



Società Italiana di Patologia Clinica e Medicina di Laboratorio

Componente della World Association of Societies of Pathology
and Laboratory Medicine www.sipmel.it

Certificazione ISO



**Commissione Nazionale Qualità
ed Accreditemento
GdS-Informatica**
Coordinatore: Marco Pradella

**GdS-Evidence Based Laboratory
Medicine**
Coordinatore: Sofia Chiatamone
Ranieri
GdS-Management sanitario
Coordinatore: Giovanni Casiraghi

Versione 0.1
**documento M2-2
relazioni**
approvato dal
Consiglio Nazionale
del 5/10/2025

**M2 Raccomandazioni per Accreditemento
ISO 15189 del laboratorio medico - requisiti
per le relazioni con pazienti e medici curanti,
parte seconda: intervalli e limiti**

*M2 Recommendations for Medical
Laboratory ISO 15189 Accreditation -
requirements for relationships with patients
and treating physicians, part two: intervals
and limits*

Codifica di questo documento:

Flusso Operativo	Elementi fondamentali del sistema qualità	
Pre-esame Richiesta di esame A	Documenti e Registri L	
Prelievo B	Organizzazione M	#
Trasporto del campione C	Personale N	
Accettazione e trattamento del campione D	Strumentazione O	
Esame Analisi E	Acquisti e gestione scorte P	
Revisione e flusso dei risultati F	Controllo del processo Q	
Interpretazione di laboratorio G	Gestione delle informazioni R	
Post-esame Trasmissione e archiviazione del risultato H	Gestione degli inconvenienti S	
Conservazione e smaltimento del campione I	Verifiche T	
	Miglioramento del processo U	
	Servizio e Soddisfazione V	
	Impianti e sicurezza Z	

Riferimenti normativi

UNI EN ISO 15189:2025. Laboratori medici - Requisiti riguardanti la qualità e la competenza.

UNI EN ISO 22367:2020. Laboratori medici - Applicazione della gestione del rischio ai laboratori medici

UNI CEI EN ISO/IEC 17025:2018. Requisiti generali per la competenza dei laboratori di prova e taratura

UNI ISO/TS 22583:2021. Guida per supervisori e operatori di dispositivi di test point-of-care (POCT)

ISO 15190:2020. Medical laboratories - Requirements for safety.

UNI EN ISO 22300:2021 Sicurezza e resilienza - Vocabolario

ISO/DIS 37009. Conflict of interest in organizations — Guidelines

UNI ISO 37001:2016 Sistemi di gestione anticorruzione - Requisiti con guida all'uso

UNI ISO 37000:2021 Governance delle organizzazioni – Guida

UNI ISO 28003:2020 Sistemi di gestione per la sicurezza della catena logistica - Requisiti per gli organismi che forniscono audit e certificazione dei sistemi di gestione per la sicurezza della catena logistica

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

indice

Codifica di questo documento:	2
Riferimenti normativi.....	3
Introduzione: la revisione di ISO 15189	5
NOTA di ortografia.....	6
Intervalli di riferimento e limiti decisionali clinici	7
NOTA. Intervalli e limiti nella pratica del laboratorio medico e le guide CLSI	7
NOTA. Costruire gli intervalli di riferimento, la guida aggiornata CLSI	8
Considerazioni sulla guida CLSI EP44.....	14
Figura 1. Criticità per il laboratorio sulla discriminazione con gli intervalli di riferimento	16
NOTA. Guida CLSI per utilizzo di intervalli di riferimento nel laboratorio medico.....	16
Metodi “indiretti”	20
Considerazioni sulla guida CLSI EP45.....	21
NOTA. Mondo reale fonte di informazioni mediche	22
ISO 18727: dati sanitari nel mondo reale.....	22
Produzione di intervalli e limiti decisionali	25
NOTA. Ripartizione per sesso e/o genere.....	34
NOTA. Intervalli di riferimento nella validazione clinica: CLSI EP49 e ISO 20916	35
CLSI EP49: le prestazioni cliniche	35
ISO 20916:2019: gli studi di prestazione clinica	36
NOTA. Prestazione clinica e validazione clinica nella letteratura scientifica.....	37
Conclusioni e Raccomandazioni	41

Introduzione: la revisione di ISO 15189

SIPMeL ha già affrontato il tema delle risorse di personale, ambiente e attrezzature, anche per i servizi POCT, i punti prelievo e i laboratori in rete, con le Raccomandazioni dei documenti Q13 (POCT)^{1,2}, Q12 (prelievi)^{3,4}. Si deve tornare sull'argomento alla luce della revisione della norma ISO per l'accreditamento dei laboratori medici, affrontando in modo sistematico i temi delle risorse di personale, ambiente, attrezzature, consumabili, accordi di servizio.

La norma ISO 15189 revisionata è stata pubblicata il 6 dicembre 2022, quasi un anno dopo il termine previsto, e recepita da UNI in Italia poco tempo dopo.⁵ Il processo di revisione è stato lungo e faticoso, iniziato già in ottobre 2018⁶ ha attraversato molteplici versioni della bozza e altrettante votazioni. Dai vertici ISO e del Comitato tecnico competente (ISO/TC 212) sono state ricevute, tra le altre, alcune importanti direttive: usare ISO/IEC 17025:2017 come modello, incorporare ISO 22870 (la norma per i Point-of-care), stabilire collegamenti con ISO 15190 (salute e sicurezza), ISO 22367 (gestione dei rischi) e ISO/TS 20658 (fase pre-esame), ridurre i requisiti prescrittivi per basarsi sul rischio per il paziente, prendere in considerazione altri documenti ISO pubblicati pertinenti, con l'obiettivo di evitare anche ripetizioni ridondanti, sincronizzando le clausole pertinenti in ISO 15190, ISO 22367, ISO TS 20658, ISO 17511 (taratura), ISO TS 20914 (incertezza di misura) e la serie di standard diagnostici molecolari sviluppato da ISO TC. Tuttavia, le innovazioni chieste da ISO si sono scontrate nel processo di revisione con non poche resistenze, con il risultato della presenza nel testo finale di diversi compromessi, potenziali difficoltà per laboratori medici e ispettori di accreditamento.⁷ Un particolare impegno a laboratori e ispettori è richiesto dalla disposizione di collegare ISO 15189 a numerosi altri documenti ISO pertinenti, al fine di evitare ridondanze.

In questa nota di raccomandazioni vengono presentati i punti salienti di UNI EN ISO 15189:2023 per le comunicazioni con pazienti e medici, in particolare il capitolo 4 (imparzialità, riservatezza, requisiti dei pazienti), i punti 7.3.5 (Intervalli biologici di riferimento e limiti decisionali clinici), 7.3.6 (Documentazione delle procedure di esame), 7.4.1 (risultati degli esami).

1. SIPMeL. Raccomandazioni per l'inclusione dei requisiti degli esami eseguiti vicino al paziente (point-of-care testing, ISO 22870) nei requisiti dei laboratori medici (ISO 15189). 07/10/2019. Documento Q13-POCT ISO 22870 in ISO 15189. <https://www.sipmel.it/lineeguida/approvate/115525>

2. Pradella M. Accredimento dei POCT con la nuova ISO 15189 e ISO 22583: le raccomandazioni SIPMeL. La Rivista Italiana della Medicina di Laboratorio 2019 settembre;15(3):225-32. DOI: 10.23736/S1825-859X.19.00037-9
3. SIPMeL. Raccomandazioni per l'accreditamento di punti di prelievo e laboratori in rete. 07/10/2019. Documento Q12-BC1 punti prelievo e laboratori in rete. <https://www.sipmel.it/lineeguida/approvate/116056>

4. Pradella M. Accredimento ISO di punti di prelievo e laboratori in rete. La Rivista Italiana della Medicina di Laboratorio 2020 marzo;16(1):50-9. DOI: 10.23736/S1825-859X.20.00046-8

5. UNI EN ISO 15189:2023. Laboratori medici - Requisiti riguardanti la qualità e la competenza. Data disponibilità: 02 marzo 2023

6. Pradella M. Requisiti dei laboratori medici, forensi, antidoping e alimentari: nuove ISO 15189 e ISO 17025. La Rivista Italiana della Medicina di Laboratorio 2019 dicembre;15(4):252-62. DOI: 10.23736/S1825-859X.19.00033-1

7. Pradella M. New ISO standards for medical biology laboratories, prescriptions and deviations. Volume 80, issue 5, September-October 2022. Annales de Biologie Clinique. 2022;80(5):451-453. doi:10.1684/abc.2022.1755

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

NOTA di ortografia.

Come descritto dalla Accademia della Crusca,⁸ la lingua italiana si comporta diversamente dall'inglese. Per i prefissati con anti-, auto-, pre-, vice- i dizionari mostrano la tendenza a mantenere il trattino solo per gli anglicismi. Il trattino si mostra più resistente nelle formazioni del tipo post- + nome, in particolare con termini specialistici. Tuttavia, Treccani⁹ elenca numerosi esempi senza trattino (postpliocene, postmoderno, postoperatorio, postbellico, postludio, postmaturo, postdiluviano, postdatore), decisamente più eleganti in italiano. Anche il Dizionario De Mauro riporta esempi di parole senza trattino: prelavaggio, prenatale, presalarario, presenile, prescientifico, prefrontale.¹⁰ Ma anche: postbellico, postindustriale, postoperatorio, postrisorgimentale, postcomunismo.¹¹

In questo documento, perciò, si preferirà l'atteggiamento prescrittivo di Treccani e De Mauro a quello descrittivo della Crusca e verranno usati i termini "preesame" e "postesame".

Si cercheranno inoltre di evitare per quanto possibile anglicismi e usi errati di falsi amici inglesi della lingua italiana. In ottemperanza alla legge vigente, dove si stabilisce che la lingua ufficiale della Repubblica è l'italiano.¹² Si prenderanno se necessario come ispirazioni le soluzioni individuate da lingue romanze come la lingua francese e quella spagnola.

Per esempio, il verbo "implement" viene sovente erroneamente ricalcato nella lingua italiana con il significato di "aumentare, migliorare". Ancora, la categoria "pre-analytical" viene talvolta ricalcata erroneamente con il significato di "trattamento preparatorio" di una fase del processo di esame, quando si tratta di una sequenza articolare, generando un'ambiguità ulteriore con i processi preesame, che invece si riferiscono al prelievo e al trasporto del campione in laboratorio. Aggiuntiva alla nota ambiguità tra attività di "misura" e quelle di "analisi", bene esposta nel Vocabolario Internazionale.

8. Raffaella Setti. Il trattino: quando usarlo? Redazione Consulenza Linguistica Accademia della Crusca 2009. <https://accademiadellacrusca.it/it/consulenza/il-trattino-quando-usarlo/249>

9. Istituto della Enciclopedia Italiana fondata da Giovanni Treccani. Post- Vocabolario on line. <https://www.treccani.it/vocabolario/post/>

10. Nuovo De Mauro. Pre- <https://dizionario.internazionale.it/parola/pre->

11. Nuovo De Mauro. Post- <https://dizionario.internazionale.it/parola/post->

12. Legge 15 dicembre 1999, n. 482. Norme in materia di tutela delle minoranze linguistiche storiche. GU n.297 del 20-12-1999. <https://www.normattiva.it/uri-res/N2Ls?urn:nir:stato:legge:1999;482>

Intervalli di riferimento e limiti decisionali clinici

Il punto 7.3.5 di ISO 15189:2022 è interamente dedicato a intervalli biologici di riferimento (IR) e limiti decisionali clinici (LD). Se necessari per l'interpretazione dei risultati degli esami, gli intervalli di riferimento biologici e i limiti decisionali clinici devono essere definiti e comunicati agli utenti. Intervalli e limiti devono riflettere la popolazione di pazienti servita dal laboratorio, tenendo conto del rischio per i pazienti. Per gli esami che identificano la presenza o l'assenza di una caratteristica, l'intervallo biologico di riferimento è la caratteristica da identificare, ad esempio gli esami genetici.

Per ISO 15189 i valori forniti dal fabbricante dei dispositivi, sistemi e reagenti devono essere verificati e solo se la base della popolazione da cui sono tratti è verificata e ritenuta accettabile dal laboratorio possono essere utilizzati. Non solo, intervalli e limiti vanno revisionati periodicamente, ogni modifica deve essere comunicata agli utenti.

Il laboratorio deve esaminare l'impatto sugli intervalli e sui limiti decisionali quando vengono apportate modifiche a un esame o a un metodo di raccolta del campione, comunicarlo agli utenti, se del caso.

NOTA. Intervalli e limiti nella pratica del laboratorio medico e le guide CLSI

Intervalli e limiti costituiscono una parte rilevante delle informazioni date dal laboratorio agli utenti, pazienti e medici. Informazione rinnovata sia nel caso del cambio del metodo di esame o prelievo che a seguito di una revisione periodica. ISO sembra ritenere il contributo dei fabbricanti una eventualità quasi residua, come se il compito di definizione degli intervalli ricadesse innanzitutto sul laboratorio.

Gli intervalli di riferimento (IR) non sono esclusivi dei risultati quantitativi numerici, ma vanno definiti e comunicati anche per quelli qualitativi nominali, come accade in genetica o in virologia. L'utilità dei limiti decisionali per l'interpretazione dei risultati è ovviamente indiscutibile. Quella degli intervalli di riferimento invece molto meno. Gli intervalli non sono collegati alle patologie che giustificano la richiesta di esami, bensì al comportamento degli individui privi di patologie. Resta tuttavia l'importante funzione degli intervalli a rinforzo di altre caratteristiche dei risultati quantitativi degli esami, come la scelta delle unità di misura e la tracciabilità del metodo dovuta alla taratura.

La revisione 2024 del documento CLSI H20 Conta differenziale dei leucociti di riferimento (proporzionale) e valutazione dei metodi strumentali¹³ (in questo momento non ancora approvata) offre al punto 3.3.8 una interessante interpretazione dell'intervallo di riferimento, molto diversa dalla precedente edizione di Koepke e collaboratori.

Le raccomandazioni di H20 prendono in considerazione solo i leucociti trovati in individui sani (non malati), ovvero neutrofili (segmentati), neutrofili (forme a banda), linfociti (normali), linfociti (forme reattive), monociti, eosinofili e basofili. Se una cellula non rientra in questi gruppi, si segnala come anormale, sospetta o non classificata, oppure come globuli rossi nucleati (NRBC). CLSI H20 si limita alle cellule che circolano nel flusso sanguigno umano di individui sani, oltre a una categoria per altre cellule che, se rilevate, vengono segnalate per un ulteriore esame o valutazione (1.2.3 Limitazioni). La segnalazione di campioni anormali per l'esame visivo è parte della valutazione delle prestazioni. Le opinioni divergono sugli intervalli di riferimento accettabili per i leucociti, complicando questa identificazione. Esiste inoltre un disaccordo sull'opportunità di

13. CLSI. Reference Leukocyte (WBC) Differential Count (Proportional) and Evaluation of Instrumental Methods. 3rd ed. CLSI Standard H20. Clinical and Laboratory Standards Institute; 2024.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

un conteggio proporzionale (percentuale) rispetto a quello assoluto. H20 2024 riprende le indicazioni di CLSI H26¹⁴, secondo cui a differenza di alcuni esami immunologici e chimici, i contaglobuli ematologici non dovrebbero ottenere intervalli di riferimento dal fabbricante, ma i fabbricanti possono fornire informazioni sugli intervalli di riferimento; quindi, ogni laboratorio deve valutare tali informazioni in base alla propria popolazione di pazienti. Gli intervalli di riferimento specifici per lo strumento, per tutti i componenti dell'esame emocromocitometrico, devono essere calcolati durante la valutazione dello strumento. Idealmente, almeno 120 campioni provenienti da individui apparentemente sani, con valori nei limiti prefissati per l'emocromo e i parametri differenziali, vengono esaminati entro quattro ore dal prelievo. Se possibile, devono essere esaminati diversi gruppi etnici e di genere. Occorre prestare particolare attenzione alla tecnica di prelievo, al tempo di applicazione del laccio emostatico e al flusso di sangue dalla vena per evitare l'emoconcentrazione e la coagulazione. Se appropriato, devono essere calcolati intervalli per neonati e bambini di età diverse. Applicare metodi statistici appropriati (basati sulla distribuzione dei dati) per determinare media, varianza e intervalli di confidenza al 95% per ciascun gruppo. I campioni restanti dopo gli esami possono essere utilizzati se tutti gli esami richiesti sono stati completati, ma deve essere richiesto il consenso informato in linea con le linee guida etiche regionali. Il punto 3.3.6 si occupa della capacità di rilevare campioni anormali. A parte l'uso errato dei termini "sensibilità" e "specificità", affetti da ambiguità nei confronti delle corrispondenti prestazioni cliniche, H20 vuole utilizzare intervalli di riferimento stabiliti (quelli del punto 3.3.8). Possiamo tuttavia notare la circolarità della procedura. Se gli intervalli sono prodotti con un criterio statistico fissato, la capacità di non rilevare gli anomali in campioni non anomali ("selettività") è implicita nel criterio statistico stesso. La sua controparte (capacità di rilevare gli anomali nei soggetti con patologie, spesso confusa con la "sensibilità clinica") richiederebbe lo studio dei gruppi con patologia, non prevista nel punto 3.3.8 di H20.

NOTA. Costruire gli intervalli di riferimento, la guida aggiornata CLSI

Il documento Clinical and Laboratory Standards Institute EP44-Establishing Reference Intervals¹⁵ è destinato a sviluppatori e ricercatori che creano intervalli di riferimento (RI) per esami di laboratorio con risultati quantitativi, escludendo quelli con risultati qualitativi ovvero classificatori, diversamente dalle indicazioni ISO. Contiene protocolli che soddisfino requisiti minimi, sia per un nuovo esame che per un nuovo metodo di misurazione.

CLSI EP44 si rivolge a ricercatori e sviluppatori (sia fabbricanti che laboratori che creano esami in casa), fornisce una guida per intervalli di riferimento (RI) di misurandi endogeni, non per farmaci o tossici esogeni, ad esempio. Descrive il disegno dello studio, la selezione dei soggetti e i metodi di analisi dei dati, con approcci non parametrici, parametrici e robusti. Infine, fornisce esempi di presentazione dei risultati.

CLSI EP44 sostituisce CLSI EP28-A3c, pubblicato nel 2008, con aggiornamento della definizione degli intervalli, rimando della guida per l'utente finale per la verifica o il trasferimento degli intervalli al documento diverso CLSI EP45¹⁶, introduzione del metodo quantile-quantile per

14. CLSI. Validation, Verification, and Quality Assurance of Automated Hematology Analyzers; Approved Standard — Second Edition. CLSI document H26-A2. Clinical and Laboratory Standards Institute; 2010.

15. CLSI. Establishing Reference Intervals. 4th ed. CLSI guideline EP44. Clinical and Laboratory Standards Institute; 2025.

16. CLSI. Implementation of Reference Intervals in the Medical Laboratory. 4th ed. CLSI guideline EP45. Clinical and Laboratory Standards Institute; 20XX.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

L'analisi parametrica dei dati.

CLSI EP44 vuole presentare le procedure per determinare intervalli di riferimento “associati alla salute”, con tecnica di campionamento “a priori” o “a posteriori”. A priori è applicazione dei criteri di selezione degli individui di riferimento prima della raccolta dei campioni. Mentre a posteriori è applicazione dei criteri di selezione degli individui di riferimento dopo la raccolta dei campioni.

EP44 ammette che la salute è una condizione relativa che manca di una definizione universale. La soluzione di EP44 è fondamentalmente quella di riferimenti di letteratura risalenti al 1987.^{17,18} CLSI EP44 non copre i metodi indiretti che utilizzano grandi insiemi di dati con poche informazioni sui pazienti che consentano inclusione o esclusione. Il comitato CLSI riconosce la letteratura correlata pertinente, tuttavia ritiene l'argomento troppo complesso per questa linea guida e rimanda a CLSI EP45 anche i metodi indiretti per gli intervalli di riferimento. EP44 adotta un approccio molto tradizionale: chiede risultati di osservazione, misurazione o calcolo dagli individui di riferimento del gruppo di campioni di riferimento, calcola l'intervallo centrale di valori (ad esempio, il 95% centrale).

EP44 afferma che quando si inizia uno studio di validazione o di verifica, è necessario un progetto che descriva in dettaglio tutti gli elementi dello studio (ad esempio, le modalità di esecuzione, il numero di campioni, il metodo di esame e i criteri di accettazione) (Capitolo 2: Considerazioni generali). Gli elementi chiave sono sia il disegno dello studio sia le analisi dei dati. Lo sviluppatore deve considerare che il reclutamento di una popolazione di riferimento sana con un metodo diretto migliora la qualità del gruppo selezionato e dei valori di riferimento risultanti, ma richiede un questionario per inclusione ed esclusione. Criteri di suddivisione, come il sesso^{NOTA} e l'età, possono portare a selezionare meglio le popolazioni di soggetti sani e per l'interpretazione clinica. La dimensione del campione deve essere presa in considerazione per ciascun sottogruppo, al fine di determinare intervalli centrali del 95% separati per ciascuno di essi, insieme a intervalli di confidenza (CI) derivati statisticamente. Infine, si deve utilizzare la tecnica di analisi statistica appropriata.

EP44 prende in considerazione gli studi continui covariati e univariati, dove le linee dei limiti di riferimento vengono smussate, graficamente e/o per calcolo, per rappresentare una funzione continua dell'età. Tuttavia, le presentazioni tabellari, per fasce d'età, si ritengono il modo quasi universale con cui i laboratori medici trasmettono i limiti di riferimento attraverso i sistemi informativi. EP44 riconosce difficoltà per questo approccio, come la scelta della covariata (sesso alla nascita, stato di gravidanza, la fase del ciclo mestruale o la massa corporea), la larghezza delle fasce d'età, la gestione dei punti isolati (outliers).

17. Solberg H. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. Clin Chim Acta. 1987;170(2-3):S13-S32.

doi:10.1016/0009-8981(87)90151-3

18. Solberg H, PetitClerc C. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. Clin Chim Acta. 1988;177(3):S3-S11.

doi: 10.1016/0009-8981(88)90074-5

NOTA Sesso e genere sono due termini diversi ma interconnessi e non vanno confusi, come accade talvolta. CLSI EP44 e molta letteratura scientifica usano la parola “sex”, mentre il vocabolario Treccani riporta per “genere” 15 diversi significati, raccolti in 5 gruppi (<https://www.treccani.it/vocabolario/genere/>), tra cui (4.b) “carattere maschile o femminile dell'individuo, anche in senso biografico, sociale, professionale”. WHO distingue sesso cromosomico e biologico dal genere costruzione sociale (https://www.who.int/health-topics/gender#tab=tab_1). Lo stesso fa il Consiglio d'Europa (<https://www.coe.int/en/web/gender-matters/sex-and-gender>). In questo documento si usa quindi la parola “sesso” nel senso cromosomico-biologico inteso dagli studi fondamentali per IR e LD, mentre degli stati di confine nonché delle conseguenze sul laboratorio del genere psico-sociale viene annotato a parte.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

Un aspetto critico del processo è il numero di soggetti. EP44 ritiene che uno studio che coinvolge 120 soggetti sia piccolo, mentre per ottenere stime di intervalli migliori vengono spesso condotti studi molto più grandi, che coinvolgono diverse centinaia di soggetti o anche di più. Per alcuni esami e/o per alcune popolazioni (per esempio, la pediatria), però, anche 120 soggetti potrebbero essere molto difficili da reclutare. EP44 ritiene che studi di piccole dimensioni, anche basati su un numero di soggetti pari a 60 o 80, potrebbero comunque essere di grande valore, almeno a scopo orientativo. (punto 3.2)

Caso particolare è quello del metodo non parametrico, dove è impossibile distinguere tra due percentili (p) di una distribuzione se non si ottengano almeno $n = (100/p) - 1$ osservazioni. Il metodo non parametrico si basa esclusivamente sui ranghi della distribuzione di valori (ordinati dal più grande al più piccolo). Ulteriore complicazione è data dalla presenza di valori isolati (outliers). Con un minimo di 120 osservazioni, si consente il calcolo dei limiti di confidenza del 90% in modo non parametrico per ciascun limite di riferimento. Per questi stessi percentili con una confidenza del 95%, sono necessari almeno 146 valori; per una confidenza del 99%, sono necessari almeno 210 valori. Per distribuzioni dei risultati molto distorte, si dovrebbero ottenere fino a 700 valori.¹⁹

Se vengono rimosse osservazioni anomale (outliers), è necessario selezionare altri soggetti fino a ottenere almeno 120 valori accettabili. Inoltre, se sono necessari intervalli separati per diverse sottoclassi (ad esempio, sesso o età), ciascuno di questi deve essere basata sul minimo raccomandato di 120 osservazioni.

Lo strumento fondamentale per la raccolta dei dati è il Questionario (punto 3.3), indispensabile per applicare i criteri di esclusione e suddivisione o ripartizione. I questionari devono essere semplici da compilare per i partecipanti. Le domande dovrebbero richiedere per lo più risposte affermative o negative o risposte semplici ed esplicative. Devono essere incluse domande volte a scoprire informazioni su stati patologici che notoriamente influenzano il metodo o i metodi di prova in esame. Il questionario può essere utilizzato insieme a semplici misurazioni, come la pressione sanguigna, l'altezza e il peso. I soggetti di riferimento devono essere intervistati quando le informazioni critiche non possono essere ottenute con altri mezzi. EP44 si basa sull'esempio IFCC pubblicato nel 2013.²⁰

Il metodo EP44 si basa su meccanismi di esclusioni e suddivisione o ripartizione. Si ammette che non è possibile reclutare individui "apparentemente sani" senza una sufficiente valutazione medica di laboratorio di ciascun individuo. La richiesta di condizioni troppo restrittive per la "salubrità" rende difficile raccogliere un numero sufficiente di volontari. Quindi, al momento del reclutamento si possono utilizzare criteri di esclusione più indulgenti, seguiti da criteri di esclusione secondari basati su informazioni individuali (ad esempio, indicatore di stato di malattia, massa corporea, abitudini di consumo, fattori genetici) o sui risultati di altri esami. La Tabella 2 di EP44 elenca ben venti criteri, più i risultati di esami correlati, tra primari e secondari. La Tabella 3 di EP44 raccoglie quindici criteri di esclusione secondaria, ciascuno basato su esami di laboratorio e su elementi anamnestici.

Ancora più sofisticato è l'approccio CLSI EP44 al meccanismo di ripartizione (punto 3.4.3). Si ipotizza l'analisi di regressione multipla (MRA) o l'analisi della varianza (ANOVA). La Tabella 4. mette in lista quindici fattori di ripartizione utilizzabili.

19. Linnet K. Two-stage transformation systems for normalization of reference distributions evaluated. Clin Chem. 1987;33(3):381-386

20. Ozarda Y, Ichihara K, Barth JH, et al. Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values. Clin Chem Lab Med. 2013;51(5):1027-1040. doi:10.1515/cclm2013-0249

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

EP44 si preoccupa nel capitolo 4 dell'effetto di fattori preesame e dei metodi di misurazione. Tutti i fattori rilevanti, tra cui la preparazione dei soggetti, la raccolta e il trattamento dei campioni, le variabili del metodo di misura e della strumentazione, devono essere accuratamente definiti. La documentazione deve essere sufficientemente approfondita per comprendere i limiti dello studio e consentirne la ripetizione. Per quantificare le differenze sulle misure EP44 rimanda al confronto dei metodi di CLSI EP09²¹ o CLSI EP45²². Così si può valutare se le differenze osservate tra i metodi siano clinicamente significative e se sia necessario stabilire intervalli separati.

La Tabella 5 di EP44 elenca nove fattori relativi al soggetto, dieci relativi al prelievo, cinque relativi alla manipolazione dei campioni. Per i fattori preesame si rimanda al documento CLSI MM13, anche se relativo solo ai metodi molecolari.²³

Il capitolo 5 di EP44 estende molto la trattazione dell'analisi statistica dei valori di riferimento. La Tabella 6 illustra bene i metodi per l'analisi statistica dei valori di riferimento, distinguendo metodo non parametrico, parametrico e robusto, dando per ciascuno caratteristiche, vantaggi e limitazioni. Il classico metodo parametrico non può essere utilizzato se i valori non possono essere trasformati in una distribuzione gaussiana. Per il metodo robusto è necessario il supporto di un biostatistico.

Un tema particolarmente ostico è quello dei valori anomali (outliers, EP44 punto 5.1), Si tratta di valori così distanti dal resto del campione da far sospettare che provenga da una popolazione diversa dal resto.²⁴ A parte i criteri per individuarli, un grosso problema è quello del loro trattamento. Si possono eliminare e proseguire con la lettura restante, oppure analizzare i dati con il metodo non parametrico, o ancora utilizzare una procedura robusta in cui vengono "addomesticati" in modo da non contaminare le stime parametriche. Se non sono stati identificati errori nei risultati, occorre purtroppo indagare sui valori anomali. Potrebbero comprendere una sottopopolazione che potrebbe richiedere una propria partizione degli intervalli (ad esempio, un certo gruppo etnico) o una modifica dei criteri di esclusione (ad esempio, i soggetti che assumono lo stesso farmaco, originariamente non considerato come criterio di esclusione).

EP44, sulla base di una ricerca IFCC del 2017²⁵ e del 2023²⁶, sostiene che i limiti di riferimento parametricamente, spesso hanno una precisione molto maggiore dei metodi non parametrici.

Lo schema della metodologia quantile-quantile è descritto nella Appendice B di EP44.

I metodi parametrici quantile-quantile (PQQ) usano la correlazione tra due quantili: i valori (o le loro trasformazioni) e i punteggi normali (z-scores) della distribuzione gaussiana standard.

Ovvero la relazione quantile-quantile (QQ) collega i valori ordinati di un campione (ordinata) con i punteggi corrispondenti alla distribuzione che si ritiene seguano i dati (ascissa).^{27,28} L'ideatore

21. CLSI. Measurement Procedure Comparison and Bias Estimation Using Patient Samples. 3rd ed. CLSI guideline EP09c. Clinical and Laboratory Standards Institute; 2018.

22. CLSI. Implementation of Reference Intervals in the Medical Laboratory. 4th ed. CLSI guideline EP45. Clinical and Laboratory Standards Institute; 20XX.

23. CLSI. Collection, Transport, Preparation, and Storage of Specimens for Molecular Methods. 2nd ed. CLSI guideline MM13. Clinical and Laboratory Standards Institute; 2020.

24. Barnett V, Lewis T. Outliers in Statistical Data. Wiley; 1994.

25. Ichihara K, Ozarda Y, Barth JH, et al. A global multicenter study on reference values: 1. Assessment of methods for derivation and comparison of reference intervals. Clin Chim Acta. 2017;467:70-82. doi:10.1016/j.cca.2016.09.016

26. Ichihara K, Yamashita T, Kataoka H, Sato S. Critical appraisal of two Box-Cox formulae for their utility in determining reference intervals by realistic simulation and extensive real-world data analyses. Comput Methods Programs Biomed. 2023;242:107820. doi:10.1016/j.cmpb.2023.107820

27. Hawkins DM, Esquivel RN. A quantile-quantile toolbox for reference intervals. J Appl Lab Med. 2024;9(2):357-370. doi:10.1093/jalm/jfad109

28. Hawkins DM. Quantile-Quantile Methodology—Detailed Results. arXiv:2303.03215. 2023. Accessed 11 September

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

del metodo QQ afferma che questi metodi parametrici sono particolarmente interessanti per i laboratori medici perché, pur essendo statisticamente rigorosi, non richiedono programmi informatici specializzati o competenze statistiche, ma possono essere attuati anche in fogli di calcolo.²⁹

EP44 procede quindi con la determinazione del coefficiente di correlazione quantile-quantile (QQCC). Se col QQCC la distribuzione appare gaussiana, eseguire i calcoli parametrici (ad esempio, media, scarto tipo). Se la distribuzione per QQCC non è gaussiana, determinare il QQCC vincolato, ovvero troncato (winsorised). Se il QQCC troncato supera il test di normalità, utilizzare la regressione QQ per ottenere media e scarto tipo e completare i calcoli parametrici. Se la QQCC vincolata non supera il test di normalità, utilizzare l'insieme originale per cercare una trasformazione ottimale. Se la trasformazione riesce a normalizzare i dati, determinare intervalli e confidenze parametricamente, quindi ritrasformare queste statistiche nella scala originale. Se le alternative precedenti non superano il test di normalità, utilizzare la regressione Box-Cox vincolata QQ per trovare una trasformazione di potenza normalizzante. La trasformazione Box-Cox ha un solo parametro: λ .³⁰ Se il valore λ è uguale a zero, viene eseguita la trasformazione logaritmica. Se il valore λ differisce da zero, la trasformazione è elevazione a potenza. Se il parametro λ è uguale a uno, la distribuzione rimane invariata.³¹

Un QQCC elevato corrisponde a una funzione che si avvicina a una linea retta nel grafico QQ, che è semplicemente un diagramma di dispersione dei RV (X) o delle loro trasformazioni (Y) rispetto ai punteggi normali (z). L'intercetta e la pendenza risultanti con l'analisi della regressione ai minimi quadrati ordinaria, possono servire rispettivamente come stime della media e della SD. I metodi robusti sono tecniche statistiche che mostrano la distribuzione dei dati senza gli effetti degli anomali. L'informazione su dove si trovano i percentili estremi (ad esempio, il 2,5° e il 97,5°) si trova nella parte dei dati del campione in cui possono esserci dati anomali. Lo stimatore robusto è progettato per stimare tali percentili da distribuzioni simmetriche, in cui le code possono essere più pesanti di quelle della distribuzione gaussiana.³²

EP44 non si ferma all'analisi statistica. Al punto 5.5 richiama le fonti di variabilità per i valori di riferimento e ne chiede l'analisi come parte integrante dello studio degli intervalli. Le fonti includono sesso, età, stato nutrizionale e le abitudini di vita, ottenuti attraverso il questionario sullo stato di salute. Il profilo delle fonti è diverso da un esame all'altro; pertanto, essi devono essere analizzati singolarmente per ciascun esame.

EP44 chiede di distinguere tra approcci univariati e multivariati, specie quando più SV sono associati tra loro. Ad esempio, l'età potrebbe essere confusa dalla massa corporea. Allo stesso modo, le differenze di sesso si confondono con le abitudini di fumo o di consumo.

Il capitolo 6 di CLSI EP44 si occupa della presentazione dei risultati dello studio sugli intervalli di riferimento. Se non sembrassero cambiare in base all'età, al sesso, all'etnia o ad altri fattori di

2024. <https://arxiv.org/pdf/2303.03215>

29. Douglas M Hawkins, Quantile–Quantile Toolbox, The Journal of Applied Laboratory Medicine, Volume 9, Issue 5, September 2024, Pages 1073–1074, <https://doi.org/10.1093/jalm/jfae068>

30. Ichihara K, Yamashita T, Kataoka H, Sato S. Critical appraisal of two Box-Cox formulae for their utility in determining reference intervals by realistic simulation and extensive real-world data analyses. *Comput Methods Programs Biomed.* 2023;242:107820. doi:10.1016/j.cmpb.2023.107820

31. Rigon, Tommaso. Trasformazione box cox: un'analisi basata sulla verosimiglianza. Facoltà di Scienze Statistiche. Corso Statistica, Economia e Finanza 2013/2014. https://thesis.unipd.it/retrieve/8ffd135e-d9b4-48db-8469-ec223cfc953a/Rigon_Tommaso.pdf

32. Horn PS. A biweight prediction interval for random samples. *J Am Stat Assoc.* 1988;83(401):249-256. doi:10.2307/2288947

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

ripartizione, potrebbe essere sufficiente una riga. Per distribuzioni complesse di dati, potrebbe essere necessaria una tabella. In altri casi, una rappresentazione grafica è un modo migliore per mostrare come il RI varia con alcune o più caratteristiche.

La tabella o i grafici per EP44 devono sempre includere informazioni sulla raccolta dei dati (ad esempio, il numero di osservazioni, i fattori di ripartizione, il percentile [se viene scelto il rango] e gli intervalli di confidenza). Inoltre, i laboratori che li usano devono fornire informazioni rilevanti per esami specifici, come il luogo o la regione di raccolta e il tipo di campione o di prelievo.

EP44 fornisce indicazioni specifiche ai fabbricanti di metodi diagnostici, che dovrebbero dare informazioni dettagliate sugli intervalli nella documentazione del prodotto. Per gli esami ben studiati e standardizzati con risultati quantitativi, è possibile giustificare l'uso di un intervallo stabilito piuttosto che riportare dati raccolti prospetticamente. Per “intervalli stabiliti” EP44 intende un libro di testo standard o a un rapporto dell'Organizzazione Mondiale della Sanità. Ma anche nel caso di esami ben studiati (ad esempio, diagnostica dell'emostasi, enzimi) potrebbe essere necessario uno studio prospettico specifico per il metodo.

In assenza di differenze per fattori di ripartizione, si può dare un solo intervallo. Tuttavia, va accompagnato con dati minimi sullo studio: strumento e metodo di esame, numero e posizione anatomica dei prelievi, numero di soggetti, criteri di inclusione ed esclusione, preparazione del soggetto, come stato di digiuno e postura, ora del giorno in cui vengono raccolti i campioni.

Frazione di soggetti di ciascun sesso, fascia di età e razza e/o etnia. EP44 riferisce che si riporta il 95° percentile centrale, ma anche il 99°, insieme agli intervalli di confidenza per i limiti dell'intervallo. Anche la mediana, il 25° percentile e il 75° percentile della distribuzione sono utili per informare sulla normalità o asimmetria della distribuzione dei dati.

EP44 fornisce diversi strumenti di supporto alle procedure raccomandate. La tabella 17 elenca alcune caratteristiche demografiche tipiche del Nord America: Indiani d'America/Nativi dell'Alaska, Asiatici, Neri, Bianchi, Altri, Ispanici, Non Ispanici. Curiosa la ripartizione delle ultime due categorie come “etnicità” e delle prime come “razza”. La tabella 18 esemplifica intervalli di riferimento per stagione, sesso e carnagione, la figura 12 rappresenta l'intervallo di riferimento per la vitamina D in funzione della stagione e della carnagione, la figura 13 e la tabella 19 per un ormone nelle donne in base all'età, la figura 15 per un biomarcatore in vari stati di progressione della malattia. L'Appendice A descrive gli intervalli di confidenza per il metodo non parametrico. L'appendice B contiene lo schema della metodologia quantile-quantile descritta in precedenza. L'appendice segnala che nonostante il tipo di analisi della regressione ai minimi quadrati richieda un programma statistico specializzato, la regressione ai minimi quadrati ordinari (OLS) potrebbe funzionare anche con una semplice approssimazione dei punteggi normali attesi. Le tabelle B1 e B2 forniscono costanti da utilizzare nel calcolo. L'appendice C descrive un esempio di metodo robusto, con la tabella C1 per il calcio, la figura C1 la funzione di ponderazione bi-peso (bi-quadrato) per dati, per illustrare la variazione regolare dei pesi man mano che le osservazioni si allontanano dal centro, la tabella C2 con i pesi durante le iterazioni del calcolo.^{33, 34}

Considerazioni sulla guida CLSI EP44.

Notiamo diversi fattori che limitano pesantemente l'utilità di CLSI EP44 per i laboratori medici,

33. Horn PS. A biweight prediction interval for random samples. J Am Stat Assoc. 1988;83(401):249-256.

doi:10.2307/2288947

34. Kafadar K. A biweight approach to the one-sample problem. J Am Stat Assoc. 1982;77(378):416-424. 3 Horn PS, Pesce AJ. Reference Intervals: A User's Guide. AACCPress; 2005.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

sia pure quelli che sviluppano metodi in casa (LDT), finanche per gli stessi fabbricanti di sistemi di misurazione di laboratorio.

La realizzazione di uno studio originale e specifico, esame per esame, per costruire intervalli di riferimento è un compito enormemente complicato, lungo, incerto, faticoso e costoso per molti motivi.

Si chiede il reclutamento di un numero molto alto di soggetti, fino a diverse centinaia, da sottoporre ad indagini approfondite anche prima del prelievo del campione. L'organizzazione ha bisogno di competenze statistiche molto elevate, accanto a quelle sulla tecnologia di misura e quelle mediche, eventualmente assicurate da esperti appositamente assunti.

Il prelievo dei campioni, così come la loro conservazione, il trasporto e la manipolazione, richiede tempo e risorse umane dedicate, competenti e dal costo rilevante.

I criteri di selezione e ripartizione dei soggetti da studiare sono tutt'altro che pacifici. Se gli intervalli sono rivolti alla condizione di "buona salute" (a differenza dei limiti decisionali), non si può ignorare l'incertezza ben nota della definizione di questa caratteristica. L'esperienza del calcolo della filtrazione glomerulare renale (eGFR) ha insegnato che il concetto di "razza" o "etnia" è abbastanza inafferrabile, tanto da renderlo inutile o addirittura dannoso.^{35,36}

La Figura 1 rappresenta qui le difficoltà incontrate dal laboratorio quando gli intervalli di riferimento si applicano a gruppi differenti per razza/etnia, età, situazione sociale e così via. La difficoltà non è solo presente nella fase di definizione, volendo applicare criteri di selezione e ripartizione, ma anche nella fase di applicazione con i risultati: a quale gruppo appartiene il singolo paziente che si sottopone ad esami di laboratorio?

Non si possono altresì trascurare alcune contraddizioni della proposta CLSI. I fabbricanti possono proporre intervalli tratti dalla letteratura, invece di imbarcarsi in uno studio lungo, faticoso e oneroso. La presentazione degli intervalli agli utenti può essere semplice ma anche molto complessa, addirittura grafica. Possono essere il 95% ma anche il 99% della distribuzione.

Le informazioni di accompagnamento richieste da EP44 sono molte, da quelle sulle modalità dello studio fino a statistiche come intervalli di confidenza, mediana e quartili della distribuzione. Certo che la presentazione è destinata ai laboratori, ma ai laboratori resta l'onere di trasferire queste informazioni ai medici e ai pazienti, sempre più spesso nella forma di messaggi digitali nei sistemi informatici. La statistica della regressione richiesta per il metodo "quantile-quantile" sarebbe sofisticata, ma di validità messa in discussione dallo stesso documento CLSI. L'approccio "robusto" descritto da CLSI EP44 (intervallo predittivo "biponderato") è raccomandato da alcune fonti^{37,38}, ma criticato da altre³⁹, che non ne riconoscono l'applicabilità nei tipici laboratori medici.

Il limite descrittivo al 99° percentile viene proposto come se fosse un limite decisionale clinico. Ad esempio, le linee guida dell'IFCC (International Federation of Clinical Chemistry and Laboratory

35. El-Khoury JM, Karger AB, Cavalier E, Kalyesubula R, Teo BW, Costa E Silva VT, Inker LA. International Perspectives on GFR Estimation and Race-Based Adjustments. Clin Chem. 2023 Aug 2;69(8):796-802. doi: 10.1093/clinchem/hvad095.

36. El-Khoury JM, Ovalle AA, Cervinski MA; Experts: Levey AS, Jones G, Eneanya ND, Lieske JC, Jha V. Is It Time to Move On? Reexamining Race in Glomerular Filtration Rate Equations. Clin Chem. 2021 Mar 31;67(4):585-591. doi: 10.1093/clinchem/hvaa333.

37. Horn PS, Pesce AJ. Reference Intervals: A User's Guide. AACC Press; 2005.

38. CLSI EP44th Ed. Establishing Reference Intervals. Projects in Progress 2025.

39. Ferruccio Ceriotti, Reference Intervals: A User's Guide. Paul S. Horn and Amadeo J. Pesce. Washington, DC: AACC Press, 2005, 115 pp., \$61.00 (\$49.00 AACC members), softcover (CD-ROM included). ISBN 1-59425-035-9., Clinical Chemistry, Volume 52, Issue 3, 1 March 2006, Pages 544–545, <https://doi.org/10.1373/clinchem.2005.058131>

Raccomandazioni M2-2

*relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025*

Versione 0.1

Medicine)⁴⁰ indicano che un dosaggio affidabile della troponina cardiaca ad alta sensibilità dovrebbe rilevare almeno il 50% degli individui sani al di sopra del limite di rilevamento e il coefficiente di variazione (CV, una misura dell'imprecisione) al 99° percentile dovrebbe essere pari o inferiore al 10%. Caratteristiche dimostrate per i metodi disponibili sul mercato, dove peraltro si è dimostrato che la funzione descrittiva del 99° percentile può oscillare (al 90% di confidenza) tra 15 e 38 ng/L, ovvero più del doppio.⁴¹

Le incertezze, le limitazioni e le ambiguità del protocollo CLSI e soprattutto i suoi costi possono scoraggiare molti fabbricanti. Non parliamo dei laboratori che scelgano di sviluppare in casa alcuni metodi di esame. Specie se si considera la relativamente ridotta rilevanza per le decisioni mediche, contrariamente ai limiti decisionali. Accompagnare i risultati delle misure quantitative con uno o più intervalli di riferimento consente di rinforzare e confermare le informazioni fornite dalle unità di misura e dall'eventuale definizione del metodo di esame. Pescare intervalli nella letteratura scientifica o in testi autorevoli, confermandoli con studi di estensione limitata, può essere un'alternativa praticabile.

Non abbiamo argomenti scientificamente fondati per rifiutare del tutto la produzione di intervalli di riferimento, anche nei casi in cui metodi statistici sofisticati cercano di superare le insufficienze di numerosità e rappresentatività della popolazione selezionata. I laboratori medici possono porsi due quesiti: se la produzione ex novo di intervalli sia alla portata delle loro risorse economiche e professionali e d'altra parte se intervalli prodotti con le limitazioni descritte e artifici computazionali non comprensibili siano realmente utili per classificare il risultato dell'esame di laboratorio per il singolo paziente. Intervalli e limiti, spogliati della funzione di classificare i risultati in categorie, possono comunque mantenere un valore per la descrizione del metodo usato per l'esame, accanto alla denominazione e alle unità di misura.



Figura 1. Criticità per il laboratorio sulla discriminazione con gli intervalli di riferimento

40. K.M. Aakre, A.K. Saenger, R. Body, P. Collinson, O. Hammarsten, A.S. Jaffe, P. Kavsak, T. Omland, J. Ordonez-Lianos, F.S. Apple, Analytical considerations in deriving 99th percentile upper reference limits for high-sensitivity cardiac troponin assays: educational recommendations from the IFCC committee on clinical application of cardiac biomarkers, Clin. Chem. 68 (8) (2022 Jul 27) 1022–1030, <https://doi.org/10.1093/clinchem/hvac092>. P

41. N. Auberger, I. Coin, L. Marillet, F. Raymond, S. Michel-Busseret, P.- G. Claret, C. Pease. Assay precision, 99th Percentile reference value and proportion of detected healthy European Adults for VIDAS® High-Sensitive Troponin I. Practical Laboratory Medicine, <https://doi.org/10.1016/j.plabm.2024.e00441>.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

NOTA. Guida CLSI per utilizzo di intervalli di riferimento nel laboratorio medico.

Il documento CLSI EP45 ED1PD:2024 descrive come utilizzare gli intervalli di riferimento nel laboratorio medico.⁴² Questa guida ha rilevante importanza per i laboratori, perché in conformità ai requisiti di ISO 15189 descrive concretamente fatti, procedure e documentazioni potenzialmente soggetti a ispezione in occasione delle visite di accreditamento.

EP45 contiene una guida per attuare gli intervalli di riferimento da parte del laboratorio medico per i metodi quantitativi di esame. I laboratori possono fare sempre più affidamento su altri laboratori o fabbricanti sul mercato per ottenere intervalli di riferimento appropriati e adeguati. EP45 sostituisce la verifica degli intervalli di riferimento contenute nella EP28-A3c⁴³.

CLSI EP45 è stato sviluppato come continuazione di CLSI EP28, che trattava la creazione, la verifica e l'attuazione degli intervalli di riferimento. Il contenuto del documento CLSI EP28 viene suddiviso in due documenti separati, uno per la definizione degli intervalli di riferimento da parte di uno sviluppatore e uno per l'attuazione da parte del laboratorio utente finale.

EP45 contiene un aggiornamento della guida per il laboratorio, ulteriori indicazioni per i diversi studi sperimentali, quale studio scegliere, il numero di campioni da raccogliere e la valutazione dei risultati, l'introduzione della verifica con metodi indiretti.

Per accettare un intervallo di riferimento candidato si considerano tre aspetti principali: la comparabilità dei metodi di esame, la comparabilità della popolazione di pazienti e la comparabilità dei fattori preesame. CLSI prevede l'esecuzione di una prova sperimentale per la verifica degli intervalli di riferimento con due approcci principali: uno studio di verifica binomiale su piccola scala o uno studio di confronto tra metodi. Lo studio su scala ridotta ha lo scopo di verificare che la popolazione locale del laboratorio sia paragonabile a quella dell'intervallo candidato quando il metodo di misura e i fattori preesame sono comparabili. Lo studio di confronto del metodo per CLSI ha lo scopo di verificare che il metodo di prova sia comparabile e può essere combinato con lo studio su piccola scala o utilizzato da solo quando la popolazione e i fattori di preesame sono noti come comparabili.

EP45 disegna diversi scenari per verificare un intervallo di riferimento, che per la maggior parte di questi scenari rientra nella fase di “verifica” dell’esame: nuovo metodo un esame esistente, nuovo esame con intervalli forniti dallo sviluppatore, valutazione del rischio durante la fase di manutenzione dell’esame con verifica degli intervalli. CLSI immagina la verifica su base regolare come parte dei requisiti di accreditamento.

CLSI descrive uno studio di verifica su scala ridotta per verificare la comparabilità tra gli individui di riferimento del laboratorio e un intervallo di riferimento candidato. Descrive anche la verifica mediante confronto del metodo con il precedente metodo di prova del laboratorio, o uno eseguito da un altro laboratorio. L'intervallo di riferimento candidato del metodo di confronto sarà verificato per l'uso con il nuovo metodo di prova in base all'analisi di regressione nello studio di confronto.

CLSI accenna anche ai metodi indiretti. Il laboratorio può scegliere di eseguire un'analisi dei dati storici. Questo metodo (“approccio indiretto”) potrebbe essere applicabile per la revisione periodica degli intervalli di riferimento esistenti senza modifiche del metodo analitico, oppure per usare un intervallo di riferimento proveniente da uno studio rigoroso della letteratura senza

42. CLSI. Implementation of Reference Intervals in the Medical Laboratory. Ed1. CLSI guideline EP45. Clinical and Laboratory Standards Institute; 202[XX].

43. 1 CLSI. Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory; Approved Guideline. 3rd ed. CLSI document EP28-A3c. Clinical and Laboratory Standards Institute; 2010.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

modificare il metodo di raccolta e di esame dei campioni.

Gli intervalli di riferimento per esami comuni sono spesso disponibili da fonti quali lo sviluppatore del metodo di esame, un altro laboratorio che esegue lo stesso metodo o metodi simili, linee guida di pratica professionale o la storia degli intervalli di riferimento del laboratorio stesso. Il laboratorio deve identificare e valutare le diverse fonti, documentare tutte le informazioni disponibili fornite dalla fonte, compresi i criteri di inclusione ed esclusione, il numero di individui nel gruppo di campioni di riferimento, le caratteristiche demografiche del gruppo di campioni di riferimento e le tecniche di analisi statistica dei dati utilizzate. Spesso l'intervallo di riferimento proviene dal fabbricante per metodi che hanno ricevuto l'autorizzazione normativa, pubblicato nelle istruzioni per l'uso. EP45 critica questa soluzione perché ritiene che spesso non siano disponibili dettagli adeguati sulla composizione demografica della popolazione di riferimento del fabbricante.

EP45 ammette intervalli di riferimento candidati provenienti da fonti pubblicate dopo revisione tra pari, purché condotti in modo rigoroso. Il laboratorio deve compilare un elenco di studi, scartando quelli che non sono stati condotti secondo i requisiti minimi di progettazione. Un'altra possibilità sono intervalli di riferimento per metodi che hanno dimostrato la comparabilità dei risultati attraverso programmi di standardizzazione. Infine, intervalli di riferimento possono venire da un metodo di prova di confronto, un precedente metodo di esame del laboratorio o un metodo eseguito da un altro laboratorio con verifica attraverso uno studio di confronto del metodo in base al documento CLSI EP09.

CLSI prevede il confronto dei dettagli del metodo del laboratorio con il metodo dell'intervallo di riferimento candidato: principio dell'esame, taratura, tracciabilità metrologica, precisione, accuratezza e interferenze. Per la tracciabilità metrologica richiama il documento CLSI EP32.⁴⁴ EP45 prevede inoltre che il laboratorio accerti se la popolazione di pazienti del proprio laboratorio è paragonabile a quella utilizzata per stabilire l'intervallo di riferimento candidato. Se le caratteristiche demografiche differiscono, come ad esempio per età o composizione etnica, può essere giustificato uno studio completo per l'intervallo di riferimento. Il laboratorio deve inoltre confrontare i criteri di esclusione per l'intervallo di riferimento candidato, nel caso influiscano sulla somiglianza della popolazione. CLSI fornisce nella tabella 2 gli esempi di criteri di esclusione, divisi in criteri di esclusione primari e criteri di esclusione secondari. Analogamente, se le ripartizioni sono applicate per lo studio dell'intervallo di riferimento candidato, il laboratorio deve considerare se queste ripartizioni si applicano alla propria popolazione di pazienti. La Tabella 3 elenca esempi di fattori di suddivisione, divisi in criteri comunemente considerati (Sesso biologico, Razza/etnia, Età, Posizione geografica, regione, altitudine) e criteri di suddivisione dipendenti dallo schema di studio.

Il laboratorio per EP45 deve confrontare tutti i processi preesame, compresa la preparazione dei soggetti, le procedure di raccolta dei campioni, le procedure di trattamento dei campioni per lo studio dell'intervallo di riferimento candidato con le procedure del laboratorio. Il laboratorio deve confrontare la preparazione dei soggetti. Compresi i cambiamenti posturali quando si confrontano i risultati di pazienti ricoverati e ambulatoriali. Infine, si devono valutare le eventuali differenze per il contenitore di raccolta del campione o il tipo di campione.

Il capitolo 4 di EP45 divide gli approcci sperimentali in tre categorie: Studio di verifica binomiale su piccola scala (sottocapitolo 4.1: Un test statistico su un numero relativamente piccolo di individui di riferimento, $n = 20$), Studio di verifica su larga scala (sottocapitolo 4.2: Valutazione di

44. CLSI. Metrological Traceability and Its Implementation; A Report. CLSI document EP32-R. Clinical and Laboratory Standards Institute; 2006.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

un numero maggiore di individui di riferimento, ad esempio, $n = 60$), Trasferimento mediante confronto/regressione del metodo (sottocapitolo 4.3). Lo studio di verifica binomiale su piccola scala è l'approccio principale raccomandato. Un numero maggiore di individui di riferimento ($n \sim 60$) può essere incluso nello studio, per rappresentare meglio l'effettiva popolazione di riferimento del laboratorio. Anche uno studio di confronto del metodo può essere un approccio rigoroso, a seconda del numero di campioni inclusi nel confronto e di altre caratteristiche di prestazione valutate. La tabella 4 di EP45 raccoglie vantaggi e svantaggi delle diverse metodologie di verifica. Lo studio piccolo non rileva la possibile varianza ridotta nella popolazione del laboratorio rispetto all'intervallo candidato, portando a una falsa accettazione di un intervallo di riferimento troppo ampio. Lo studio più largo richiede più campioni e un trattamento statistico più avanzato. La comparazione tra metodi non rileva le differenze tra la popolazione del laboratorio e quella utilizzata per stabilire l'intervallo candidato. Infine, il metodo indiretto si basa su un'analisi statistica avanzata e la disponibilità di dati storici compatibili.

L'Appendice A di EP45 contiene la simulazione della potenza per alcune procedure statistiche. In breve, lo studio piccolo consiste nel selezionare 20 individui, che devono rappresentare ragionevolmente la popolazione sana del laboratorio e soddisfare adeguatamente i criteri di esclusione e di partizione, altrimenti si deve verificare ogni partizione ($N = 20$ per partizione). Esaminare quindi i campioni, i campioni di controllo e verificare i criteri di accettazione delle misure. Analizzare i risultati per assicurarsi che rappresentino un gruppo statisticamente omogeneo e che i valori anomali sospetti siano studiati per individuare eventuali cause. Con il test binomiale, se non più del 10% dei risultati si trova al di fuori dei limiti stabiliti dell'intervallo di riferimento candidato, l'esito è positivo. Non solo, la distribuzione dei 20 (o 40) risultati deve essere confrontata con i limiti, per giudicare se appaiono troppo lontani, poiché il test binomiale non rileva una varianza ridotta. La tabella 5 contiene i limiti di accettazione per studi su piccola scala di intervalli al 95%.

In alternativa, utilizzando l'analisi parametrica, insieme di dati più ampi, comunque inferiori ai 120 individui, possono consentire di determinare i limiti dell'intervallo di riferimento con intervalli di confidenza accettabili. Questi "mini" studi di determinazione possono essere utilizzati per verificare un intervallo di riferimento esterno, confrontando direttamente i limiti e gli intervalli di confidenza. Questo tipo di analisi richiede metodologie statistiche più avanzate. Nella Appendice E (Approcci di verifica parametrica) CLSI segnala importanti difficoltà dell'uso dei parametri media e scarto tipo (SD). Sebbene l'analisi parametrica consenta di ottenere intervalli di confidenza relativamente ristretti, comporta il rischio di adattare una distribuzione errata e di fare un'inferenza errata sul limite di riferimento. Inoltre, richiede strumenti statistici più complessi sia per l'adattamento che per la valutazione delle distribuzioni di densità ed è quindi consigliata solo a personale con competenze statistiche sufficienti per eseguire i passaggi necessari e giudicare l'affidabilità delle stime risultanti. Un passaggio critico è quello della valutazione della "normalità" della distribuzione, dove si possono utilizzare diverse metodologie. CLSI raccomanda il coefficiente di correlazione quantile-quantile (QQCC), ma cita altri approcci includono come i test di Shapiro-Wilks e Shapiro-Francia. I dati possono essere sottoposti a trasformazione prima di eseguire la stima parametrica dei limiti dell'intervallo di riferimento. Nell'Appendice D il documento EP45 presenta una procedura euristica non parametrica applicata a dati di 60 soggetti senza analisi statistiche avanzate. L'intervallo centrale del 95% può essere stimato come le osservazioni classificate al 2° e 59° posto, quindi confrontato con i limiti dell'intervallo di riferimento candidato. Qualsiasi differenza osservata viene giudicata in base al

contesto clinico e alla conoscenza della distribuzione sottostante.

CLSI EP45 raccomanda uno studio di confronto tra metodi per valutare in modo rigoroso un intervallo di riferimento tra un metodo di esame e un altro, cosa che diventa effettivamente uno “studio di trasferimento”. In alcuni scenari, sembra l’unica alternativa praticabile. Ad esempio, se il laboratorio decide di cambiare metodo di esame, per non dover raccogliere campioni da individui di riferimento, si avvale del confronto tra i due metodi utilizzando il documento CLSI EP09⁴⁵. Tuttavia, EP45 segnala alcune criticità. Se non è rappresentato un intervallo sufficiente di valori, la correlazione può sembrare peggiore di quanto non sia, mentre se si utilizza un intervallo troppo ampio di campioni si può sovrastimare la qualità della correlazione. L’equazione di regressione derivata dall’intero intervallo di misurazione di entrambi i metodi può differire da quella negli individui sani. L’intervallo di valori più appropriato per la popolazione di pazienti, che può variare a seconda dell’esame. È inoltre importante considerare eventuali cambiamenti nella popolazione di pazienti. Se il nuovo metodo presenta un’imprecisione simile e interferenze note, utilizza standard o calibratori uguali o comparabili e fornisce valori accettabilmente comparabili, l’intervallo di riferimento può essere trasferito. EP45 fornisce a corredo due esempi: Esempio 1: i metodi sono completamente comparabili; esempio 2, i risultati degli esami sono altamente correlati, ma uno fornisce risultati che sono proporzionalmente distorti verso l’alto o verso il basso (ad esempio, misurare l’attività enzimatica a 37°C invece che a 30°C). Per CLSI il processo attuativo deve includere alla fine un impegnativo “riassunto” delle motivazioni utilizzate dal laboratorio per verificare e trasferire l’intervallo di riferimento e qualsiasi citazione della letteratura pertinente o altro materiale utilizzato durante il processo di valutazione. Nella sintesi si descriveranno l’utilità clinica prefissata dell’esame, le procedure specifiche preesame, dettagli del metodo di misura, comprese reazione chimica, taratura, tracciabilità e le influenze, le caratteristiche delle prestazioni di misura, come imprecisione e distorsione, le caratteristiche demografiche della popolazione del laboratorio, compresa la distribuzione per sesso, età, etnia e posizione geografica, la tecnica di analisi statistica dei dati. CLSI EP45 non dice nulla sulla gestione degli intervalli nella presentazione dei risultati degli esami, né tantomeno sulla loro gestione nei messaggi digitalizzati dei sistemi informatici.

Metodi “indiretti”

CLSI EP45 introduce il concetto di verifica attraverso metodi indiretti, fornisce i riferimenti pertinenti^{46, 47, 48, 49, 50, 51, 52} e descrive gli approcci disponibili. Non vengono fornite indicazioni sul

45. CLSI. Measurement Procedure Comparison and Bias Estimation Using Patient Samples. 3rd ed. CLSI guideline EP09c. Clinical and Laboratory Standards Institute; 2018.

46. Zierk J, Metzler M, Rauh M. Data mining of pediatric reference intervals. Lab Med. 2021;45:311-317. doi:10.1515/labmed-2021-0120

47. Jones GRD, Haeckel R, Loh TP, et al. Indirect methods for reference interval determination - review and recommendations. Clin Chem Lab Med. 2018;57(1):20-29. doi:10.1515/cclm-2018-0073

48. Haeckel R, Wosniok W, Streichert T. Review of potentials and limitations of indirect approaches for estimating reference limits/intervals of quantitative procedures in laboratory medicine. Lab Med. 2021;45:35-53. doi:10.1515/labmed-2020-0131

49. Sikaris K. Separating disease and health for indirect reference intervals. Lab Med. 2021;45:55-68. doi:10.1515/labmed-2020-0157

50. 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. 2023;61(3):402-406. doi:10.1515/cclm-2022-1061

51. Doyle K, Bunch DR. Reference intervals: past, present, and future. Crit Rev Clin Lab Sci. 2023;60(6):466-482. doi:10.1080/10408363.2023.2196746

52. Ma S, Yu J, Qin X, Liu J. Current status and challenges in establishing reference intervals based on real-world data.

Raccomandazioni M2-2
relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

disegno dello studio, sulla valutazione e sulla documentazione. Nel capitolo 5 si afferma che metodi indiretti sono quelli che utilizzano risultati di laboratorio ottenuti in precedenza, per uno scopo diverso (come il controllo di routine del paziente, la diagnosi e il monitoraggio del trattamento), non per la definizione o alla verifica di un intervallo di riferimento. Non richiedono l'arruolamento di individui sani, la raccolta di campioni o la determinazione del misurando, sovente costosa. Per EP45 questo approccio può essere utilizzato per verificare un intervallo di riferimento, ma solo dopo averlo attuato. Gli svantaggi individuati da EP45 sono il rischio di includere popolazioni eterogenee, la mancanza di accesso a informazioni mediche complete o rilevanti, il necessario l'accesso ai dati storici, un'alta probabilità di includere soggetti inappropriati, analisi dei dati che può essere complessa, dati "sani" per esami specializzati come gli ormoni potrebbero essere molto pochi.⁵³

*EP45 suddivide i metodi indiretti in due gruppi: quelli di prima generazione (come i metodi di Hoffmann⁵⁴ e Bhattacharya⁵⁵) e quelli di seconda generazione (una lista nutrita⁵⁶, comprendente Massima verosimiglianza troncata kosmic⁵⁷ - Chi-quadro minimo troncato⁵⁸ - refineR⁵⁹)
Ulteriori dettagli sono aggiunti da EP45 nella Appendice F (Considerazioni sulla verifica), come lista di controllo che elenca fattori del paziente (Popolazione, Demografia, Sesso, Distribuzione dell'età, Razza, Etnia, Posizione geografica), fattori preesame (Preparazione del paziente, Tipo di campione, Procedure di trattamento del campione), fattori della misurazione (Procedura di misurazione, Metodologia, Reazione chimica, Approccio di taratura e tracciabilità, Interferenze/influenze Prestazioni del sistema di misura validato), fattori postesame (Tecnica di analisi statistica, analisi dei dati per ricavare l'intervallo di riferimento, Criticità del limite di riferimento dei misurandi nel processo decisionale clinico).*

Considerazioni sulla guida CLSI EP45.

CLSI EP45 riferisce che in molti casi i laboratori usano intervalli di riferimento proposti dal fabbricante per metodi che hanno ricevuto l'autorizzazione normativa, pubblicati nelle istruzioni per l'uso. Le guide CLSI sul tema (EP28) sono disponibili da molto tempo, eppure la prassi quotidiana dei laboratori sembra essere andata in una direzione diversa. CLSI insiste a proporre il suo approccio, organizzativamente oneroso e concettualmente sofisticato, senza avviare una riflessione critica sulla sua praticabilità. CLSI pone molta fiducia sul confronto tra metodi, sia all'interno del laboratorio che tra laboratori, senza considerare la rarità, la difficoltà e il costo di questo tipo di studi. Tuttavia, non fornisce riferimenti precisi e non tiene conto della vasta

Crit Rev Clin Lab Sci. 2023;60(6):427-441. doi:10.1080/10408363.2023.2195496

53. 29 Ammer T, Schützenmeister A, Prokosch HU, Zierk J, Rank CM, Rauh M. RIBench: A proposed benchmark for the standardized evaluation of indirect methods for reference interval estimation. Clin Chem. 2022;68(11):1410-1424. doi:10.1093/clinchem/hvac142

54. Hoffmann RG. Statistics in the practice of medicine. JAMA. 1963;185:864-873.

55. Bhattacharya CG. A simple method of resolution of a distribution into gaussian components. Biometrics. 1967;23(1):115-135

56. Ammer T, Schützenmeister A, Prokosch HU, Zierk J, Rank CM, Rauh M. RIBench: A proposed benchmark for the standardized evaluation of indirect methods for reference interval estimation. Clin Chem. 2022;68(11):1410-1424. doi:10.1093/clinchem/hvac142

57. 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(1):1704.

58. Wosniok W, Haeckel R. A new indirect estimation of reference intervals: truncated minimum chi-square (TMC) approach. Clin Chem Lab Med. 2019;57(12):1933-1947.

59. Ammer T, Schützenmeister A, Prokosch HU, Rauh M, Rank CM, Zierk J. refineR: A novel algorithm for reference interval estimation from real-world data. Sci Rep. 2021;11(1):16023.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

letteratura che evidenzia le criticità dei confronti. Aggiungendo peraltro che nei confronti l'analisi della regressione, richiamata da EP44, ha un ruolo decisamente marginale. EP44 segnala alcune criticità dei confronti, ma si dilunga molto sul trattamento dei dati, trascurando del tutto le criticità organizzative, soprattutto la necessità di avere, tarare, controllare e mantenere un metodo di confronto, nonché addestrare il personale. Oneri organizzativi e costi insostenibili per la maggioranza dei laboratori.⁶⁰

EP45 tocca l'argomento dei metodi indiretti, trattandolo però in modo poco approfondito, come se fosse un tema di interesse marginale, forse trascurabile. Al contrario, i risultati nella copiosa letteratura scientifica in merito inducono a rivalutare questo approccio, praticabile almeno dai laboratori più grandi, utile per confermare o meno gli intervalli in uso e persino per aggiustarne i limiti al fine di avvicinarsi al ruolo decisionale di cui medici e laboratori hanno bisogno.⁶¹

Indubbiamente interessante, forse curioso, è l'ultimo punto della lista di controllo nell'Appendice F, ovvero il valore del limite di riferimento come limite decisionale.

Infine, CLSI EP45 non dice nulla sulla gestione degli intervalli nella presentazione dei risultati degli esami, né tantomeno sulla loro gestione nei messaggi digitalizzati dei sistemi informatici.

60. Pradella M, Cesana BM. 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. DOI: 10.13140/RG.2.2.14477.55526

61. Patel S, Verma N, Padhi P, et al. (September 12, 2023) An Approach to Re-evaluate the Reference Cutoff of the Parameters of Newborn Screening: An Observational Study. Cureus 15(9): e45139. DOI 10.7759/cureus.45139

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

NOTA. Mondo reale fonte di informazioni mediche

ISO 15189:2022 (7.3.5) chiede ai laboratori la definizione e la revisione periodica di IR, in relazione alla popolazione servita dal laboratorio stesso. Per soddisfare questi requisiti l'approccio "indiretto" è molto più adatto di quello tradizionale "diretto", purché la base di dati analizzata sia sufficientemente ampia, come avviene nei laboratori centralizzati più grandi.

ISO 18727: dati sanitari nel mondo reale

ISO ha approvato (giugno 2025) il progetto 18727.⁶²

Le iniziative aziendali basate sui dati del mondo reale (RWD) stanno diventando sempre più critiche. Le aspettative di utilizzo dei RWD per migliorare l'efficienza della ricerca clinica e il processo di sviluppo di nuovi farmaci sono in aumento e prese in seria considerazione da agenzie come FDA.^{63,64,65} ISO 18727 definisce (3.1.1) dati dal mondo reale, RWD come dati relativi allo stato di salute del paziente e/o all'erogazione di assistenza sanitaria raccolti di routine da una varietà di fonti

La norma ISO 18727 documenta un caso di studio di un processo di gestione della qualità per supportare l'integrazione dell'archiviazione di informazioni mediche all'interno di una struttura di archiviazione gerarchica, tra più siti medici.

ISO TR 18727 descrive l'esperienza giapponese nell'ottenere la RWD attuando la gestione della qualità, nell'ambito del progetto MID-NET® sponsorizzato dal Ministero della Salute, del Lavoro e del Welfare e dall'Agenzia per i prodotti farmaceutici e i dispositivi medici.⁶⁶ Questa esperienza ha dimostrato una riduzione dei costi e dei tempi e un netto miglioramento dell'affidabilità dei dati. Il progetto ha dimostrato che l'implementazione della gestione della qualità può ridurre i costi e i tempi, mantenere l'affidabilità dei dati ed essere attuata in modo standardizzato. Le lezioni apprese in Giappone possono essere applicate ovunque nel mondo e questo documento è stato concepito per fornire una guida globale.⁶⁷

Lo strumento RWD si sta progressivamente affermando come alternativa valida agli approcci tradizionali, formali ma molto costosi e spesso impraticabili. Un esempio ben noto da molti anni è quello degli intervalli di riferimento per esami di laboratorio (RI) ottenuti con metodi cosiddetti "indiretti".

Per utilizzare metodi indiretti per verifica, conferma o modifica di IR è necessario prescindere dal concetto primordiale che con gli IR si debba descrivere la "salute" o "l'assenza di malattia". È necessario disporre di una grande base di dati e di una capacità sofisticata di elaborazione statistica dei dati. Solo i laboratori di maggiori dimensioni possono utilizzare i metodi indiretti. Gli altri laboratori non possono far altro che curare la tracciabilità dei propri metodi per la confrontabilità dei risultati con quelli del fabbricante o dei grandi laboratori. Affiancando il metodo proposto da CLSI EP45 come studio di verifica su scala ridotta. Gli IR andrebbero presentati come descrizione della prevalenza dei risultati nella popolazione servita dal laboratorio, conferma delle unità di misura usate e della taratura del metodo, base per individuare limiti per la decisione clinica.

Lo studio eseguito a Seoul del gruppo di Kang fornisce un buon esempio di metodo indiretto.⁶⁸ Si avvale di 1.392.356 misurazioni di PCR per 353.340 pazienti, 951.497 di VES per 206.894 pazienti), e 4.219 misurazioni di presepsina PSEP per 1.086 pazienti. Per determinare la curva di distribuzione dell'istogramma, è stata eseguita la stima della densità kernel (nucleo), un approccio non parametrico per stimare la funzione di densità di probabilità di una variabile. La trasformazione Box-Cox⁶⁹ è stata eseguita per normalizzare la distribuzione dei dati asimmetrici. L'analisi del profilo latente (LPA) è un metodo analitico che fornisce una variabile latente

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

categoriale da dati continui. L'intersezione della curva di distribuzione è stata considerata il limite di riferimento per le popolazioni non malate e malate. Sulla base di questi dati, la stima dell'intervallo può essere effettuata con un metodo indiretto. Questo metodo richiede le seguenti assunzioni: (1) la procedura di analisi può essere eseguita per dati con skewed a destra; (2) la dimensione della misura deve essere superiore a 2.500; (3) due classi della variabile latente sono di dimensioni simili, il che è sconosciuto. Il limite principale ammesso dagli Autori è che l'esame LPA è stato eseguito sulla base dell'ipotesi che esistano classi non malate e malate all'interno di una popolazione mista. Tuttavia, potrebbero esserci più classi all'interno delle classi non malate o malate. Il processo statistico è stato applicato a dati con deviazione a destra, per cui sono necessari ulteriori studi per i dati con deviazione a due lati. Noi possiamo aggiungere che individuare il limite in base all'intersezione di due distribuzioni statistiche non offre informazioni sull'utilizzo a scopo diagnostico, ovvero sensibilità e specificità per determinate patologie. Tutt'altro. Si allontana molto dal criterio del 95% dei tradizionali intervalli di riferimento. Quindi, può dare una descrizione approssimativa della distribuzione di valori nella popolazione servita dal laboratorio, rinunciando però all'utilizzo come criterio decisionale per la diagnosi, nemmeno di esclusione.

Un esempio simile è stato prodotto a Kochi in Giappone.⁷⁰ Sono stati analizzati più di 166 milioni di risultati per 315mila soggetti, dividendo quelli ripetuti a breve termine, supposti "anormali", dagli altri, supposti "normali". La differenza ottenuta tra i due insiemi di dati è stata utilizzata per stimare la distribuzione dei risultati degli esami "ritenuti validi dal medico". Dalla distribuzione stimata è stata stimata anche una funzione di densità di probabilità, con il 2,5° e il 97,5° percentile come limiti inferiore e superiore. La distribuzione dei valori dei risultati è stata convertita mediante la trasformazione Box-Cox. Sono stati calcolati il 2,5° e il 97,5° percentile. La trasformazione inversa Box-Cox è stata applicata ai percentili calcolati e riportata alla scala dei risultati originali dell'esame. Inoltre, è stata calcolata e confrontata la divergenza di Kullback-Leibler (misura non simmetrica della differenza tra due distribuzioni di probabilità⁷¹) delle distribuzioni di frequenza. Gli intervalli dei valori dei test calcolati nello studio sono informazioni del mondo reale che riflettono il giudizio globale dei medici basato su informazioni dai pazienti nella pratica. Il metodo dovrebbe essere applicato alla visualizzazione delle decisioni del medico,

62. ISO/CD TR 18727. Health Informatics – Quality management of health data in the real world – A case study. <https://www.iso.org/standard/85414.html>

63. Framework for FDA's Real-World Evidence Program. 2018. <https://www.fda.gov/media/120060/download>

64. U.S. Food and Drug Administration. Real-World Evidence. 2025. <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>

65. Klonoff DC. The New FDA Real-World Evidence Program to Support Development of Drugs and Biologics. J Diabetes Sci Technol. 2020 Mar;14(2):345-349. doi: 10.1177/1932296819832661. Epub 2019 Mar 12. PMID: 30862182; PMCID: PMC7196856.

66. ISO/CD TR 18727 Quality management of health data in the real world - A case study

67. Framework for FDA's Real-World Evidence Program. 2018.

<https://www.fda.gov/media/120060/download>

68. Kang T, Yoo J, Jekarl DW, Chae H, Kim M, Park YJ, Oh EJ, Kim Y. Indirect Method for Estimation of Reference Intervals of Inflammatory Markers. Ann Lab Med. 2023 Jan 1;43(1):55-63. doi: 10.3343/alm.2023.43.1.55. Epub 2022 Sep 1. PMID: 36045057; PMCID: PMC9467833.

69. Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. Journal of the Royal Statistical Society, B, 26(211-234)

70. Hyohdoh Y, Hatakeyama Y, Okuhara Y. A simple method to identify realworld clinical decision intervals of laboratory tests from clinical data. Inform Med Unlocked 2021;23:100512.

71 S. Kullback. R. A. Leibler. "On Information and Sufficiency." Ann. Math. Statist. 22 (1) 79 - 86, March, 1951. <https://doi.org/10.1214/aoms/1177729694>

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

per la fenotipizzazione in malattie specifiche e la rivalutazione di intervallo e limiti decisionali. Anche in questo caso possiamo osservare come la descrizione efficace della realtà comporti l'allontanamento dal valore diagnostico dei limiti individuati. Un altro studio suggestivo è stato recentemente condotto nei laboratori ARUP.⁷² Sono stati estratti dal sistema informativo del laboratorio i dati per l'esame di interesse, elaborabili con l'applicazione refineR, ad esempio con frazione patologica prevista sotto il 50%. Il materiale è quindi stato costituito ad esempio da trentamila risultati circa per alfa1 antitripsina, il doppio per ceruloplasmina, circa diecimila per transaminasi, quasi cinquantamila per FT3. Per la pulizia dei dati, ad esempio, sono rimossi i risultati non validi, come quelli di alcuni reparti (ad esempio, oncologia, unità di terapia intensiva), eventuali ripetizioni dello stesso paziente. Quindi si ripartiscono per sesso ed età e si garantisce la stabilità dei risultati con il controllo della qualità. Si applica l'algoritmo refineR con la trasformazione Box-Cox a 1 parametro predefinita per stimare un modello della distribuzione "non patologica" (per ogni partizione) utilizzando i dati puliti e la tecnica "bootstrapping per ottenere intervalli di confidenza [CI]. "bootstrapping" letteralmente significa tirarsi su per gli stivali. Si tratta di un approccio numerico per generare intervalli di confidenza con dati ricampionati o simulati.^{73,74} Se la distribuzione appare visivamente distorta e spostata rispetto allo zero, si ripete il passaggio della trasformazione Box-Cox. Si verificano poi la plausibilità dei risultati, ad esempio confrontando gli intervalli stimati con altre fonti, come quelli forniti dal fabbricante con le istruzioni per l'uso. ARUP ha visto la corrispondenza della maggior parte degli intervalli, ad eccezione di alcuni come TSH e prealbumina. ARUP ritiene il metodo indiretto un'alternativa conveniente per rispondere ai requisiti ISO 15189. Anche il gruppo di Karachi ha sperimentato il metodo indiretto, in collaborazione con Erlangen.⁷⁵ Ha analizzato quasi centomila risultati di creatinina, come aveva fatto prima per fosfatasi alcalina. L'analisi statistica è stata eseguita utilizzando Microsoft Excel 2010 e l'algoritmo indiretto proposto e pre-validato dal gruppo di lavoro della Società tedesca di chimica clinica e medicina di laboratorio (GSCCLM), liberamente disponibile online come pacchetto software [5,7,12].^{76,77,78} Per modellare la distribuzione dei campioni "non patologici" nei dati è stata eseguita una distribuzione normale di potenza, definita come distribuzione gaussiana dopo la trasformazione BoxCox.

72 Ammer T, Schützenmeister A, Rank CM, Doyle K. Estimation of Reference Intervals from Routine Data Using the refineR Algorithm-A Practical Guide. J Appl Lab Med. 2023;8(1):84-91. doi:10.1093/jalm/jfac101

73 EconomyUp. Bootstrapping – definizione <https://www.economyup.it/glossario/bootstrapping-definizione/>

74 Nelson WA. Statistical methods. In S.E. Jørgensen, B.D. Fath (Eds.), Encyclopedia of Ecology, Academic Press, Elsevier Science (2008), pp. 3350-3362, 10.1016/B978-008045405-4.00661-3

<https://www.sciencedirect.com/topics/earth-and-planetary-sciences/bootstrapping>

75 Ahmed S, Zierk J, Siddiqui I, Khan AH. Indirect determination of serum creatinine reference intervals in a Pakistani pediatric population using big data analytics. World J Clin Pediatr. 2021;10(4):72-78. Published 2021 Jul 9. doi:10.5409/wjcp.v10.i4.72

76 Bohn MK, Higgins V, Adeli K. CALIPER paediatric reference intervals for the urea creatinine ratio in healthy children & adolescents. Clin Biochem 2020; 76: 31-34 [PMID: 31838019 DOI: 10.1016/j.clinbiochem.2019.12.001]

77 Ahmed S, Zierk J, Khan AH. Establishment of Reference Intervals for Alkaline Phosphatase in Pakistani Children Using a Data Mining Approach. Lab Med 2020; 51: 484-490 [PMID: 31860088 DOI: 10.1093/labmed/lmz096]

78 Arzideh F, Brandhorst G, Gurr E, Hinsch W, Hoff T, Roggenbuck L, Rothe G, Schumann G, Wolters B, Wosniok W, Haeckel R. An improved indirect approach for determining reference limits from intra-laboratory data bases exemplified by concentrations of electrolytes. Laboratoriums Medizin 2009; 33: 52-66 [DOI: 10.1515/JLM.2009.015]

Ancora dalla Turchia proviene lo studio di più di 120 mila risultati di esami emocromocitometrici, deliberatamente non selettivi, con un numero di patologie fino a cinque, che ha dimostrato in un numero sorprendente di casi una distribuzione normale, verificata con metodi statistici.⁷⁹ Confermando così l'applicabilità dello studio del mondo reale per fornire l'immagine dei risultati tipici attesi in laboratori di grandi dimensioni. Senza pretendere di ricavarne soglie decisionali. Utilizzando un approccio "grandi numeri" ("big-data") ricercatori olandesi sono stati in grado di determinare intervalli di riferimento tracciabili per 18 esami di chimica generale.⁸⁰ Osservazioni confermate da altri ricercatori spagnoli e olandesi⁸¹, secondo cui i risultati di esami medici che seguono una distribuzione normale hanno mostrato intervalli di riferimento comparabili e coerenti tra gli studi. Pertanto, un semplice metodo indiretto è un approccio fattibile ed efficiente in termini di costi per il calcolo degli intervalli di riferimento. Tuttavia, gli stessi Autori precisano che gli sforzi dovrebbero concentrarsi anche sulla standardizzazione dei metodi e sulla valutazione della differenza sistematica utilizzando materiali commutabili.

Produzione di intervalli e limiti decisionali

La letteratura scientifica continua a proporre anche recentemente approcci molto diversi ai valori di riferimento, che dovrebbero essere attentamente valutati non solo sui risultati e sugli aspetti statistici, ma anche e soprattutto sulle modalità di esecuzione. Ancora oggi si utilizzano donatori di sangue per ottenere valori di riferimento.⁸² Il gruppo internazionale di Sansoni e collaboratori si è esercitato sulla variabilità biologica di marcatori ossei come proteina 1 Dikkopf-relata, leptina, osteoprotegerina, sclerostina, lipocalina2 (Lcn2) e resistina.⁸³ I campioni sono stati prelevati una volta alla settimana, per 10 settimane consecutive, da 91 soggetti appartenenti a coorti di 5 Paesi europei. Nonostante le risorse a disposizione, il numero di soggetti selezionati è rimasto ben al di sotto di quello richiesto per gli intervalli di riferimento. Tuttavia, gli autori hanno ritenuto di fornire stime per i limiti degli intervalli di riferimento, anche se con qualche discrepanza nelle tabelle, usando la tecnica "intervallo predittivo" (senza riferimento) al posto della statistica descrittiva, senza fornire i percentili impostati nella procedura (si presume del pacchetto R⁸⁴), mentre nelle figure descrivono intervalli da 5% a 95%, ovvero al 90%, non 95%.

79 Sayili, U., & Gunver, M. G. (2025). A novel modification approach for the one sample Kolmogorov-Smirnov test in large sample size. *Scandinavian Journal of Clinical and Laboratory Investigation*, 85(4), 287–298.

<https://doi.org/10.1080/00365513.2025.2512384>

80 den Elzen, Wendy P.J., Brouwer, Nannette, Thelen, Marc H., Le Cessie, Saskia, Haagen, Inez-Anne and Cobbaert, Christa M.. "NUMBER: standardized reference intervals in the Netherlands using a 'big data' approach" *Clinical Chemistry and Laboratory Medicine (CCLM)*, vol. 57, no. 1, 2019, pp. 42-56. <https://doi.org/10.1515/cclm-2018-0462>

81 Martínez-Sánchez L, Cobbaert CM, Noordam R, Brouwer N, Blanco-Grau A, Villena-Ortiz Y, Thelen M, Ferrer-Costa R, Casis E, Rodríguez-Frias F, den Elzen WPJ. Indirect determination of biochemistry reference intervals using outpatient data. *PLoS One*. 2022 May 19;17(5):e0268522. doi: 10.1371/journal.pone.0268522. PMID: 35588100; PMCID: PMC9119462.

82 Ameratunga R, Lim D, Longhurst H, Mehrtens J, Leung E, Lehnert K, Steele R, Woon ST. Reference intervals for functional lymphocyte proliferation studies using 3H thymidine uptake in adults. *J Immunol Methods*. 2025 May 27;542:113884. doi: 10.1016/j.jim.2025.113884. Epub ahead of print. PMID: 40441410.

83 Sansoni V, Lombardi G, Díaz-Garzón J, Calle PF, Bartlett WA, Coşkun A, Itkonen O, Jonker N, Sandberg S, Aarsand AK, Banfi G, Carobene A; 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 Mar 15;570:120213. doi: 10.1016/j.cca.2025.120213. Epub 2025 Feb 25. PMID: 40010661.

84 Contini ME, Spence JR, Stanley DJ. Introducing prediction intervals for sample means. *Biochem Med (Zagreb)*.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

La tecnica dell'intervallo predittivo bponderato è ritenuta del tipo "robusto", purtuttavia è una metodologia parametrica fondata sulla distribuzione Gaussiana.⁸⁵

La popolazione studiata dal gruppo di Sansoni proviene dal progetto European Biological Variation Study (EuBIVAS)^{86, 87}. EuBIVAS ha coinvolto 91 volontari presumibilmente sani (53 femmine e 38 maschi; età compresa tra 21 e 69 anni) provenienti da 6 Paesi europei, che hanno compilato un questionario per fornire informazioni sul loro stile di vita e sul presunto stato di salute, che è stato ulteriormente verificato da una serie di esami di laboratorio di routine eseguiti durante ogni prelievo. I soggetti sono 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, la popolazione studiata da EuBIVAS ha fornito stime approssimate della variabilità biologica, ma l'estrazione dagli stessi dati di percentili utilizzabili come intervalli di riferimento utilizzabili dai laboratori medici, non sembra sostenuta dalla numerosità dei soggetti e dalle loro caratteristiche. Anche applicando eventuali stimatori statistici robusti come l'intervallo predittivo bponderato, raccomandato da alcune fonti^{88, 89}, ma criticato da altre⁹⁰, che infatti non ne ammettono l'applicabilità pratica nei tipici laboratori medici. L'esperienza EuBIVAS rimane in apparenza un esercizio speculativo utile per simulazioni, ma di scarsa utilità nel mondo reale dei laboratori. Ricercatori in Turchia hanno studiato la variabilità biologica di HbA1c.⁹¹ Il reclutamento ha incluso inclusi 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 personale di laboratorio. Lo stesso studio ammette importanti limitazioni: si tratta di volontari presumibilmente sani la cui variabilità biologica rappresenta la variazione in condizioni fisiologiche e può quindi differire da quella dei pazienti con diabete. Inoltre, si segnala la mancanza di partecipanti giovani e anziani. Da ricercatori di diverse università negli USA e Turchia abbiamo ricevuto un'analisi della coerenza e dell'ampiezza degli intervalli di riferimento calcolati da sei procedure e la valutazione di un protocollo per la fusione dei dati di riferimento di diverse istituzioni.⁹² Usando lo stesso

2024 Oct 15;34(3):030101. doi: 10.11613/BM.2024.030101. Epub 2024 Aug 5. PMID: 39171086; PMCID: PMC11334197.

85 Horn PS. A biweight prediction interval for random samples. J Am Stat Assoc. 1988;83(401):249-256. doi:10.2307/2288947.

86. Carobene, Anna, Campagner, Andrea, Ucheddu, Christian, Banfi, Giuseppe, Vidali, Matteo and Cabitza, Federico. "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" Clinical Chemistry and Laboratory Medicine (CCLM), vol. 60, no. 4, 2022, pp. 556-568. <https://doi.org/10.1515/cclm-2021-0599>.

87. Carobene A, Strollo M, Jonker N, Barla G, Bartlett WA, Sandberg S, Sylte MS, Røraas T, Sølvi UØ, Fernandez-Calle P, Díaz-Garzón J, Tosato F, Plebani M, Coşkun A, Serteser M, Unsal I, Ceriotti F; 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 Oct 1;54(10):1599-608. doi: 10.1515/cclm-2016-0035. PMID: 27169681.

88. Horn PS, Pesce AJ. Reference Intervals: A User's Guide. AACC Press; 2005.

89. CLSI EP444th Ed. Establishing Reference Intervals. Projects in Progress 2025.

90. Ferruccio Ceriotti, Reference Intervals: A User's Guide. Paul S. Horn and Amadeo J. Pesce. Washington, DC: AACC Press, 2005, 115 pp., \$61.00 (\$49.00 AACC members), softcover (CD-ROM included). ISBN 1-59425-035-9., Clinical Chemistry, Volume 52, Issue 3, 1 March 2006, Pages 544-545, <https://doi.org/10.1373/clinchem.2005.058131>

91. Keleş, M., & Ünver Şeker, G. (2024). Within- and between-subject biological variation data for whole blood HbA1c from 38 apparently healthy Turkish subjects. Scandinavian Journal of Clinical and Laboratory Investigation, 84(7-8), 535-539. <https://doi.org/10.1080/00365513.2024.2439394>

92. George G Klee, Kiyoshi Ichihara, Yesim Ozarda, Nikola A Baumann, Joely Straseski, Sandra C Bryant, Christina M

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

gruppo di soggetti: 92 uomini e 148 donne raccolti a Rochester, mentre 125 uomini e 125 donne sono stati raccolti a Salt Lake City. Lo studio ha considerato 28 esami di chimica clinica, da albumina ad acido urico. Mayo-SAS quintile, LAVE(-)NP e EP Evaluator hanno generato intervalli simili, ma spesso più ampi di quelli ottenuti con procedure parametriche. Le procedure LAVE hanno generato intervalli più stretti per i marcatori nutrizionali e infiammatori. La trasformazione con parametri di regressione non ha garantito l'omogeneità dei dati riuniti. I metodi parametrici sono da loro consigliati quando non è possibile escludere valori inappropriati. Per ottenere stime robuste, si raccomandano insiemi di dati superiori a 200. Occorre fare attenzione quando si vogliono fondere dati di istituzioni diverse.

Secondo ricercatori in Turchia, è vero che il processo di sviluppo di intervalli di riferimento è spesso al di là delle capacità di un singolo laboratorio perché complesso, costoso e richiede molto tempo.⁹³ Pertanto, un laboratorio può alternativamente trasferire e verificare intervalli stabilite da una fonte esterna (ad esempio, foglietti illustrativi dei fabbricanti, pubblicazioni). L'IFCC C-RIDL negli ultimi anni ha condotto studi multicentrici per ottenere RI comuni. Tuttavia, poiché la responsabilità più ampia del Comitato, dal suo nome, include i "limiti decisionali", il 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. I limiti decisionali sono associati a un rischio di esiti avversi 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.⁹⁴

Åsberg e Mikkelsen da Trondheim portano un esempio di elaborazione bivariata.⁹⁵ L'intervallo bivariato è migliore dei due intervalli univariati per TSH e fT4, ma al laboratorio si presenta la difficoltà di rappresentarlo accanto ai risultati degli esami.

Una possibilità è quella di rappresentare graficamente la coppia s-TSH e s-FT4 come un punto in un sistema di coordinate in grafico. In questo caso, il sistema informativo dei dati di laboratorio deve essere in grado di creare e inviare il grafico e i sistemi di cartelle cliniche dei medici devono essere in grado di ricevere, memorizzare e riprodurre il grafico. Questo è più facile a dirsi che a farsi. Un altro modo è quello di calcolare la distanza di Mahalanobis della coppia di s-TSH e s-FT4 come se fosse derivata dalla popolazione di riferimento, e quindi riferire se la distanza è all'interno dell'intervallo di riferimento bivariato del 95% o meno. Questa dicotomizzazione sarebbe analoga all'"etichettatura" secondo l'intervallo di riferimento univariato. Un terzo modo è quello di riportare il percentile nella popolazione di riferimento corrispondente alla distanza Mahalanobis delle concentrazioni misurate. Se il quadrato di queste distanze ha una distribuzione chi-quadro con 2 gradi di libertà, si potrebbe usare una funzione chi-quadro inversa per calcolare i percentili. In caso contrario (come nel caso in esame), si potrebbe invece utilizzare una tabella di

Wood-Wentz, Reference Intervals: Comparison of Calculation Methods and Evaluation of Procedures for Merging Reference Measurements From Two US Medical Centers, American Journal of Clinical Pathology, Volume 150, Issue 6, December 2018, Pages 545–554, <https://doi.org/10.1093/ajcp/aqy082>

93. Ozarda, Yesim. Establishing and using reference intervals. Turkish Journal of Biochemistry, vol. 45, no. 1, 2020, pp. 1-10. <https://doi.org/10.1515/tjb-2017-0299>

94. 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 Sep;55(6):420-431. doi: 10.1080/10408363.2018.1482256. Epub 2018 Jul 26. PMID: 30047297.

95. Åsberg, A., & Mikkelsen, G. (2024). A bivariate reference interval for TSH and free thyroxine. Scandinavian Journal of Clinical and Laboratory Investigation, 1–6. <https://doi.org/10.1080/00365513.2024.2437613>

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

percentili per fornire un valore approssimativo. Per esempio, una coppia di s-TSH pari a 0,640 mIU/L e s-FT4 pari a 18,4 pmol/L avrebbe i valori trasformati Box-Cox di -0,451 e 2,90, rispettivamente, e una distanza Mahalanobis di 2,22 che corrisponde approssimativamente al 90° percentile nella popolazione di riferimento (Tabella 1). In conclusione, secondo i 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.

Gli stessi Autori norvegesi mettono in luce che se si utilizzano intervalli di riferimento per calcolare il quoziente di probabilità (likelihood ratio) e l'area sotto la curva caratteristica operativa (Receiver operating characteristic, ROC).⁹⁶ La debolezza dello strumento degli intervalli può inficiare la validità di strumenti raffinati e potenti come il quoziente di probabilità. I norvegesi si chiedono cosa si può fare per eliminare i problemi legati all'utilizzo di standard di riferimento imperfetti. Un precedente articolo di Umemneku et al.⁹⁷ portava esempi, tuttavia ristretti a biomarcatori dicotomici e quelli per biomarcatori quantitativi non ideali. Alla fine, i ricercatori non hanno soluzioni e indirettamente confermano l'utilità di alcuni strumenti interpretativi solo per risultati di tipo qualitativo o classificati con limiti decisionali.

Le difficoltà della produzione degli intervalli di riferimento sono riconosciute da Higgins e collaboratori.⁹⁸ Poiché si tratta di processo è lungo, costoso e spesso al di là delle capacità dei singoli laboratori, gli intervalli di riferimento possono essere trasferiti da un laboratorio (da cui è stato stabilito l'intervallo di riferimento) a un altro laboratorio. Per Higgins, il trasferimento degli intervalli di riferimento comporta il calcolo di un nuovo intervallo di riferimento per un laboratorio o uno strumento di misura diverso, utilizzando l'equazione di regressione di un metodo di confronto. Dopo il trasferimento, la verifica viene utilizzata per valutare se l'intervallo di riferimento trasferito o un intervallo di riferimento preesistente può essere accettato per l'uso in un laboratorio ricevente. Garantire la comparabilità e la concordanza dei dati ottenuti da diversi laboratori medici è stato un obiettivo centrale della medicina di laboratorio e può essere raggiunto attraverso l'armonizzazione. L'armonizzazione del processo totale dell'esame (total testing, TTP), compresa l'armonizzazione degli intervalli di riferimento, standardizzerà l'interpretazione dei risultati degli esami, aumenterà la trasferibilità dei risultati tra i laboratori e, in ultima analisi, porterà ad attuare con successo gli intervalli di riferimento armonizzati nella pratica clinica.

Nathan E. Timbrell ha ragionato sulle limitazioni degli intervalli di riferimento in chimica clinica.⁹⁹ Le differenze sistematiche dei risultati dei metodi e le differenze nelle popolazioni possono influenzare notevolmente gli intervalli e impedirne il trasferimento tra laboratori. Le differenze nelle caratteristiche individuali, come l'età, l'etnia e il sesso, possono determinare ampie variazioni in alcuni esami. Alcuni pazienti, come quelli il cui sesso differisce da quello presunto alla nascita, possono richiedere intervalli propri, oppure sarà necessario decidere quale

96. Åsberg, A., & Åsberg, A. E. (2025). Imperfect reference standards cause biased likelihood ratios. *Scandinavian Journal of Clinical and Laboratory Investigation*, 1–7. <https://doi.org/10.1080/00365513.2025.2528214>

97. Umemneku Chikere CM, Wilson K, Graziadio S, et al. Diagnostic test evaluation methodology: a systematic review of methods employed to evaluate diagnostic tests in the absence of gold standard—an update. *PLOS One*. 2019;14(10):e0223832. doi: 10.1371/journal.pone.0223832.

98. Higgins V, Nieuwesteeg M, Adeli K. Reference intervals: theory and practice. In *Contemporary Practice in Clinical Chemistry*. Academic Press, 2020: 37-56 [PMID: 32151500 DOI: 10.1016/B978-0-12-815499-1.00003-X]

99. Timbrell NE. The Role and Limitations of the Reference Interval Within Clinical Chemistry and Its Reliability for Disease Detection. *Br J Biomed Sci*. 2024 Feb 28;81:12339. doi: 10.3389/bjbs.2024.12339.

utilizzare. In generale, il problema comune a questi fattori è l'interpretazione. Per questo motivo, serve migliore formazione su intervalli e loro uso, unita a una migliore comprensione delle influenze che li differenziano e a una comunicazione più trasparente da parte dei laboratori su come sono stati ricavati intervalli e limiti. Timbrell raffronta ai metodi tradizionali quelli cosiddetti "indiretti", che utilizzano i dati dei pazienti già raccolti e archiviati in laboratorio. Probabilmente il vantaggio più importante dell'approccio indiretto è che la RI stabilita dovrebbe essere derivata da dati specifici per il metodo del laboratorio e per la popolazione servita. Per verificare gli intervalli con l'approccio indiretto, rimane l'indicazione IFCC di misurare i campioni di 20 individui con uno stato di malattia noto o di monitorare la percentuale di risultati anormali e confrontarli.

Timbrell considera che gli intervalli creano un'interpretazione dicotomica basata sui dati di una popolazione di riferimento, ma i risultati possono differire di una sola cifra significativa e cadere ai lati opposti del limite inferiore o superiore. Pertanto, ci si può chiedere se la variazione dei risultati sia clinicamente significativa. Superano questa debolezza i limiti decisionali clinici, limiti o valori di soglia che sono stati determinati da studi clinici per essere associati a un rischio clinico o a un esito specifico. Sono comunemente stabiliti attraverso studi di esito clinico e comunicati attraverso linee guida pubblicate. La scelta di un limite decisionale è determinata da sensibilità e specificità cliniche di un esame a quel livello. Ovvero, probabilità di un risultato positivo in un individuo con la malattia in esame e probabilità di un risultato negativo in un individuo senza la malattia.

Secondo Westgard, i livelli decisionali medici (come quelli tratti da Statland¹⁰⁰) forniscono possibili livelli decisionali critici per valutare le prestazioni (precisione, scostamento sistematico) e determinare la metrica Sigma e le procedure di controllo di qualità appropriate.¹⁰¹

Individuare i livelli decisionali medici non è facile.¹⁰²

Dobbiamo registrare casi di ambiguità nell'uso del termine "limite decisionale clinico". Un gruppo finlandese e svedese, ad esempio, lo usa per indicare la differenza critica, calcolata sulla variabilità biologica e su quella di misura.¹⁰³ Per la proposta, la variazione combinata, definita come variazione diagnostica, CVD, comprende la variazione dell'esame, CVA, la variazione preesame, CVPRE, e la variazione biologica intra-individuale, CVI [15,28]. Per modellare le differenze clinicamente significative, CD, sono state utilizzate due distribuzioni gaussiane sovrapposte. La CD tra due valori di misurazione è stata utilizzata per calcolare la CVD clinicamente accettabile di un risultato di laboratorio utilizzando la formula convenzionale del valore di variazione relativo, RCV.

Un'altra possibilità di fraintendimento è quella con i livelli critici per il controllo di qualità.¹⁰⁴ "Limiti decisionali medici" (MDL) è il termine utilizzato da Tetrault e Steindel nel Q-Probe del 1994 del CAP, che esamina le pratiche di eccezione del controllo di qualità.¹⁰⁵ Gli MDL sono

100. Statland BE. Clinical Decision Levels for Laboratory Tests, Second Edition [Oradell NJ;Medical Economics Books, 1987].

101. Westgard JS. Medical Decision Levels. <https://westgard.com/resources/resources/decision.html>

102. F. Carle. Decision limits and diagnostic algorithms. RIMeL - IJLaM, Vol. 2, N. 1, 2006. 57-61

103. 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 Feb 1;540:117233. doi: 10.1016/j.cca.2023.117233. Epub 2023 Jan 21. PMID: 36693582.

104. Westgard JO. The Myth of Medical Decision Limits. Created on 15 February 2009.

<https://westgard.com/essays/trends/essay8.html>

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

Raccomandazioni M2-2

*relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025*

Versione 0.1

descritti come una “seconda serie di limiti fissati per i valori di controllo ... intesa come una serie più ampia di limiti che indicano la gamma di risultati accettabili dal punto di vista medico”. Il concetto di MDL fa riferimento a un precedente documento di Steindel¹⁰⁶, che a sua volta fa riferimento a un precedente abstract di un gruppo CAP¹⁰⁷. MDL non deve essere confuso con l'uso del livello decisionale medico (Xc), che si riferisce al livello o alla concentrazione a cui un esame viene interpretato criticamente per la cura e il trattamento del paziente. Per Westgard invece si deve definire un cambiamento importante dal punto di vista medico (o intervallo di decisione clinica, Dint) a un livello di decisione medica (Xc), quindi utilizzare un modello di pianificazione della qualità clinica per tradurre il requisito di utilità medica in specifiche operative per l'imprecisione e l'inesattezza consentite e il controllo di qualità necessario.

Dall'esperienza del gruppo di Cesana e Ferraro si ricavano interessanti indicazioni.¹⁰⁸ Tre dei quattro metodi indagati darebbe soglie decisionali derivate dal limite inferiore dell'intervallo di riferimento, affermandone la corrispondenza con le soglie raccomandate dalle guide cliniche vigenti (NICE), che in realtà non hanno alcun collegamento con gli intervalli.¹⁰⁹ Anzi, la guida NICE mette in guardia sul fatto che l'etnia africana avrebbe livelli di B12 più elevati dei bianchi e degli asiatici. Inoltre, lo studio dei ricercatori ha dimostrato scarti sistematici tra metodi sufficienti per invalidare l'applicazione delle soglie raccomandate dalla guida NICE. Colpisce l'approssimazione con cui fabbricanti e laboratori trattano intervalli e soglie decisionali e la scarsa attenzione verso l'armonizzazione dei risultati.

Sono stati studiati in Sri Lanka¹¹⁰ diversi indici calcolati a partire dai parametri eritrocitari per selezionare soggetti a rischio di tratto β -Talassemico, applicando ai risultati soglie fisse (cut-off) in base alle probabilità (valori predittivi, sensibilità specificità, curve operative ROC, indice di Youden derivato da sensibilità e specificità), non ai percentili delle rispettive distribuzioni. I risultati hanno dimostrato il vantaggio della separazione tra maschi e femmine, che sono stati denominati dai ricercatori “popolazioni”. Naturalmente, altri fattori sono influenti su queste misure (fattori etnici, nutrizionali e ambientali^{111, 112} [51, 52]), ma sono difficilmente isolabili in pratica nel mondo reale.

CSCC (Canadian Society of Clinical Chemists) ha ultimamente prodotto raccomandazioni per gli intervalli di riferimento che combinano armonizzazione dei metodi, dati dal mondo reale e verifica su piccoli campioni.¹¹³ Le raccomandazioni CSCC derivano da precedenti esperienze degli stessi

106. Steindel SJ. New directions in quality control: Part I. New QC systems. Lab Med 1986;17:463-466.

107. Howanitz PJ, Kafka MT, Steindel SJ, et al. Quality control run acceptance and rejection using fixed and medically useful limits for QAS Today. Clin Chem 1985;31:1016 (abstract).

108. Cesana, Bruno Mario, da Molin, Simona, Jassam, Nuthar, Barth, Julian H., Buoro, Sabrina, Tosi, Martina, Zuccotti, Gianvincenzo, Marcovina, Santica and Ferraro, Simona. "Assessing the harmonization of current total vitamin B12 measurement methods: relevance and implications" Clinical Chemistry and Laboratory Medicine (CCLM), 2025. <https://doi.org/10.1515/cclm-2025-0939>

109. Vitamin B12 deficiency in over 16s: diagnosis and management. London: National Institute for Health and Care Excellence (NICE); 2024 Mar 6. NICE guideline. 6 March 2024. www.nice.org.uk/guidance/ng239

110. Isuru Aravinda, Shashini Dharmasiri, Chathurika Sewwandi, Sinthu Karunaitas, Sonali Goonetilleke, Karunaitas Rasaratnam PII: S0009-8981(25)00471-1. Population-adjusted cut-off: A new approach for enhancing the diagnostic efficacy of hematological discrimination formulae for screening β -Thalassemia trait. DOI: <https://doi.org/10.1016/j.cca.2025.120592> Reference: CCA 120592 To appear in: Clinica Chimica Acta

111. A. Bogdanova, L. Kaestner, G. Simionato, A. Wickrema, A. Makhro, Heterogeneity of Red Blood Cells: Causes and Consequences, Front Physiol 11 (2020) 392.

112. A.T. Taher, K.M. Musallam, M.D. Cappellini, β -Thalassemias, N Engl J Med 384(8) (2021) 727- 743.

113. M.K. Bohn, D. Nyholt, C. Balion, G. Cembrowski, C. Collier, V.D. Guire, V. Higgins, B. Jung, O. Landon, Z. Mohammed-Ali, D. Seccombe, J. Taher, A.K.Y. Tsui, A.A. Venner, N.W. Al-Habeab, K. Adeli, Best practice guidelines

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

canadesi.¹¹⁴ Gli esami candidati all'armonizzazione sono stati selezionati in base alla tracciabilità e ai risultati della VEQ. Due anni di risultati dei pazienti per 16 esami sono stati estratti da quattro laboratori. Tutti gli intervalli di riferimento armonizzati raccomandati sono stati verificati in nove laboratori canadesi che includevano tutti i principali produttori utilizzando campioni di siero e plasma raccolti da 60 volontari sani, approccio che corrisponde allo studio di scala media proposto da CLSI. I canadesi accennano ai limiti decisionali, senza però distinguerne l'utilizzo dagli intervalli. La loro attenzione è tutta concentrata sulle differenze tra laboratori. Tuttavia, raccomandano che intervalli e limiti non siano imposti autonomamente dai laboratori, ma vengano discussi con i clinici.

La rassegna di tre ricercatori, uno austriaco, uno statunitense e uno svedese, ha recentemente esaminato pregi, difetti e alternative degli intervalli di riferimento.¹¹⁵ Il criterio del 95% viene sconfessato, perché produce inutilmente un 5% di falsi positivi. Difetto presente anche utilizzando statistiche non parametriche (percentili), approcci raffinati come l'intervallo predittivo, ovvero costruito con la distribuzione t di Student, oppure intervalli di tolleranza utilizzando il chi-quadrato¹¹⁶. 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. I falsi positivi sono riconosciuti un problema sia clinico che economico. Tra le alternative, i ricercatori raccolgono l'ampliamento degli intervalli a 1% centrale, la sostituzione con punteggi z, ovvero lo scarto normale standardizzato, o con percentili, come dimostrato per l'emoglobina.¹¹⁷ Ulteriore alternativa è riconosciuta all'approccio bayesiano del quoziente di probabilità, ben noto da tempo. Infine, vengono riconosciuti gli importanti limiti degli intervalli forniti dai fabbricanti. La rassegna non giunge a conclusioni definitive, ma attribuisce al laboratorio la possibilità di decidere quale sia l'approccio migliore per selezionare un intervallo di riferimento per un particolare esame, con l'obiettivo di migliorare l'interpretazione dei risultati riducendo l'ansia dei pazienti e gli sprechi nel sistema sanitario. In definitiva, non viene raccomandato un approccio "unico" agli intervalli di riferimento per tutti gli esami, ma uno specifico per ciascun esame, in base alla sua variabilità biologica e all'utilità clinica. Si afferma che il tema è maturo per un cambiamento di paradigma catalizzato da intelligenza artificiale con grandi numeri di dati provenienti da gruppi di pazienti sani e diagnosticati.

In sostanza, i ricercatori dei tre Paesi hanno evidenziato i difetti degli intervalli, senza fornire soluzioni alternative praticabili per i laboratori di dimensioni anche non molto grandi, senza considerare l'unica alternativa più affidabile: i limiti clinici decisionali, che presuppongono di abbandonare del tutto il valore decisionale degli intervalli di riferimento. Abbiamo poi esempi di situazioni "ibride". ISTH (International Society on Thrombosis and

on reference interval harmonization in Canada: Evidence-based recommendations from the CSCC working group on reference interval harmonization (CSCC WG-hRI), Clinical Biochemistry (2025), doi:

<https://doi.org/10.1016/j.clinbiochem.2025.110986>

114. Bohn MK, Bailey D, Balion C, Cembrowski G, Collier C, De Guire V, Higgins V, Jung B, Ali ZM, Secombe D, Taher J. Reference interval harmonization: harnessing the power of big data analytics to derive common reference intervals across populations and testing platforms. Clinical Chemistry. 2023 Sep;69(9):991-1008.

115. Badrick, T., El-Khoury, J. M., & Theodorsson, E. (2025). Laboratory reference intervals - history and modern approaches for improved utility. Scandinavian Journal of Clinical and Laboratory Investigation, 85(4), 229–241. <https://doi.org/10.1080/00365513.2025.2512995>

116. Young DS. tolerance: an R package for estimating tolerance intervals. J Stat Soft. 2010;36(5):1–39. doi: 10.18637/jss.v036.i05.

117. Cheng CY, Hsu TH, Yang YL, et al. Hemoglobin and Its Z score reference intervals in febrile children: a cohort study of 98,572 febrile children. Children. 2023;10(8):1–14. doi: 10.3390/children10081402.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

Haemostasis) ha pubblicato i criteri aggiornati per la diagnosi di coagulazione intravascolare disseminata (DIC).¹¹⁸ I punteggi per il conteggio delle piastrine sono basati su valori fissi (50 o 100 $\times 10^9/L$), come pure per PT-INR (3 e 6 secondi), fibrinogeno (100 mg/dL), mentre quelli per D-dimero sono multipli (3 o 7) del limite superiore degli intervalli di riferimento. In questo caso, si rifiuta la statistica degli IR, ma si mantiene il concetto di IR per definire compiutamente il metodo in confronto tra gli altri metodi. Il precedente lavoro del gruppo di Suzuki, invece, forniva limiti fissi a 3.0 e 7.0 $\mu g/mL$, con la prescrizione di armonizzare i metodi di misura utilizzati.¹¹⁹ Infine, Ceriotti ha riportato i risultati di una indagine italiana sugli intervalli per la fosfatasi alcalina.¹²⁰ Si vede un numero significativo di laboratori che non riportano i limiti inferiori corretti degli intervalli di riferimento dell'ALP pediatrica, ostacolando così l'identificazione dell'ipofosfatasia. Per sospettare questa patologia, è essenziale che i laboratori comunichino correttamente i bassi valori di ALP e che i limiti inferiori degli intervalli di riferimento siano definiti con precisione. Tra i laboratori che trattano pazienti pediatrici e che hanno intervalli specifici per età e sesso (N.=70), il 65% fornisce un LRL corretto per i maschi di 2 anni, il 51% per le femmine di 12,5 anni, il 60% per i maschi di 13,5 anni e il 77% per i maschi di 17,5 anni. Nonostante la selettività del campione studiato, gli intervalli hanno dimostrato anche in questo caso tutte le loro debolezze.

Uno studio di Boston si è dedicato all'esame albuminuria¹²¹, partendo dall'applicazione pedissequa delle regole IFCC¹²², ovvero il 95° percentile come limite superiore di "normalità" (sic), con un minimo di 120 partecipanti. Raccolti 1623 adulti sani partecipanti a un'indagine nazionale, hanno escluso quelli di età inferiore ai 20 anni e con una lunghissima serie di patologie, dalle malattie cardiovascolari, all'uso pesante di alcol, droghe iniettabili, indice di massa corporea troppo basso o troppo alto, malattie virali, fino a carenza alimentare e fumo. Inoltre, sono stati esclusi i partecipanti con valori superiori al 97,5° percentile specifico per sesso (ritenuti anomali o "outlier"). Infine, hanno riscontrato differenze tra maschi e femmine, nonostante il ritardo della medicina nel definire soglie specifiche di maschi e femmine per l'albuminuria. Trovati i limiti dell'intervallo, i ricercatori hanno notato che questi trascurano lo spettro continuo di rischio al di sotto dei 30 mg/g, nonostante la relazione continua provata tra albuminuria apparentemente "normale" e rischio di eventi cardio-renali¹²³. Hanno fatto così emergere l'enigma di definire un intervallo di riferimento basato sulla frequenza dei valori in popolazioni "sane" (un approccio puramente statistico) rispetto alla definizione di intervalli di riferimento basati su esiti clinicamente significativi, concludendo che nel caso dell'albuminuria, il limite superiore statistico della normalità sarebbe di 20-30 mg/g, mentre il limite superiore clinico potrebbe essere di 7-10 mg/g.

118. 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*. 2025 Aug 19. doi: 10.1055/a-2675-6068. Epub ahead of print. PMID: 40829630.

119 Suzuki K, Wada H, Imai H, Iba T, Thachil J, Toh CH Subcommittee on Disseminated Intravascular Coagulation. A re-evaluation of the Ddimer 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(07):1442–1444

120 Ferruccio Ceriotti. Risultati del questionario relativo ai limiti inferiori degli intervalli di riferimento pediatrici della fosfatasi alcalina. *Biochimica Clinica* 2025 Settembre;49(3):226-30 DOI: 10.23736/S0393-0564.25.00024-X

121 Sophie E Claudel, Sushrut S Waikar, Ashish Verma, Albuminuria Reference Ranges in Healthy US Adults, *The Journal of Applied Laboratory Medicine*, Volume 10, Issue 3, May 2025, Pages 767–771, <https://doi.org/10.1093/jalm/jfaf015>

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

123 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.

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

La produzione di limiti decisionali è un lavoro molto delicato. Ad esempio, le linee guida raccomandano di utilizzare la troponina T ad alta sensibilità (hsTnT) e il peptide natriuretico N-terminale di tipo pro-B (NT-proBNP) per stratificare il rischio dei pazienti emodinamicamente stabili con embolia polmonare acuta (PE).¹²⁴ Tuttavia, non sono stati definiti valori soglia basati sulle prove (“evidence based”) per questa applicazione clinica. Un gruppo di Boston ha verificato l’uso in pratica dei limiti raccomandati.¹²⁵ I risultati indicano che la soglia ottimale della hsTnT è circa tre volte superiore al cutoff raccomandato dalle linee guida. Tuttavia, i risultati dicono anche che le prestazioni diagnostiche dei due marcatori (AUC, sensibilità e specificità) sono scarse, suggerendo che le attuali linee guida potrebbero richiedere una revisione. Ricercatori di Salt Lake City e San Diego¹²⁶ hanno studiato il 99° percentile di esami come anticorpi anticardiolipina (aCL) e anti-β₂-glicoproteina I (αβ₂GP1) IgG/IgM per la sindrome antifosfolipidi proposti dai criteri dell’American College of Rheumatology¹²⁷. Hanno dimostrato che tutto dipende dal metodo utilizzato, concludendo che, quando gli studi clinici non sono fattibili, le soglie raccomandate dal fabbricante offrono un’alternativa affidabile. Infatti, le linee guida sono poco seguite dai laboratori: un’indagine condotta dall’ECAT (External Quality Control of Diagnostic Assays and Tests) ha rilevato che circa il 45% dei laboratori ha adottato i valori limite del fabbricante e solo il 38% li ha verificati¹²⁸. I limiti proposti dal fabbricante sono tipicamente derivati da ampi studi di validazione e considerando popolazioni con la patologia. Le recenti osservazioni di ricercatori di Yale hanno evidenziato l’influenza della diffusione dei metodi LC-MS/MS accanto ai precedenti immunometrici sulle soglie decisionali per il testosterone (TT)¹²⁹. La soglia per il testosterone basso a 300 ng/dL (10.4 nmol/L) si basa su studi che correlano i livelli misurati con i precedenti metodi immunometrici con la gravità dei sintomi dei pazienti. Tuttavia, esistono differenze nelle misurazioni di TT tra i metodi LC-MS/MS e immunometria, con una variabilità osservata all’estremità inferiore dell’intervallo di riferimento di TT. La guida dell’American Urological Association del 2018 per la carenza di testosterone raccomanda un limite di 300 ng/dL¹³⁰. Ma il limite basso di 264 ng/dL proposto dalla Endocrine

124. Konstantinides SV, Meyer G, Becattini C, Bueno H, Geersing G-J, Harjola V-P, et al. 2019 ESC guidelines for the diagnosis and management of acute pulmonary embolism developed in collaboration with the European Respiratory Society (ERS): the task force for the diagnosis and management of acute pulmonary embolism of the European Society of Cardiology (ESC). *Eur Heart J* 2020;41:543–603

125. Timothy M Matthews, Gregory A Peters, Grace Wang, Nora Horick, Kyle E Chang, Savannah Harshbarger, Christiana Prucnal, Drew A Birrenkott, Karsten Stannek, Eun Sang Lee, Isabel Dhar, Jesse O Wrenn, William B Stubblefield, Christopher Kabrhel, Optimal Cutoff Values and Utility of High-Sensitivity Troponin T and NT-proBNP for the Risk Stratification of Patients with Acute Pulmonary Embolism, *Clinical Chemistry*, Volume 71, Issue 5, May 2025, Pages 559–566, <https://doi.org/10.1093/clinchem/hvae212>

126. Alexis Dadelahi, Brandon S Walker, Dipanwita Banerjee, Michael Mahler, Abdulrahman Saadalla, Vijayalakshmi Nandakumar, Is the 99th Percentile Cutoff Still Relevant? A Single-Center Assessment of Different Thresholds for Diagnosing Antiphospholipid Syndrome, *The Journal of Applied Laboratory Medicine*, Volume 10, Issue 3, May 2025, Pages 605–618, <https://doi.org/10.1093/jalm/jfae149>

127. Barbhैया M, Zuily S, Naden R, Hendry A, Manneville F, Amigo MC, et al. The 2023 ACR/EULAR antiphospholipid syndrome classification criteria. *Arthritis Rheumatol* 2023;75:1687–702.

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

129. Adith S Arun, Thomas J S Durant, Joe M El-Khoury, Harlan M Krumholz, Reevaluating the Threshold for Low Total Testosterone, *Clinical Chemistry*, Volume 71, Issue 5, May 2025, Pages 609–611, <https://doi.org/10.1093/clinchem/hvaf025>

130. 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

Raccomandazioni M2-2
relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

Society potrebbe essere una soglia più accurata in futuro, rispetto a quella storicamente utilizzata di 300 ng/dL¹³¹, perché i nuovi metodi immunometrici sono generalmente allineati con la LC-MS/MS.

NOTA. Ripartizione 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. Non fornisce alcuna indicazione sui criteri di ripartizione. CLSI EP45 in Tabella 3 elenca esempi di fattori di suddivisione, divisi in criteri comunemente considerati, iniziando da “Sesso biologico”. La stessa cosa fa CLSI EP44. CLSI non affronta la controversa questione delle condizioni intermedie tra i sessi biologici e nemmeno delle conseguenze sugli esami di laboratorio variazioni di genere psico-sociale.

La questione ha attirato l'attenzione di alcuni ricercatori.

Due australiani hanno riferito che la terapia ormonale per l'affermazione del genere comporta cambiamenti nei parametri di laboratorio con intervalli di riferimento specifici per il sesso.¹³²

Quattro statunitensi hanno 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.¹³³ Altri negli USA hanno mostrato negli individui transgenere sottoposti a terapia ormonale mascolinizante o femminilizzante cambiamenti significativi in alcuni esami che presentano variazioni sesso-specifiche.¹³⁴ Irwig da Boston ritiene che molti sistemi informativi di laboratorio e sanitari debbano essere aggiornati per rilevare l'identità di genere.¹³⁵

Davis di ASCLS-Maryland consiglia ai laboratori di accompagnare i risultati con una nota tipo “Alcuni studi hanno dimostrato che questo intervallo di riferimento potrebbe non essere applicabile alle minoranze sessuali/di genere”.¹³⁶

La dissertazione forse più aggiornata, completa e obiettiva sul tema si trova nella rassegna australiana di quest'anno¹³⁷, dove si ammette che l'uso di intervalli derivati dalla popolazione transgenere è complesso, gli studi hanno generalmente una bassa dimensione del campione e includono adulti con trattamenti ormonali consolidati da tempo. Esistono significative disparità

131. 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.

132. Brendan J Nolan, Ada S Cheung, Laboratory Monitoring in Transgender and Gender-Diverse Individuals, Clinical Chemistry, Volume 71, Issue 3, March 2025, Pages 358–377, <https://doi.org/10.1093/clinchem/hvaf001>.

133. 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 Apr 30;36:100350. doi: 10.1016/j.jcte.2024.100350.

134. Robert M Humble, Dina N Greene, Robert L Schmidt, Gabrielle Winston McPherson, Jessica Rongitsch, Katherine L Imborek, Nicole Nisly, Nancy J Dole, Susan K Dane, Janice Frerichs, Matthew D Krasowski, Reference Intervals for Clinical Chemistry Analytes for Transgender Men and Women on Stable Hormone Therapy, The Journal of Applied Laboratory Medicine, Volume 7, Issue 5, September 2022, Pages 1131–1144, <https://doi.org/10.1093/jalm/jfac025>.

135. Michael S Irwig, Which reference range should we use for transgender and gender diverse patients?, The Journal of Clinical Endocrinology & Metabolism, Volume 106, Issue 3, March 2021, Pages e1479–e1480, <https://doi.org/10.1210/clinem/dgaa671>.

136. Davis D. ASCLS Today. Volume 36 Number 6 | December 2022. The Importance of Accurate Laboratory Reference Intervals for Transgender Patients. <https://ascls.org/the-importance-of-accurate-laboratory-reference-intervals-for-transgender-patients/>

137. Kevin Jessen, Nilika Wijeratne & Ailie Connell (2025) The intersection of the laboratory and transgender care, Critical Reviews in Clinical Laboratory Sciences, 62:5, 347–362, DOI: 10.1080/10408363.2025.2488839

Raccomandazioni M2-2
relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

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 intervalli e limiti decisionali del sesso dichiarato è appropriata 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 (Ministry of Health—Gender Medicine team¹³⁸) non sostituiscono sesso con genere o viceversa, ma li tengono affiancati.

Il gruppo del Ministero afferma che va considerato, 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.

In conclusione, tutta la materia non sembra abbastanza matura per esprimere Raccomandazioni solide da indirizzare ai laboratori medici.

NOTA. Intervalli di riferimento nella validazione clinica: CLSI EP49 e ISO 20916

CLSI produce un documento sulla validità clinica degli esami di laboratorio.¹³⁹

ISO dispone di uno standard sugli studi delle prestazioni cliniche degli esami di laboratorio.¹⁴⁰ è utile esaminare le due proposte per collocare esattamente il ruolo eventualmente svolto dagli intervalli di riferimento.

CLSI EP49: le prestazioni cliniche

CLSI sostiene che “validazione clinica” non abbia una definizione accettata a livello globale; quindi, la proposta di CLSI EP49 è processo che fornisce prove oggettive della conformità dei risultati del metodo di prova alle esigenze dell'utente e all'uso prefissato. Processo che comprende l'esecuzione di esami in condizioni d'uso reali o simulate per dimostrare la capacità di un metodo di prova di fornire risultati associati a una particolare condizione clinica o a un processo o stato fisiologico o patologico in conformità con la popolazione di destinazione e gli utenti previsti. La validazione clinica si basa sulla validità scientifica. La validità scientifica stabilisce l'associazione biologica sottostante tra un misurando e una condizione clinica o uno stato fisiologico, mentre la validazione clinica valuta la capacità del metodo di analisi di rilevare o prevedere il disordine o la condizione di interesse in una popolazione bersaglio. Esempi di misure delle prestazioni cliniche sono le coppie sensibilità clinica e specificità clinica, rapporto di verosimiglianza (o quoziente di probabilità^{141, 142, 143}) positiva (PLR) e rapporto di verosimiglianza negativa (NLR); e la tripletta valore predittivo positivo (PPV), valore predittivo negativo (NPV) e prevalenza. CLSI intende per

138. Cattaneo A, Bellenghi M, Ferroni E, Mangia C, Marconi M, Rizza P, Borghini A, Martini L, Luciani MN, Ortona E, Carè A, Appetecchia M, 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 Aug 27;14(9):908. doi: 10.3390/jpm14090908.

139. CLSI. Framework for Developing Evidence of Clinical Validity of Medical Laboratory Test Methods. 1st ed. CLSI guideline EP49. Clinical and Laboratory Standards Institute, USA, 2025.

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

141. Associazione Medici Endocrinologi. Quoziente di probabilità. 2022.

<https://www.associazionemediciendocrinologi.it/index.php/glossario-statistico/quoziente-di-probabilita>

142. Giavarina D. – Evidence Based Medicine: impiego dei test di laboratorio – Caleidoscopio Italiano n.229 Medical Systems 2014 Genova ISSN 0394 3291

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

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

“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 fase di valutazione (denominato “metodo di prova candidato”) si ottiene mediante il confronto con un “comparatore” che dovrebbe essere il miglior metodo disponibile per determinare se la condizione è presente o assente, oppure potrebbe essere uno di quelli tradizionalmente utilizzati come comparatore. EP49 attualmente, tuttavia, non risolve l’ambiguità tra intervalli di riferimento e limiti o soglie decisionali (punto 3.4.2), anzi. Ma le modalità descritte da EP49 per stimare le coppie e la tripletta della validità clinica restringono le possibilità ai soli risultati qualitativi o a quelli quantitativi classificati con limiti decisionali.

ISO 20916:2019: gli studi di prestazione clinica

ISO 20916 è stata adottata e ripubblicata da UNI, in inglese, solo nel 2024.¹⁴⁴ Nel frattempo, è stata riproposta e confermata in ISO con votazione e ampia maggioranza nel settembre 2024. ISO 20916 definisce le buone prassi di studio per gli studi di prestazione clinica condotti per valutare le prestazioni cliniche e la sicurezza dei dispositivi medico-diagnostici in vitro (IVD) a fini regolatori.

La maggior parte degli studi sui dispositivi medici IVD viene eseguita utilizzando campioni derivanti da resti di campioni prelevati per scopi di cura (avanzi o archiviati). In questi studi, per ISO non vi è alcun rischio per i soggetti derivante dalle informazioni fornite dal dispositivo medico IVD o dalla procedura di raccolta del campione. Tuttavia, quando non vengono utilizzati questi campioni, è necessario considerare che, nel caso i campioni siano raccolti specificamente per lo studio, si presenta un rischio aggiuntivo di danno diretto per il soggetto (ad es. puntura lombare o biopsia tissutale, prelievo di sangue da neonati o pazienti gravemente malati). Inoltre, accade che le informazioni ottenute dai dispositivi medici IVD durante lo studio vengano utilizzate per prendere decisioni sulla gestione del paziente (ad esempio, studi interventistici), con il rischio di risultati falsi negativi o falsi positivi.

ISO 20916 nasce per le buone pratiche di studio per la pianificazione, la progettazione, la conduzione, la registrazione e la rendicontazione degli studi per valutare le prestazioni cliniche e la sicurezza dei dispositivi medici diagnostici in vitro (IVD) a fini regolatori. Per lo scopo di questa nota, sono interessanti concetti e definizioni relativi alle prestazioni dei dispositivi, quindi degli esami di laboratorio.

Ad esempio, la definizione 3.10 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. Le prestazioni cliniche possono includere i valori attesi, la sensibilità e la specificità diagnostica in base alla condizione clinica nota o al processo/stato fisiologico/patologico dell’individuo e i valori predittivi negativi e positivi in base alla prevalenza della malattia. Definizione raccolta da GHTF/SG5/N6:2012.¹⁴⁵

Gli intervalli di riferimento vengono citati nel capitolo 5 (Pianificazione dello studio di prestazione clinica), dove si riferisce che in alcune circostanze, potrebbe essere opportuno

144. UNI EN ISO 20916:2024. Dispositivi medico-diagnostici in vitro - Studi di prestazione clinica che utilizzano campioni provenienti da soggetti umani - Buone pratiche di studio.

145. Global Harmonization Task Force, Clinical Evidence for IVD medical devices — Key Definitions and Concepts [GHTF/SG5/N6:2012], <https://www.imdrf.org/sites/default/files/docs/ghtf/final/sg5/technical-docs/ghtf-sg5-n6-2012-clinical-evidence-ivd-medical-devices-121102.pdf>

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

eseguire gli esami solo presso il sito del fabbricante, motivandolo. Ad esempio, uno studio per determinare i valori di riferimento può spesso essere eseguito interamente presso il sito del fabbricante. Al punto 5.3 (Disegno dello studio di prestazione clinica) lettera j), si includono nelle caratteristiche di prestazione clinica prefissate sensibilità, specificità, valore predittivo positivo, valore predittivo negativo, intervalli di riferimento, soglie decisionali). La natura degli studi è rivalutata al punto 5.5.3.8 (Dispositivo medico IVD in esame e comparatore/i), dove si chiede di elencare il/i comparatore/i. Se il comparatore è un metodo commerciale, includere il nome e il fabbricante e, se applicabile, il numero di versione o di catalogo. Se il comparatore è un metodo di riferimento o un “gold standard”, vanno forniti adeguati riferimenti pubblicati a sostegno della metodologia. Il metodo di riferimento è definito al punto 3.35 (procedura di misurazione di riferimento) come la procedura di misurazione accettata in quanto fornisce risultati idonei ad essere utilizzati per valutare l'esattezza di altre procedure di misurazione per grandezze dello stesso tipo, o nella taratura o nella caratterizzazione di materiali di riferimento. Come definito da ISO 15193¹⁴⁶, a sua volta derivato dal vocabolario di ISO/IEC Guide 99:2007 2.7.¹⁴⁷ Nella sua neutralità integrale, ISO 20916 comprende negli studi di prestazione quelli per intervalli di riferimento, come se la diagnosi di “salute” o di “normalità” fosse un obiettivo di pari rango rispetto alla diagnosi, a fini prognostici e terapeutici, di una condizione patologica.

NOTA. Prestazione clinica e validazione clinica nella letteratura scientifica

Il gruppo “Rete internazionale per la qualità in patologia” (IQN Path, International Quality Network for Pathology) si è recentemente espresso proprio sulla prestazione clinica di un biomarcatore in oncologia.¹⁴⁸ Per IQN Path la definizione ISO 9000 di validazione è il punto di partenza: “la conferma, attraverso la fornitura di prove oggettive, che i requisiti per un uso o un'applicazione specifica sono stati soddisfatti”. La definizione ISO 9000 ha tre componenti che, tradotti nel linguaggio della medicina di laboratorio, vengono applicati come “livello di certezza” (conferma), “caratteristiche di prestazione dell'esame” (evidenza oggettiva) e “valore per lo scopo” (uso prefissato). Dopo il successo di una sperimentazione clinica, la procedura utilizzata per rilevare il biomarcatore può assumere la forma di un esame diagnostico complementare (CDx) o di un esame sviluppato in laboratorio (LDT). I laboratori clinici possono anche modificare CDx o scegliere un LDT piuttosto che un CDx. È meno noto che qualsiasi modifica di un saggio CDx approvato dall'agenzia regolatoria trasformi automaticamente un CDx in un LDT. È necessaria una validazione clinica indiretta che sia adatta allo scopo per gli LDT. La necessità di una validazione clinica è rafforzata dalla FDA e dalle normative europee con l'IVDR in Europa. Quindi si deve considerare il tipo di biomarcatore predittivo, distinti in tre gruppi (Indirect Clinical Validation, ICV groups). 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 che sono direttamente o indirettamente informativi sulle risposte immunologiche al tumore e sui potenziali benefici dell'immunoterapia (ad esempio, tumor

146. ISO 15193:2009, In vitro diagnostic medical devices — Measurement of quantities in samples of biological origin — Requirements for content and presentation of reference measurement procedures .

147. ISO/IEC Guide 99:2007, International vocabulary of metrology — Basic and general concepts and associated terms (VIM)

148. Torlakovic, E.E., Al Dieri, R., Badrick, T. et al. Indirect clinical validation for predictive biomarkers in oncology: International Quality Network for Pathology (IQN Path) Position Paper. Virchows Arch (2025).
<https://doi.org/10.1007/s00428-025-04169-4>

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

mutation burden (TMB), instabilità microsatellitare, sovraespressione della proteina c-MET, PD-L1 o PD1 e altri). Il terzo è un esame di selezione, sviluppato per ridurre il costo degli esami e i tempi di esecuzione per escludere i pazienti che non hanno alcuna possibilità di essere positivi con un biomarcatore definitivo (ad esempio, ROS1 e pan-TRK IHC). Per il Gruppo 1, la validazione clinica indiretta è ottenuta seguendo le linee guida/raccomandazioni pubblicate per la validazione della misurazione. Gli esami del gruppo 2 sono quasi sempre di tipo qualitativo; pertanto, si applicano le linee guida CLSI EP-12 ED3:2023¹⁴⁹. La validazione clinica indiretta per il gruppo 2 richiede un “gold standard” /di riferimento (di solito CDx) della stessa metodologia. Gli esami del gruppo 3 sono di selezione (“screening”). Questo è in contrasto con i gruppi 1 e 2 dell'ICV, che sono determinazioni di biomarcatori definiti, i cui risultati verrebbero utilizzati per decidere il trattamento dei pazienti. Secondo IQN Path, per le ricerche di biomarcatori in cui la lettura viene eseguita da patologi, specialisti o tecnici (P/S/T) e/o dall'analisi delle immagini o intelligenza artificiale (IA/AI), ci si deve regolare in modo che la sensibilità rimanga del 100%. IQN Path dimostra la stessa limitazione di CLSI EP49: il cosiddetto “uso prefissato” prescinde dall'utilità dell'uso. In ISO si sono preoccupati di questo aspetto gli informatici in ISO/IEC/IEEE 24765 3.4468¹⁵⁰, dove utilità viene definita (dalla precedente ISO/IEC 25010:2011¹⁵¹) “grado di soddisfazione dell'utente per il raggiungimento percepito di obiettivi pratici, compresi i risultati dell'uso e le conseguenze dell'uso. Ma il concetto rientra nella recente definizione di utilità clinica di ISO 5649:2024, 3.9, 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.¹⁵² Definizione ripresa ancora una volta da GHTF/SG5/N6:2012, 4.7.¹⁵³ Curioso che il concetto di utilità venga valorizzato più da rappresentanti delle agenzie di regolamentazione dei dispositivi medici e dell'industria regolamentata (come Global Harmonization Task Force) che da esponenti della professione specialistica del laboratorio medico, con l'eccezione della norma su esami fatti in casa ISO 5649:2024. Evidente uno squilibrio di prospettiva, dove il laboratorio peserebbe più degli utilizzatori, medici e pazienti, a cui gli operatori sul mercato sono più attenti. A dimostrazione di questo fenomeno troviamo il documento “MDCG 2022-2 Linee guida sui principi generali dell'evidenza clinica per i dispositivi medici diagnostici in vitro (IVD) Gennaio 2022”¹⁵⁴. il Gruppo di coordinamento per i dispositivi medici (MDCG) 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. Il Gruppo di coordinamento dei dispositivi medici (MDCG) si occupa di questioni chiave del settore dei dispositivi medici, dalla

149. CLSI (2023) Evaluation of qualitative, binary output examination performance. CLSI guideline EP12

150. ISO/IEC/IEEE 24765:2017. Systems and software engineering — Vocabulary

151. ISO/IEC 25010:2011 Systems and software engineering - Systems and software Quality Requirements and Evaluation (SQuaRE) - System and software quality models

152. ISO 5649:2024. Medical laboratories — Concepts and specifications for the design, development, implementation and use of laboratory-developed tests

153. Global harmonisation task force (GHTF): GHTF /SG5/N6:2012: Clinical Evidence for IVD medical devices – Key Definitions and Concepts, November 2nd, 2012.

<https://www.imdrf.org/sites/default/files/docs/ghf/final/sg5/technical-docs/ghf-sg5-n6-2012-clinical-evidence-ivd-medical-devices-121102.pdf>

154. MDCG 2022-2 Guidance on general principles of clinical evidence for In Vitro Diagnostic medical devices (IVDs) January 2022. https://health.ec.europa.eu/medical-devices-dialogue-between-interested-parties/medical-device-coordination-group-working-groups_en

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

supervisione degli organismi notificati o dalla standardizzazione alla sorveglianza del mercato, passando per questioni internazionali, nuove tecnologie e indagini cliniche. La sua competenza deriva dalla suddivisione in 13 sottogruppi, che rispettivamente forniscono consulenza e redigono linee guida sul proprio settore di competenza. I membri dei sottogruppi sono nominati dagli Stati membri per un periodo di 3 anni. Le parti interessate e le associazioni con sede in Europa partecipano alle riunioni a seguito di richieste di manifestazione di interesse. Per l'Italia partecipano come delegati del Ministero della Salute Achille Iachino e Antonella Colliardo.¹⁵⁵ Il documento MDCG 2022-2 delinea i principi generali delle evidenze cliniche e fornisce indicazioni sul processo continuo di valutazione delle prestazioni dei dispositivi medico-diagnostici in vitro (IVD), come stabilito dal Regolamento (UE) 2017/746 – Regolamento sui dispositivi medico-diagnostici in vitro (IVDR).¹⁵⁶ Il pubblico di riferimento di MDCG 2022-2 è costituito da produttori di dispositivi medici in vitro, ricercatori e finanziatori di studi clinici. Il documento intende inoltre informare le autorità di regolamentazione, gli organismi notificati e altre parti interessate nella valutazione delle prove cliniche fornite dai fabbricanti. L'articolo 56 di IVDR sottolinea che il fabbricante deve specificare e giustificare il livello di evidenza clinica in considerazione delle caratteristiche del dispositivo e della sua destinazione d'uso. Pertanto, la definizione della destinazione d'uso di un IVD deve essere considerata un fattore chiave alla base della valutazione complessiva. MDCG 2022-2 afferma che lo scopo dovrebbe essere chiaramente espresso nella dichiarazione, riportato nelle istruzioni per l'uso (IFU) e dovrebbe includere gli elementi specifici delineati nell'allegato I, 20.4.1 (c.): (i) cosa viene rilevato e/o misurato; (ii) la sua funzione (ad esempio selezione, monitoraggio, diagnosi o ausilio alla diagnosi, prognosi, previsione, diagnostica di accompagnamento); (iii) le informazioni specifiche che si intendono fornire nel contesto di: — uno stato fisiologico o patologico; — menomazioni fisiche o mentali congenite; — la predisposizione a una condizione medica o a una malattia; — la determinazione della sicurezza e della compatibilità con i potenziali destinatari; — la previsione della risposta o delle reazioni al trattamento; — la definizione o il monitoraggio delle misure terapeutiche; (iv) se è automatizzato o meno; (v) se è qualitativo, o quantitativo; (vi) il tipo di campione/i richiesto/i; (vii) ove applicabile, la popolazione sottoposta all'esame; e (viii) per gli esami diagnostici di accompagnamento, la denominazione comune internazionale (DCI) del medicinale associato per il quale si tratta di un esame di accompagnamento. MDCG 2022-2 definisce “Beneficio clinico” l'impatto positivo di un dispositivo in relazione alla sua funzione, come quella di selezione (screening), monitoraggio, diagnosi o ausilio alla diagnosi dei pazienti, o un impatto positivo sulla gestione dei pazienti o sulla salute pubblica. (UE 2017/746 (IVDR), articolo 2 punto 37 e considerando 64). “Prove cliniche” 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, quando utilizzato come previsto dal fabbricante (UE 2017/746 (IVDR), articolo 2 punto 36). “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

155. Medical Device Coordination Group (MDCG) members– MD & IVD.

https://health.ec.europa.eu/document/download/3913729e-a2a6-44e0-a231-abc06406c49a_en?filename=mdcg_md-ivd_members_en.pdf

156. Document 32017R0746. Legal act. 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. <https://eur-lex.europa.eu/eli/reg/2017/746/oj>

Raccomandazioni M2-2

relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025

Versione 0.1

base alla popolazione target e all'utilizzatore previsto (UE 2017/746 (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. (UE 2017/746 (IVDR), articolo 2 punto 38).

A parte l'uso di terminologia obsoleta e non conforme con gli standard, ma purtroppo fissata nel Regolamento, 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 (UE 2017/746 (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 (UE 2017/746 (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 rilevato da uno specifico metodo di esame.

Il concetto di “Prestazione clinica” è ripreso nel paragrafo 6.6 di MDCG, dove si descrive come le prestazioni cliniche possono essere caratterizzate dalla dimostrazione e dalla valutazione di 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 alle caratteristiche di metodi di confronto per biomarcatori.

Analizzando scientificamente la realtà degli intervalli e dei limiti dei risultati di esami di laboratorio, si deve ammettere che intervalli non coerenti con le differenze sistematiche causano dannose differenze tra i laboratori nel processo decisionale clinico. Intervalli e limiti armonizzati possono essere adottati immediatamente se la differenza osservata è accettabile o eliminata, oppure dopo la standardizzazione/armonizzazione dei misurandi dove manca una completa tracciabilità metrologica.¹⁵⁷ Non solo. Anche l'incertezza di misura compromette le prestazioni diagnostiche del risultato proveniente dal laboratorio.¹⁵⁸ Per fortuna, raramente i medici si basano esclusivamente sui risultati degli esami per prendere una decisione clinica, ma tengono conto dello stato del paziente, della loro esperienza clinica e di altri fattori durante questo processo.

157. Marith van Schroyen Lantman, Miranda van Berkel, Philip Kuijper, Marloes Langelaan, Nannette Brouwer, Marc Thelen. Clinical Decision-Making Suffers from Inequivalent Measurement Results and Inadequate Reference Intervals. *Clinical Chemistry*, Volume 70, Issue 11, November 2024, Pages 1383–1392, <https://doi.org/10.1093/clinchem/hvae129>

158. Schroyen Lantman, M. van. How measurement uncertainty impedes clinical decision-making. 2024, Dissertation Version of the following full text: Publisher's version. <https://hdl.handle.net/2066/311846> DOI: 10.13140/RG.2.2.31007.37285.

Conclusioni e Raccomandazioni

Questo documento di Raccomandazioni non sostituisce la norma originale UNI EN ISO. I vertici ISO chiedevano a ISO 15189 di ridurre i requisiti prescrittivi ma basarsi sul rischio per il paziente, prendere in considerazione altri documenti ISO pubblicati pertinenti, con l'obiettivo di evitare anche ripetizioni ridondanti. Ne risulta la presenza nel testo della norma di alcuni riferimenti ad altri documenti ISO, la cui consultazione diventa ineludibile. La direttiva non è stata però applicata in modo perfetto, alcuni riferimenti mancano e vanno integrati. Ad esempio, mancano sufficienti riferimenti a ISO 22583 per gli strumenti dei servizi POCT e al Vocabolario di metrologia ISO/IEC Guide 99:2007. Vanno inoltre considerati diversi documenti CLSI, per la declinazione applicativa dei principi poco prescrittivi di ISO 15189. Senza i riferimenti ad altri documenti ISO e i documenti CLSI i laboratori possono incontrare difficoltà ad adeguarsi ai requisiti 15189 e gli ispettori di accreditamento possono a loro volta trovarsi in imbarazzo per la verifica delle evidenze.

Queste Raccomandazioni 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.

Si possono quindi esprimere le seguenti Raccomandazioni:

- | | |
|--|--|
| <p>1. Si raccomanda ai laboratori di avvicinarsi ai requisiti di ISO 15189:2022 anche prima di avviare un percorso di accreditamento con un organismo che operi in conformità alla norma ISO/IEC 17011.</p> <p>2. Si può altresì raccomandare di considerare i requisiti di ISO 9001 come utile predisposizione ai requisiti di ISO 15189 in tutti i capitoli, non solo nel capitolo 8 “Sistema di gestione”.</p> <p>3. Per evitare le insidie dei testi normativi in lingua inglese, si raccomanda di utilizzare le versioni tradotte o le raccomandazioni nazionali in lingua italiana (ISO 15189 7.3.6 lettera b).</p> <p>4. Si può altresì raccomandare di considerare i requisiti di ISO 9001 come utile predisposizione ai 26 requisiti di ISO 15189</p> | <p>1. Laboratories are recommended to approach the requirements of ISO 15189:2022 even before starting an accreditation process with a body operating in accordance with ISO/IEC 17011.</p> <p>2. It is also recommended to consider the requirements of ISO 9001 as a useful predisposition to the requirements of ISO 15189 in all chapters, not only in Chapter 8 'Management System'.</p> <p>3. In order to avoid the pitfalls of English-language regulatory texts, it is recommended to use the translated versions or national recommendations in Italian (ISO 15189 7.3.6 letter b).</p> <p>5. It can also be recommended to consider the requirements of ISO 9001 as a useful predisposition to the 26 requirements of ISO 15189 in all chapters, not only in Chapter 8</p> |
|--|--|

Raccomandazioni M2-2

*relazioni con medici e pazienti in ISO 15189 – parte seconda
approvato dal Consiglio Nazionale del 5/10/2025*

Versione 0.1

- in tutti i capitoli, non solo nel capitolo 8 “Sistema di gestione”.
5. anche se intervalli di riferimento e limiti decisionali clinici sono riuniti nel punto ISO 15189 7.3.5, devono restare distinti. Gli intervalli descrivono in modo approssimativo la popolazione servita dal laboratorio, non sono collegabili a sensibilità e specificità, mentre i limiti decisionali sono legati ad almeno una condizione patologica.
 6. per costi e difficoltà (numero di soggetti, criteri di selezione e ripartizione, metodi statistici) gli studi tradizionali (come quelli descritti da CLSI EP44) e i metodi “verifica su larga scala” e “trasferimento mediante confronto”, descritti da CLSI EP45, non sono consigliabili al tipico laboratorio medico ma sono appropriati 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 descritto da CLSI EP45.
 9. ai laboratori medici che eseguono un volume elevato di esami e dispongono di strumenti informatici adeguati, si raccomandano i metodi “indiretti” descritti da CLSI EP45, ovvero di dati ottenuti dal cosiddetto “mondo reale”, come descritto da ISO 18727
 10. al laboratorio medico si raccomanda di accompagnare i risultati degli esami preferenzialmente con limiti decisionali di ISO 15189 7.3.5, proposti da guide riconosciute e autorevoli.
 11. Per utilizzare limiti decisionali di ISO 15189 7.3.5, i laboratori curano l’armonizzazione
- ‘Management System’.
5. although reference intervals and clinical decision limits are brought together in ISO 15189 7.3.5, they should remain distinct. Intervals roughly describe the population served by the laboratory and cannot be linked to sensitivity and specificity, while decision limits are linked to at least one disease condition.
 6. because of cost and difficulty (number of subjects, selection and allocation criteria, statistical methods) traditional studies (such as those described by CLSI EP44) and the “large-scale verification” and “transfer by comparison” methods described by CLSI EP45 are not advisable to the typical medical laboratory but are appropriate for manufacturers and large institutions.
 7. the medical laboratory may acquire proposed intervals from other qualified laboratories or manufacturers and verify them by recognized methods.
 8. laboratories with fewer resources are recommended the small-scale binomial method described by CLSI EP45.
 9. medical laboratories performing a high volume of examinations, equipped with appropriate computer tools, are recommended the “indirect” methods described by CLSI EP45, i.e., of data obtained from the so-called “real world,” as described by ISO 18727
 10. the medical laboratory is recommended to accompany test results preferentially with decision limits of ISO 15189 7.3.5, proposed by recognized and authoritative guides.
 11. To use decision limits of ISO 15189 7.3.5, laboratories take care to harmonize the results with the methods considered in the

Raccomandazioni M2-2 *relazioni con medici e pazienti in ISO 15189 – parte seconda* *approvato dal Consiglio Nazionale del 5/10/2025*

Versione 0.1

-
- dei risultati con i metodi considerati nelle guide.
12. Le stime delle prestazioni cliniche degli esami (ISO 15189 3.31 validazione: rilevanza, sensibilità, specificità) richiede soglie o limiti decisionali definiti (ISO 15189 7.3.5).
13. Per utilizzare limiti decisionali di ISO 15189 7.3.5, i laboratori concordano con i medici curanti di riferimento le prestazioni diagnostiche (sensibilità, specificità, valori predittivi) dell'esame.
- guides.
12. Estimating clinical performance of examinations (ISO 15189 3.31 validation: significance, sensitivity, specificity) requires defined thresholds or decision limits (ISO 15189 7.3.5).
13. To use decision limits of ISO 15189 7.3.5, laboratories agree with key user physicians on the diagnostic performance (sensitivity, specificity, predictive values) of the test.