

Gli intervalli di riferimento nel nuovo millennio

Ferruccio Ceriotti

Coordinatore IFCC Committee on Reference Intervals and Decision Limits (C-RIDL),
Diagnostica e Ricerca San Raffaele S.p.A., Milano

ABSTRACT

Reference intervals in the new millenium. The review presents an update on the reference intervals subject. The theoretical aspects are presented and the most critical issues are discussed. New facts are contributing to enhance the interest on this relatively old topic. Particularly, the requirements of ISO 15189, requesting that "biological reference intervals shall be periodically reviewed", and the in vitro diagnostic (IVD) directive asking the manufactures to provide detailed information on reference intervals, drew renewed attention to it. The traditional approach to the definition of reference intervals, proposed by the IFCC documents remains valid. The use of data mining to obtain reference data from existing databases is discussed and essentially rejected as a mean to develop new reference intervals. New statistical approaches to discard outliers and to compute the reference limits are briefly introduced. The perspectives opened by the improvement in standardization through the implementation of the traceability concept are discussed and a model to define common reference intervals via multicenter experiments is presented and described in detail. The development of common reference intervals, obtained with traceable methods, that can be easily transferred and adopted by different clinical laboratories, can be an effective solution for the improvement of the present situation of proliferation of reference intervals not justified by differences in population characteristics or in analytical methodology.

INTRODUZIONE

Il concetto di "intervalli di riferimento" come li intendiamo oggi è stato sviluppato da Gräsbeck e Saris al termine degli anni '60 del secolo scorso e presentato in un congresso nel 1969 (1). In precedenza si parlava unicamente di "valori normali", ma senza una chiara definizione del significato di questo termine. Potremmo domandarci come mai ne parliamo ancora oggi ad oltre 30 anni dalla pubblicazione del testo di Galen e Gambino "Beyond normality" (2), in un'epoca di sviluppo della medicina basata sulle prove di efficacia che dovrebbe essere caratterizzata dall'uso dei "limiti decisionali". Di questi e della loro differenza dagli intervalli di riferimento parleremo in seguito. Il motivo per cui ancora parliamo di intervalli di riferimento 20 anni dopo la pubblicazione del documento ufficiale che ha sistematizzato in modo definitivo il concetto (3) è che, anche se la teoria è molto chiara, il trasferimento di questa nella pratica di tutti i giorni è ben lungi dall'essere realizzato, se non in una minoranza di centri e per un limitato numero di analiti. L'argomento è stato inoltre riportato alla ribalta da tre recenti avvenimenti: la pubblicazione del documento dell'International Organization for Standardization (ISO) 15189: 2003 (4), l'entrata in vigore della direttiva europea 98/79 CE sui diagnostici in vitro (IVD) (5) e, collegato alla direttiva, lo sviluppo delle attività del Joint Committee for Traceability in Laboratory Medicine (JCTLM) (6). In che modo questi tre elementi, tutti di origine essenzialmente europea, ma con risonanza mondiale, hanno a che fare con gli intervalli di riferimento? Il primo documento, al paragrafo 5.5.5 specificamente recita che "biological reference intervals shall be periodically reviewed" ed in

particolare richiede che siano verificati ogniqualvolta ci sono variazioni di tecnologia analitica o preanalitica. Questo requisito, considerato il numero enorme di tipologie di analisi che anche solo un laboratorio di media entità ha da gestire e la loro variazione quasi quotidiana, appare molto impegnativo da soddisfare. Sul versante dei produttori la direttiva IVD crea un fardello non trascurabile richiedendo, tra i requisiti essenziali (Allegato I punto B, paragrafo 8.7, comma I), di porre nel manuale di istruzione "gli intervalli di riferimento per le quantità da analizzare, compresa una descrizione della popolazione di riferimento da prendere in considerazione". È vero che permette "ove necessario", ma in quali casi gli intervalli di riferimento possono non essere considerati necessari? Il terzo elemento è il requisito della riferibilità, sempre presente nella direttiva IVD, che rappresenta effettivamente la novità reale e che potrebbe aiutare a rendere la definizione e la gestione degli intervalli di riferimento più semplice e, nello stesso tempo, più corretta ed efficace.

In questa rassegna, dopo l'aggiornamento sulla teoria nel corso del quale saranno presi in esame gli aspetti critici relativi agli intervalli di riferimento, verranno analizzate le prospettive che si aprono in conseguenza della migliorata standardizzazione e, per concludere, sarà analizzato il concetto dei limiti decisionali e i possibili sviluppi futuri dell'uso degli intervalli di riferimento nella pratica clinica.

LA TEORIA DEGLI INTERVALLI DI RIFERIMENTO

I valori di riferimento sono nati assieme agli esami di laboratorio; nessun risultato ha infatti un significato "in assoluto", ma solo in quanto riferito ad un particolare

contesto e quindi, in genere, in rapporto ad una situazione fisiologica. Fino agli anni '70, prima dei lavori di Gräsbeck e dell'Expert Panel della IFCC, la dicitura "valori normali" era quella correntemente utilizzata. Una ricerca Medline relativa all'uso del termine "normal values" nel titolo di pubblicazioni relative ad esseri umani nei 5 anni compresi fra marzo 2002 e marzo 2007 fornisce 175 titoli; fortunatamente solo pochissimi fra questi (ad es. il riferimento 7) sono relativi ad esami di laboratorio, segno che le raccomandazioni della IFCC hanno lasciato il segno. Il termine "normale" può avere infatti vari significati. Murphy ne elenca addirittura sette (8). Esso risulta quindi soggettivo e ambiguo; inoltre implica che ciò che cade al di fuori dell'intervallo di normalità sia "anormale", mentre questo, data anche la modalità di calcolo degli intervalli, non è necessariamente vero. Per questi motivi il documento della IFCC (3) propone il termine "valori di riferimento". Questi valori possono essere associati ad una condizione di buona salute oppure a specifiche situazioni fisiopatologiche: fondamentale è la definizione della popolazione dei soggetti che sono stati utilizzati per ottenerli. Da qui deriva la necessità di definire chiaramente i criteri di scelta per gli *individui di riferimento* che rappresentano una popolazione di riferimento da cui è selezionato un *gruppo campione di riferimento*; su questi soggetti si effettuano le misure che permettono di ottenere dei valori di riferimento che avranno una certa distribuzione (*distribuzione di riferimento*). Questa distribuzione va analizzata con metodi statistici in modo tale da ottenere una stima dei *limiti di riferimento*. Per

convenzione questi limiti sono definiti in modo tale da comprendere non il 100% dei valori del gruppo campione di riferimento utilizzato, ma una frazione di essi, in genere pari al 95%. I limiti di riferimento superiore e inferiore definiscono un *intervallo di riferimento* (Figura 1).

Da questa serie di definizioni derivano importanti conseguenze:

1. gli intervalli di riferimento per un determinato analita dipendono dalla variabilità biologica intra- ed interindividuale del campione di soggetti del gruppo di riferimento scelto, quindi la numerosità del campione e l'eventuale suddivisione in sottogruppi può acquistare grande rilevanza;
2. gli aspetti preanalitici devono essere controllati, in modo da potere essere riprodotti sulla popolazione generale a cui gli intervalli ottenuti andranno applicati;
3. gli aspetti analitici sono fondamentali, risultano quasi sempre l'aspetto più critico nel momento in cui gli intervalli ottenuti su un certo gruppo di soggetti devono essere confrontati con quelli ottenuti su soggetti con caratteristiche diverse oppure devono essere applicati in luoghi e tempi diversi da quelli in cui sono stati ottenuti;
4. le modalità di elaborazione dei dati ed il calcolo dei limiti di riferimento possono condizionare in modo rilevante la definizione dei limiti (ad es. se non vengono esclusi dati aberranti o se sono applicati modelli statistici inappropriati).

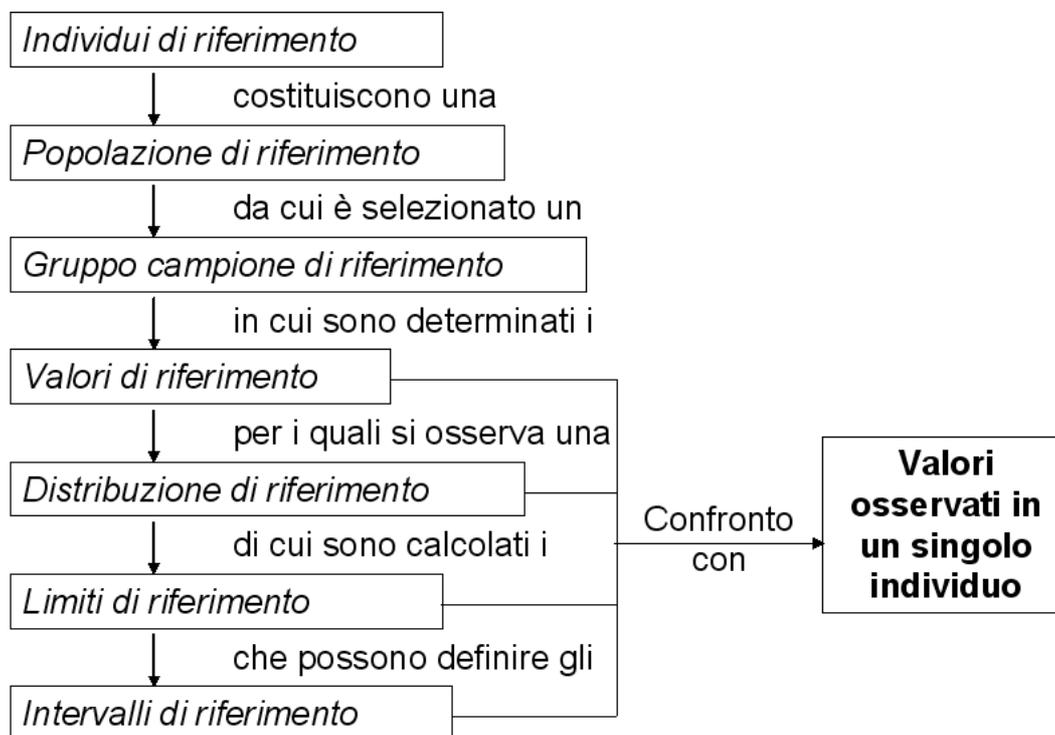


Figura 1

Schematizzazione del concetto degli intervalli di riferimento e rapporti fra i termini raccomandati.

Scelta degli individui di riferimento

Rappresenta il punto di avvio del processo. Normalmente si tratta di selezionare soggetti sani, ma a questo punto il problema che si pone è la definizione di "salute": cos'è esattamente e come la si può riconoscere. La definizione della WHO come "stato di completo benessere fisico, mentale e sociale e non solamente assenza di malattia o di infermità" non costituisce un punto di partenza utilizzabile. Inoltre il concetto di salute può essere differente nelle varie nazioni. Il comitato scandinavo per gli intervalli di riferimento nel 1975 ha provato a stendere una lista delle condizioni patologiche da escludere per considerare un individuo "sano" (9), ma, all'atto pratico, questa raccomandazione si è rilevata pressoché inapplicabile (10), a maggior ragione se ci si riferisce a popolazioni anziane. Infatti, elaborando i dati ottenuti dal terzo "National Health and Nutrition Examination Survey" (NHANES III), Horn e Pesce mostrano come non più del 10% della popolazione con età superiore ai 70 anni cada nella categoria dei "più sani" (11). È quindi necessario accontentarsi di un approccio più pragmatico, considerando la salute in base a sensazioni soggettive ed all'assenza di segni obiettivi di patologie rilevanti ed in particolare relative al componente in esame (12). Come ben chiarito da Solberg (3) è fondamentale definire lo scopo per cui vanno utilizzati gli intervalli in corso di definizione ed in base a questo decidere le modalità di selezione degli individui. Un pratico esempio di questionario è disponibile nel documento del Clinical and Laboratory Standards Institute (CLSI) C28-A2 (13). I principali stati patologici da escludere, indicati dal documento IFCC (14), sono riportati nella Tabella 1. Le modalità per effettuare questa esclusione vanno dalla visita medica, al questionario anamnestico, ad altri esami di laboratorio, a esami strumentali (ecografie, ecc.).

Nel selezionare gli individui è importante tenere presente tutte le variabili che possono influire sulle concentrazioni di un analita (sesso, età, ambiente, abitudini di vita, ecc.) per poter eventualmente definire intervalli di riferimento specifici. Il vantaggio di questa operazione è intuitivo, ma se ne può definire un razionale basato sulla biologia. Come detto in precedenza, infatti, l'ampiezza della distribuzione degli intervalli di riferimento dipende da tre fonti di variabilità: variabilità biologica intra- ed interindividuale dei soggetti di riferimento selezionati e variabilità analitica del metodo analitico utilizzato per la misura. Trascurando per il momento gli aspetti analitici, che, escludendo la misura degli elettroliti, in genere hanno effetti limitati sull'ampiezza degli intervalli di riferimento, focalizziamo l'attenzione sulla variabilità biologica. Ciascun analita è caratterizzato da specifici livelli di variabilità biologica intra- ed interindividuale, che sono legati tra loro in modo inestricabile, ma che contribuiscono in modo sostanziale ad aumentare o ridurre l'utilità di un intervallo di riferimento al fine di valutare il livello di salute di un dato individuo. Già nel 1974 Eugene Harris (15) dimostrò che, solo nel caso in cui la variabilità intraindividuale (CV_i) sia nettamente superiore alla varia-

bilità interindividuale (CV_G), la distribuzione dei risultati ottenibili su un certo individuo coprirà più o meno tutto l'intervallo di riferimento e quindi l'intervallo di riferimento stesso costituirà uno strumento utile per valutare il suo stato di salute. Sfortunatamente questa è una situazione rara. Nel caso, molto più comune, in cui il CV_i è nettamente inferiore al CV_G la distribuzione dei valori relativa ad un dato individuo coprirà solo un ambito limitato della distribuzione dei valori della popolazione e quindi, nella maggior parte dei casi, l'intervallo di riferimento risulterà di limitata utilità per valutare i dati relativi ad un individuo e la sensibilità nel rilevare alterazioni sarà scarsa. Harris dimostrò che solo se il rapporto CV_i/CV_G (indice di individualità) è >1,4 l'intervallo di riferimento è uno strumento utile e sensibile; se invece il rapporto è <0,6 l'utilità dell'intervallo di riferimento è assai relativa. Per migliorare l'utilità dell'intervallo di riferimento non resta che far crescere questo rapporto e, dato che la variabilità intraindividuale non è modificabile, l'unico modo di intervenire è quello di ridurre la variabilità interindividuale stratificando i valori di riferimento in gruppi il più possibile omogenei (16,17).

Scelta "a priori" o "a posteriori"

A questo punto si pone il dilemma se operare la scelta "a priori" o "a posteriori", cioè se decidere in anticipo quali individui scegliere o se selezionare un numero elevato di soggetti e quindi, in base ai risultati ottenuti, decidere se e come suddividere e stratificare i soggetti. In pratica, se le caratteristiche biologiche dell'analita in questione sono ben conosciute, l'approccio "a priori" è nettamente più economico; se invece si tratta di un componente nuovo, su cui si hanno poche informazioni, l'approccio "a posteriori" è quasi obbligatorio. Dato che la seconda situazione descritta è piuttosto rara, il documento della IFCC (14) propende decisamente per l'approccio "a priori". Il documento inoltre non indica alcuna differenza fra le informazioni necessarie per classificare i soggetti con i due approcci e quindi presuppone in entrambi i casi la presenza dei dati anamnestici e clinici, necessari per definire le caratteristiche della popolazione. La disponibilità di questi dati per un gran numero di soggetti appare un requisito difficile da soddisfare nell'approccio "a posteriori". Però lo stesso documento non esclude esplicitamente la possibilità di ottenere intervalli di riferimento attraverso la rielaborazione di dati esistenti, non raccolti specificamente allo scopo di definire gli

Tabella 1

Principali patologie da escludere nella selezione di individui di riferimento

- Insufficienza renale
- Insufficienza cardiaca congestizia
- Malattie respiratorie croniche
- Patologie epatiche
- Sindrome da malassorbimento
- Anemia

intervalli di riferimento, citando, senza commenti, il lavoro di Martin et al. (18).

Valori di riferimento "indiretti"

L'uso di banche dati esistenti, contenenti numeri molto elevati di dati di pazienti, per estrarre intervalli di riferimento è sempre stato un approccio molto invitante in quanto rappresenta una via più rapida e meno costosa per arrivare al risultato, ma non va confuso con l'approccio "a posteriori" a meno che nella banca dati siano presenti anche dati clinici affidabili. I primi lavori di questo tipo risalgono agli anni sessanta (19, 20) e sono sostanzialmente basati sull'assunto che la maggior parte dei risultati prodotti da un laboratorio è "normale" e si può quindi applicare una elaborazione statistica in grado di distinguere la distribuzione dei valori dei soggetti "sani" da quella dei soggetti "non sani". Una vasta letteratura è disponibile sull'argomento (18, 21-27), ma i limiti di questo approccio sono evidenti: 1) a meno che non ci siano dati clinici attendibili (come nel riferimento 24), non rispetta i criteri definiti dalla IFCC (28); 2) i risultati ottenuti dipendono fortemente dal modello di calcolo applicato; se non si conoscono a priori le caratteristiche della popolazione di riferimento non si riesce ad applicare un modello matematico corretto e, se la distribuzione della popolazione è scodata, in genere i modelli falliscono (29); se invece già si conosce la distribuzione della popolazione di riferimento di fatto questa attività non è più necessaria; 3) gli aspetti pre-analitici in genere non sono noti o comunque non sono controllati; 4) è estremamente difficile, se non impossibile, fornire prove della tracciabilità metrologica dei dati ottenuti, in questo modo gli intervalli sono applicabili solo dal centro che li ha definiti, ma difficilmente sono generalizzabili. In conclusione, anche se oggi, con lo sviluppo dell'informatica, appare piuttosto semplice accedere a grandi quantità di informazioni ed eseguire elaborazioni matematiche complesse, a meno che ai modelli matematici non si abbinino anche dati clinici affidabili e che il metodo analitico usato non abbia una dimostrata tracciabilità al metodo di riferimento, questo approccio, per quanto invitante, non è considerato di prima scelta per la definizione degli intervalli di riferimento e potrà al più essere utilizzato come metodo di verifica o di validazione di un intervallo definito in modo convenzionale.

Aspetti preanalitici

Questo punto è semplice da enunciare: i campioni per la produzione degli intervalli di riferimento dovrebbero essere raccolti nelle condizioni più simili possibile a quelle esistenti nella pratica clinica (30). Il problema è che, nella pratica clinica, le condizioni sono poco standardizzate. Quindi è fondamentale, nel momento in cui si raccolgono campioni per produrre intervalli di riferimento, codificare e descrivere in modo accurato le condizioni preanalitiche adottate. In questo modo sarà possibile ad altri mettersi nelle medesime condizioni operative oppure comprendere l'influenza di fattori preanalitici (ad es. la postura) sugli intervalli di riferimento ottenuti. Una lista delle principali condizioni preanalitiche da considerare è riportata nella Tabella 2. Naturalmente, per liquidi diversi dal sangue andranno considerati anche altri aspetti così come analiti particolari potrebbero necessitare di specifiche più dettagliate (ad es. il livello di stress emotivo nel caso del prelievo di ormoni).

Aspetti analitici

Questa rappresenta la parte probabilmente più sottovalutata nella maggior parte dei lavori che trattano di intervalli di riferimento. Il documento della IFCC che affronta specificamente l'argomento (31) dà una serie di raccomandazioni finalizzate essenzialmente a documentare in modo dettagliato le modalità operative e focalizza l'attenzione sulle pratiche di controllo di qualità. Tutto questo è molto utile in caso di intervalli che andranno applicati all'interno dello stesso laboratorio, in quanto consente di definire una situazione di partenza e quindi di comprendere se, al cambiamento di qualche aspetto analitico, è necessario o meno modificare anche gli intervalli di riferimento. Non è invece particolarmente efficace nel dare indicazioni che permettano di definire intervalli di riferimento "trasferibili" con facilità da un laboratorio all'altro. Il motivo fondamentale per questa mancanza di concretezza è la non chiara definizione, all'epoca (1991), del concetto di sistema di riferimento e di riferibilità. Molti metodi di riferimento per gli analiti più comuni erano già disponibili, come anche il concetto della gerarchia dei metodi e dei materiali (32), ma alcuni tasselli pratici per l'implementazione del sistema non erano ancora stati posti. Solo alcuni anni dopo il concetto di sistema di riferimento, basato su metodi di riferimento, materiali di riferimento e laboratori di riferimento è stato ufficializzato (33). Il sistema di riferimento costituisce la

Tabella 2

Principali fattori preanalitici da prendere in considerazione nella produzione di intervalli di riferimento

Preparazione del soggetto	Fattori metodologici	
	Prelievo	Gestione del campione
• Diggiuno	• Ora della giornata	• Trasporto
• Assunzione di farmaci	• Con o senza laccio	• Tempo prima della centrifugazione
• Attività fisica	• Postura	• Modalità e tempi di centrifugazione
	• Tipo di anticoagulante	• Modalità e tempi di conservazione prima dell'analisi
	• Materiali usati per il prelievo	

base per l'accuratezza e la messa a punto di metodi di routine che garantiscano la riferibilità a questo sistema dei risultati ottenuti sui pazienti consente di trasferire questa accuratezza ai laboratori clinici. La riferibilità ("traceability") è definita dal Vocabolario Internazionale di Metrologia (VIM) come "property of a measurement result relating the result to a stated metrological reference through an unbroken chain of calibrations of a measuring system or comparisons, each contributing to the stated measurement uncertainty"¹ (34). Due specifici documenti sono stati sviluppati dall'ISO sull'argomento: EN ISO 17511: 2003 (35) e EN ISO 18153: 2003 (36).

Solo intervalli di riferimento ottenuti con modalità analitiche che consentano di dimostrare la riferibilità dei risultati ottenuti al corrispondente sistema di riferimento possono essere facilmente trasferiti tra laboratori diversi. Il raggiungimento di un elevato livello di standardizzazione attraverso l'uso attento di metodi riferibili permetterà lo sviluppo di intervalli di riferimento comuni (vedi oltre).

Modalità di calcolo dei limiti di riferimento

Questo è l'aspetto che ha suscitato le maggiori attenzioni ed ha stimolato lo sviluppo di abbondante letteratura. Fondamentalmente possiamo evidenziare tre tipi di problematiche: 1) quali siano gli strumenti statistici più corretti per estrapolare all'intera popolazione i risultati ottenuti su un campione della stessa; 2) come decidere se assegnare limiti diversi a gruppi di soggetti differenti (per età, sesso o altro); 3) come riconoscere (e quindi essere autorizzati ad eliminare) dati aberranti. Per chi fosse interessato ad approfondire l'argomento ci sono almeno due testi interessanti: *Statistical bases of reference values in laboratory medicine* di Harris e Boyd (37) ed il recente *Reference intervals. A user's guide* di Horn e Pesce (38). Qui di seguito sono riassunti i concetti principali.

Metodi statistici

Il testo di Harris e Boyd (37) attribuisce a Wootton et al. nel 1951 (39) la prima applicazione della statistica gaussiana alla definizione degli intervalli di riferimento. Ben presto gli stessi autori si resero conto che in realtà questo modello statistico era applicabile solo in una minoranza di casi e già nel 1953 proposero la trasformazione logaritmica dei dati per ottenere una distribuzione gaussiana (40). La pratica scorretta di definire gli intervalli di riferimento come $\text{media} \pm 2 \text{ DS}$, senza una verifica preliminare della distribuzione dei dati ottenuti, è però proseguita per anni. Numerosi lavori sulle modalità di calcolo dei quantili per la definizione degli intervalli di riferimento furono prodotti negli anni '70-'80 (41-44), ma la pietra miliare sull'argomento è costituita dal lavoro della IFCC pubblicato nel 1987 (45) che lucidamente definisce una serie di punti che oggi, a distanza di 20 anni conservano a pieno la loro validità.

- a) Ufficializza la scelta (arbitraria) che l'intervallo di riferimento debba essere rappresentato dalla frazione 0.95 centrale della distribuzione di riferimento (ossia contenente il 95% dei dati). A parte alcune voci contrarie, che proporrebbero di ampliare questo intervallo al 99,8% al fine di ridurre le "false positività" quando sono richiesti un gran numero di tipi di esami (46), questo dato è oggi comunemente accettato.
- b) I limiti di riferimento ottenuti dovrebbero sempre essere accompagnati dall'intervallo di confidenza al 90%. Questo intervallo di confidenza è tanto più ristretto quanto maggiore è il numero dei soggetti utilizzati per la definizione degli intervalli.
- c) Raccomanda l'uso di metodi non parametrici per la stima dei limiti di riferimento. Anche se riconosce che i sistemi di stima dei limiti di riferimento basati su metodi parametrici hanno teoricamente una maggiore affidabilità, soprattutto se i campioni sono piccoli, l'incertezza sulla reale "gaussianità" della distribuzione (di partenza o dopo trasformazione) aggiunge incertezza alla stima.
- d) Propone un metodo di calcolo dei limiti di riferimento semplice ed efficace. Il metodo è basato sulla stima dei quantili 0.025 e 0.975. Il quantile α non è calcolabile a meno che il suo valore non sia superiore a $1/N$, dove N è il numero dei soggetti (dimensione del campione). Quindi il numero minimo di soggetti necessari per applicare questo metodo è 40 ($\alpha = 1/40 = 0.025$). Però, se si utilizza questo approccio, con solo 40 soggetti il valore minimo e massimo ottenuti rappresentano i limiti inferiore e superiore dell'intervallo e non è possibile stimarne l'incertezza. Per poter calcolare l'incertezza attorno ai limiti di riferimento è necessario che i soggetti siano almeno 120 (in questo caso, ordinando i valori in modo crescente, dal più basso al più alto, il centile 2,5 è costituito dal terzo valore della serie, mentre il 97,5° è rappresentato dal 118° valore e gli intervalli di confidenza al 90% del limite inferiore sono il 1° ed il 7° valore della serie, mentre quelli del limite superiore sono il 114° ed il 120°). Da qui nasce la raccomandazione sul numero minimo di individui necessari per ciascuna classe. Per calcolare l'incertezza attorno ai limiti di riferimento si può utilizzare la tabella riportata nel lavoro (45). Solberg ha in seguito anche sviluppato un software per implementare queste raccomandazioni (47,48).

Come si vede l'approccio proposto dalla IFCC appare semplice da applicare, ma il limite del numero di soggetti da arruolare costituisce una difficoltà rilevante, soprattutto nei casi in cui si ha a che fare con bambini, con analisi a costo elevato o con analiti che hanno intervalli di riferimento che variano con l'età, come ad es. la creatinina, la fosfatasi alcalina o gli ormoni sessuali. Questo limite sarebbe facilmente superabile con l'uso di

¹Proprietà del risultato di una misura che mette il risultato stesso in relazione con un riferimento metrologico definito per mezzo di una catena ininterrotta di calibrazioni o confronti, ciascuno dei quali contribuisce all'incertezza.

statistiche parametriche, ma per poterle applicare è necessario trasformare i dati in modo da fare assumere loro una distribuzione gaussiana. Ottenere questa trasformazione in modo efficiente è difficile perché la forma della distribuzione dei valori non è conoscibile a priori ed è diversa da analita ad analita e, per lo stesso analita, in gruppi diversi di individui. Per superare il problema Horn et al. (49,50) hanno proposto un metodo definito "robusto", basato su una trasformazione dei dati originali secondo l'algoritmo proposto da Box e Cox (51), seguita da una elaborazione complessa basata sull'uso di indicatori robusti, cioè in grado di fornire risposte corrette anche in situazioni che si discostano dalle condizioni ideali. L'uso di questi complessi algoritmi dovrebbe permettere di ottenere stime accurate degli intervalli di riferimento anche con campioni costituiti da un numero limitato di soggetti (anche solo 20). Per calcolare l'intervallo di confidenza dei limiti calcolati con questo metodo è possibile applicare tecniche di "bootstrap" in cui le osservazioni sono ricampionate, con sostituzione, dai dati originali creando uno "pseudo-campione" da cui derivare gli intervalli. Questo processo è ripetuto un gran numero di volte (1000 – 2000) creando una distribuzione di limiti di riferimento; prendendo il 5° ed il 95° centile di questa distribuzione è possibile definire l'intervallo di confidenza al 90%. Un ulteriore approccio per superare il problema dei piccoli numeri, nel caso di intervalli di riferimento dipendenti dall'età, è quello di utilizzare l'analisi della regressione per calcolare i limiti, senza la necessità di suddividere i soggetti in classi di età (52,53).

Criteri di partizione

Come già ampiamente discusso nella sezione relativa alla scelta degli individui di riferimento, la decisione se separare o meno i vari gruppi di soggetti ha una rilevanza notevole. Inoltre, in base a quanto discusso al punto precedente, suddividere in gruppi richiede un gran numero di individui oppure metodi statistici sofisticati. Gli approcci statistici per definire se i dati appartenenti a due gruppi di soggetti sono significativamente differenti o possono essere combinati sono quindi di fondamentale importanza. Un approccio intuitivo è quello basato sulle medie delle sottoclassi: se la loro differenza è statisticamente significativa ciascuna sottoclasse merita un intervallo specifico. Però questo approccio, soprattutto se il numero di soggetti in ciascuna sottoclasse è elevato, può portare alla definizione di classi diverse anche per differenze piccole, non rilevanti dal punto di vista clinico. Sinton et al. (54) suggeriscono di non separare due classi a meno che la differenza fra le medie rappresenti almeno il 25% dell'intervallo calcolato sul 95% degli individui dell'intera popolazione. Questo criterio, proposto per specifici analiti (calcio, fosfato e fosfatasi alcalina), quando applicato ad altri si rivela troppo conservativo e non consente separazioni adeguate (55). Il metodo di partizione più noto è quello proposto da Harris e Boyd (37,56) e adottato dal documento C28-A2 del CLSI (13). Gli autori partono dall'idea che creare dei sottogruppi dovrebbe portare ad una riduzione della variabilità inter-individuale all'interno del sottogruppo rispetto a quella

del gruppo generale, ma in realtà, anche se si ottiene una riduzione, questa in genere non risulta statisticamente significativa, anche a fronte di importanti differenze fra le medie. Per questo motivo gli autori si focalizzano invece sulla frazione di individui nel sottogruppo che risulterebbero al di fuori dei limiti degli intervalli di riferimento comuni. La domanda fondamentale che si fanno è la seguente: "una singola coppia di limiti, derivante dalla combinazione delle varie sottopopolazioni, è in grado di soddisfare in modo sufficientemente buono il criterio di escludere il 2,5% dei soggetti all'estremo superiore ed il 2,5% all'estremo inferiore di ogni singola sottopopolazione?". A questo punto il problema che si pone è come definire la soglia di accettabilità per questo scostamento ed Harris e Boyd indicano che se la percentuale al di fuori di uno dei due limiti supera il 4% è il caso di definire limiti separati. Il criterio appare valido in quanto tiene conto non solo della media, ma anche della DS della sottoclasse che può modificare i limiti anche a parità di media. Il test quindi si compone di 2 parti, una che valuta la significatività della differenza fra le medie ed una che confronta le DS (per maggiori dettagli fare riferimento a (13) dove il procedimento è spiegato in modo dettagliato e con esempi pratici di calcolo). Questo modello presenta però una serie di limiti, che si possono sostanzialmente riassumere nel fatto che esso si applica bene quando le popolazioni hanno distribuzioni gaussiane (o facilmente trasformabili), sono di uguale numerosità ed hanno DS simile; in tutti gli altri casi può dare problemi (55). Per superarli Lahti propone un metodo basato sui concetti che hanno ispirato Harris e Boyd, ma che permette di valutare specificamente la percentuale dei soggetti di una sottoclasse che cade al di fuori degli intervalli comuni (57-59). L'autore, sulla base dei criteri utilizzati per accettare gli effetti della variabilità analitica esposti da Gowans et al. (60) propone come limite per la suddivisione dei sottogruppi la presenza di almeno 4,1% dei soggetti del sottogruppo al di fuori di uno dei limiti del gruppo comune. Se la percentuale è <3,2% suggerisce di unificare i gruppi, per le situazioni intermedie, che chiama "marginali", indica di basare la scelta se suddividere o meno su considerazioni di tipo clinico.

Esclusione dei dati aberranti

Qualsiasi sia il metodo di calcolo degli intervalli, la presenza di dati aberranti può indurre alterazioni significative, anche in quelli basati su metodi statistici e matematici complessi (49,50). Un primo metodo per individuare i dati aberranti è quello della osservazione della distribuzione dei dati; il problema che si pone poi, se non ci sono motivi di esclusione legati alle condizioni del soggetto o a evidenti errori analitici, di calcolo o di trascrizione, è quello di identificare ragioni di tipo statistico. Il metodo statistico più popolare è quello proposto da Dixon (61) basato sul rapporto D/R, dove D è il valore assoluto della differenza fra l'aberrante ed il primo valore successivo e R è il valore relativo all'intero ambito delle osservazioni (massimo – minimo), aberrante incluso. In accordo con Reed et al. (41), il documento del CLSI propone 1/3 come limite per il valore di questo rap-

porto (13). Il test però non risulta molto sensibile, soprattutto se gli aberranti sono più di uno e quindi la presenza di un aberrante un po' meno estremo maschera la presenza dell'altro (o degli altri). In questo caso il suggerimento è di esaminare per primo l'aberrante meno estremo; se il test lo identifica come tale, allora è possibile eliminare anche gli altri più estremi.

Horn et al. propongono un algoritmo più sofisticato, a 2 stadi (62). La prima fase prevede la trasformazione, attraverso il metodo di Box e Cox (51), dei dati del campione per renderli di tipo gaussiano, la seconda fase prevede l'identificazione degli eventuali aberranti attraverso l'approccio robusto proposto da Tukey (63). Questo metodo prevede l'identificazione dei valori estremi utilizzando il 50% centrale del campione, eliminando in questo modo il possibile effetto confondente dato dalla presenza di più aberranti. Si calcolano il 25° ed il 75° percentile (Q_1 e Q_3) e l'intervallo interquartile (IQR) ($Q_3 - Q_1$). I limiti di accettabilità dei dati sono quindi calcolati come segue: limite inferiore = $Q_1 - 1,5 \times IQR$, limite superiore = $Q_3 + 1,5 \times IQR$. Qualsiasi dato che cada al di fuori dei limiti è considerato un aberrante.

SITUAZIONE ATTUALE E SVILUPPI FUTURI

Nel lontano 1960, nelle conclusioni di un lavoro intitolato "Some thoughts on normal, or standard, values in clinical medicine" (64), Schneider scriveva: "...practical medicine is basically founded on comparison. If medicine is to be scientific, we must not only understand the

structural, functional and chemical relations operating in individuals, but we must also understand the bases of our comparisons"². Da quanto detto fino ad ora potrebbe sembrare che la prospettiva disegnata da Schneider nel suo lucidissimo lavoro sia stata raggiunta. In realtà, se gli aspetti teorici sono ben delineati, non altrettanto si può affermare relativamente alla loro traduzione nella pratica quotidiana, che appare molto lontana dall'ottimale.

Infatti, i laboratori utilizzano spesso intervalli di riferimento tra loro diversi, ma senza un razionale dovuto alla metodologia in uso o alla popolazione di riferimento. Un esempio eclatante è riportato nella Figura 2, che mostra gli intervalli di riferimento in uso nel 2005 in un gruppo di 93 laboratori italiani, tutti utilizzando lo stesso metodo (IFCC con piridossal fosfato) per la misura dell'attività catalitica della alanina aminotransferasi (ALT) (anche se con diverse apparecchiature). Come si può vedere, almeno la metà dei laboratori non fornisce intervalli di riferimento distinti per maschi e femmine e non fornisce un limite inferiore di riferimento. Il limite superiore di riferimento va da 40 U/L (supponendo che i pochi laboratori che hanno riportato valori inferiori a 40 U/L abbiano commesso un errore) a 72 U/L, sia per i maschi che per le femmine. C'è da sottolineare inoltre che, nonostante alcuni produttori come Dade-Behring e Vitros suggeriscano intervalli di riferimento decisamente superiori al resto del gruppo, questi non sono applicati da tutti gli utilizzatori di queste tecnologie analitiche, ma solo da una percentuale, che nel caso del Vitros è inferiore al 50%. Situazioni di questo tipo sono comuni per le attività enzi-

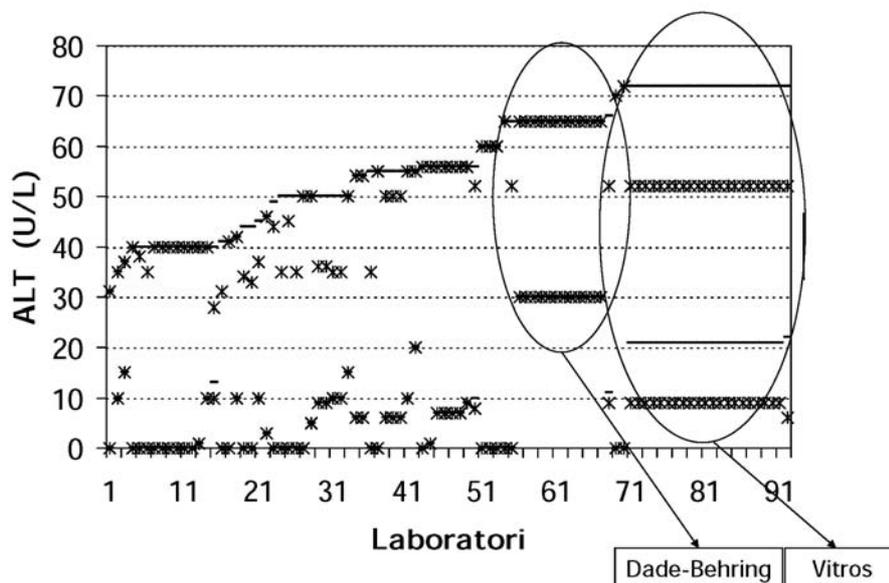


Figura 2

Intervalli di riferimento per la alanina aminotransferasi (ALT) in 93 laboratori italiani partecipanti al programma di VEQ PROLA-RIT. Tutti i laboratori adottano un metodo con piridossal fosfato che segue le raccomandazioni IFCC, applicato su varie piattaforme analitiche. I limiti superiore ed inferiore per il sesso maschile sono indicati da ----, quelli per il sesso femminile da ***.

²La medicina pratica è basata fondamentalmente sul confronto. Se la medicina deve essere scientifica non ci è sufficiente comprendere le relazioni strutturali, funzionali e chimiche che avvengono all'interno degli individui, ma dobbiamo anche comprendere le basi dei nostri confronti".

matiche, ma non sono rare anche per altri analiti. Questo genera una situazione confusa, sia per il clinico che per il paziente. Comporta la difficoltà di confrontare risultati ottenuti in diversi laboratori (lo stesso valore può essere considerato fisiologico o patologico a secondo del laboratorio in cui è stato ottenuto) e l'impossibilità di combinare risultati ottenuti da diversi laboratori in una banca dati comune. Queste differenze fra laboratori non sembrano essere giustificate né dalla diversa metodologia analitica (vedi Figura 2), né tantomeno da reali differenze di popolazione, ma sono soprattutto dovute all'adozione di intervalli proposti dal produttore o tratti dalla letteratura. Infatti, una effettiva attività di definizione degli intervalli di riferimento fatta sulla popolazione afferente al laboratorio risulta impegnativa, sia come costi che come tempi, anche solo per un singolo analita e diventa proibitiva se allargata alle centinaia di analiti misurati e alle continue variazioni imposte dalla tecnologia.

Una possibile prospettiva di superamento di questa situazione piuttosto confusa è quella dello sviluppo di intervalli di riferimento "comuni".

Intervalli di riferimento comuni

Il concetto è semplice: se il metodo analitico è lo stesso o fornisce risultati identici perché è correttamente standardizzato e la popolazione ha le stesse caratteristiche, o è noto che un certo analita non è influenzato da razza o ambiente, perché non adottare intervalli di riferimento comuni? Sfortunatamente l'applicazione pratica di questo semplice concetto non è così facile da realizzare. Per essere applicato è necessario che si realizzino contemporaneamente una serie di condizioni indicate nella

Tabella 3.

Produzione degli intervalli di riferimento

L'importanza degli aspetti analitici ed il concetto di sistema di riferimento sono stati trattati in precedenza. Ammesso quindi che un sistema di riferimento esista, il problema è quello di avere dati attendibili che possano essere adottati come intervalli comuni. In particolare, è importante avere indicazioni sull'esistenza di reali effetti di componenti etniche o ambientali che possano rendere più difficile oppure addirittura impedire l'adozione di un intervallo comune, a meno di non allargare i limiti, riducendone così l'utilità.

Il modo migliore per ottenere queste informazioni è quello di organizzare un esperimento multicentrico che coinvolga vari laboratori. Questo approccio è stato presentato inizialmente da un gruppo spagnolo (65-69) ed è stato sviluppato al meglio in un progetto dei paesi nordici (70-74)

Uno studio multicentrico, per poter produrre risultati adottabili da tutti i laboratori in grado di operare con condizioni analitiche e preanalitiche simili, deve essere organizzato con cura e richiede:

- Selezione "a priori" dei soggetti secondo quanto riportato nello specifico paragrafo. Il numero dei centri partecipanti ed il numero di soggetti per centro deve essere rapportato al numero complessivo di soggetti necessari per permettere la possibile suddivisione per sesso, età, razza, abitudini di vita, ecc. Per ottenere intervalli di confidenza ristretti attorno ai limiti degli intervalli il numero ottimale di individui in ciascun gruppo dovrebbe essere di circa 500 (72) o

Tabella 3

Requisiti per l'uso di intervalli di riferimento comuni

Categoria del requisito	Requisito	Responsabilità
Analitico	Esistenza di un sistema di riferimento base per l'accuratezza	IFCC, JCTLM, Istituti metrologici nazionali e internazionali
	Esistenza di metodi commerciali riferibili al sistema di riferimento	Produttori
	Utilizzo corretto del metodo nei laboratori clinici	Laboratori clinici
	Controllo delle prestazioni dei metodi in modo da mantenerle entro limiti definiti di imprecisione e bias	Laboratori clinici, Organizzatori di VEQ
Clinico	Definizione accurata degli intervalli di riferimento con informazioni relative all'influenza di fattori biologici e ambientali.	Collaborazione fra IFCC, Produttori e Laboratori clinici
	Fase preanalitica compatibile con le condizioni utilizzate per la definizione degli intervalli di riferimento	Laboratori clinici
	Validazione dell'applicabilità degli intervalli comuni alla popolazione afferente al laboratorio	Laboratori clinici
	Adozione degli intervalli comuni	Laboratori clinici

JCTLM, Joint Committee on Traceability in Laboratory Medicine.

al minimo di 120, a meno di non utilizzare tecniche statistiche particolari (tipo bootstrap, vedi sopra). I criteri per la partizione fra sottogruppi dovrebbero essere quelli indicati da Lahti (57-59).

- Dettagliata e accurata definizione della fase preanalitica. Teoricamente le analisi andrebbero eseguite su campioni freschi, in analogia con l'operatività usuale dei laboratori, ma, per ridurre la variabilità analitica, è accettabile il congelamento dei campioni, a patto che sia dimostrato che il congelamento non comporti problemi. È altamente consigliata la conservazione di almeno un'aliquota dei ciascun campione per usi futuri.
- Estrema attenzione agli aspetti analitici in modo da garantire la riferibilità dei dati al sistema di riferimento e la confrontabilità dei laboratori partecipanti. La standardizzazione e la riferibilità dovrebbero essere garantite attraverso l'uso di almeno 2 materiali commutabili (sieri congelati) con valori assegnati mediante metodi di riferimento, meglio se da una rete di laboratori di riferimento. La distribuzione di questi materiali ai laboratori partecipanti ne garantisce la standardizzazione anche se non tutti utilizzano la stessa tecnologia, a condizione che i metodi in uso abbiano specificità comparabile a quella del metodo di riferimento. Un esempio tipico è la necessità del piridossal fosfato nei metodi per la misura delle transaminasi. La confrontabilità tra laboratori deve anche essere controllata attraverso un programma di CQI comune con chiari criteri per l'accettazione o il rifiuto delle serie analitiche.
- Una analisi dei dati accurata. I risultati dei diversi centri devono essere confrontati per escludere la presenza di bias analitici (attraverso i dati del controllo di qualità) o di atipica distribuzione dei dati della popolazione. Queste differenze dalla popolazione generale possono essere verificate con test statistici (per es. Mann-Whitney o Kruskal-Wallis) (69,71). Dopo un controllo per la presenza di aberranti il calcolo degli intervalli potrà essere effettuato con uno dei metodi descritti in precedenza.

Uso degli intervalli comuni

Per poter adottare gli intervalli di riferimento comuni un laboratorio clinico deve verificare tre aspetti: le condizioni preanalitiche, il metodo in uso e le sue prestazioni e le caratteristiche della popolazione che a lui affinisce.

- Aspetti preanalitici. Gli intervalli comuni possono essere utilizzati solo se si adottano condizioni preanalitiche identiche a quelle impiegate nell'esperimento che ha portato alla loro definizione o se è possibile dimostrare che le differenze introdotte non hanno effetti (ad es. equivalenza tra siero e plasma eparinato).
- Aspetti analitici. Il metodo in uso deve essere riferibile al sistema di riferimento. Con l'attuazione della direttiva europea 98/79/CE sui diagnostici in vitro (5), questa dovrebbe essere la regola per tutti gli analiti per cui un sistema di riferimento è stato effettivamente attivato, ma recenti esperienze nel campo della

misura di attività enzimatiche hanno dimostrato che questo oggi non è sempre vero (75). Fondamentale è inoltre che il metodo sia implementato nel modo corretto e che abbia le necessarie caratteristiche di specificità (ad es. l'uso di un diverso substrato per l'alfa-amilasi o la mancanza di piridossal fosfato nella misura delle transaminasi sono alla base di diversa specificità analitica). La qualità analitica del metodo deve essere infine adeguatamente controllata, sia in termini di imprecisione che di bias. I traguardi per la massima imprecisione possono essere basati sul principio della variabilità biologica (76) ed, in particolare, i requisiti del massimo bias accettabile per poter adottare un intervallo di riferimento comune dovrebbero essere quelli proposti da Gowans et al. (77). Il bias dovrebbe essere $<0,25 (CV_I^2 + CV_G^2)^{1/2}$ dove CV_I e CV_G sono la variabilità intra- e interindividuale e la loro somma costituisce quindi la variabilità della popolazione. Questo criterio è equivalente agli intervalli di confidenza per un campione di 120 soggetti. Queste specifiche sono state in seguito graduate a 0,125, 0,25 e 0,375 volte la variabilità della popolazione per ottenere un criterio ottimale, desiderabile o minimo (78). Una lista aggiornata delle specifiche di qualità è presente sul sito <http://www.westgard.com/bio-database1.htm>. La presenza e le dimensioni del bias possono essere valutate dai programmi di VEQ e dai risultati dei controlli interni allargati.

- Caratteristiche della popolazione. Si possono verificare due condizioni in base agli analiti:
 - a) se è noto che razza o abitudini di vita non hanno effetti sugli intervalli di riferimento è sufficiente la verifica degli aspetti analitici e preanalitici;
 - b) se l'appartenenza a gruppi etnici o le abitudini di vita sicuramente influenzano gli intervalli di riferimento o se mancano informazioni attendibili su questo aspetto è utile che il laboratorio validi l'intervallo su un piccolo gruppo di soggetti di riferimento presi dalla sua popolazione. Questa validazione può essere attuata seguendo le indicazioni del documento CLSI C28-A2 al paragrafo 8.2 (13). In breve, il documento suggerisce di esaminare 20 soggetti di riferimento e confrontare i risultati ottenuti con l'intervallo di riferimento da adottare. Questi 20 individui dovrebbero rappresentare la popolazione sana che affinisce al laboratorio ed essere scelti secondo i criteri di selezione definiti. Dopo aver scartato eventuali aberranti, se non più di 2 dati su 20 cadono al di fuori dell'intervallo questo può essere adottato. Se tre o più dati cadono al di fuori dell'intervallo l'esperimento dovrebbe essere ripetuto con altri 20 soggetti. Se al secondo tentativo la regola del 2 su 20 è rispettata l'intervallo si può adottare, altrimenti è probabile (ammesso che non ci siano bias di tipo analitico o preanalitico) che la popolazione sia differente e quindi è necessario definire un intervallo di riferimento specifico. Questo tipo di test, definito binomiale, funziona abbastanza bene se i valori hanno distribuzione di tipo gaus-

siano, se invece la distribuzione è molto scodata il test è poco sensibile (J. Boyd, comunicazione personale). In questi casi, se tutti i dati relativi all'intervallo di riferimento comune sono disponibili, è possibile applicare test statistici come quello di Kolmogorov-Smirnov. Un altro possibile approccio è quello di elaborare i dati relativi ai 20 soggetti con un metodo robusto tipo quello proposto da Horn et al. (49,50) e confrontare i limiti ottenuti con quelli dell'intervallo di riferimento comune: se i valori ottenuti sono entro l'intervallo di confidenza per quelli comuni, l'intervallo comune può essere adottato. Infine si possono elaborare, con un metodo fra quelli descritti in precedenza (19, 20, 23), i valori ottenuti sulla popolazione in esame e confrontare l'intervallo ottenuto con quello comune: se non più del 3,2% della popolazione cade al di fuori di uno dei due limiti, l'intervallo comune può essere adottato.

Tutto questo processo però non è né semplice né rapido. La realizzazione del sistema di riferimento è un processo lungo e la definizione di intervalli di riferimento comuni è una operazione difficile e costosa. Ci si potrebbe chiedere se valga veramente la pena di cimentarsi in un'impresa così lunga, complessa e costosa. La risposta è che, se ben definiti, gli intervalli di riferimento possono davvero costituire un pilastro del sistema di riferimento e non richiedere ulteriori continui controlli da parte di migliaia di laboratori sparsi per il mondo.

Adozione di intervalli di riferimento "validati"

Non solamente intervalli di riferimento ottenuti con esperimenti multicentrici possono essere adottati come intervalli di riferimento comuni, ma anche intervalli ottenuti in un singolo centro, purché seguendo tutte le indicazioni date in precedenza, sia sotto l'aspetto analitico che organizzativo. Il Comitato della IFCC sugli Intervalli di Riferimento e Limiti Decisionali (C-RIDL) ha in corso un lavoro di validazione di intervalli di riferimento per la creatinina plasmatica. Nel caso dell'adozione di intervalli sviluppati in un singolo centro acquista naturalmente un maggior rilievo la necessità di verifica preliminare dell'intervallo stesso sulla popolazione del proprio laboratorio.

Tabella 4

Principali differenze fra intervalli di riferimento e limiti decisionali

	Intervalli di riferimento	Limiti decisionali
Dipendenza	<ul style="list-style-type: none"> • Popolazione • Fascia di età • Sesso 	<ul style="list-style-type: none"> • Problema clinico • Categoria di pazienti
Informazioni ottenute	Appartenenza o meno alla popolazione di riferimento	Eleggibilità del paziente per un determinato intervento
Statistica	95% centrale della distribuzione	Nessuna
Numero	Coppie di dati	Dato singolo oppure più dati in base al livello di rischio o alla probabilità di una data classificazione o al quesito clinico

Intervalli di riferimento e limiti decisionali

I due concetti sono simili per alcuni aspetti, ma per altri sono completamente differenti e non vanno confusi. Nella Tabella 4 sono schematizzate le differenze fondamentali. La confusione deriva dal fatto che, in alcune situazioni, i limiti di riferimento possono essere utilizzati come limiti decisionali (ad es. nell'ambito di un'attività di screening si decide che tutti gli individui che hanno risultati al di fuori del limite superiore o inferiore di riferimento siano sottoposti ad un esame strumentale di conferma o a qualche tipo di profilassi preventiva), ma concettualmente le due cose sono completamente differenti. Infatti, mentre i primi, come abbiamo visto, descrivono le caratteristiche biologiche di una popolazione e da esse dipendono, i secondi dipendono da scelte o quesiti clinici e quindi, invece di essere calcolati con metodi statistici, sono definiti in genere attraverso un processo di consenso, che non richiede in genere statistiche particolari. I limiti decisionali identificano una soglia di rischio, come nel caso dei valori di lipidi o di emoglobina glicata, oppure una soglia di classificazione per una data patologia. Mentre nel primo caso i limiti sono definiti in modo arbitrario, in base al livello di rischio ad essi collegato ottenuto nel corso di studi clinici, nel secondo caso dovrebbero essere definiti, sempre in modo arbitrario, ma in base al confronto fra i valori ottenuti sulla popolazione sana (i valori di riferimento appunto) e quelli ottenuti sulla popolazione affetta da una specifica patologia. In base alla sovrapposizione fra le due popolazioni ed alla percentuale di falsi negativi e di falsi positivi (sensibilità e specificità cliniche) che si decide di accettare, si scelgono i limiti decisionali. Come si può vedere quindi, anche se è auspicabile la definizione di limiti decisionali per la maggior parte delle patologie, per definire questi non si può prescindere dall'identificazione degli intervalli di riferimento. Il secondo passaggio poi, con la popolazione dei malati, avrà le stesse caratteristiche di quanto descritto per i sani, ma con tutti i problemi aggiuntivi legati all'accurata definizione della presenza di una data patologia.

Intervalli di riferimento individuali

Rappresentano la sfida per il futuro, anche se Harris (79), già 30 anni fa, teorizzava questa possibilità e presentava possibili approcci statistici per una sua imple-

mentazione. Oggi lo sviluppo dell'informatica fornisce la possibilità sia di archiviare dati individuali per un lungo periodo di tempo sia di eseguire rapidamente elaborazioni complesse ed il miglioramento della standardizzazione delle misure consente di ridurre la variabilità analitica sia nel tempo che tra laboratori. Questi presupposti tecnici possono far sì che quella che era una teoria affascinante diventi una prospettiva concreta. Il modello sperimentale prevede la raccolta di più campioni dallo stesso individuo in un periodo di salute stabile. I risultati di queste misure costituiscono una serie temporale, un dato pregresso da utilizzare per valutare i risultati futuri. Il problema fondamentale è costituito ovviamente dal numero di campioni necessari per stabilire una "linea di base" valida e questo dipende sia dalle caratteristiche della variabilità biologica di un dato analita, sia dai modelli matematici applicati. Una trattazione approfondita dell'argomento è fatta da Queraltó (80).

CONCLUSIONI

Anche se sull'argomento è stato scritto e prodotto molto, appare chiaro che c'è ancora molto da fare per raggiungere una situazione ottimale. Il miglioramento della standardizzazione analitica e lo sviluppo di progetti multicentrici per la definizione degli intervalli di riferimento è sperabile che possa consentire, in un futuro non troppo lontano, di ridurre drasticamente l'eterogeneità degli intervalli di riferimento, fornendo al clinico informazioni più congruenti, accurate ed efficaci ed ottenendo quindi quanto auspicava Schneider (64) quasi 50 anni fa.

L'intervento di modifica degli intervalli di riferimento è comunque un'operazione sempre delicata, che richiede contatti con i clinici ed i pazienti, tempo, impegno e adeguata cultura. È necessario che i laboratori clinici si impegnino sull'argomento superando le difficoltà e l'inerzia che spesso li hanno caratterizzati. Alternativamente, per quanti sforzi si facciano per far progredire la teoria, i progressi reali ed i conseguenti benefici per i pazienti saranno insignificanti.

RINGRAZIAMENTI

Al Prof. Carlo Franzini a cui devo l'ispirazione per molte delle idee espresse nell'articolo e in particolare la Tabella 4 ed al Prof. Mauro Panteghini per l'aiuto ed il supporto ricevuti.

BIBLIOGRAFIA

- Gräsbeck R, Saris NE. Establishment and use of normal values. *Scand J Clin Lab Invest* 1969;26(Suppl 110):62-3.
- Galen RS, Gambino SR. *Beyond normality: the predictive value and efficiency of medical diagnoses*. New York: John Wiley and Sons, 1975.
- Solberg HE. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values. Approved recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem* 1987;25:337-42; *Clin Chim Acta* 1987;165:111-8; *Labmedica* 1987;4:27-31; *Ann Biol Clin* 1987;45:237-41.
- International Organization for Standardization. *Medical laboratories — Particular requirements for quality and competence*. ISO 15189. Geneva: ISO, 2003.
- Direttiva 98/79/CE del Parlamento Europeo e del Consiglio del 27 Ottobre 1998 relativa ai dispositivi medico-diagnostici in vitro. *Gazzetta Ufficiale delle Comunità Europee* 07.12.1998, L331:1-37.
- <http://www.bipm.fr/en/committees/jc/jctlm/>
- Hulsmann M, Berger R, Mortl D, et al. Incidence of normal values of natriuretic peptides in patients with chronic heart failure and impact on survival: a direct comparison of N-terminal atrial natriuretic peptide, N-terminal brain natriuretic peptide and brain natriuretic peptide. *Eur J Heart Fail* 2005;7:552-6.
- Murphy EA. The normal, and the perils of the sylleptic argument. *Prespect Biol Med* 1972;15:566-82.
- Alström T, Gräsbeck R, Hjelm M, et al. Committee on Reference Values, Scandinavian Society for Clinical Chemistry and Clinical Physiology. Recommendations concerning the collection of reference values in clinical chemistry and activity report. *Scand J Clin Lab Invest* 1975;35:(Suppl 144):1-74.
- Berg B, Nilsson J-E, Solberg HE, et al. Practical experience in the selection and preparation of reference individuals: empirical testing of the provisional Scandinavian recommendations. In: Gräsbeck R, Alström T, eds. *Reference values in laboratory medicine*. Chichester: J. Wiley & Sons, 1981;55-64.
- Horn PS, Pesce AJ. Reference intervals: an update. *Clin Chim Acta* 2003;334:5-23.
- Gräsbeck R. The evolution of the reference value concept. *Clin Chem Lab Med* 2004;42:692-7.
- CLSI document C28-A2. How to define and determine reference intervals in the clinical laboratory; Approved Guideline — 2nd ed. Clinical and Laboratory Standards Institute, Wayne PA, 2000.
- PetitClerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987;25:639-644. *Clin Chim Acta* 1987;170:S3-12.
- Harris EK. Effects of intra- and interindividual variation on the appropriate use of normal ranges. *Clin Chem* 1974;20:1535-42.
- Hyltoft Petersen P, Fraser CG, Sandberg S, et al. The index of individuality is often a misinterpreted quantity characteristic. *Clin Chem Lab Med* 1999;37:655-61.
- Fraser CG. Inherent biological variation and reference values. *Clin Chem Lab Med* 2004;42:758-64.
- Martin HF, Hoggatas JV, Driscoll J, et al. Reference values based on populations accessible to hospitals. In: Gräsbeck R, Alström T, eds. *Reference values in laboratory medicine*. Chichester: Wiley, 1981:233-62.
- Hoffmann HG. *Statistics in the practice of medicine*. *J Am Med Assoc* 1963;185:864-73.
- Bhattacharya CG. A simple method of resolution of a distribution into Gaussian components. *Biometrics* 1967;23:115-35.
- Naus AJ, Borst A, Kuppens PS. The use of patient data for the calculation of reference values for some haematological parameters. *J Clin Chem Clin Biochem* 1980;18:621-5.
- Baadenhuijsen H, Smit JC. Indirect estimation of clinical chemical reference intervals from total hospital patient data: application of a modified Bhattacharya procedure. *J*

- Clin Chem Clin Biochem 1985;23:829–39.
23. Kairisto V, Hänninen KP, Leino A, et al. Generation of reference values for cardiac enzymes from hospital admission laboratory data. *Eur J Clin Chem Clin Biochem* 1994;32:789–96.
 24. Kouri T, Kairisto V, Virtanen A, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. *Clin Chem* 1994;40:2209–15.
 25. Krøll J, Saxtrup O. On the use of patient data for the definition of reference intervals in clinical chemistry. *Scand J Clin Lab Invest* 1998;58:469–73.
 26. Bock BJ, Dolan CT, Miller GC, et al. The data warehouse as a foundation for population-based reference intervals. *Am J Clin Pathol* 2003;120:662–70.
 27. Grossi E, Colombo R, Cavuto S, et al. The REALAB project: a new method for the formulation of reference intervals based on current data. *Clin Chem* 2005;51:1232–40.
 28. Solberg HE. Using a hospitalized population to establish reference intervals: pros and cons. *Clin Chem* 1994;40:2205–6.
 29. Ferre-Masferrer M, Fuentes-Arderiu X, Puchal-Ane R. Indirect reference limits estimated from patients' results by three mathematical procedures. *Clin Chim Acta* 1999;279:97–105.
 30. Solberg HE, Petitclerc C. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem* 1988;26:593–8; *Clin Chim Acta* 1988;177:S1–12.
 31. Solberg HE, Stamm D. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values. Approved recommendation (1991) on the theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. *Eur J Clin Chem Clin Biochem* 1991;29:531–5; *Clin Chim Acta* 1991;202:S5–12.
 32. Tietz NW. A model for a comprehensive measurement system in clinical chemistry. *Clin Chem* 1979;25:833–9.
 33. Müller MM. Implementation of reference systems in laboratory medicine. *Clin Chem* 2000;46:1907–9.
 34. International vocabulary of basic and general terms in metrology (VIM) ISO VIM (DGUIDE 99999). Geneva: International Organization for Standardization, 2004.
 35. International Organization for Standardization. In vitro diagnostic medical devices - Measurement of quantities in biological samples - Metrological traceability of values assigned to calibrators and control materials. EN ISO 17511:2003. Geneva ISO, 2003.
 36. International Organization for Standardization. In vitro diagnostic medical devices — Measurement of quantities in samples of biological origin — Metrological traceability of values for catalytic concentration of enzymes assigned to calibrators and control materials. EN ISO 18513:2003. Geneva ISO, 2003.
 37. Harris EK, Boyd JC. Statistical bases of reference values in laboratory medicine. New York, NY: Marcel Dekker, 1995.
 38. Horn PS, Pesce AJ. Reference intervals. A user's guide. Washington DC: AACC Press, 2005.
 39. Wootton IDP, King EJ, Maclean Smith J. The quantitative approach to hospital biochemistry: normal values and the use of biochemical determinations for diagnosis and prognosis. *Br Med Bull* 1951;7:307–11.
 40. Wootton IDP, King EJ. Normal values for blood constituents. Interhospital differences. *Lancet* 1953;1:470–1.
 41. Reed AH, Henry RJ, Manson WB. Influence of statistical method used on the resulting estimate of normal range. *Clin Chem* 1971;17:275–84.
 42. Hanie EK, Demets DL. Estimation of normal ranges and cumulative proportions by transforming observed distributions to gaussian form. *Clin Chem* 1972;18:605–12.
 43. Boyd JC, Lacher DA. A multi-stage Gaussian transformation algorithm for clinical laboratory data. *Clin Chem* 1982;28:1735–41.
 44. Shultz EK, Willard KE, Rich SS, et al. Improved reference-interval estimation. *Clin Chem* 1985;31:1974–8.
 45. Solberg HE. International Federation of Clinical Chemistry (IFCC). Scientific Committee, Clinical Section. Expert Panel on Theory of Reference Values and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values. Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem* 1987;25:645–56; *Clin Chim Acta* 1987;170:S13–32.
 46. Jørgensen LGM, Brandslund I, Hyltoft Petersen P. Should we maintain the 95 percent reference intervals in the era of wellness testing? A concept paper. *Clin Chem Lab Med* 2004;42:747–51.
 47. Solberg HE. RefVal: a program implementing the recommendations of the International Federation of Clinical Chemistry on the statistical treatment of reference values. *Comput Meth Prog Biomed* 1995;48:247–56.
 48. Solberg HE. The IFCC recommendation on estimation of reference intervals. The RefVal Program. *Clin Chem Lab Med* 2004;42:710–4.
 49. Horn PS, Pesce AJ, Copeland BE. A robust approach to reference interval estimation and evaluation. *Clin Chem* 1998;44:622–31.
 50. Horn PS, Pesce AJ, Copeland BE. Reference interval computation using robust vs. parametric and nonparametric analyses. *Clin Chem* 1999;45:2284–5.
 51. Box GEP, Cox DR. An analysis of transformations. *J R Stat Soc* 1964;B26:211–52.
 52. Virtanen A, Kairisto V, Irijala K, et al. Regression-based reference limits and their reliability: example on hemoglobin during the first year of life. *Clin Chem* 1998;44:327–35.
 53. Virtanen A, Kairisto V, Uusipaikka E. Parametric methods for estimate covariate-dependent reference limits. *Clin Chem Lab Med* 2004;42:734–8.
 54. Sinton TJ, Crowley D, Bryant SJ. Reference values for calcium, phosphate, and alkaline phosphatase as derived on the basis of multichannel-analyzer profiles. *Clin Chem* 1986;32:76–9.
 55. Lahti A. Partitioning biochemical reference data into subgroups: comparison of existing methods. *Clin Chem Lab Med* 2004;42:725–33.
 56. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clin Chem* 1990;36:265–70.
 57. Lahti A, Hyltoft Petersen P, Boyd JC, et al. Objective criteria for partitioning Gaussian-distributed reference values into subgroups. *Clin Chem* 2002;48:338–52.
 58. Lahti A, Hyltoft Petersen P, Boyd JC. Impact of subgroup prevalences on partitioning Gaussian-distributed reference values. *Clin Chem* 2002;48:1987–99.
 59. Lahti A, Hyltoft Petersen P, Boyd JC, et al. Partitioning of nongaussian distributed biochemical reference data into subgroups. *Clin Chem* 2004;50:891–900.
 60. Gowans EMS, Hyltoft Petersen P, Blaabjerg O, et al. Analytical goals for acceptance of common reference intervals for laboratories throughout a geographical area.

- Scand J Clin Lab Invest 1988;48:757-64.
61. Dixon WJ. Processing data for outliers. *Biometrics* 1953;9:74-89.
 62. Horn PS, Feng L, Li Y, et al. Effect of outliers and non-healthy individuals on reference interval estimation. *Clin Chem* 2001;47:2137-45.
 63. Tukey JW. *Exploratory data analysis*. Reading, MA: Addison-Wesley, 1977.
 64. Schneider AJ. Some thoughts on normal, or standard, values in clinical medicine. *Pediatrics* 1960;26:973-84.
 65. Fuentes-Arderiu X, Ferré-Masferrer M, Gonzales-Alba JM, et al. Multicentric reference values for some quantities measured with Tina-Quant reagents systems and RD/Hitachi analysers. *Scand J Clin Lab Invest* 2001;61:273-6.
 66. Ferré-Masferrer M, Fuentes-Arderiu X, Alvares Funes V, et al. Multicentric reference values: shared reference limits. *Eur J Clin Chem Clin Biochem* 1997;35:715-8.
 67. Ferré-Masferrer M, Fuentes-Arderiu X, Gomà-Llongueras M, et al. Regional reference values for some quantities measured with the ADVIA Centaur analyser. A model of co-operation between the in vitro diagnostic industry and clinical laboratories. *Clin Chem Lab Med* 2001;39:166-9.
 68. Fuentes-Arderiu X, Ferré-Masferrer, González-Alba JM, et al. Multicentric reference values for some quantities measured with the Elecsys 2010 analyser. *Clin Chim Acta* 2001;304:143-6.
 69. Fuentes-Arderiu X, Mas-Serra R, Alumá-Trullás A et al. Guideline for the production of multicentre physiological reference values using the same measurement system. A proposal of the Catalan Association for Clinical Laboratory Sciences. *Clin Chem Lab Med* 2004;42:778-82.
 70. Bäck S-E, Nilsson J-E, Fex G, et al. Towards common reference intervals in clinical chemistry. An attempt at harmonization between three hospital laboratories in Skåne, Sweden. *Clin Chem Lab Med* 1999;37:573-92.
 71. Rustad P, Felding P, Lahti A. Proposal for guidelines to establish common biological reference intervals in large geographical areas for biochemical quantities measured frequently in serum and plasma. *Clin Chem Lab Med* 2004;42:783-91.
 72. Hyltoft Petersen P, Rustad P. Prerequisites for establishing common reference intervals. *Scand J Clin Lab Invest* 2004;64:285-92.
 73. Rustad P, Felding P, Franzson I, et al. The Nordic Reference Interval Project 2000: recommended reference intervals for 25 common biochemical properties. *Scand J Clin Lab Invest* 2004;64:271-84.
 74. Rustad P, Felding P, Lahti A, et al. Descriptive analytical data and consequences for calculation of common reference intervals in the Nordic Reference Interval Project 2000. *Scand J Clin Lab Invest* 2004;64:343-70.
 75. Jansen R, Schumann G, Baadenhuijsen H, et al. Trueness verification and traceability assessment of results from commercial systems for measurement of six enzyme activities in serum: an international study in the EC4 framework of the Calibration 2000 project. *Clin Chim Acta* 2006;368:160-7.
 76. Cotlove E, Harris EK, Williams GZ. Biological and analytic components of variation in long-term studies of serum constituents in normal subjects. III. Physiological and medical implications. *Clin Chem* 1970;16:1028-32.
 77. Gowans EMS, Hyltoft Petersen P, Blaabjerg O, et al. Analytical goals for acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988;48:757-64.
 78. Fraser CG, Hyltoft Petersen P, Libeer J-C, et al. Proposals for setting generally applicable quality goals solely based on biology. *Ann Clin Biochem* 1997;34:8-12.
 79. Harris EK. Some theory of reference values. II. Comparison of some statistical models of intraindividual variation in blood constituents. *Clin Chem* 1976;22:1343-50.
 80. Queraltó JM. Intraindividual reference values. *Clin Chem Lab Med* 2004;42:765-77.