

I. JCCLS Section

臨床化学における定量検査の精密さ・正確さ評価法指針 (改訂版)

(GC-JAMT1-1999)

Evaluation of precision and accuracy performance for quantitative routine methods in clinical chemistry

(社)日本臨床衛生検査技師会

定量検査の精密さ・正確さ評価法標準化ワーキンググループ

Committee for standardization division, Japan Association of Medical Technologists

Summary

This guideline was prepared to evaluate the precision and accuracy of a routine method in clinical chemistry. It was designed for the evaluation of current routine method in comparison with some reference methods.

The precision is estimated by using the control sera and patient's specimens as a test sample and the analysis of variance (ANOVA) is evaluated statistically. Allowable limits of the precision is judged from within-subject biological variation of healthy persons. The accuracy is evaluated by serum reference materials and comparative methods. In the serum reference materials method, the evaluation depends on the number of reference serum with different concentration. When one or two concentrations of serum is applied, it is compared with the reference value of serum reference materials. If more than three kinds of concentration serum are applied, it is estimated by the slope and the intercept derived from linear regression procedure. Comparison method using patient's specimens is estimated by the slope and the intercept derived from linear structural relationship

procedure. Allowable limits of the accuracy will be judged as the bias lower than 5 per cent (or 2 per cent for the measurement of sodium and chloride concentrations) of the expected value. To be available for a personal computer, a computer program has been already provided and one can easily calculate the statistics of evaluation.

序文

臨床化学検査法の信頼性に関する特性の評価法については、従来から多くの勧告または勧告案が提示されている。なかでもIFCC (International Federation of Clinical Chemistry)¹⁾、NCCLS (National Committee for Clinical Laboratory Standards)^{2), 3)}、日本臨床化学会^{4), 5)}などによる精密さと正確さの評価法は、日常検査法や自動分析装置の性能評価などに利用されてきた。しかし、各評価法の詳細は不明確な部分もあり、それらを利用する実践の場において少なからぬ混乱が生じている。本指針は臨床化学定量検査法の精密さと正確さを適正に評価するための方法について、従来の提示内容を再検討し、精密さと正確さの両者を体系的かつ具体的に評価する方法を新たに提示するものである。

1. 適用範囲

この評価法は、日常検査に用いる臨床化学定量検査法の精密さと正確さを評価する方法について定める。

精密さは管理試料および患者試料を用いて評価する。正確さは血清標準物質の測定、あるいは、比較対照法を用いる比較実験により評価する。

2. 用語^{6,7)}

2. 1 真の値 (true value)

測定量の正しい値。実際には求められないから標準物質の標準値、または、基準法で得られた値を真の値とみなす。

2. 2 誤差 (error)

測定値から真の値を引いた値。

2. 3 系統誤差 (systematic error)

測定結果にかたよりを与える原因によって生じる誤差。系統誤差には比例系統誤差 (proportional systematic error) と一定系統誤差 (constant systematic error) がある。

2. 4 偶然誤差 (random error)

突き止められない原因によって起こり、測定値のばらつきとなって現れる誤差。

2. 5 かたより (bias)

測定値の母平均から真の値を引いた値。

2. 6 ばらつき (dispersion)

測定値の大きさがそろっていないこと。ばらつきをの大きさを表すには、例えば標準偏差を用いる。

2. 7 不確かさ (uncertainty)

測定値の真の値が存在する範囲を示す推定値。

2. 8 精密さ (precision)

ばらつきの小さい程度。ばらつきを標準偏差や変動係数で表した値を精密度という。

2. 9 正確さ (accuracy)

かたよりの小さい程度。推定したかたよりの限界の値で表した値を正確度という。

2. 10 血清標準物質 (serum reference material)

血清そのものを組成とした標準物質で、特性値が基準法、実用基準法、常用基準法のいずれかによって決定されたもの。

2. 11 標準値 (reference value)

標準物質の特性または組成等を調製者が表示した値。

2. 12 個体内生理的変動の標準偏差 (standard deviation of within-subject biological variation, SD_w)

健康な個人が生理的に示す平均的な変動の大きさを標準偏差で表したもの。

2. 13 医学的意思決定濃度 (medical decision level)

臨床医が診断、病態鑑別、治療の意思決定を行う上で重要な (物質や成分の) 濃度。

2. 14 許容誤差限界 (allowable limit of error)

医学的に許容できると判断できる誤差の限界。

2. 15 基準法 (definitive method)

測定原理的に誤差が最低の水準に抑えられる測定法で、これは理論的にも実験的にも証明されている方法。

2. 16 実用基準法 (reference method)

測定体系上基準法に次ぐ測定法であり、十分に研究され、必要条件ならびに手順が明確に記述され、目的とする用途に相応じた正確さと精密さをもった値が得られる測定法。

2. 17 日常一般法 (field method)

多数の測定試料に応じることが出来、日常的に用いる精密さを有する測定法。

2. 18 被検法 (test method)

評価の対象となる日常一般法。

2. 19 比較対照法 (comparative method)

正確さを評価するときの基準になる方法。一般にこれは基準法あるいは実用基準法であるが、基準法から正確さが適正に伝達された日常一般法もこれに準ずる。

3. 評価法の概要

3. 1 実施手順

評価法の実施手順を流れ図として図1に示す。

3. 2 精密さの評価

精密さは、管理試料の反復測定値を用いて日間変動と日内変動を評価するとともに、多数の患者

試料の二重測定値を用いて実検体測定時の精密さを評価する。

精密さの許容誤差限界は、個体内生理的変動の標準偏差 (standard deviation, SD_w) の1/2 ($SD_w/2$) またはそれに対応する変動係数 (coefficient of variation, CV_w) の値とする⁹⁾ (表1)。

3. 3 正確さの評価

正確さは、1 (または2) 種類濃度の血清標準物質のみしか得られない場合は、その反復測定値から求めた平均値を標準値からのかたよりで評価する。3種類以上の濃度の血清標準物質が得られる場合は、それらの標準値に対する反復測定値の関係を直線回帰式によって評価する。血清標準物

質が入手できない場合は、多数の患者試料を用いて比較対照法に対する被検法の測定値の関係を直線関係式⁹⁾によって評価する。

3種類以上の血清標準物質を用いる評価、あるいは、比較対照法との比較評価においては、正確さは直線関係を表す式 $Y = a + b x$ の一定系統誤差と比例系統誤差 (それぞれ $a = 0$, $b = 1$ からのかたよりの) 検定によって評価する。有意な系統誤差を認めた場合は医学的意思決定濃度におけるかたよりで評価し、かたよりの許容誤差限界は、一般的な成分で5%、血清Na, Cl測定法で2%とする (表1)。

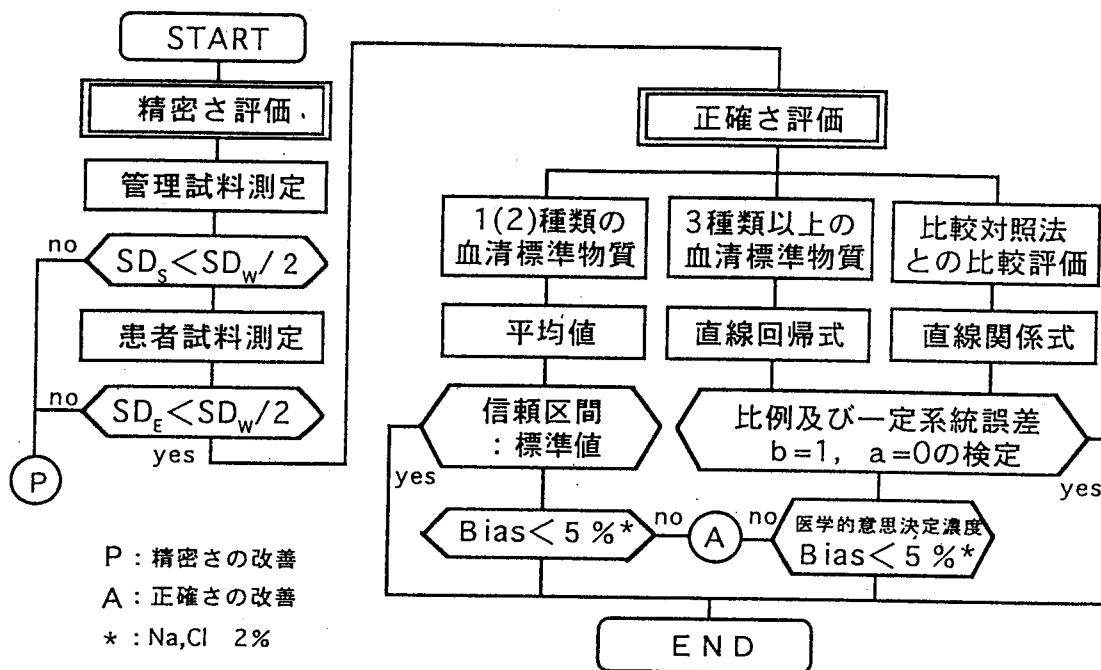


図1 精密さ・正確さ評価法の流れ図

表1 精密さと正確さの評価指標と許容誤差限界

評価特性	試料	評価実験	評価指標	許容誤差限界
精密さ	管理試料	日間・日内の反復測定	分散分析の標準偏差(変動係数) ^{*1}	$SD_w / 2^{*2}$
	患者試料	ランダムイズ2回測定	分散分析の標準偏差	$SD_w / 2$
正確さ	標準物質	1種類を反復測定	かたより $B = XB - x_r$	5% (Na, Cl 2%)
	標準物質	3種類以上を反復測定	直線回帰式 $a = 0, b = 1$ の検定	$\alpha = 0.05$
			医学的意思決定濃度でのかたより	5% (Na, Cl 2%)
	患者試料	比較対照法との比較実験	直線関係式 $a = 0, b = 1$ の検定	$\alpha = 0.05$
医学的意思決定濃度でのかたより			5% (Na, Cl 2%)	

*1: 管理試料の値が基準範囲上限以上のときは変動係数による

*2: 健康人の個体内生理的変動の1/2 ($SD_w / 2$)

4. 試料

4.1 患者試料

患者試料を用いて精密さ・正確さを評価するときには、患者検体の中から溶血や混濁また高ビリルビン血清などマトリックス効果による影響が認められる試料¹⁰⁾を除外して、50例以上を準備する。

目的成分の濃度は、日常測定濃度の全範囲にわたることが必要である。望ましい一般的な試料濃度の分布を、試料総数に対する比率として表2に示す。日常検査の濃度範囲が狭い検査項目(血清Na, Cl濃度など)では、日常検査で遭遇する異常高・低濃度の患者試料も準備する必要がある。

異常濃度の試料を短期間に準備できないときは、凍結保存した患者試料の中から選択する。また、量が不足するときは、近似した濃度のもの2本を混和してもよい。

準備した試料は、評価実験の期間中密栓凍結保

存(例えば-20°C以下)する。

4.2 管理試料

日間変動と日内変動の評価に使用する管理試料は、入手可能であれば3種類以上(基準範囲内濃度1種類と異常濃度2種類など)準備し、評価実験期間中密栓凍結保存する。なお、試料は比較対照法と被検法の間で目的成分に対する反応性が患者検体と同様であることを事前に確認¹⁰⁾しておく必要がある。

4.3 血清標準物質

正確さ評価に用いる血清標準物質は、入手可能ならば基準範囲内から異常濃度域を概ね網羅する3種類(できれば4種類)以上を準備する。

5. 精密さの評価

5.1 管理試料による日間精密度と日内精密度の評価

1) 測定

被検法が安定な状態にあるときに、管理試料を毎日2本以上、日常検査検体の中にランダムに挿入して測定する。これを20日以上繰り返す。

2) 結果の解析

毎日 $n (\geq 2)$ 本の管理試料を $k (\geq 20)$ 日間測定したデータから、表3の分散分析表の各パラメータを計算した後、精密度 (SDとCV) を求める。

$$x_{ij} : i \text{ 日目の } j \text{ 本目測定値 } (i=1,2,\dots,k; j=1,2,\dots,n)$$

XB_i : i 日目の平均値

XBB : 総平均値

$$\text{日間精密度 } SD_A = \{(V_A - V_E) / n\}^{1/2} \quad (1)$$

$$\text{日内精密度 } SD_E = [S_E / \{k(n-1)\}]^{1/2} \quad (2)$$

$$\text{総合精密度 } SD_S = (SD_A^2 + SD_E^2)^{1/2} \quad (3)$$

$$\text{変動係数 } CV_S = 100 \times SD_S / XB (\%) \quad (4)$$

XBB が基準範囲上限以下のときは SD_S を $SD_W/2$ と比較し、上限以上のときは CV_S で評価する。

$$CV_S < 100 \times (SD_W/2) / \text{健常者基準値の平均値} \quad (5)$$

また、次式により日内変動に対する日間変動の有意性を、自由度 $f_1=f_A, f_2=f_E$ により検定する。

$$F_0 = V_A / V_E \quad (6)$$

$$F_0 > F_{\alpha\beta}(f_A, f_E) \quad (7)$$

表2 望ましい患者試料の濃度分布 (NCCLS 指針³⁾ より要約)

患者試料濃度域	基準範囲が低濃度域の場合*1	基準範囲が中濃度域の場合*2
基準範囲の下限値以下	15~30%	10~30%
基準範囲の下限値から平均値		20~25%
基準範囲の平均値から上限値	20~30%	20~25%
基準範囲の上限値から上限値の2倍	20~40%	20~40%
基準範囲の上限値の2倍から4倍	10~20%	
基準範囲の上限値の4倍以上	約10%	

*1 : 各種血清酵素、BUNなど *2 : グルコース、Na、K、Cl など

表3 日間・日内精密度を求めるための分散分析表

要因	平方和	自由度	分散	期待値
日間	$S_A = \sum_i n (XB_i - XBB)^2$	$f_A = k - 1$	$V_A = S_A / (k - 1)$	$\sigma_E^2 + n \sigma_A^2$
日内	$S_E = \sum_i \sum_j (x_{ij} - XB_i)^2$	$f_E = k(n - 1)$	$V_E = S_E / \{k(n - 1)\}$	σ_E^2
計	$S_T = \sum_i \sum_j (x_{ij} - XBB)^2$	$f_T = kn - 1$		

5. 2 患者試料による評価

1) 測定

被検法が安定な状態にあるときに、患者試料（試料数：n）に順番を付け、それらを試料として1回目の測定を実施する。2回目の測定は、試料の順番をランダムに変更するか、あるいは、全く逆の順番にし、同日内に実施する。

2) 結果の解析

(2. 1) 差のプロット図

それぞれの患者試料の2回測定値 ($x_{i1}, x_{i2}; i=1, 2, \dots, n$) の平均値 ($XB_i = (x_{i1} + x_{i2}) / 2$) を横軸にとり、1回目と2回目の差 ($x_{i1} - x_{i2}$) を縦軸にとって、図2の差のプロット図を作成する。

差のプロット図は、精密度の濃度に対する関係を表す精密さのプロファイル (precision profile) の把握が目的であり、差のプロットが XB_i の大きさにかかわらずほぼ一様であることを確認する。差が測定値によって一様でない場合（値の上昇に

伴い増大するなど）は、ほぼ一様と考えられる濃度群（例えば低濃度、中濃度、高濃度群の別）に分けて、それぞれの濃度群について評価する。また、極端に大きな差がある測定値は外れ値でないかを検討する。

(2. 2) 外れ値の検出

1回目と2回目の測定値の差から、次のように外れ値を検出する。

$$2 \text{ 回測定} \text{ の 範 圍 } d_i = |x_{i1} - x_{i2}| \quad (8)$$

$$\text{範 圍} \text{ の 平 均 値 } dB = (\sum_i d_i) / n \quad (9)$$

このとき各 d_i が平均値 dB の4倍以上ある測定値は、その原因を検討し外れ値のときは除外する。

(2. 3) 統計計算

$$\text{群内平方和 } S_E = \sum_i \sum_j (x_{ij} - XB_i)^2 = \sum_i (x_{i1} - x_{i2})^2 / 2 \quad (10)$$

$$\text{誤差分散 } V_E = S_E / n \quad (11)$$

$$\text{標準偏差 } SD_E = (V_E)^{1/2} \quad (12)$$

この SD_E を $SD_w/2$ と比較する。

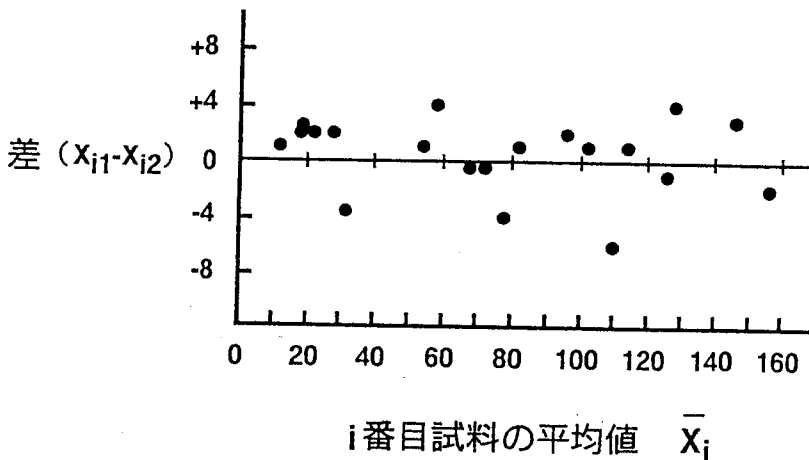


図2 差のプロット図による濃度と精密さの関係の検討

6. 正確さの評価

6. 1 1種類の濃度の血清標準物質による評価

1) 測定

血清標準物質を添付されている認証書の指示に従い調製し、被検法が安定な状態にあるときに10回以上繰り返し測定する。

2) 結果の解析

10回以上の繰り返し測定値 (x_1, x_2, \dots, x_n) から、次式によりかたよりと平均値の95%信頼区間を求める。なお、2種類の濃度の血清標準物質がある場合は、濃度ごとに評価する。

x_r : 血清標準物質の標準値

n : 測定回数

XB : 平均値
SD : 標準偏差

$t_{0.05(n-1)}$: t分布の自由度(n-1)における
両側確率0.05のt値

$$\text{平均値の95\%信頼区間の幅 } XB \pm t_{0.05(n-1)} \cdot SD / n^{1/2} \quad (13)$$

$$\text{かたより } B = XB - x_r \quad (14)$$

このとき、平均値の95%信頼区間に x_r が含まれていれば、平均値付近の濃度域における有意なかたよりはないとみなし、信頼区間から外れた場合は $100 \times |B| / x_r$ が5%以下(N_a, C_I は2%以下)かどうかで正確さを評価する。

6. 2 3種類以上の濃度の血清標準物質による評価

3種類以上の濃度の血清標準物質を用いることが可能な場合は、血清標準物質の標準値に対する被検法測定値の関係を直線回帰式 $Y = a + b x$ で表し、被検法の正確さを一定系統誤差($a = 0$ からのかたより)と比例系統誤差($b = 1$ からのかたより)の大きさで評価する。

1) 測定

3種類以上の濃度の血清標準物質を添付されている認証書の指示に従い調製し、被検法が安定な状態にあるときにそれぞれを5回以上繰り返し測定する。

2) 結果の解析

(2. 1) 散布図の作成

血清標準物質の標準値を横軸、被検法による測定値を縦軸にとり、図3のような散布図を作成する。散布図から測定法の直線性を観察し、直線性から外れたときはまずその原因を追究する。正確さの評価は直線性が得られる範囲で行う。また、後述する残差プロット図で、全濃度域にわたるばらつきの一様性を評価し、また、極端に乖離する測定値は外れ値であるかどうかを検討し外れ値のときは除外する。

(2. 2) 統計計算

(2. 2. 1) 直線回帰式の推定

n種類の濃度の血清標準物質の標準値(x_i)

に対する被検法によるq回繰り返し測定値(y_{ij} ; $i = 1, 2, \dots, n$; $j = 1, 2, \dots, q$)の関係を表す直線回帰式 $Y = a + b x$ の回帰係数および残差標準偏差を次式により求める。

$$\text{傾き } b = S_{xy} / S_{xx} \quad (15)$$

$$\text{切片 } a = YB - bXB \quad (16)$$

ただし、 S_{xx} 、 S_{yy} 、 S_{xy} はxとyの変動と共変動で次のようになる。

$$\text{xの変動 } S_{xx} = q \sum_i (x_i - XB)^2, \text{ ただし、} \\ XB = (\sum_i x_i) / n \quad (17)$$

$$\text{yの変動 } S_{yy} = \sum_i \sum_j (y_{ij} - YB)^2, \text{ ただし、} \\ YB = (\sum_i \sum_j y_{ij}) / (nq) \quad (18)$$

$$\text{xとyの共変動 } S_{xy} = \sum_i \sum_j (x_i - XB)(y_{ij} - YB) \quad (19)$$

求めた回帰式の適合性を、標準値(x_i)に対する回帰式の予測値($Y_i = a + b x_i$)と測定値(y_i)との差を意味する残差($= y_i - Y_i$)によって検討する。それには、測定値の全変動 S_{yy} を特性値の大きさに依存する変動 S_B と直線性以外の要因による変動 S_R 、および、純誤差変動 S_E に分解する。

$$S_B = \sum_i q (a + b x_i - YB)^2 = S_{xy}^2 / S_{xx} \quad (20)$$

$$S_R = \sum_i q (YB_i - a - b x_i)^2 = \sum_i q (YB_i - YB)^2 - S_{xy}^2 / S_{xx} \quad (21)$$

$$S_E = \sum_i \sum_j (y_{ij} - YB_i)^2 = S_{yy} - \sum_i q (YB_i - YB)^2 \quad (22)$$

$$S_{yy} = S_B + S_R + S_E \quad (23)$$

ただし、 $YB_i = (\sum_j y_{ij}) / q$ である。これらより残差標準偏差 $s_{y \cdot x}$ を求める。

$$s_{y \cdot x} = \{S_E / (nq - n)\}^{1/2} \quad (24)$$

回帰式の測定値に対する適合性は、血清標準物質の標準値を横軸にとり、各残差を残差標準偏差で割った標準化残差($e_{ij} = (y_{ij} - Y_i) / s_{y \cdot x}$)の値を縦軸にとった残差プロット図から、極端に外れた点がないか、残差に片寄った傾向がないかなどを検討する(図3)。

また、被検法による測定値の直線性からのズレの検討は、次の統計量をF検定すればよい。

$$F_0 = \{S_R / (n - 2)\} / \{S_E / (nq - n)\} \quad (25)$$

(2. 2. 2) 系統誤差の検定

傾きが $b = 1$ (比例系統誤差なし) から有意な
 かたよりをもちかどうかを、次のように検定する。

$$t_b = \{ |b - 1| (S_{xx})^{1/2} \} / s_{y \cdot x} \quad (26)$$

この値を自由度 $(nq - n)$ の t 分布の両側確率 5
 % 点 $t_{\alpha/05(nq-n)}$ と比較し

$$t_b \leq t_{\alpha/05(nq-n)} \quad (27)$$

であれば、有意水準 5 % で有意な比例系統誤差を
 認めないとみなし、この場合は、次の一定系統誤
 差の評価を行う。

切片が $a = 0$ (一定系統誤差なし) から有意な
 かたよりをもちかどうかを、次のように検定する。

$$t_a = (|a| / s_{y \cdot x}) \{ (n S_{xx}) / (\sum |x_i|^2) \}^{1/2} \quad (28)$$

$$t_a \leq t_{\alpha/05(nq-n)} \quad (29)$$

であれば、有意水準 5 % で有意な一定系統誤差を
 認めないとみなす。

(2. 2. 3) 医学的意思決定濃度における評価
 $a = 0$ あるいは $b = 1$ から有意なかたよりが認め
 られたときは、直線回帰式の医学的意思決定濃度
 (x_0) における予測値 $(Y_0 = a + b x_0)$ のかた
 より B の x_0 に対する割合を求める。

$$|B| / x_0 = |Y_0 - x_0| / x_0 \quad (30)$$

この値が 5 % 以下 (Na, Cl は 2 % 以下) であるか
 どうかで正確さを評価する。

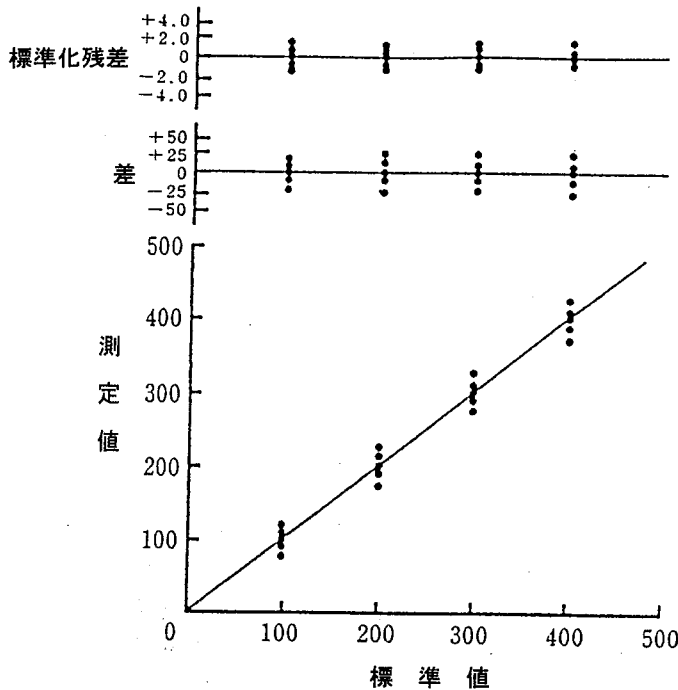


図3 3種類以上の濃度の血清標準物質を用いる正確さの評価

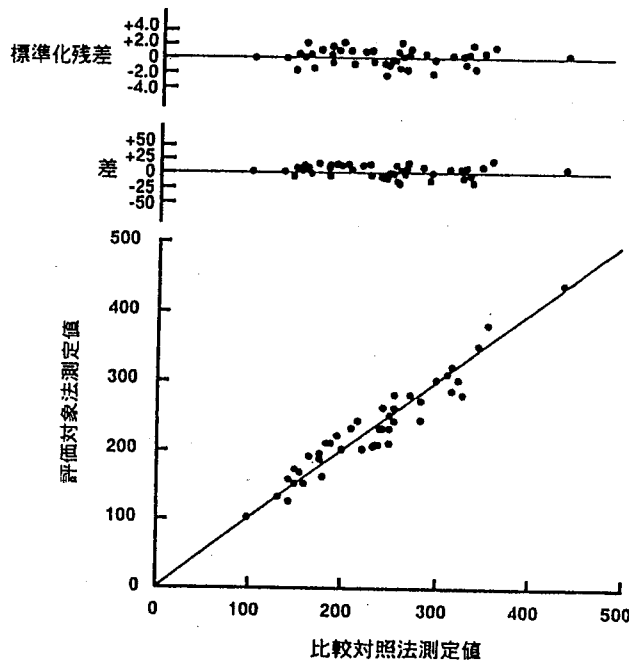


図4 比較対照法との比較実験による正確さの評価

6. 3 患者試料による比較対照法との比較実験による評価

多数の患者試料の比較対照法による測定値に対する被検法による測定値の関係を直線関係式 $Y = a' + b'x$ で表し、被検法の正確さを一定系統誤差 ($a' = 0$ からのかたより) と比例系統誤差 ($b' = 1$ からのかたより) の大きさで評価する。

1) 測定

比較対照法と被検法が安定な状態にあるときに、用意した50例以上 (n例) の患者試料のそれぞれを両法で測定する。なお、比較対照法あるいは被検法に有意な日間変動が存在する場合は、患者試料の全部を1日で測定せず、毎日5~10例ずつ5日以上に分けて測定する。実験期間中は適当な精度管理試料を被検試料と同様に測定し精度管理を行う。

2) 結果の解析

(2. 1) 散布図と差のプロット図

比較対照法による測定値 ($x_i; i=1, 2, \dots, n$) を横軸、被検法による測定値 (y_i) を縦軸にと

り、図4のような散布図と測定法間の差のプロット図を作成する。

散布図から測定法間の直線性を観察し、直線性から外れたときはその原因を追究するとともに、正確さの評価は直線性が得られる範囲で行う。また、差のプロット図と後述する残差プロット図で、全濃度にわたってばらつきの一様性を評価し、極端に乖離する測定値は外れ値であるかどうかを検討する。

(2. 2) 外れ値の検出

次のように外れ値を検出し、その原因を究明し外れ値であることが確認された場合は除外する。

d'_i : 比較対照法の値に対する対照法と被検法の差の絶対値の割合 ($= |y_i - x_i| / x_i$)

$$(31)$$

相対的な差の平均値 $dB' = (\sum_i d'_i) / n$

$$(32)$$

このとき、各 d'_i がその平均値 dB' の4倍以上ある測定値は、その原因を究明し外れ値のときは除外する。

(2. 3) 統計計算

(2. 3. 1) 直線関係式の推定

あらかじめ両測定法で同一試料を20 (m) 回以上繰り返し測定し、それらの測定値から測定法の誤差分散を求める。

x_{Pj} : 同一試料の比較対照法による繰り返し測定値 ($j=1,2,\dots,m$)

y_{Pj} : 同一試料の被検法による繰り返し測定値
比較対照法の誤差分散 $V_x = \{\sum_j (x_{Pj} - \bar{X}_P)^2\} / m$ (33)

被検法の誤差分散 $V_y = \{\sum_j (y_{Pj} - \bar{Y}_P)^2\} / m$ (34)

V_x 、 V_y のかわりに、両測定法で前述の精密さの評価を実施し(3)あるいは(12)式で得られる誤差標準偏差を2乗した値を用いてもよい。

次に、次式により誤差分散比 λ と直線関係式の傾き b' と切片 a' を求める。

誤差分散比 $\lambda = V_y / V_x$ (35)

傾き $b' = [S_{yy} - \lambda S_{xx} + \{(S_{yy} - \lambda S_{xx})^2 + 4 \lambda S_{xy}^2\}^{1/2}] / (2 S_{xy})$ (36)

切片 $a' = \bar{Y} - b' \bar{X}$ (37)

ただし、 S_{xx} 、 S_{yy} 、 S_{xy} は x と y の変動と共変動で次のようになる。

x の変動 $S_{xx} = \sum_i (x_i - \bar{X})^2$, ただし、
 $\bar{X} = (\sum_i x_i) / n$ (38)

y の変動 $S_{yy} = \sum_i (y_i - \bar{Y})^2$, ただし、
 $\bar{Y} = (\sum_i y_i) / n$ (39)

x と y の共変動 $S_{xy} = \sum_i (x_i - \bar{X})(y_i - \bar{Y})$ (40)

なお、比較対照法と被検法による測定値の直線関係の程度を表す相関係数は次式で得られる。

相関係数 $r = S_{xy} / (S_{xx} S_{yy})^{1/2}$ (41)

直線関係式の残差プロットは、次式で得られる横軸座標 x_i^* に対して、上記で求めた V_x 、 λ 、 b' 、 a' から次式により得られる標準化残差(e'_i)を縦軸にとってプロットする(図4)。

$x_i^* = (b' y_i + \lambda x_i - a' b') / (\lambda + b'^2)$ (42)

$e'_i = (y_i - a' - b' x_i) / \{V_x^{1/2} (\lambda + b'^2)^{1/2}\}$ (43)

(2. 3. 2) 系統誤差の検定

直線関係式の傾き b' の標準誤差と95%信頼区間を分布型によらないブートストラップ(Bootstrap)法⁶⁾を利用して求め、信頼区間の中に $b'=1$ が入っていれば有意な比例系統誤差を認めないとみなす。

直線関係式の切片 a' についても同様にブートストラップ法によりその信頼区間を求め、一定系統誤差の検定を行う。

(2. 3. 3) 医学的意思決定濃度における評価
 $a'=0$ あるいは $b'=1$ から有意なかたよりが認められたときは、直線関係式の医学的意思決定濃度(x_0)における予測値($Y_0 = a' + b' x_0$)のかたよりの割合を B の x_0 に対する割合を求める。

$|B| / x_0 = |Y_0 - x_0| / x_0$ (44)

この値が5%以下(Na, Cl は2%以下)であるかどうかで正確さを評価する。

(2. 3. 4) 正確さ評価グラフ

比較対照法に対する被検法の正確さを、図4のように直線式、差のプロット、残差プロットにより表示する。

文 献

- 1) Buttner J, et al: Approved recommendation on quality control in clinical chemistry. J Clin Chem Clin Biochem, 18:78-88, 1980
- 2) NCCLS Approved Guideline (EP5-A): Evaluation of Precision Performance of Clinical Chemistry Devices. 1999
- 3) NCCLS Approved Guideline (EP9-A): Method Comparison and Bias Estimation Using Patient Samples. 1995
- 4) 日本臨床化学会分析部会近畿支部機器委員会: 自動分析装置性能評価法試案(Step 1, 1984-12-1). 生物試料分析, 7: 79-92, 1984
- 5) 日本臨床化学会分析部会関東支部機器委員会: 測定値評価の方法-濃度測定における精密さと正確さ-勧告法試案(Step 1, Draft 1, 1987-11-15). 臨床化学, 16(補冊): 287-298, 1987
- 6) 日本規格協会: 品質管理用語 Z 8101, 計測用語 Z 8103

7)日本臨床化学会標準品情報専門委員会：標準に関する用語(Ver.2.4). 臨床化学, 25:126-134, 1996

8)Fraser CG, et al: Proposed Quality Specifications for the Imprecision and Inaccuracy of Analytical System for Clinical Chemistry. Eur J Clin Chem Clin Biochem, 30:311-317, 1992

9)丹後俊郎：測定誤差を考慮に入れた線形関係式；測定法の比較のための統計学的方法. 臨床病理, 36:1101-1108, 1988

10)NCCLS Proposed Guideline (EP14-P): Evaluation of Matrix Effects. 1998

解 説

1. 定量検査の評価特性

I F C C¹⁾によれば日常検査法の信頼性に関する特性の評価は、精密さはwithin-run imprecision, between-run imprecision, between-day imprecisionについて、正確さは calibration standard, linearity, recovery, blanks, specificity, interference, comparison of methodsなど、および、detectabilityの各項目を勧告している。これら特性の評価法について、従来からいくつかの組織により勧告（または勧告案）が提示されている¹⁻⁹⁾。今回、本指針が扱っている内容は、精密さの評価と、正確さについての標準物質の測定と比較対照法との比較評価に関する部分である。したがって、被検法の直線性、特異性、妨害物質の影響、最低検出限界などについては、満足されていることを前提としており、本指針とは別に評価することが必要である。その背景には、今回取り上げた信頼性に関する特性の評価法が、臨床検査データの施設間差解消や精度保証体制確立を目的とする標準化活動において重要な手段となっていながら、実践の場において少なからぬ混乱がみられるからである¹⁰⁻¹³⁾。一方、本指針の適用範囲は臨床化学検査法に限定しているが、それは最も代表的な定量検査の評価法をまず共通の認識としたいためであり、免疫反応を含めることによる特殊性などを問

題としなくてもよい測定条件下では、広く一般的な定量検査にも適用が可能である。

2. 許容誤差限界

許容誤差限界に対する従来の考え方は4つに大別される。第1は臨床的な有用性に基づき医学的意思決定濃度における基準を経験的に定める方法¹⁴⁾であるが、立場や疾患によって基準が異なるという点が懸念される。第2は現在の技術水準 (the state-of-the-art) に基づいて定める方法で、優れた検査室における分析法の実際の測定誤差の大きさに目標を求めるものである。第3はTonksによって提唱された基準であり¹⁵⁾、基準範囲の1/4を基準範囲の中央値で割った値を用いその最大値を10%と定めている。これに対し第4は北村¹⁶⁾、Cotloveら¹⁷⁾、Fraserら¹⁸⁾による考え方であり、測定誤差を生体の生理的変動幅と比較し、個体内変動幅の半分を基準値の平均値で割った値を誤差限界としており、現在最も広く利用されている。この値はTonksの基準より厳しいが、個人の時系列データの観察という検査データの臨床的利用の観点からは妥当と考えられる¹⁸⁾。

これら背景に基づき、本指針では精密さの許容誤差限界として個体内生理的変動の半分の値 ($S_{Dw}/2$) を用いた。個体内変動幅の推定値としては北村¹⁶⁾やCotloveら¹⁷⁾の報告があるが、その後20年以上も経過し提示項目数も少ないため、最近の報告例^{19, 20)}を参考に暫定的な限界値を表4に示した。ただし、より信頼性の高い個体内変動幅の推定値が得られれば今後改訂する必要がある。また、これらの値は基準値付近の濃度に適用されるものであり、異常域の誤差限界についてはCV%値をその参考として用いる。また、個体内生理的変動が大きい成分については、許容誤差限界としてCV5%を上限值とする。

一方、正確さの許容誤差限界については、1（または2）種類の血清標準物質を用いる評価では平均値の信頼区間と標準値を比較し、3種類以上の血清標準物質を用いる評価あるいは比較対照法との比較実験においては、比例および一定系統

誤差 (b=1, a=0からのかたより) の統計学的検定によることとした。この条件を満足することが望ましいが、棄却された場合にはもう一段あまい基準として、バイアスの相対値の経験的な限界を5% (Na,Clは2%) 以下であることとした。評価のための濃度水準は医学的意思決定濃度とし、

具体的には基準範囲上限値や治療方針決定値また既報の文献¹⁴⁾からの引用値などを用いる。

また、比較対照法との比較実験においては、基準法あるいは実用基準法を用いることが望ましいが、これら基準法から正確さが適正に伝達された自動分析装置による方法などを用いてもよい。

表4 健康者の個体内生理的変動幅と許容誤差限界

成分	基準範囲 単位	個体内変動(SD _w)	許容誤差限界(SD _w /2)	許容誤差限界(CV%)
Na	134-147 mEq/l	1.8	0.9	0.6
K	3.4-4.8 mEq/l	0.23	0.11	2.8
Cl	96-110 mEq/l	1.8	0.9	0.9
Ca	8.9-10.4 mg/dl	0.27	0.13	1.4
無機リン	3.3-5.5 mg/dl	0.43	0.21	5.5(5.0)
Fe	M 48-185 μg/dl	16.0	8.0	10.0(5.0)
総蛋白	5.8-8.5 g/dl	0.26	0.13	1.7
アルブミン	4.1-5.2 g/dl	0.17	0.08	1.8
ZTT	1.9-9.5 KU	0.91	0.45	7.2(5.0)
TTT	0.0-4.0 MU	0.29	0.14	9.9(5.0)
BUN	8-20 mg/dl	2.1	1.0	7.5(5.0)
クレアチニン	M 0.56-1.10 mg/dl	0.06	0.03	4.9
尿酸	2.0-6.0 mg/dl	0.50	0.25	5.2(5.0)
T. Bil	0.3-1.3 mg/dl	0.13	0.06	9.4(5.0)
D. Bil	0.0-0.3 mg/dl	0.05	0.02	17.5(5.0)
グルコース	60-110 mg/dl	4.1	2.0	2.0
中性脂肪	40-170 mg/dl	25.9	12.9	14.4(5.0)
T. Chol	126-251 mg/dl	12.3	6.1	3.4
F. Chol	31-75 mg/dl	5.1	2.5	5.3(5.0)
リン脂質	142-267 mg/dl	15.2	7.6	3.7
総胆汁酸	0-10 μmol/l	1.55	0.77	15.1(5.0)
SIAL	44-73 mg/dl	4.11	2.05	3.6
AST	10-32 U/l, 37°C	1.8	0.9	5.9(5.0)
ALT	2-31 U/l, 37°C	2.4	1.2	6.1(5.0)
LD	118-213 U/l, 37°C	17.1	8.5	3.2
ALP	M 121-320 U/l, 37°C	11.1	5.5	4.3
γGT	0-50 U/l, 37°C	2.6	1.3	7.6(5.0)
LAP	80-190 U/l, 37°C	6.0	3.0	2.6
ChE	170-420 U/l, 37°C	15.0	7.5	2.5
GUA	0.0-1.7 U/l, 37°C	0.12	0.06	7.9(5.0)
CK	M 60-263 U/l, 37°C	16.8	8.4	7.6(5.0)
AMY	45-150 U/l, 37°C	8.6	4.3	5.2(5.0)
LP	8-50 U/l, 37°C	2.37	1.18	8.2(5.0)

3. 精密さの評価

本指針では精密さの評価を、管理試料による評価と患者試料による評価の2種類を実施し、いずれの場合も精密度の推定には分散分析法を用いた。

管理試料による評価の主目的は、長期間安定な試料を使用する日間精密度と日内精密度の推定である。繰り返し測定日数は20日間以上とした。従来精密度の推定に際しては10日間以下の繰り返し測定が主であったが、信頼性の高い日間精密度を求めるためには最低でも16日以上20日程度の期間が必要であり²⁰⁾、従来に比べるとより長い日数が指定されている。

患者試料による評価は、多数の実検体の二重測定値を用いることにより、実際の患者試料の精密さの状況を観察するとともに精密度と濃度の関係の把握を目的とする。この場合の二重測定は同日内に実施するため日間変動を含まず、管理試料から得られる総合精密度に比較して小さな推定値が得られ日内精密度に近い値となる可能性が高い。ここで、多数の患者試料の測定値は、正確さの評価に用いるデータと共有すればよく、試料の濃度分布や測定上の注意点などは正確さの評価の項に準ずる。ただし、異常高値域を多く含む試料を用いた評価は、健康者の基準値を基に設定した許容誤差限界が適用できない場合があるので注意を要する。また、精密度と濃度の関係を表す精密さのプロファイルにおいて、濃度の上昇とともに精密度が明らかに大きくなる場合は、ほぼ一定と考えられる濃度域群（例えば低濃度域、中濃度域、高濃度域別など）に分けそれぞれの濃度群別に評価する。なお、精密さのプロファイルは、多数の患者試料の二重測定値を用い、平均値を横軸に精密度の指標である範囲（または標準偏差や変動係数）を縦軸にプロットしたものであり^{22, 29)}、測定範囲全域の精密度の状況を評価する方法である。

また、これらデータに分散分析法を適用することにより、従来曖昧であった日間精密度と日内精密度のそれぞれを適切に推定することが可能となる²⁰⁾。精密度の評価法として過去に種々の報告があったが⁵⁾、統計学的な不適切性を指摘されてい

るものもあり^{12, 23, 33)}、本指針では上記評価法を採用した。

4. 正確さの評価

正確さの評価は、1（または2）種類の血清標準物質による評価、3種類以上の血清標準物質を用いる評価、および、多数の患者試料を用いた比較対照法との比較評価の3つに大別され、それぞれ評価のための解析手法も異なる。ここで、1種類の血清標準物質による評価は、分析法の直線性が満足され一定系統誤差がないという前提がなければ、必ずしも測定範囲全域の正確さが保証できない。その点、3種類以上の血清標準物質を用いた評価や比較対照法との比較評価は、全域の正確さを比例系統誤差と一定系統誤差で評価することができ、より望ましい評価法といえる。ただし、その前提として分析法の直線性を事前に確認しておく必要があり、また、系統誤差を高い信頼性で推定するためには、血清標準物質は4種類以上の濃度を用いることが望ましい²⁰⁾。一方、血清標準物質を入手できない場合は、実用基準法で値付けしたプール血清を、血清標準物質の代わりに用いることもできる。また、実際に得られる標準物質は1種類であることが多いが、それを希釈調製して複数濃度の標準物質を作製する場合、希釈用試料は一般的な患者血清と物理化学的性質が同様なものを用い、希釈操作は重量法などを用いて極力正確に行う必要がある³⁰⁾。

血清標準物質の繰り返し測定は、ある日に集中的に実施するようになっているが、その前提には日間誤差が無視できることを仮定しており、精度管理が十分な状況下で測定を実施する必要がある。したがって、日間誤差が無視できない分析法の場合には、5日間程度に繰り返し測定を分けて実施し、得られた測定値を用いて結果の解析を行う必要がある。

多数の患者試料の測定も、日間誤差を相殺できるように、分析法が安定状態にあるときに精度管理を行いながら、比較対照法と被検法の両法で毎日5～10例ずつ5日以上に分けて測定を実施する。

患者試料数は、信頼性の高い推定を行うために50検体以上を用いる²⁹⁾。試料分布についてはNCCLSの指針³¹⁾に準じて表2を示したが、分析法の測定可能範囲を考慮し低値や高値側に極端なデータよりを示したり2相性分布にならないような試料を集める必要がある。また、全ての患者試料をそれぞれ分析法で二重測定しておけば、正確さと精密さの両評価に利用することができる。

測定値の解析については、3種類以上の血清標準物質を用いる場合は直線回帰式を適用し、比較対照法に対する比較評価には直線関係式を適用した。その背景には、比較対照法に無視できない誤差が含まれる場合に、通常回帰式を適用することの統計学的な問題があるためである^{10, 26-33)}。その際、直線関係式の算出において、二つの測定法の誤差分散比を事前に検討することなく、 λ を1と見なして計算することは避けるべきである。

ただし、直線関係式で得られる傾きと切片の検定には通常の仮説検定が行えないため、標準誤差と信頼区間の推定には分布型によらないブートストラップ (Bootstrap) 法³⁵⁾を利用した。

本法のアルゴリズムは次のようになる¹⁰⁾。

① n 個の患者試料測定値 (x_i, y_i) , $i=1, 2, \dots, n$ の各組を Z_i とする。

② n 個の Z_i の中から乱数を利用して重複を許し独立に無作為に n 個のブートストラップ標本を抽出し、標本から直線関係式の係数 b' と a' を求めることを500回繰り返す。

③ 500個の b' と a' から両分析法間の関係を表す直線式の傾きと切片の標準誤差および95%信頼区間を求める。

④ それぞれの95%信頼区間の中に $b = 1$ あるいは $a = 0$ があれば有意な系統誤差はないと判断する。

また、本指針で提示した直線関係式を求める方法以外にも標準主軸回帰法³⁶⁾やPassing-Bablok法³⁷⁾による方法も報告されている。

5. コンピュータプログラム

本評価法指針で用いる統計計算を支援するため

のコンピュータプログラムを別に用意する。プログラムは情報システム研究班の臨床検査マイコンライブラリーに登録されている。

文 献

- 1) Buttner J, et al: Approved recommendation on quality control in clinical chemistry. J Clin Chem Clin Biochem, 18:78-88, 1980
- 2) NCCLS Approved Guideline (EP5-A): Evaluation of Precision Performance of Clinical Chemistry Devices. 1999
- 3) NCCLS Approved Guideline (EP9-A): Method Comparison and Bias Estimation Using Patient Samples. 1995
- 4) 日本臨床化学会分析部会近畿支部機器委員会: 自動分析装置性能評価法試案(Step 1, 1984-12-1). 生物試料分析, 7: 79-92, 1984
- 5) 日本臨床化学会分析部会関東支部機器委員会: 測定値評価の方法—濃度測定における精密さと正確さ—勧告法試案(Step 1, Draft 1, 1987-11-15). 臨床化学, 16(補冊): 287-298, 1987
- 6) 日本規格協会: 品質管理用語 Z 8101, 計測用語 Z 8103
- 7) 日本規格協会: 分析・試験の許容差通則 Z 8402
- 8) 細萱茂実, 芳賀利一: 測定値の評価の方法. 日本臨床化学会第9回夏期セミナー資料集, 10-16, 1989
- 9) 宮原洋一: 定量検査の精密さ・正確さ評価法—日臨技標準法試案(1995-9-1). 第12回臨床検査情報システム研修会テキスト, 29-51, 日本臨床衛生検査技師会情報システム研究班, 1995
- 10) 丹後俊郎: 測定誤差を考慮に入れた線形関係式; 測定法の比較のための統計学的方法. 臨床病理, 36: 1101-1108, 1988
- 11) 細萱茂実, 久米章司: 測定法の比較評価に用いられる直線関係式のMonte Carlo simulationによる検討. 衛生検査, 39: 828-834, 1990
- 12) 飯田暢子, 丹後俊郎: 精密度及び正確度評価法の提案. 臨床化学, 21: 280-281, 1992
- 13) 宮原洋一: Monte Carlo Simulationによる

- 精密度評価法の比較検討. 医学検査, 42 : 1824-1829, 1993
- 14) Barnett RN: Medical significance of laboratory results. *Am J Clin Path*, 50:671-676, 1968
- 15) Tonks D: A study of the accuracy and precision of clinical chemistry determinations in 170 Canadian laboratories. *Clin Chem*, 9:217-223, 1963
- 16) 北村元仕: 精度管理の方法と問題点—臨床化学検査. 医学のあゆみ, 59(別冊): 383-394, 1966
- 17) Cotlove E, et al: Biological and analytic components of variation in long term studies of serum constituents in normal subjects: III. Physiological and medical implications. *Clin Chem*, 16:1028-1032, 1970
- 18) Fraser CG, et al: Proposed quality specifications for the imprecision and inaccuracy of analytical systems for clinical chemistry. *Eur J Clin Chem Biochem*, 30:311-317, 1992
- 19) 飯塚儀明, 堀越晃: 臨床化学成分における個人の生理的変動幅. 医学検査, 41 : 1708-1714, 1992
- 20) 細萱茂実, 久米章司: 生理的変動幅に基づいた許容誤差限界. 臨床化学, 23(Suppl) : 118a-123a, 1994
- 21) 細萱茂実, 他: 施設間誤差の要因分析を目的としたコントロール・サーベイの方法論に関する検討. 臨床検査自動化学会誌. 19 : 739-743, 1994
- 22) 細萱茂実: 精度管理と基準値. 臨床検査法提要改訂第31版(金井正光編), p79-116, 金原出版, 東京, 1998
- 23) 細萱茂実, 他: 患者試料を用いた測定法の精密さと正確さに関する比較・評価法. 臨床検査自動化学会誌. 15 : 705-709, 1990
- 24) 宮原洋一: 実試料を用いる精密度評価法の検討. 医学検査, 41 : 1178-1183, 1992
- 25) 宮原洋一: 患者試料による精密性評価法について. 臨床化学, 22 : 133-135, 1993
- 26) 細萱茂実, 尾崎由基男: 精密さ・正確さ評価に関する統計的側面. 臨床化学, 28(Supp.1) : 23a-24a, 1999
- 27) Wackers PJM et al: Applications of statistics in clinical chemistry, a critical evaluation of regression line. *Clin Chem acta*, 64:173-184, 1975
- 28) Cornbleet PJ and Gochman N : Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem*, 25 : 432-438, 1979
- 29) Hathaway NR : Some comments on linear regression analysis by the method of W.E. Deming. *Clin Chem*, 26 : 1511, 1980
- 30) Parvin CA: A direct comparison of two slope-estimation techniques used in method-comparison studies. *Clin Chem*, 30 : 751-754, 1984
- 31) 宮原洋一: 精密性、正確性、相関性の統計学的評価法. 衛生検査, 39 : 817-822, 1990
- 32) 中 甫: 測定法の比較の仕方. *Med Tech*, 19 : 29-34, 1991
- 33) 大橋靖雄: 医学統計ソフトを使う前に第3回相関と回帰. 医療とコンピュータ, 2 : 78-86, 1989
- 34) Baudner S, et al: The certification of a matrix reference material for immunochemical measurement of 14 human serum protein-CRM 470. Community of Bureau of Reference, Commission of the European Communities, Brussels, 1-172, 1993
- 35) 小西貞則(村上征勝, 田村義保編): プートストラップ法による推定量の誤差評価. パソコンによるデータ解析, 123-142, 朝倉書店, 東京, 1992
- 36) 市原清志: 臨床化学検査の分析能の比較評価法. 臨床化学, 27 : 21-49, 1998
- 37) Passing H, Bablok W: A new biometrical procedure for testing the equality of measurement. *J Clin Chem Clin Biochem*, 21:709-720, 1983

(ワーキンググループメンバー)

- 飯塚 儀明 (筑波大学附属病院)
 雲類 鷺雄一 (水戸協同総合病院)
 大澤 進 (千葉大学附属病院) * 責任者
 木村 清隆 (秋田県立脳血管研究所)
 中 甫 (日本福祉総合医療研究所)
 細萱 茂実 (山梨医科大学附属病院)
 宮原 洋一 (鳥取大学医療技術短大)

(検討経緯)

- 第1回 平成6年12月16日
 第2回 平成7年2月10日
 第3回 平成7年3月25日
 第4回 平成7年9月1日
 第5回 平成7年10月20日
 第6回 平成7年11月22日
 第7回 平成8年3月8日
 第8回 平成9年2月21日
 第9回 平成9年2月22日
 第10回 平成9年3月28日
 第11回 平成11年7月24日

適用事例

1. 方法および試料

グルコース測定法(Glucose oxidase法)の精密さと正確さを本指針に従い評価する。なお、精密さの評価は、基準範囲(60~100mg/dl)濃度における許容誤差限界をSD=2.0mg/dl、CV=2.0%とする。また、正確さの評価は、1種類の濃度の血清標準物質(標準値100mg/dl)による評価、4種類の濃度の血清標準物質(標準値10、60、110、160mg/dlの4種類)による評価、および、比較対照法をHexokinase法として50例の患者試料を両法でそれぞれ2重測定した測定値を用いる評価の3種類を実施し、医学的意思決定濃度として140mg/dlを用いる。

2. 精密さの評価

2. 1 管理試料による日間精密度と日内精密度の評価

3種類の管理試料のそれぞれを、1日2回20日間測定した測定値を表5に示す。この間は、精度管理上特別な管理異常を認めず、測定法は安定した状態にあるとみなされた。

表5 3種類の管理試料の1日2回20日間の測定値

測定日	管理試料1の測定値		管理試料2の測定値		管理試料3の測定値	
1	41	41	91	90	143	141
2	40	41	93	96	144	141
3	39	42	92	89	144	142
4	41	40	91	92	142	143
5	41	42	92	92	146	143
6	41	41	93	91	141	140
7	40	41	95	94	142	138
8	40	39	93	91	142	145
9	41	40	90	91	143	140
10	40	41	90	89	145	143
11	41	40	94	94	142	140
12	40	40	92	90	143	142
13	40	40	90	89	143	145
14	40	39	89	89	146	143
15	39	40	91	91	142	142
16	39	40	92	94	141	143
17	42	42	91	92	145	147
18	40	40	93	93	141	141
19	40	40	92	92	140	142
20	41	40	92	92	141	138

管理試料2の分散分析表は表6のようになる。
これより、管理試料2の日間精密度、日内精密度、

表6 管理試料2の分散分析表

要因	平方和	自由度	分散	F ₀
日間	94.275	19	4.9618	4.84*
日内	20.500	20	1.0250	
計	114.775	39	F _{0.05} (19, 20)=2.14	

総合精密度またその変動係数は次のようになる。

日間精密度 SD_A=1.40、日内精密度 SD_E=1.01、
総合精密度 SD_S=1.73、変動係数(%) CV_S=1.9
この値を、グルコースの許容誤差限界と比較する

と2.0mg/dl以下であり、精密度は満足な状態にあるとみなせる。

なお、管理試料1と管理試料3の日間精密度、日内精密度、総合精密度またその変動係数は表7のようになる。

表7 管理試料1と管理試料3の日間・日内・総合精密度

試料	日間精密度	日内精密度	総合精密度	変動係数(%)
管理試料1	0.43	0.72	0.84	2.1
管理試料3	1.22	1.62	2.03	1.4

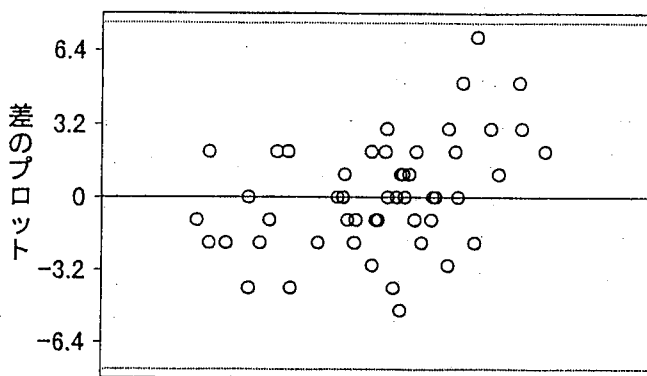


図5 50例の患者試料を用いた差のプロット図による濃度と精密さの関係の検討

2. 2 患者試料による精密さの評価

患者試料50例を被検法でランダムに2重測定し表8の測定値を得た。差のプロット図(図5)により濃度と精密さの関係を観察すると、ばらつきはほぼ一様とみなされる。そこで、このデータから

2重測定値の群内平方和、誤差分散、標準偏差のそれぞれを求める。

群内平方和 S_E=149.50、誤差分散 V_E=2.99、標準偏差 SD_E=1.73

この値を、グルコースの許容誤差限界と比較する

と2.0mg/dl以下であり、精密度は満足な状態にあるとみなせる。

なお、後に表8のデータを正確さの評価にも使

するため、50例の患者試料を被検法とともに比較対照法で2重測定した結果を同時に示す。

表8 患者試料50例の比較対照法と被検法のそれぞれによる2重測定値

患者試料	比較対照法 (x _i)		被 検 法 (y _i)		患者試料	比較対照法 (x _i)		被 検 法 (y _i)	
1	39	38	40	38	26	127	125	132	127
2	104	105	109	109	27	101	104	104	101
3	44	46	44	46	28	118	116	118	119
4	37	38	38	40	29	52	53	51	55
5	123	123	128	126	30	84	82	87	87
6	79	82	85	85	31	53	54	53	53
7	117	117	119	119	32	117	118	120	120
8	138	137	141	138	33	106	104	109	108
9	76	77	77	79	34	99	96	98	99
10	102	101	105	110	35	87	85	88	87
11	90	90	91	92	36	122	125	126	123
12	104	102	103	107	37	154	160	160	158
13	96	96	96	99	38	98	96	99	100
14	34	34	34	35	39	66	67	66	70
15	128	130	133	135	40	111	113	112	113
16	131	131	138	131	41	94	94	98	96
17	65	65	64	62	42	60	60	60	61
18	55	59	56	58	43	151	146	152	147
19	85	83	88	89	44	110	110	114	112
20	106	106	108	107	45	138	143	143	142
21	107	106	111	110	46	113	110	114	116
22	120	118	123	126	47	99	102	103	101
23	67	65	68	66	48	103	102	103	103
24	146	146	152	149	49	88	91	90	92
25	106	104	106	106	50	125	127	128	128

3. 正確さの評価

3. 1 1種類の血清標準物質による評価

標準値100mg/dlの血清標準物質を10回繰り返し測定した結果、以下の測定値が得られた。

[104 101 106 102 103 100 103 102 101 102]

10回繰り返し測定値の平均値は102.4mg/dl、標用

標準偏差は1.7mg/dlとなる。 $t_{n(n-1)} = 2.262$ より、平均値の95%信頼区間は [101.2~103.6] となり、この範囲に標準物質の標準値100mg/dlが含まれておらず、有意なかたよりが認められる。ただし、かたよりの大きさが $B=2.4$ であり、これは $100 \times 2.4 / 100 = 2.4(\%)$ より、5%以下という限界は満足する。

3. 2 4種類の血清標準物質による評価

4種類の血清標準物質を被検法で5重測定し表9の測定値が得られた。これより、標準値に対す

る測定値の関係を直線回帰式で表し、比例系統誤差と一定系統誤差の大きさを評価する。

表9 4種類の血清標準物質の被検法による5重測定値

試料	標準値	5重繰り返し測定値				
標準物質1	10 mg/dl	11	10	10	11	10
標準物質2	60 mg/dl	61	60	62	62	59
標準物質3	110 mg/dl	109	111	113	113	114
標準物質4	160 mg/dl	166	162	163	165	165

散布図を図6に示す。残差プロット図より極低濃度域ではばらつきが小さい傾向があるものの基準値以上の濃度ではほぼ同様と思われた。また、外れ値は認められなかった。直線回帰式 $Y = a + bx$ の回帰係数と残差標準偏差は次のようになった。

$$b = 1.025, a = -0.29, s_{y \cdot x} = 1.47$$

$b = 1$ および $a = 0$ の検定の結果、有意な比例系統誤差は認められないが、一定系統誤差を認めるとみなされた。そこで、医学的意思決定濃度140 mg/dlにおける予測値を求めると $Y_0 = 143.24$ となり、予測値のかたより3.24の140mg/dlに対する割合は2.3%であるから、これは5%以下という限界は満足する。

3. 3 患者試料による比較対照法との比較実験による評価

表8の50例の患者試料に対する比較対照法と被検法による2重測定値の1回目だけの測定値を用いて、本文6.3の方法(線形関係式による比例系統誤差と一定系統誤差の検定)に従い正確さを評価する。ただし、2つの測定法の誤差分散比(λ)は、上記の2.1患者試料の2重測定値による精密さの評価、の方法により求めた両測定法の精密度を用いる。

散布図を図7に示す。残差プロット図より、ばらつきはほぼ同様と認められ、外れ値は認められ

なかった。直線関係式 $Y = a' + b'x$ の傾き b' と切片 a' は、

比較対照法の誤差分散 $V_x = 2.37$ 、被検法の誤差分散 $V_y = 2.99$ 、誤差分散比 $\lambda = 1.262$ より、つぎのようになった。

$$\text{傾き } b' = 1.043, \text{ 切片 } a' = -1.97$$

次に、試行回数を500回としたブートストラップ法により b' と a' の信頼区間を求めると、 b' [1.030~1.059]、 a' [-3.37~-0.71]となり、有意な比例系統誤差と一定系統誤差の両者を認めた。そこで、医学的意思決定濃度140mg/dlにおける予測値を求めると $Y_0 = 144.02$ となり、予測値のかたより4.02の140mg/dlに対する割合は2.9%であるから、これは5%以下という限界は満足する。

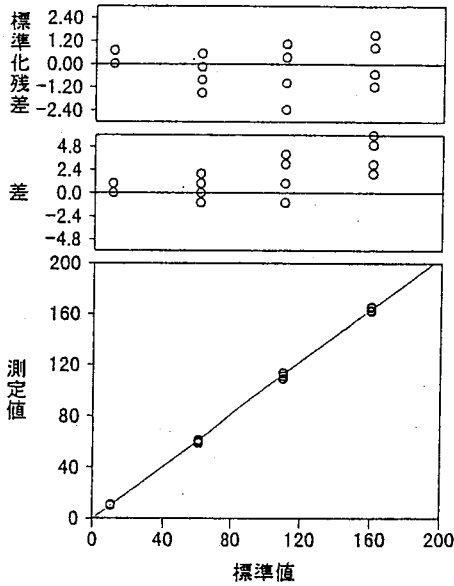


図6 4種類の濃度の血清標準物質を用いた正確さの評価

試料数	p	4
データ数	n	20
残差標準偏差	$S_{v,x}$	1.4748
直線性からのズレの検討	F_0	0.1037
ズレは認められない		
系統誤差の検定 直線回帰式 $y=a+bx$		
傾き	係数 b	1.0252
	t_b	0.4865
	比例系統誤差は認められない	
切片	係数 a	-0.2920
	t_a	4.2718
	一定系統誤差が認められる	
医学的意志決定濃度による評価		
意志決定濃度	X_0	140.00
予測値	Y_0	143.24
かたより	B	3.2360
(基準値 5.0%)	B%	2.31
正確である		

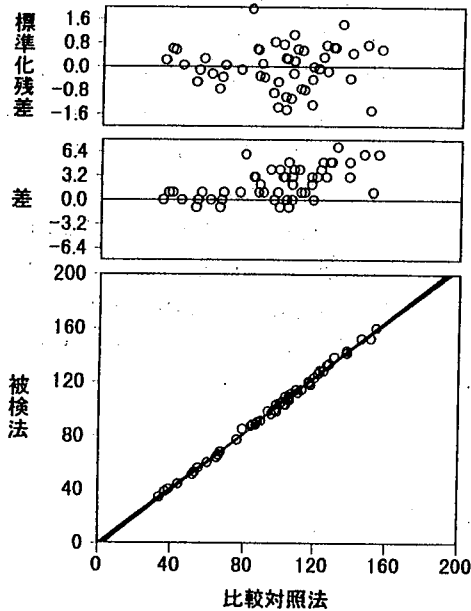


図7 比較対照法との比較実験による正確さの評価

	比較対照法	被検法
平均値	97.500	99.700
標準偏差	30.2960	31.5939
データ数		50
相関係数		0.9984
誤差分散比		1.2616
系統誤差の検定 直線関係式 $y=a'+b'x$		
傾き	係数 b'	1.0427
	標準誤差	0.0073
	信頼区間	1.030 ~ 1.059
	比例系統誤差が認められる	
切片	係数 a'	-1.9652
	標準誤差	0.6761
	信頼区間	-3.369 ~ -0.708
	一定系統誤差が認められる	
医学的意志決定濃度による評価		
意志決定濃度	X_0	140.00
予測値	Y_0	144.02
かたより	B	4.0156
(基準値 5.0%)	B%	2.87
正確である		

第1版へのコメント および改訂版への修正点

定量検査の精密さ・正確さ評価法指針(第1版 J AMT-CEP1-97)を「医学検査 46:1130-1142,1997」に公表し、内容に関する意見を広く求めた。また、1997年から1998年にかけて、国内の関連団体および学会に評価法指針(第1版)の内容を照会し、コメントおよび内容承認に関する意見を求めた。その結果、日本臨床化学会、日本臨床病理学会、日本臨床検査自動化学会、日本分析機器工業会の各団体・学会、および、多くの個人・団体から貴重なコメントをいただいた。また、社団法人日本小児科学会、社団法人日本肝臓学会、社団法人日本糖尿病学会、財団法人予防医学事業中央会、財団法人化学検査協会の各団体・学会からは内容承認をいただき、日本臨床化学会、日本臨床検査自動化学会からは条件付き承認をいただいた。なお、日本医科器械学会、社団法人日本腎臓学会、社団法人日本血液学会、社団法人日本分析機器工業会の各団体・学会からは、コメントを控える、または、日本臨床検査標準協議会(JCCLS)での検討を求めるとの回答をいただいた。

以下に、それらコメントの中で主要なものについて要約を示す。また、それらコメントを検討し、さらに、評価法指針(第1版)公表後にワーキンググループ内で検討を重ねた結果、下記に示す主要な点を含めて内容に修正を加え、本評価法指針(改訂版 JAMT-CEP2-99)とした。

1. 全体的事項

1) タイトルには「臨床化学定量検査の…」と明記し、対象を「臨床化学検査と免疫検査の大半」などと規定する必要がある。この基準は血液・凝固検査では使用できない。

・本評価法指針が対象とする測定法について解説に説明を記述したが、より明確にするためタイトルに「臨床化学」を明記した。

2) 実際に利用する際の助けとなるよう、例題を付記したほうがよい。

・この点については、第1版検討時から必要性が議論されており、改訂版に追加した。

3) ガイドラインの構成が分かりやすくなるよう、見出しに工夫が必要である。

・見出しの分類番号を簡素化し、全体の構成を分かりやすくした。

4) 一般向けに、出来得る限り平易な用語と表現を用いたほうがよい。

・極力平易な表現を心がけたが、統計学的記述に配慮する必要もあり、従来の経緯や背景の補足について必要な部分は「解説」で説明を加えた。

5) JIS Z8402「分析・試験の許容差通則」付属書6、ASTM 6-178-94(外れ値の扱い)を参考に再検討するとともに、直線性などを含めた検討が必要である。

・ご指摘の資料については第1版作成段階で参考とし検討した。臨床検査の領域に直接応用するには困難と思われる側面もあり、その基本的な考え方は本指針に反映させている。また、精密さ・正確さ評価に当たり、直線性以外にも妨害物質の影響や最低検出限界など考慮すべき特性があるが、解説でふれたようにそれら特性を含めると内容が複雑となり、別の評価事項として分けて扱うこととした。

6) 引用文献は国際的に評価されている学術雑誌からの引用が適当である。

・改訂版(特に本文)では極力国際的な学術誌からの引用に努めた。

2. 適用範囲について

1) 「外部精度評価における測定法評価にも適用できる」、また「従来法と新法の比較評価にも適用できる」の2点は、「精密さ・正確さ評価」の観点からは混乱を招き問題である。

・混乱を避けるため、「外部精度評価における測定法評価にも適用できる」、また「従来法と新法の比較評価にも適用できる」の2点を削除し、日常検査法の精密さ・正確さ評価を適用範囲とした。

2) 臨床化学検査分野での標準物質が存在し、標準測定法のある項目に関して限定したほうがよい。

・正確さの評価について、ご指摘の点は「1. 適用範囲」と「3. 評価法の概要」の中で明記した。

3. 用語について

1) 「かたよりの小さい程度」の表現は文法的に正しくない。

・用語「正確さ」の定義は、JIS Z8103 計測用語をそのまま引用した。

2) 「トレーサブル」の意味と使用法が不明である。

・「正確さが適正に伝達された」の表現に変更した。

3) 用語では"precision, accuracy"とされているが、Summaryでは"imprecision, inaccuracy"と記載されており、一致する必要がある。

・Summaryの用語を"precision, accuracy"に変更した。

4) 標準値はassigned valueの表現でよいのか。

・日本臨床化学会「標準に関する用語(Ver.2.4)」に従い、"reference value"に変更した。

5) 生理的個体内変動の日本語表現は、個体内生理的変動のほうがよくはないか。

・最近の文献では"within-subject biological variation"の表現が多くなっており、これに合わせて日本語も「個体内生理的変動」に変更した。

4. 試料について

1) 管理試料について「患者検体と同様な反応性を有すること」の判断手法と基準が不明確である。また、「溶血、混濁、高ビリルビン血清などを除外し」の表現について、具体的判定法と基準が曖昧である。

・両者とも、NCCLSのガイドライン“Evaluation of Matrix Effects (EP14-P)”に従うべく、文献として引用した。

2) 評価に用いる患者検体を50例とする根拠が典拠を含めて不明確である。また、「基準範囲が低(または、中)濃度域の場合」について、具体的な解説が必要である。

・評価に用いる患者検体の必要最低数については、

NCCLSのガイドラインで40例の提示があるが、今回新たにコンピュータ・シミュレーション実験により検討した結果、50例以上でないといふ比例および一定系統誤差の推定精度が劣る結果が得られた。また、「基準範囲が低(または、中)濃度域の場合」の具体例を表の補足説明に加えた。

3) 評価に用いる試料を「評価期間中安定な条件で密栓凍結保存する」および「評価試験中安定に保管する」の表現がおかしい。

・具体的な表現形式に変更した。

4) 「血清性状が異常でない患者試料」とは具体的にどのようなものか。

・表現が曖昧であったため、削除した。

5. 精密さの評価について

1) 管理試料による日間・日内精密度の評価において、反復測定日数の20日は多いように思われる。また、分散分析による精密度の推定は、反復数を任意としたほうが有用である。

・反復測定日数については、NCCLSのガイドラインにおいても20日間の指定があり、統計学的な検討からも一定の信頼性のもとに標準偏差の推定を行うには20日程度が必要である。また、反復数を任意とすることによって応用範囲は広がるが、評価手順が複雑となるため、本指針では一定数とした。

2) (5)式の「基準範囲の平均値」は表現が不明確である。

・「健常者基準値の平均値」の表現に変更した。

3) 表4について、測定法間差を考慮すると測定原理を明記したほうがよい。また、SI単位を併記するほうがよい。

・表4は解説に記述したようにあくまでも暫定的な参考データである。ただし、現状において測定法間差は無視できず、逆に一方では測定原理を記述することが必ずしも参考にならない状況があるため、表4の引用文献における基準範囲を併記することとした。また、SI単位については省略させていただいた。

4) 精密さの許容限界を個体内生理的変動の標準

偏差の1/2とする根拠について、個体内生理的変動を求める母集団の定義、また、略語を s_p とする理由なども不明である。

・引用文献と解説を参照いただきたい。また、略語 s_p はCotloveの文献より引用したものであるが、標準偏差の表現として他の箇所はすべてSDを用いているため、個体内生理的変動の標準偏差として新たに SD_p の表現に変更して用いることにした。

5) 許容誤差限界に対する従来の考え方で、Tonksの提唱した基準は必ずしも生理的変動を前提としたものではなく、これを分けて4つに分類したほうがよい。

・指摘に従い修正した。

6) 患者検体による精密さ評価で、2重測定の間にはキャリブレーションを実施する必然性が不明である。また、精密さのプロファイルに関する表現が曖昧である。

・第1版では、2重測定値に日間誤差を反映させたほうがよいとの判断で間にキャリブレーションを実施するよう指定したが、1回のキャリブレーションで日間誤差が十分に反映されない危険性もあり、改訂版ではキャリブレーション実施の記述をなくした。また、精密さのプロファイルに関する表現を修正した。

7) 精密さが許容できないとき正確さの評価を中止し精密さを改善するとなっているが、この内容は評価法指針の範囲を超えるものである。

・「精密さが許容できないとき正確さの評価を中止し精密さを改善する」の記述を削除した。

6. 正確さの評価について

1) 患者試料による評価において、直線関係式を求める際に2つの測定法の誤差分散比(λ)を算出せずに $\lambda = 1$ と仮定することには問題がある。また、標準主軸回帰法やPassing-Bablok法などを用いる方法がよい。

・指摘内容を解説中に追記した。

2) 正確さの評価結果判断の表現について、「かたより(または、比例系統誤差、一定系統誤差)はないと判断する」を「有意なかたより(または、

比例系統誤差、一定系統誤差)は認めないとみなす」としたほうがよい。

・指摘内容に従い修正した。

3) 標準物質を「10回以上繰り返し測定」とあるが、10回の根拠はなにか、また、機器計測の場合1回目の測定値は安定しないことがあり、そのような場合への対処法が必要である。

・平均値の推定精度は、10回の繰り返し測定値を用いることで統計学的に安定することを確認した。また、測定法の安定状態は機器や試薬の組み合わせでさまざまであり一概に指定できないため、「測定法が安定な状態にあるときに」という記述に追加修正した。

4) 1種類の血清標準物質のみによる評価であっても、必ずしも対象濃度が標準値付近だけとはいえないのではないかと。

・解説の中で記したように、1種類の血清標準物質のみによる評価は、測定法の直線性が満足されかつ一定系統誤差がないという前提がなければ、必ずしも測定範囲全域の正確さは保証できない。

4) 「血清標準物質は4濃度以上を用いる」について、高価で少量しかいない標準物質を濃度調製する際に希釈誤差が大きく影響されるが、この点に関する記述がない。

・多濃度の血清標準物質を希釈調製して用いる際の注意点を解説の中に追記した。

5) 電解質の2%という許容誤差限界が適用される項目の範囲が曖昧であり、カリウムでは厳し過ぎ、また、電解質以外の成分について一律に5%とせず、成分(クレアチニン、総ビリルビン、グルコースなど)によっては異なるレベルが妥当な場合がある。また、これらの根拠が示されていない。

・2%という許容誤差限界の適用を「電解質」から「血清Na,Cl測定法」に変更した。また、一般的な成分についての5%の限界値はあくまでも経験的な基準である。個々の成分のより適切な限界値についてさまざまな設定が考えられるが、本指針では一般的な限界値としての5%のみを示した。

6) 医学的意思決定濃度について、具体的な数値

か参照すべき文献を示す必要がある。

・解説の中で文献を引用した。

7) 50例以上の患者試料を毎日5～10例ずつ、5日間以上測定とすると、10日以上かかることもあり、本法が精密性評価に適しているのか出典も含めて不明である。

・日間誤差がない測定法は1日で全試料を測定すればよいが、日間誤差がある場合は、たまたまある日の測定値が高い（または、低い）と正確さの評価にその誤差がそのまま反映されてしまう。これら日間誤差の影響を避けるために複数日測定を指定している。