariato è migliore dei due intervalli univariati per TSH e fT4, ma al laboratorio si presenta la difficoltà di rappresentarlo comunicando il risultato degli esami. In conclusione, secondo ricercatori norvegesi l'intervallo di riferimento bivariato di s-TSH e s-FT4 potrebbe essere clinicamente più accurato degli intervalli di riferimento univariati combinati, non è molto difficile da stimare e potrebbe essere utilizzato riportando il percentile nella popolazione di riferimento che corrisponde alle concentrazioni misurate. Tuttavia, gli stessi Autori norvegesi segnalano che volendo calcolare il quoziente di probabilità (*likelihood ratio*) e l'area sotto la curva caratteristica operativa (*receiver operating characteristic,* ROC),[55] la debolezza dello strumento degli intervalli può inficiare la validità di strumenti raffinati e potenti. Alla fine, i ricercatori non hanno soluzioni nette e indirettamente confermano l'utilità solo per risultati di tipo qualitativo o classificati con limiti decisionali.

Le difficoltà della produzione degli intervalli di riferimento è riconosciuta da molti.[56] É un processo lungo, costoso e spesso al di là delle capacità dei singoli laboratori, perciò gli intervalli possono essere trasferiti da un soggetto (laboratorio o produttore) a un altro. Tuttavia, il trasferimento richiede l'equazione di regressione ottenuta da un confronto di metodi e la verifica successiva.

La discussione sulle limitazioni degli intervalli di riferimento non è cessata[57] e riguarda sia metodi tradizionali che quelli cosiddetti "indiretti", che utilizzano i dati dei pazienti già raccolti e archiviati in laboratorio. I limiti decisionali clinici sembrano superare almeno alcune debolezze degli intervalli.

Si segnalano casi di ambiguità nell'uso del termine "limite decisionale clinico". C'è chi lo usa per indicare la differenza critica, calcolata sulla variabilità biologica e su quella di misura.[58] Un'altra possibilità di fraintendimento è quella con i livelli critici per il controllo di qualità.[59] "Limiti decisionali medici" (MDL) è il termine utilizzato nel Q-Probe del 1994 del CAP, che riguarda appunto il controllo di qualità, non le decisioni.[60] Gli MDL sono descritti come una seconda linea di limiti per i controlli per ampliare l'intervallo dei risultati accettabili. MDL non vanno confusi con il livello decisionale medico che viene usato per la cura e il trattamento del paziente.

Infine, nei risultati di una recente indagine italiana sugli intervalli per la fosfatasi alcalina (ALP)[61] si vede un numero alto di laboratori che non riportano i limiti inferiori corretti degli IR per ALP pediatrica, ostacolando così l'identificazione dell'ipofosfatasia. Altre anomalie sono evidenti per i maschi di varie età e le femmine di 12,5 anni.

Il Comitato IFCC *Reference Intervals and Decision Limits* (C-RIDL) negli ultimi anni ha condotto studi multicentrici per ottenere RI comuni tra laboratori. Tuttavia, poiché la responsabilità più ampia del Comitato, dal suo nome, include i "limiti decisionali", C-RIDL sottolinea anche l'importanza dell'uso corretto sia degli intervalli che dei limiti decisionali e di incoraggiare i laboratori a specificare le informazioni appropriate ai medici secondo le necessità.[62]

Questi esempi e quelli riportati nelle Raccomandazioni SIPMeL non sono una raccolta esaustiva, ma evidenze della natura variabile degli intervalli di riferimento nel mondo reale, spesso adattata a necessità locali contingenti in deroga ai principi standard.

Inoltre si deve evidenziare che intervalli non coerenti con le differenze sistematiche tra metodi e tra laboratori causano errori nel processo decisionale clinico. Intervalli e limiti armonizzati possono essere adottati immediatamente se la differenza osservata è accettabile o eliminata, oppure dopo la standardizzazione o armonizzazione dei misurandi dove manca una completa tracciabilità metrologica.[63] Non solo. Anche l'incertezza di misura compromette le prestazioni diagnostiche del risultato proveniente dal laboratorio.[64]

Produzione di limiti decisionali

I limiti decisionali sono associati a rischi specifici e sono comunemente utilizzati per interpretare i risultati degli esami di laboratorio, compresi i parametri lipidici, il glucosio, l'emoglobina A1c (HbA1c) e i marcatori tumorali, per determinare il rischio di malattia, per diagnosticare o trattare.[65]

Da uno studio recente si ricavano interessanti indicazioni sulle soglie decisionali.[66] Tre dei quattro metodi indagati darebbero soglie decisionali corrispondenti al limite

inferiore dell'intervallo di riferimento, simili alle soglie raccomandate da *National Institute for Health and Care Excellence* (NICE). Tuttavia, le soglie NICE non hanno alcun collegamento con gli intervalli.[67] Addirittura, la guida NICE avverte sul fatto che l'etnia africana di riferimento avrebbe livelli di B12 più elevati dei bianchi e degli asiatici. Lo studio del gruppo di Cesana ha dimostrato scarti sistematici tra metodi sufficienti per invalidare l'applicazione delle soglie raccomandate dalla guida NICE. Ne emergerebbe una certa approssimazione di fabbricanti e laboratori nel uso di intervalli e soglie decisionali e la necessità di armonizzazione dei risultati.

Un esempio di limiti decisionali è stato descritto in Sri Lanka,[68] usando parametri eritrocitari per il rischio di tratto β-Talassaemico, applicando probabilità (valori predittivi, sensibilità, specificità, curve operative ROC, indice di Youden derivato da sensibilità è specificità), non i percentili delle distribuzioni. Si possono separare maschi e femmine, ma altri fattori influenti (fattori etnici, nutrizionali e ambientali[69, 70]), sono difficilmente isolabili in pratica nel mondo reale.

La *Canadian Society of Clinical Chemists* (CSCC) ha ultimamente prodotto raccomandazioni per gli intervalli di riferimento che combinano armonizzazione dei metodi, dati dal mondo reale e verifica su piccoli campioni,[71] come già sperimentato degli stessi canadesi.[72] É significativo riconoscere l'origine dei dati: i "laboratori comunitari" dello studio sono strutture diagnostiche, pubbliche o private, radicate nel territorio, che si occupano della maggior parte dei prelievi e degli esami dei campioni dei pazienti al di fuori degli ospedali. Sono stati estratti da quattro laboratori di questo tipo due anni di risultati dei pazienti per 16 esami. Gli intervalli di riferimento sono stati quindi stimati utilizzando l'algoritmo refineR. Gli intervalli sono stati verificati in nove laboratori con un metodo che corrisponde allo studio di scala media (60 soggetti) proposto da CLSI.

Pregi, difetti e alternative degli intervalli di riferimento sono stati recentemente sottoposti a revisione nell'articolo già citato.[33] Nella revisione si critica il criterio del 95%, perché produce inutilmente un 5% di falsi positivi, qualunque sia la statistica utilizzata. Si riconosce che i metodi indiretti hanno il vantaggio di provenire dalla stessa popolazione che normalmente si serve del laboratorio che li produce e ridurrebbero il numero di falsi positivi, uno svantaggio sia clinico che economico. Tra le alternative al 95%, i ricercatori elencano l'ampliamento degli intervalli a 99% centrale, la sostituzione con punteggi z, ovvero lo scarto normale standardizzato, o con percentili, come dimostrato utile per l'emoglobina.[73] Ulteriore alternativa è riconosciuta al ben noto approccio bayesiano del quoziente di probabilità. Infine, vengono riconosciuti gli importanti limiti degli intervalli forniti dai fabbricanti. La rassegna non giunge a conclusioni nette, ma attribuisce al laboratorio la possibilità di decidere quale sia la soluzione migliore caso per caso, per contenere l'ansia dei pazienti e gli sprechi nel sistema sanitario. In definitiva, si sconsiglia un approccio "unico" agli intervalli di riferimento per tutti gli esami. In fondo il tema sarebbe maturo per un cambiamento di paradigma catalizzato da intelligenza artificiale con grandi numeri di dati provenienti da gruppi di pazienti sia sani che con patologie diagnosticate. La revisione ha evidenziato i difetti degli intervalli, senza scegliere soluzioni alternative, ma non considera l'alternativa dei limiti clinici decisionali, che finalmente porterebbero al completo abbandono del ruolo decisionale dicotomico degli intervalli di riferimento.

Si conoscono esempi di situazioni "ibride". Ad esempio, l'*International Society on Thrombosis and Haemostasis* (ISTH) ha pubblicato criteri per la diagnosi di coagulazione intravascolare disseminata (DIC),[74] rappresentati da valori fissi per il conteggio delle piastrine (50 o 100 $x10^9$/L), come pure per PT-INR (3 e 6 secondi) e fibrinogeno (100 mg/dL), mentre quelli per D-dimero sarebbero multipli (3 o 7) del limite superiore degli IR. Un precedente studio, invece, forniva per D-dimero limiti fissi a 3,0 e 7,0 µg/mL, però con la prescrizione di armonizzare i metodi utilizzati.[75]

Partendo dall'applicazione pedissequa delle regole IFCC,[76, 77] ovvero il 95° percentile come limite superiore di "normalità" (*sic*), con un minimo di 120 soggetti, uno studio di Boston si è dedicato all'esame albuminuria[78] con dati di 1623 adulti presunti sani con varie esclusioni e ripartizioni. Trovati i limiti dell'intervallo, i ricercatori hanno notato lo spettro continuo di rischio anche al di sotto del limite di riferimento, data la relazione provata tra albuminuria non elevata e rischio cardio-renale.[79] Emerge così il dilemma tra intervallo di riferimento basato sulla frequenza dei valori in popolazioni presunte "sane" (un approccio puramente statistico) rispetto alla definizione di intervalli basati su esiti clinicamente significativi, arrivando a dire che nel caso dell'albuminuria il limite superiore statistico sarebbe di 20-30 mg/g, mentre il limite superiore clinico potrebbe essere di 7-10 mg/g.

La produzione di limiti decisionali non è facile. Ad esempio, le linee guida utilizzano la troponina T ad alta sensibilità (hsTnT) e il peptide natriuretico N-terminale di tipo pro-B (NT-proBNP) per stratificare il rischio dei pa-

zienti con embolia polmonare acuta (PE).[80] Tuttavia, non sono stati definiti valori soglia basati sulle prove ("*evidence based*") per questa applicazione clinica. Uno studio avrebbe verificato che la soglia ottimale di hsTnT è circa tre volte superiore alla soglia raccomandata dalle linee guida.[81]

Un altro esempio è lo studio[82] sul 99° percentile di anticorpi anticardiolipina (aCL) e anti-β2-glicoproteina I (aβ2GP1) IgG/IgM per la sindrome antifosfolipidi, proposto dai criteri dell'*American College of Rheumatology*,[83] dove si dimostra che tutto dipende dal metodo utilizzato, concludendo che, quando gli studi clinici indipendenti non sono realizzabili, l'unica alternativa sono le soglie raccomandate dal fabbricante.

Le linee guida sono poco seguite dai laboratori: un'indagine condotta da *External Quality Control of Diagnostic Assays and Tests* (ECAT) ha rilevato che circa il 45% dei laboratori ha adottato i valori limite del fabbricante per aCL e aβ2GP1, e solo il 38% li ha verificati.[84]

La diffusione dei metodi cromatografici con spettrometria di massa (LC-MS/MS) accanto ai precedenti immunometrici ha influenzato le soglie decisionali per il testosterone (TT).[85] La soglia per il testosterone basso a 300 ng/dL (10,4 nmol/L) si basa su correlazioni dei metodi immunometrici con la gravità dei sintomi dei pazienti. Anche se la guida della *American Urological Association* del 2018 raccomanda un limite di 300 ng/dL,[86] il limite basso di 264 ng/dL proposto dalla *Endocrine Society* potrebbe essere una soglia più accurata,[87] perché i nuovi metodi immunometrici sono di solito correlati con LC-MS/MS.

Le soglie dei marcatori tumorali possono essere un problema complicato. I pazienti con CA19-9 negativo rappresentano un sottogruppo ad alto rischio, come quelli con CA19-9 >200 U/mL. Quando la genotipizzazione non è disponibile, si arriva a un modello a doppia soglia (≤7 e >200 U/mL) per la stratificazione del rischio.[88]

IR e LD possono persino entrare in conflitto tra loro. Il fenotipo ematologico Duffy-null [Fy(a−b−)] è associato a valori basali più bassi della conta dei globuli bianchi (WBC) e della conta assoluta dei neutrofili (ANC), nonostante un decorso clinico benigno. Uno studio su 122 soggetti sani con fenotipo Fy(a−b−)[89] ha definito IR utilizzando metodi non parametrici per il 95% centrale (WBC: 2,43–7,05 × 10^3/μL; ANC: 1,01–4,34 × 10^3/μL) inferiori alle soglie convenzionali, con il limite inferiore dell'ANC al di sotto del valore soglia per la neutropenia.

Anche questa raccolta di esempi, non certo esaustiva, testimonia la varietà di approcci adottati nel mondo reale per fornire risposte a quesiti specifici.

Ripartizione di intervalli e limiti per sesso e/o genere

Il punto 7.3.5 di ISO 15189:2022 stabilisce che intervalli e limiti devono riflettere la popolazione di pazienti servita dal laboratorio, tenendo conto del rischio per i pazienti e non fornisce alcuna indicazione sui criteri di ripartizione. CLSI EP45 nella sua tabella 3 elenca esempi di fattori di suddivisione, divisi in criteri comunemente considerati, iniziando da "sesso biologico". La stessa cosa fa CLSI EP44 nella sua tabella 4. CLSI non affronta la controversa questione delle condizioni intermedie tra i sessi biologici[90] e nemmeno delle conseguenze sugli esami di laboratorio delle variabilità di genere di origine psico-sociale.

La questione ha attirato l'attenzione di molti ricercatori.

Abbiamo evidenze che la terapia ormonale per l'affermazione del genere comporta cambiamenti nei risultati di laboratorio con IR specifici per il sesso.[91] É stato raccomandato ai medici di essere consapevoli del potenziale impatto della terapia di conferma del genere sugli esami di laboratorio e del sesso/genere utilizzato per assegnare gli intervalli.[92] Negli individui transgenere sottoposti a terapia ormonale mascolinizzante o femminilizzante si vedono cambiamenti specifici significativi in alcuni esami.[93]

É stato proposto che molti sistemi informativi di laboratorio e sanitari debbano essere aggiornati per rilevare l'identità di genere.[94] Ad esempio, si consiglia di accompagnare i risultati di laboratorio con una nota tipo questa: "*Alcuni studi hanno dimostrato che questo intervallo di riferimento potrebbe non essere applicabile alle minoranze sessuali/di genere*".[95]

Un aggiornamento completo e obiettivo sul tema si trova forse nella rassegna australiana del 2025,[96] dove si ammette che l'uso di intervalli derivati dalla popolazione transgenere è complicato, gli studi hanno generalmente una dimensione ridotta del campione e includono adulti con trattamenti ormonali consolidati da tempo. Esistono significative disparità sanitarie per le persone transgenere, di genere diverso e non binarie e questo gruppo di popolazione ha una serie di esigenze sanitarie uniche rispetto ad altri gruppi. Dare IR e LD del sesso dichiarato è appropriato in alcuni casi, ma non in tutti, e in letteratura ci sono incertezze sull'approccio migliore da adottare. In generale, non esiste un approccio "unico" adatto a tutti, e personalizzare le cure è fondamentale.

Le raccomandazioni italiane (Ministero della Salute — Gruppo di lavoro sulla medicina di genere)[97] non sostituiscono sesso con genere o viceversa, ma li tengono affiancati. Il gruppo del Ministero afferma che vanno considerate, sia nella ricerca preclinica che in quella clinica, le diverse caratteristiche cliniche e biologiche associate al

sesso e/o al genere, dove le differenze di sesso sono principalmente influenzate da determinanti biologiche e quelle di genere da questioni socioculturali ed economiche.

Sembra che tutta la materia non sia abbastanza matura per esprimere Raccomandazioni solide da indirizzare ai laboratori medici.

Intervalli di riferimento e prestazioni cliniche degli esami: CLSI EP49

Trovare gli IR nel capitolo delle prestazioni cliniche degli esami non è per nulla ovvio e scontato, anzi. Alla luce del modo in cui si producono IR, delle criticità della statistica applicata sui dati, ritrovarne il concetto nel processo di validazione clinica o di prestazioni cliniche è certamente inaspettato.

Invece, abbiamo qualcosa su cui discutere. CLSI produce un documento sulla cosiddetta "validità clinica" degli esami di laboratorio.[98] ISO da parte sua dispone di uno standard sugli studi delle prestazioni cliniche degli esami di laboratorio.[99] Il confronto tra ISO e CLSI fornisce informazioni di grande interesse.

CLSI sostiene che "validazione clinica" non abbia una definizione accettata a livello globale. CLSI EP49 propone un processo per dare prove oggettive della conformità dei risultati del metodo di esame alle esigenze dell'utente e all'uso prefissato. Il processo prevede l'esecuzione di esami in condizioni d'uso reali o simulate e di associare i risultati a una particolare condizione clinica, un processo o stato fisiologico o patologico. La validazione clinica per CLSI si baserebbe sulla validità scientifica. La validità scientifica consiste nell'associazione tra un risultato e una condizione clinica o uno stato fisiologico. La validità clinica invece sarebbe la capacità del metodo di laboratorio di rilevare o prevedere il disordine o la condizione di interesse in una popolazione bersaglio. Differenza sottile ma non trascurabile. Esempi di misure delle prestazioni cliniche sono le coppie sensibilità clinica e specificità clinica, rapporto di verosimiglianza (o quoziente di probabilità)[100-102] positiva (PLR) e rapporto di verosimiglianza negativa (NLR), nonché la tripletta valore predittivo positivo (PPV), valore predittivo negativo (NPV) e prevalenza. CLSI intende per "condizione bersaglio" una particolare malattia, uno stadio della malattia, lo stato di salute o qualsiasi altra condizione o caratteristica identificabile di interesse in un soggetto. Secondo la proposta di EP49, la validità clinica di un metodo di prova in valutazione (denominato "metodo di prova candidato") si ottiene mediante il confronto con un "comparatore" che dovrebbe essere il miglior metodo disponibile, o tradizionalmente utilizzato, per determinare se la condizione è presente o assente.

Fino a questo punto, non sono ancora comparsi IR. Vengono introdotti da CLSI EP49 al punto 3.4.2 (esami con risultati quantitativi) e al successivo 3.4.3 (esami con risultati "semi-quantitativi"), dove si afferma che stabilendo un intervallo di riferimento di concentrazione l'esame può essere utilizzato per determinare la presenza o l'assenza di una condizione bersaglio, quindi i risultati vengono interpretati in base ai limiti degli intervalli di riferimento. Addirittura, i risultati "semi-quantitativi" possono essere confrontati con un intervallo di riferimento per fornire un'interpretazione qualitativa. Nella fase di consultazione, SIPMeL ha espresso le proprie critiche su questi punti, senza tuttavia ottenere in questo caso (a differenza di altri) l'attenzione del Comitato CLSI guidato da Michelle R. Campbell. I riferimenti scientifici portati a sostegno da CLSI sono un articolo datato 2019 (che non fa cenno agli intervalli di riferimento)[103] e un documento del Consorzio per l'innovazione nel settore dei dispositivi medici (MDIC, autorevole ente statunitense con rappresentanti dell'industria, delle istituzioni pubbliche come CDC, FDA, NIH e NIST, delle associazioni di tutela dei pazienti, nonché fornitori e distributori del settore delle tecnologie mediche) prodotto nel 2024.[104] Il documento MDIC cita di sfuggita i risultati "semi-quantitativi" ma non accenna mai a intervalli di riferimento, solo a soglie (*cutoff*).

Le perplessità su CLSI EP49 non finiscono qui.

Lo stesso titolo contiene un termine anomalo: "validità clinica", assente nella terminologia standard, sicuramente da intendere con "prestazioni cliniche", lo stesso termine usato da numerosi riferimenti sia nelle norme tecniche che nella letteratura scientifica. CLSI EP49 al punto 1.4 (terminologia) conferma di voler usare "validità clinica" per descrivere la capacità di un metodo di esame di identificare o prevedere la classificazione clinica di campioni o soggetti, mentre "validazione clinica" sarebbe il processo volto a determinare tale classificazione.

Il comitato CLSI riconosce che tali termini potrebbero non essere di uso comune in alcune regioni. Si tratta di un errore grossolano: la "regionalizzazione" dei vocaboli affermata da CLSI non ha nulla a che vedere con il rigore scientifico e il processo di armonizzazione dei glossari da più parti auspicato. In CLSI EP49 sono trattate solo le prestazioni del risultato per individuare positivi e negativi, non viene gestito in alcun punto il concetto di "utilità clinica". Utilità necessaria per rendere valido un esame. Diversamente da quanto affermato da CLSI, il documento MDIC non contiene affatto la "*descrizione approfondita dell'utilità clinica dei metodi di analisi di laboratorio,*

nonché delle strategie per lo sviluppo di prove a sostegno di tale utilità clinica".

Inoltre, CLSI EP49 rifiuta i vocaboli standard ISO per i laboratori medici (esame, pre- e post-esame) e inserisce una terminologia errata secondo criteri metrologici come "semi-quantitativo".

Il termine "semi-quantitativo" non è di natura metrologica e non è previsto da ISO 15189:2022 3.8 tra le possibili alternative di esame. Non è un caso. Le definizioni in vari documenti CLSI relative al termine 'semiquantitativo' (o "quasiquantitativo") sono incoerenti tra loro (POCT04, I/LA02, I/LA33, EP26, H62). Il vocabolario standard di metrologia,[105, 106] dispone al punto 1.26 del lemma "quantità ordinale", usato anche da varie fonti[103, 107] che riconoscono come i dosaggi chiamati (impropriamente) "semiquantitativi" sono di fatto qualitativi.

Gli errori concettuali e persino nelle citazioni bibliografiche di CLSI EP49 sono decisamente sconcertanti e impediscono di includere EP49 nelle Raccomandazioni SIPMeL.

Studi di prestazione clinica: ISO 20916:2019 e GHTF

Il citato documento ISO 20916 definisce le buone prassi di studio per gli studi condotti per valutare le prestazioni cliniche e la sicurezza dei dispositivi medico-diagnostici in vitro (IVD) a fini regolatori.[99] ISO 20916 è stata confermata in ISO nel settembre 2024.

La maggior parte degli studi sui dispositivi medici IVD viene eseguita utilizzando campioni derivanti da resti di prelievi per scopi di cura (avanzi o archiviati). In questi studi, per ISO non vi è alcun rischio per i soggetti. In altri casi può esserci un rischio aggiuntivo per il soggetto (ad es. puntura lombare o biopsia tissutale, prelievo di sangue da neonati o pazienti gravemente malati). Oppure, se i risultati degli esami venissero utilizzati per prendere decisioni sul paziente (ad esempio, studi interventistici), c'è il rischio dovuto a eventuali falsi negativi o falsi positivi.

La definizione 3.10 di ISO 20916 attribuisce alle "prestazioni cliniche di un dispositivo medico IVD" il significato di "capacità di fornire risultati correlati a una particolare condizione clinica o a un processo/stato fisiologico/patologico in conformità allo scopo dell'esame, popolazione bersaglio e utente previsto". Sono inclusi i valori attesi, sensibilità e specificità diagnostica e i valori predittivi negativi e positivi in base alla prevalenza della malattia. Definizione raccolta dal documento del Gruppo di lavoro per l'armonizzazione globale (*Global Harmonization Task Force)* GHTF/SG5/N6:2012.[108] GHTF è un gruppo internazionale volontario composto da rappresentanti delle autorità di regolamentazione dei dispositivi medici e delle associazioni di categoria provenienti dall'Europa, dagli Stati Uniti d'America (USA), dal Canada, dal Giappone e dall'Australia. ISO 20916 non ha una definizione di "utilità clinica, ma GHTF invece si: utilità dei risultati ottenuti da esami effettuati con il dispositivo medico IVD, valore delle informazioni da IVD per la persona o per la popolazione in generale.

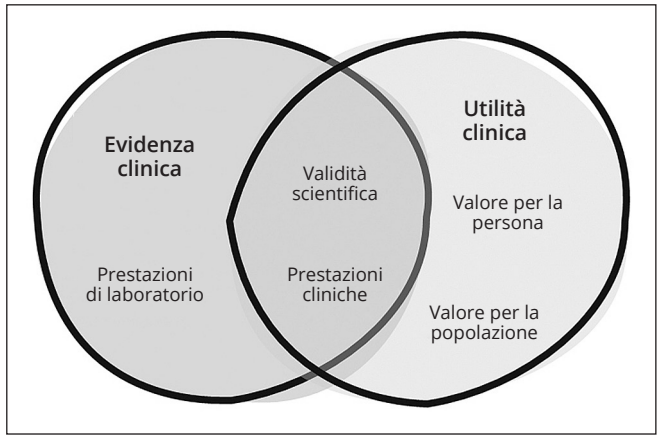


Figura 6.—Relazione tra evidenza clinica, utilità clinica e prestazioni (disegnata in base alle definizioni del GHTF/SG5/N6:2012.[108] Non sono presenti gli intervalli di riferimento).

GHTF sconsiglia il termine "validità clinica", preferisce "prestazioni cliniche". Contiene invece il termine "validità scientifica", riferita all'associazione di un esame a una condizione clinica o a uno stato fisiologico. Si tratta di concetti collegati ma ben distinti, come rappresentato in Figura 6. A connettere i concetti sta quello di "evidenza clinica", ovvero tutte le informazioni che sostengono la validità scientifica e le prestazioni di un dispositivo medico IVD nell'ambito dell'uso previsto dal fabbricante.

IR non compaiono nelle definizioni di ISO 20916, nemmeno nella definizione 3.10 di prestazione clinica. Esattamente lo stesso che viene fatto da GHTF. IR viene citato da ISO 20916 solo nel capitolo 5, dove si ipotizza uno studio di IR presso la sede del fabbricante. Al punto 5.3 IR viene incluso nelle caratteristiche di prestazione clinica, insieme a sensibilità, specificità, valore predittivo positivo, valore predittivo negativo, soglie decisionali), ma non viene fornito un razionale.

Dall'esame delle fonti originali risulta evidente la discrepanza con CLSI EP49 su IR e prestazioni cliniche e per il concetto di "validità clinica".

Si deve ammettere la presenza in alcuni settori della comunità scientifica di tendenze conservative di vecchi luoghi comuni, sottovalutazione dell'effetto delle improprietà di linguaggio sulla buona pratica dei servizi sanitari.

Prestazione clinica: validazione clinica, uso prefissato e utilità dell'uso

Per non lasciare il tema poco approfondito, la Raccomandazione SIPMeL M2-2 raccoglie altri contributi scientifici sul tema della prestazione clinica.

Il gruppo "Rete internazionale per la qualità in patologia" (*International Quality Network for Pathology, IQN Path*) nasce nel 2014 per migliorare l'uso clinico dei biomarcatori tissutali con lo scambio di competenze tra i principali esperti del settore, per definire rapidamente raccomandazioni relative all'adozione di nuovi biomarcatori, stabilire parametri di riferimento e migliori pratiche, promuovere VEQ/PT rafforzando la voce dei fornitori di servizi VEQ.

IQN Path si è recentemente espresso proprio sulla prestazione clinica di un biomarcatore in oncologia.[109] IQN Path muove dalla definizione ISO 9000 di validazione: "la conferma, attraverso la fornitura di prove oggettive, che i requisiti per un uso o un'applicazione specifica sono stati soddisfatti". Vediamo tre componenti che possiamo tradurre con "livello di certezza" (conferma), "caratteristiche di prestazione dell'esame" (evidenza oggettiva) e "valore per lo scopo" (uso prefissato). IQN Path fornisce molti dettagli sulla concretizzazione del principio.

IQN Path suddivide i marcatori in tre categorie: presenza o la sovraespressione di marcatori tumorali (ad esempio, eventi di fusione o amplificazione genica di diverse tirosin-chinasi, come ALK, NTRK e HER2/EGFR). Il primo gruppo rileva un evento biologico specifico che innesca la presenza o la sovraespressione di marcatori tumorali (ad esempio, eventi di fusione o amplificazione genica di diverse tirosin-chinasi, come ALK, NTRK e HER2/EGFR). Il secondo rileva altri eventi molecolari informativi sulle risposte immunologiche al tumore (ad esempio, carico mutazionale tumorale o *tumor mutation burden,* TMB), instabilità microsatellitare, sovraespressione delle proteine come c-MET, PD-L1 o PD1 e altri, il terzo è un esame di selezione (*screening*), per ridurre costo e tempi di esecuzione degli esami escludendo i pazienti che non hanno alcuna possibilità di essere positivi con un biomarcatore definitivo (ad esempio, ROS1 e pan-TRK IHC). Nel Gruppo 1, la validazione clinica consiste nella validazione della misurazione con i metodi standard, quelli del gruppo 2 vanno riferiti a linee guida CLSI EP-12[110] perché sono quasi sempre di tipo qualitativo, con un *gold standard*/di riferimento della stessa metodologia. Gli esami del gruppo 3, a differenza degli altri, si deve fare semplicemente in modo che la sensibilità rimanga del 100%.

Un punto critico di IQN Path, come di CLSI EP49, è la distinzione tra "uso prefissato" e "utilità dell'uso". In ISO si sono preoccupati di questo aspetto gli informatici in ISO/IEC/IEEE 24765,[111] dove al punto 3.4468 utilità (ereditata dalla precedente ISO/IEC 25010[112]) è "grado di soddisfazione dell'utente per il raggiungimento percepito di obiettivi pratici". Stessa definizione comparsa in 3.9 di ISO 5649[113] (ereditata a sua volta da 4.7 di GHTF/SG5/N6:2012,[108] ovvero utilità dei risultati ottenuti e il valore dell'informazione per l'individuo sottoposto all'esame o per la popolazione in generale, o per entrambi.

Si osserva uno squilibrio e forse una differenza di prospettiva. Il concetto di utilità verrebbe valorizzato più da agenzie di regolamentazione dei dispositivi medici e dell'industria regolamentata che da specialisti del laboratorio medico, con l'eccezione di ISO 5649:2024 (esami di laboratorio sviluppati in casa).

Un contributo importante sul tema viene dal documento "MDCG 2022-2 Linee guida sui principi generali dell'evidenza clinica per i dispositivi medici diagnostici in vitro (IVD)"[114], ripreso dalle Raccomandazioni SIPMeL M2-2.

Il Gruppo di coordinamento per i dispositivi medici (MDCG) è istituito dall'articolo 103 del Regolamento (UE) 2017/745. L'MDCG è composto da rappresentanti di tutti gli Stati membri ed è presieduto da un rappresentante della Commissione europea. MDCG si occupa di questioni chiave del settore dei dispositivi medici, dalla supervisione degli organismi notificati o dalla standardizzazione alla sorveglianza del mercato, passando per questioni internazionali, nuove tecnologie e indagini cliniche.

Il documento MDCG 2022-2 parla di evidenze cliniche e prestazioni dei dispositivi medico-diagnostici in vitro (IVD), come stabilito dal Regolamento (UE) 2017/746.[115] L'articolo 56 di IVDR attribuisce al fabbricante il dovere di specificare e giustificare il livello di evidenza clinica in considerazione della destinazione d'uso. Pertanto, la definizione della destinazione d'uso di un IVD deve essere considerata un fattore chiave alla base della valutazione complessiva.

Per MDCG 2022-2 lo scopo d'uso dovrebbe essere chiaramente espresso nella dichiarazione, riportato nelle istruzioni per l'uso (IFU) e dovrebbe includere gli elementi specificati in un lungo elenco derivato da allegato I, 20.4.1 (c.) di IVDR.

MDCG 2022-2 ricava da IVDR, articolo 2 punto 37 e considerando 64 il concetto di "Beneficio clinico" come l'impatto positivo di un dispositivo in relazione alla sua funzione, se selezione (*screening*), monitoraggio, diagnosi o ausilio alla diagnosi dei pazienti, o sulla gestione dei pazienti o sulla salute pubblica. "Prove cliniche" (IVDR

articolo 2 punto 36) sono dati clinici e risultati della valutazione delle prestazioni relativi a un dispositivo, in quantità e qualità sufficienti a consentire una valutazione qualificata della sicurezza del dispositivo e del conseguimento dei benefici clinici previsti. "Prestazione clinica" è la capacità di un dispositivo di produrre risultati correlati a una particolare condizione clinica o a un processo o stato fisiologico o patologico, in base alla popolazione bersaglio e all'utilizzatore previsto (IVDR articolo 2 punto 41). "Validità scientifica di un esame" è l'associazione di un risultato di esame con una condizione clinica o uno stato fisiologico (IVDR articolo 2 punto 38).

A parte l'uso di una certa terminologia obsoleta e non conforme con gli standard, ma purtroppo fissata nel Regolamento europeo quindi difficile da correggere, MDCG riprende un errore dello stesso Regolamento, notato da soggetti non europei come CLSI: la definizione di Sensibilità diagnostica e Specificità diagnostica. Per IVDR e MDCG, "Sensibilità diagnostica" sarebbe la capacità di un dispositivo di identificare la presenza di un marcatore bersaglio associato a una particolare malattia o condizione (IVDR Articolo 2 punto 50), mentre "Specificità diagnostica" sarebbe la capacità di un dispositivo di riconoscere l'assenza di un marcatore bersaglio associato a una particolare malattia o condizione (IVDR Articolo 2 punto 49). La norma europea confonde palesemente il marcatore con la malattia; quindi, fornisce una definizione adatta per il confronto tra metodi ma non per la prestazione clinica, associata dal punto 41 alla condizione clinica, non al marcatore.

Il concetto di "Prestazione clinica" è ripreso nel paragrafo 6.6 di MDCG, dove se ne descrive la composizione costituita da sensibilità diagnostica, specificità diagnostica, valore predittivo positivo, valore predittivo negativo, numero medio di pazienti che devono essere trattati/diagnosticati per avere un impatto su una persona, numero di pazienti che devono essere diagnosticati/trattati per avere un effetto avverso su un paziente, quoziente di probabilità (rapporto di verosimiglianza) positivo, quoziente di probabilità negativo, rapporto di disparità (*odds*), usabilità o interfaccia utente. Tutti fattori legati alla condizione di malattia, non certamente al confronto di metodi.

Ma soprattutto, il documento MDCG non prende mai in considerazione gli IR.

Conclusioni e raccomandazioni

Il documento SIPMeL M2-2 di Raccomandazioni non sostituisce la norma originale UNI EN ISO.

ISO 15189:2022 per evitare i requisiti prescrittivi e collegarsi ad altri documenti ISO pubblicati pertinenti, per evitare ripetizioni ridondanti, necessita di integrazioni per l'applicazione pratica. Ad esempio, servono riferimenti a ISO 22583 per gli strumenti dei servizi POCT e al Vocabolario di metrologia ISO/IEC Guide 99:2007. Sono inoltre molto utili diversi documenti CLSI, inseriti in punti specifici delle Raccomandazioni SIPMeL. Con riferimenti ad altri documenti ISO e documenti CLSI i laboratori sono aiutati ad adeguarsi ai requisiti 15189 e gli ispettori di accreditamento troveranno più facile la verifica delle evidenze.

Le Raccomandazioni SIPMeL M2-2 mettono in evidenza alcuni aspetti critici: intervalli di riferimento e limiti decisionali clinici richiedono netta distinzione tra scopo descrittivo e utilizzo inferenziale per la diagnosi, produzione e verifica degli intervalli non prescindono da dimensioni e caratteristiche del laboratorio, nelle condizioni opportune i metodi "indiretti" sono convenienti per lo scopo descrittivo. Le prestazioni cliniche degli esami di laboratorio (sensibilità, specificità, valori predittivi) richiedono la definizione di soglie o limiti decisionali. I limiti di riferimento non sono utili per la valutazione delle prestazioni cliniche.

Il documento SIPMeL M2 parte seconda esprime così alcune raccomandazioni fondamentali (Tabella I). Non si devono confondere intervalli di riferimento e limiti decisionali clinici. I primi non sono collegabili a sensibilità e specificità diagnostiche, nonostante qualcuno erroneamente lo faccia, prescindendo dalla loro struttura statistica. Dal punto di vista pratico, gli studi tradizionali e i metodi "verifica su larga scala" e "trasferimento mediante confronto", sono adatti per fabbricanti e grandi istituzioni, non per la generalità dei laboratori medici, molti di piccole dimensioni e con risorse ridotte. Il laboratorio medico invece può acquisire intervalli proposti da altri laboratori qualificati o da fabbricanti e verificarli con metodi riconosciuti. Tra i metodi disponibili, ai laboratori con meno risorse si raccomanda il metodo binomiale su piccola scala. Invece, ai laboratori medici che eseguono un volume elevato di esami e dispongono di strumenti informatici adeguati, si raccomandano i metodi "indiretti" con dati dal mondo reale. Per la comunicazione con medici e pazienti, si raccomanda di accompagnare i risultati degli esami preferenzialmente con limiti decisionali, purché chiaramente identificati e non confusi con intervalli di riferimento. Per confrontare risultati, intervalli e limiti ottenuti da laboratori diversi, rimane insostituibile l'armonizzazione dei metodi a partire da taratura e riferibilità metrologica. Per stimare le prestazioni cliniche degli esami sono necessarie soglie o LD, mentre gli IR sono assolutamente inadatti. Le prestazioni diagnostiche (sensibilità, specificità, valori

predittivi, quindi i LD) dell'esame dovrebbero essere concordate con i medici curanti di riferimento, non imposte unilateralmente dal laboratorio.

Lo studio di temi apparentemente settoriali come gli IR mette in luce aspetti forse inaspettati dei processi normativi, sia a livello di leggi e regolamenti, che nelle norme tecniche, che infine nelle guide di pratica professionali. Tra questi, la persistenza in alcuni attori, già noti per affidabilità e autorevolezza, di forze di attrazione magnetica verso la tradizione, il luogo comune, la prassi comune, in antitesi al progresso in direzione del rigore scientifico, dell'interoperabilità dei concetti tra discipline, dell'armonizzazione dei linguaggi. Fenomeno ben rappresentato dalle note poste nell'introduzione al documento CLSI EP49. Ne ricaviamo l'insegnamento alla valutazione critica di ogni fonte, anche di alto livello presunto, al confronto puntuale di fonti diverse, alla verifica delle conseguenze nel mondo reale di ogni singola soluzione.

Bibliografia

1. SIPMeL M2 Raccomandazioni per Accreditamento ISO 15189 del laboratorio medico - requisiti per le relazioni con pazienti e medici curanti, parte seconda: intervalli e limiti. [Internet]. Disponibile alla pagina: https://www.sipmel.it/it/lineeguida/approvate/123225 [citato 21 agosto 2026].

2. SIPMeL M2 Recommendations for Medical Laboratory ISO 15189 Accreditation - requirements for relationships with patients and treating physicians, part two: intervals and limits; 2026 [Internet]. Disponibile all apagina: https://www.sipmel.it/en/lineeguida/approvate/123225 [citato 21 agosto 2026].

3. Pradella M. Requisiti dei laboratori medici, forensi, antidoping e alimentari: nuove ISO 15189 e ISO 17025. Riv Ital Med Lab 2019;15:252–62.

4. Pradella M. Decision rules for measurement results in testing and medical laboratories with ISO accreditation requirements. Metrology 2026;6:40.

5. Pradella M. Nuova definizione di incertezza delle misure e conseguenze per il laboratorio medico. Riv Ital Med Lab 2026;22:51–7.

6. ISO/IEC 17025:2017. General requirements for the competence of testing and calibration laboratories. Geneva, Switzerland: International Organization for Standardization; 2017.

7. ISO 15189:2022. Medical laboratories — Requirements for quality and competence. Geneva, Switzerland: International Organization for Standardization; 2022.

8. ISO/IEC Guide 98-4:2012. Uncertainty of measurement — Part 4: Role of measurement uncertainty in conformity assessment. Geneva Switzerland: International Organization for Standardization; 2012.

9. ILAC-G8. 09. Guidelines on Decision Rules and Statements of Conformity. Newton SA 5074, Australia; 2019 [Internet]. Disponibile alla pagina: https://www.iasonline.org/wp-content/uploads/2021/03/ILAC_G8_09_2019.pdf [citato 21 agosto 2026].

10. Williams A, Magnusson B. Eurachem/CITAC Guide: Use of uncertainty information in compliance assessment - 2nd ed; 2021 [Internet]. Disponibile alla pagina: https://www.eurachem.org/images/stories/Guides/pdf/MUC2021_P1_EN.pdf [citato 21 agosto 2026].

11. CLSI. Establishing Reference Intervals. CLSI document EP44-A4. Malvern, PA: Clinical & Laboratory Standards Institute; 2026.

12. CLSI. Implementation of Reference Intervals in the Medical Laboratory. CLSI document EP45-A4. Malvern, PA: Clinical & Laboratory Standards Institute; 2026.

13. CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory CLSI document EP28A3C. Malvern, PA: Clinical & Laboratory Standards Institute; 2010.

14. El-Khoury JM, Karger AB, Cavalier E, Kalyesubula R, Teo BW, Costa E Silva VT, *et al.* International Perspectives on GFR Estimation and Race-Based Adjustments. Clin Chem 2023;69:796–802.

15. El-Khoury JM, Ovalle AA, Cervinski MA, Levey AS, Jones G, Eneanya ND, *et al.*; Moderators; Experts. Is It Time to Move On? Re-examining Race in Glomerular Filtration Rate Equations. Clin Chem 2021;67:585–91.

16. Horn PS, Pesce AJ. Reference Intervals: A Users Guide. AACC Press; 2005.

17. Ceriotti F. Reference Intervals: A User's Guide. Paul S. Horn and Amadeo J. Pesce-. Clin Chem 2006;52:544–5.

18. Meyer A, Streichert T. When laboratory values lose their frame of reference: missing reference intervals in layperson questions in an online health forum. J Lab Med 2026;50:162–6.

19. Barry PL, Westgard JO. MV - The Transference of Reference Intervals. Created on 12 January 2009. [Internet]. Disponibile alla pagina: https://westgard.com/lessons/basic-method-validation/lesson30.html [citato 21 agosto 2026].

20. Zierk J, Metzler M, Rauh M. Data mining of pediatric reference intervals. J Lab Med 2021;45:311–7.

21. 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.

22. 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.

23. Sikaris K. Separating disease and health for indirect reference intervals. Lab Med 2021;45:55–68.

24. Haeckel R, Adeli K, Jones G, Sikaris K, Wosniok W. Definitions and major prerequisites of direct and indirect approaches for estimating reference limits. Clin Chem Lab Med 2022;61:402–6.

25. Doyle K, Bunch DR. Reference intervals: past, present, and future. Crit Rev Clin Lab Sci 2023;60:466–82.

26. Ma S, Yu J, Qin X, Liu J. Current status and challenges in establishing reference intervals based on real-world data. Crit Rev Clin Lab Sci 2023;60:427–41.

27. Ammer T, Schützenmeister A, Prokosch HU, Zierk J, Rank CM, Rauh M. RIbench: A poposed benchmark for the standardized evaluation of indirect methods for reference interval estimation. Clin Chem 2022;68:1410–24.

28. Hoffmann RG. Statistics in the practice of medicine. JAMA 1963;185:864–73.

29. Bhattacharya CG. A simple method of resolution of a distribution into gaussian components. Biometrics 1967;23:115–35.

30. 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.

31. 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.

32. Ammer T, Schützenmeister A, Prokosch HU, Rauh M, Rank CM, Zierk J. A novel algorithm for reference interval estimation from real-world data. Sci Rep 2021;11:16023.

33. Badrick T, El-Khoury JM, Theodorsson E. Laboratory reference in-

tervals - history and modern approaches for improved utility. Scand J Clin Lab Invest 2025;85:229–41.

34 Pradella, M, Cesana, B.M. Rischi, errori e costi dei confronti tra metodi --- Risks, errors and costs of method comparisons. Poster n° 190. 9° Congresso Nazionale SIPMeL Riva del Garda, 28-29-30 Ottobre 2024.

35. Pradella M, Cesana BM. Evitare rischi, errori e costi dei confronti. Tutto Misure 2024;26:65–7.

36. Patel S, Verma N, Padhi P, Shah S, Nanda R, Mohapatra E. An approach to re-evaluate the reference cutoff of the parameters of newborn screening: an observational study. Cureus 2023;15:e45139.

37. Jones GR, Koetsier SD. RCPAQAP First Combined Measurement and Reference Interval Survey. Clin Biochem Rev 2014;35:243–50.

38. Koerbin G, Tate JR. Harmonising Adult Reference Intervals in Australia and New Zealand - the Continuing Story. Clin Biochem Rev 2016;37:121–9.

39. Standardardised Pathology Informatics RC. in Australia (SPIA) Chemical Pathology Harmonised Reference Intervals. [Internet]. Disponibile alla pagina: https://www.rcpa.edu.au/Manuals/RCPA-Manual/General-Information/IG/Table-6-Harmonised-reference-intervals-for-chem [citato 21 agosto 2026].

40. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev 2007;28:115–21.

41. ISO/TR 18727:2026. Health Informatics – Quality management of health data in the real world – A case study. [Internet]. Disponibile alla pagina: https://www.iso.org/standard/85414.html [citato 21 agosto 2026].

42. Framework for FDA's Real–World Evidence Program. 2018. [Internet]. Disponibile alla pagina: https://www.fda.gov/media/120060/download [citato 21 agosto 2026].

43. U.S. Food and Drug Administration. Real-World Evidence; 2025. [Internet]. Disponibile alla pagina: https://www.fda.gov/science-research/science-and-researchspecial-topics/real-world-evidence [citato 29 giugno 2026].

44. Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. J Diabetes Sci Technol 2020;14:345–9.

45. Ozarda Y, Higgins V, Adeli K. Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations. Clin Chem Lab Med 2018;57:30–7.

46. Vidali M, Dittadi R, Ceriotti F, Brugnoni D, Giavarina D. Metodi indiretti per la determinazione e la verifica degli intervalli di riferimento. Bioc Clin 2023;47:221–35.

47. Ameratunga R, Lim D, Longhurst H, Mehrtens J, Leung E, Lehnert K, *et al.* Reference intervals for functional lymphocyte proliferation studies using 3H thymidine uptake in adults. J Immunol Methods 2025;542:113884.

48. Sansoni V, Lombardi G, Díaz-Garzón J, Calle PF, Bartlett WA, Coşkun A, *et al.*; European Federation of Clinical Chemistry and Laboratory Medicine Committee on Biological Variation. Novel biomarkers in bone pathophysiology: establishing reference intervals and biological variations estimates for serum leptin, sclerostin, lipocalin-2, osteoprotegerin, resistin and Dickkopf-related protein-1 from the European biological variation study (EuBIVAS) populations. Clin Chim Acta 2025;570:120213.

49. Carobene A, Campagner A, Uccheddu C, Banfi G, Vidali M, Cabitza F. The multicenter European Biological Variation Study (EuBIVAS): a new glance provided by the Principal Component Analysis (PCA), a machine learning unsupervised algorithms, based on the basic metabolic panel linked measurands. Clin Chem Lab Med 2021;60:556–68.

50. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, *et al.*; Biological Variation Working Group, European Federation of Clinical Chemistry and Laboratory Medicine. Sample collections from healthy volunteers for biological variation estimates' update: a new project undertaken by the Working Group on Biological Variation established by the European Federation of Clinical Chemistry and Laboratory Medicine. Clin Chem Lab Med 2016;54:1599–608.

51. Keleş M, Ünver Şeker G. Within- and between-subject biological variation data for whole blood HbA1c from 38 apparently healthy Turkish subjects. Scand J Clin Lab Invest 2024;84:535–9.

52. Klee GG, Ichihara K, Ozarda Y, Baumann NA, Straseski J, Bryant SC, *et al.* Reference Intervals: Comparison of Calculation Methods and Evaluation of Procedures for Merging Reference Measurements From Two US Medical Centers. Am J Clin Pathol 2018;150:545–54.

53. Ozarda Y. Establishing and using reference intervals. Turk J Biochem 2020;45:1–10.

54. Åsberg A, Mikkelsen G. A bivariate reference interval for TSH and free thyroxine. Scand J Clin Lab Invest 2024;84:521–6.

55. Åsberg A, Åsberg AE. Imperfect reference standards cause biased likelihood ratios. Scand J Clin Lab Invest 2025;85:361–7.

56. Higgins V, Nieuwesteeg M, Adeli K. Chapter 3 - Reference intervals: theory and practice. In Clarke W, Marzinke MA, editors. Contemporary Practice in Clinical Chemistry (Fourth Edition). Cambridge MA, USA: Academic Press; 2020. p. 37–56.

57. Timbrell NE. The Role and Limitations of the Reference Interval Within Clinical Chemistry and Its Reliability for Disease Detection. Br J Biomed Sci 2024;81:12339.

58. Rotgers E, Linko S, Theodorsson E, Kouri TT. Clinical decision limits as criteria for setting analytical performance specifications for laboratory tests. Clin Chim Acta 2023;540:117233.

59. Westgard JO. The Myth of Medical Decision Limits; 2009. [Internet]. Disponibile alla pagina: https://westgard.com/essays/trends/essay8.html [citato 29 giugno 2026].

60. Tetrault GA, Steindel SJ. Q-Probe 94-08. Daily quality control exception practices. Chicago: College of American Pathologists; 1994.

61. Ceriotti F. Risultati del questionario relativo ai limiti inferiori degli intervalli di riferimento pediatrici della fosfatasi alcalina. Bioch Clin 2025;49:226–30.

62. IFCC Reference Intervals and Decision Limits (C-RIDL); 2025 [Internet]. Disponibile alla pagina: https://ifcc.org/ifcc-scientific-division/sd-committees/c-ridl/ [citato 21 agosto 2026].

63. van Schrojenstein Lantman M, van Berkel M, Kuijper P, Langelaan M, Brouwer N, Thelen M. Clinical decision-making suffers from inequivalent measurement results and inadequate reference intervals. Clin Chem 2024;70:1383–92.

64. van Schrojenstein Lantman M. How measurement uncertainty impedes clinical decision-making; 2024 [Internet]. Disponibile alla pagina: https://hdl.handle.net/2066/311846 [citato 21 agosto 2026].

65. Ozarda Y, Sikaris K, Streichert T, Macri J. IFCC Committee on Reference intervals and Decision Limits (C-RIDL). Distinguishing reference intervals and clinical decision limits - A review by the IFCC Committee on Reference Intervals and Decision Limits. Crit Rev Clin Lab Sci 2018;55:20–431.

66. Cesana BM, da Molin S, Jassam N, Barth JH, Buoro S, Tosi M, *et al.* Assessing the harmonization of current total vitamin B12 measurement methods: relevance and implications. Clin Chem Lab Med 2025;63:2433–41.

67. NICE guideline. 6 March 2024. Vitamin B12 deficiency in over 16s: diagnosis and management. London: National Institute for Health and Care Excellence (NICE); 2024 [Internet]. Disponibile alla pagina: https://www.nice.org.uk/guidance/ng239 [citato 21 agosto 2026].

68. Aravinda I, Dharmasiri S, Sewwandi C, Karunaithas S, Goonetilleke S, Rasaratnam K. Population-adjusted cut-off: A new approach for enhancing the diagnostic efficacy of hematological discrimination formulae for screening β-Thalassemia trait. Clin Chim Acta 2026;579:120592.

69. Bogdanova A, Kaestner L, Simionato G, Wickrema A, Makhro A. Heterogeneity of Red Blood Cells: causes and Consequences. Front Physiol 2020;11:392.

70. Taher AT, Musallam KM, Cappellini MD. β-Thalassemias. N Engl J Med 2021;384:727–43.

71. Bohn MK, Nyholt D, Balion C, Cembrowski G, Collier C, Guire

V, *et al.* Best practice guidelines on reference interval harmonization in Canada: evidence-based recommendations from the CSCC working group on reference interval harmonization (CSCC WG-hRI). Clin Biochem 2025;139:110986.

72. Bohn MK, Bailey D, Balion C, Cembrowski G, Collier C, De Guire V, *et al.* Reference Interval Harmonization: Harnessing the Power of Big Data Analytics to Derive Common Reference Intervals across Populations and Testing Platforms. Clin Chem 2023;69:991–1008.

73. Cheng CY, Hsu TH, Yang YL, Huang YH. Hemoglobin and Its Z Score Reference Intervals in Febrile Children: A Cohort Study of 98,572 Febrile Children. Children (Basel) 2023;10:1402.

74. Iba T, Maier CL, Scarlatescu E, Levy JH. Introducing the New Definition and Diagnostic Criteria of Disseminated Intravascular Coagulation Released by the International Society on Thrombosis and Haemostasis in 2025. Semin Thromb Hemost 2026;52:10–7.

75. Suzuki K, Wada H, Imai H, Iba T, Thachil J, Toh CH; Subcommittee on Disseminated Intravascular Coagulation. A re-evaluation of the D-dimer cut-off value for making a diagnosis according to the ISTH overt-DIC diagnostic criteria: communication from the SSC of the ISTH. J Thromb Haemost 2018;16:1442–4.

76. Ceriotti F, Hinzmann R, Panteghini M. Reference intervals: the way forward. Ann Clin Biochem 2009;46:8–17.

77. Solberg HE. The theory of reference values Part 5. Statistical treatment of collected reference values. Determination of reference limits. J Clin Chem Clin Biochem 1983;21:749–60.

78. Claudel SE, Waikar SS, Verma A. Albuminuria References Ranges in Healthy US Adults. J Appl Lab Med 2025;10:767–71.

79. Grams ME, Coresh J, Matsushita K, Ballew SH, Sang Y, Surapaneni A, *et al.*; Writing Group for the CKD Prognosis Consortium. Estimated Glomerular Filtration Rate, Albuminuria, and Adverse Outcomes: An Individual-Participant Data Meta-Analysis. JAMA 2023;330:1266–77.

80. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing GJ, Harjola VP, *et al.*; ESC Scientific Document Group. 2019 ESC Guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS). Eur Heart J 2020;41:543–603.

81. Matthews TM, Peters GA, Wang G, Horick N, Chang KE, Harshbarger S, *et al.* Optimal cutoff values and utility of high-sensitivity troponin T and NT-proBNP for the risk stratification of patients with acute pulmonary embolism. Clin Chem 2025;71:559–66.

82. Dadelahi A, Walker BS, Banerjee D, Mahler M, Saadalla A, Nandakumar V. Is the 99th Percentile Cutoff Still Relevant? A Single-Center Assessment of Different Thresholds for Diagnosing Antiphospholipid Syndrome. J Appl Lab Med 2025;10:605–18.

83. Barbhaiya M, Zuily S, Naden R, Hendry A, Manneville F, Amigo MC, *et al.*; ACR/EULAR APS Classification Criteria Collaborators. The 2023 ACR/EULAR Antiphospholipid Syndrome Classification Criteria. Arthritis Rheumatol 2023;75:1687–702.

84. Vanoverschelde L, Kelchtermans H, Musial J, de Laat B, Devreese KM. Influence of anticardiolipin and anti-β2 glycoprotein I antibody cut-off values on antiphospholipid syndrome classification. Res Pract Thromb Haemost 2019;3:515–27.

85. Arun AS, Durant TJ, El-Khoury JM, Krumholz HM. Reevaluating the threshold for low total testosterone. Clin Chem 2025;71:609–11.

86. Mulhall JP, Trost LW, Brannigan RE, Kurtz EG, Redmon JB, Chiles KA, *et al.* Evaluation and Management of Testosterone Deficiency: AUA Guideline. J Urol 2018;200:423–32.

87. Bhasin S, Brito JP, Cunningham GR, Hayes FJ, Hodis HN, Matsumoto AM, *et al.* Testosterone Therapy in Men With Hypogonadism: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab 2018;103:1715–44.

88. Yeh CM, Yu CC, Lee AF, Cheng SM, Hsu CW, Chiang NJ, *et al.* A New CA19-9 Cutoff Value Identifies Lewis Antigen Status and Refines Prognostic Stratification in PDAC. Clin Cancer Res 2026;32:2722–31.

89. Barouqa M, Ashfaq MA, Jiang KJ, Kilic H, Dela Cruz N, Khalid A, *et al.* Establishing Reference Intervals for White Blood Cell and Absolute Neutrophil Counts in Duffy Null Individuals. Hematol Rep 2026;18:38.

90. Štrkalj G, Pather N. Beyond the Sex Binary: toward the inclusive anatomical sciences education. Anat Sci Educ 2021;14:513–8.

91. Nolan BJ, Cheung AS. Laboratory monitoring in transgender and gender-diverse individuals. Clin Chem 2025;71:358–77.

92. Krasowski MD, Hines NG, Imborek KL, Greene DN. Impact of sex used for assignment of reference intervals in a population of patients taking gender-affirming hormones. J Clin Transl Endocrinol 2024;36:100350.

93. Humble RM, Greene DN, Schmidt RL, Winston McPherson G, Rongitsch J, Imborek KL, *et al.* Reference Intervals for Clinical Chemistry Analytes for Transgender Men and Women on Stable Hormone Therapy. J Appl Lab Med 2022;7:1131–44.

94. Irwig MS. Which reference range should we use for transgender and gender diverse patients? J Clin Endocrinol Metab 2021;106:e1479–80.

95. Davis D. The Importance of Accurate Laboratory Reference Intervals for Transgender Patients. ASCLS Today; 2022 [Internet]. Disponibile alla pagina: https://ascls.org/the-importance-of-accurate-laboratory-reference-intervals-for-transgender-patients/ [citato 21 agosto 2026].

96. Jessen K, Wijeratne N, Connell A. The intersection of the laboratory and transgender care. Crit Rev Clin Lab Sci 2025;62:347–62.

97. Cattaneo A, Bellenghi M, Ferroni E, Mangia C, Marconi M, Rizza P, *et al.*; Ministry Of Health-Gender Medicine Team. Recommendations for the Application of Sex and Gender Medicine in Preclinical, Epidemiological and Clinical Research. J Pers Med 2024;14:908.

98. CLSI. Framework for Developing Evidence of Clinical Validity of Medical Laboratory Test Methods. 1st ed. CLSI guideline EP49. Clinical and Laboratory Standards Institute; 2026.

99. ISO 20916:2019. In vitro diagnostic medical devices – Clinical performance studies using specimens from human subjects – Good study practice. Geneva, Switzerland: International Organization for Standardization; 2019.

100. Associazione Medici Endocrinologi. Quoziente di probabilità. 2022. [Internet]. Disponibile alla pagina: https://www.associazionemediciendocrinologi.it/index.php/glossario-statistico/quoziente-di-probabilita [citato 21 agosto 2026].

101. Giavarina D. Evidence Based Medicine: impiego dei test di laboratorio – Caleidoscopio Italiano n.229. Genova: Medical Systems; 2014.

102. Burlina AB. Medicina di laboratorio – Fondamenti di diagnostica. Torino: CG Edizioni MedicoScientifiche, 1992.

103. Pum JK. Evaluation of analytical performance of qualitative and semi-quantitative assays in the clinical laboratory. Clin Chim Acta 2019;497:197–203.

104. Medical Device Innovation Consortium. Real-World Clinical Evidence Generation: Advancing Regulatory Science and Patient Access for In Vitro Diagnostics (IVDs). Medical Device Innovation Consortium; 2024. [Internet]. Disponibile alla pagina: https://mdic.org/resources/real-world-clinical-evidence-generation-advancing-regulatory-science-and-patient-access-for-in-vitro-diagnostics-ivds/ [citato 21 agosto 2026].

105. ISO/IEC Guide 99:2007. International vocabulary of metrology — Basic and general concepts and associated terms (VIM). Geneva, Switzerland: International Organization for Standardization; 2007.

106. JCGM 200:2012. International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition 2008 version with minor corrections; 2008 [Internet]. Disponibile alla pagina: https://www.bipm.org/documents/20126/2071204/JCGM_200_2012.pdf [citato 21 agosto 2026].

107. Possolo A, Hibbert D, Stohner J, Bodnar O, Meija J. A brief guide to measurement uncertainty (IUPAC Technical Report). Pure Appl Chem 2024;96:113–34.

108. Global Harmonization Task Force. Clinical Evidence for IVD medical devices — Key Definitions and Concepts [GHTF/SG5/N6:2012]. [Internet]. Disponibile alla pagina: https://www.imdrf.org/sites/default/files/

docs/ghtf/final/sg5/technical-docs/ghtf-sg5-n6-2012-clinical-evidence-ivd-medical-devices-121102.pdf [citato 21 agosto 2026].

109. Torlakovic EE, Al Dieri R, Badrick T, Chen ZE, Cheung CC, Deans Z, *et al.* Indirect clinical validation for predictive biomarkers in oncology: International Quality Network for Pathology (IQN Path) Position Paper. Virchows Arch 2025;487:565–72.

110. CLSI. Evaluation of qualitative, binary output examination performance. CLSI guideline EP12. Malvern, PA: Clinical and Laboratory Standards Institute; 2023.

111. ISO/IEC/IEEE 24765:2017. Systems and software engineering — Vocabulary. Geneva, Switzerland: International Organization for Standardization; 2017.

112. ISO/IEC 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models. Geneva, Switzerland: International Organization for Standardization; 2011.

113. ISO 5649:2024. Medical laboratories — Concepts and specifications for the design, development, implementation and use of laboratory-developed tests. Geneva, Switzerland: International Organization for Standardization; 2024.

114. European Commission; Medical Device Coordination Group. MDCG 2022-2 Guidance on general principles of clinical evidence for In Vitro Diagnostic medical devices (IVDs); 2022 [Internet]. Disponibile alla pagina: https://health.ec.europa.eu/latest-updates/mdcg-2022-2-guidance-general-principles-clinical-evidence-vitro-diagnostic-medical-devices-ivds-2022-01-27_en [citato 29 giugno 2026].

115. Unione europea. Regolamento (UE) 2017/746 del Parlamento europeo e del Consiglio, del 5 aprile 2017, relativo ai dispositivi medico-diagnostici in vitro e che abroga la direttiva 98/79/CE e la decisione 2010/227/UE della Commissione. Document 32017R0746. Legal act; 2017 [Internet]. Disponibile alla pagina: https://eur-lex.europa.eu/eli/reg/2017/746/oj [citato 29 giugno 2026].

Conflitti di interesse
L'autore dichiara di non avere conflitti di interesse con alcuna ditta legata al contenuto del manoscritto.

Studi condotti su esseri umani e animali
Per questo tipo di studio non è richiesto l'inserimento di alcuna dichiarazione relativa agli studi effettuati su esseri umani e animali.

Consenso informato
Per questo tipo di studio non è richiesto il consenso informato.

Cronologia
Pubblicato online: 11 settembre 2026. - Accettato: 21 agosto 2026. - Ricevuto: 2 luglio 2026.