

866 機関の 9.1%に相当した。しかしそのうち 5 年間で 1 項目のみが 60 点未満であった機関数は 56 機関で全体の 6.5%で最も多い。一方、外部機関では過去 5 年間に 60 点未満の成績が 1 項目以上存在した機関数は 20 機関で自機関より明らかに少ないが、これら外部機関は複数の機関から受託している例があり、60 点未満の成績が影響を与えた機関数から計算すると合計 58 機関となり、全体としては過去 5 年間の機関数総計 765 機関の 7.6%であり自機関測定に比較して僅かに少ない結果となっている。しかし外部機関は健診機関の事業の一部として測定している自機関とは異なり、測定を専門的事業としているにもかかわらず 7.6%も存在することは、受託機関にそれぞれ多数の健診データが含まれていることを考慮すると自機関以上に重大な問題であると考えられた。60 点未満の成績を示した項目についてみると自機関では 5 年間に最大 3 項目で、4 項目以上は見られなかったのに対し、外部機関では 1 項目のみ；10 機関、2 項目：5 機関、3 項目：3 機関、5 項目：2 機関が 60 点未満であり、5 項目が 2 機関も存在していた。しかし、大きな影響を与えた外部機関 3 箇所を除き、その他の機関は複数項目であっても影響を与えている機関が 1 機関でありしかも同一機関であることは、問題がありながらも健診機関への影響は最小限に抑えられている

と考えることができた。精度管理調査の成績が 60 点未満を示す理由として第一に測定誤差の存在が考えられるが、その内、比例系統誤差は最も起こりやすい誤差の一つで、今回の調査でも自機関、外部機関ともに最も多く見られる誤差であった。一定系統誤差はごくまれに見られる誤差であるが、今回の調査では、後に述べる特定の項目と外部機関においてそれを示唆するデータが存在した。その他、通常の測定誤差とは異なるが、測定後のデータを転記する際の転記ミスや測定の際の試料の取り違えなども少なくなかった。また、測定法の性能に関連する誤差である直線性の欠如が示唆される問題も存在した。最後に調査試料の送付に関連する問題も残っていた。測定誤差はこれら種々の原因により生じるが、各機関内における内部精度管理の普及と実践により測定の精密さすなわち偶然誤差はきわめて小さい範囲内に管理・維持できるようになり、測定値の施設間差は、確認ミスのような偶発的な誤差を除けば、測定値の正確さに関連するかたよりが誤差原因の主流を占めるようになっている。調査結果からもかたよりの中でも比例系統誤差が大部分を占め、その原因が主として測定の標準となる検量用試料の正確さを含むキャリブレーション時に生じる誤差であることが判明している。これらのかたよりの原因となる比例系統誤差を縮小する

には、正確さが明確な高位標準物質を日常一般法で繰り返し測定を行い、その平均値とこれら高位標準物質の標準値の一致性を確認するトレーサビリティの確認を実施し、もしかたよりが存在したときにはそれを校正する作業を行うことにより達成できる。しかし、現状では高位標準物質を用いてトレーサビリティの確認を実施している機関は著しく少なく、全く実施していない機関が70%も存在することは憂うべき状況であると考えられた。この状況は日本医師会が年1回実施している臨床検査精度管理調査の結果においても同様である。測定値の施設間差の原因の大部分が校正可能な比例系統誤差であることは、一方では測定値の信頼性の保証は容易な状態にあることを意味している。現在、定量検査が行われている全ての検査項目について高位標準物質の開発が実現しているわけではないが、比較的件数の多い日常的な検査項目については既にこれらの高位標準物質が存在するので、これらの項目については各機関において少なくとも年1、2回の頻度でトレーサビリティの確認と校正を実施すべきであると考えられた（櫻井、中）。

健康診断等のあり方について産業医及び臨床検査研究者との面接により意見を聴取した結果は主観的、定性的なものであり、特に専属産業医の面接者が限られていることによる

バイアスは十分考慮しなければならない。しかし、全国的かつ国際的企業の産業医であるのでその視点は将来を洞察するにふさわしく、示唆に富む討議であった。健診の有効性や有用性の評価指標として、死亡率や寿命がどの程度妥当であるか、他の指標、例えば有所見に対する保健・生活指導など早期対応による就労期間と退職後のQOLの向上も指標になり得ないか、など評価軸に関する提案と疫学研究の必要性が指摘された。労働安全衛生の他のインフラとして、作業環境管理、作業管理、労働安全衛生マネジメントシステム、健康づくり・1次予防、などがあるが、これらは作業環境管理を除いて未整備の状態で実効性が担保されていないという指摘もあった（徳永）。

全衛連の現行の労働衛生検査項目に関する外部精度評価については、その回数、評価方法、評価項目数、検体数について、現行の方法を支持する機関が非常に多数を占めたが、これは、現行の方法により参加機関において一定の満足が得られていることを意味しており、やはり全衛連において長年検討され改良されてきたものを超える方法を見出すのは相当難しいことがわかった。しかしながら、今回提案された簡略法については、賛成、反対の両意見が各10数%ずつあり、多様な意見が見られたが、外部精度評価の回数を増やし、それぞれについては簡略化も導入す

るという方向性については、概ね4割の機関で支持された。今後この試案をもとに、改良された方法を、より明確な形で提示することができれば、現行法に加えることにより改善が可能と考えられた。情報交換については、多くの機関で懸念されており、早期に解決すべき問題であると考えられた。内部精度管理については、8~9割の機関で適切に行われていることがわかったが、数%から1割程度の機関では、そうでない可能性があることが推察された。一方、外部精度評価において、若干の問題を含む機関も数%から1割程度存在する。今回のアンケート調査では、内部精度管理と外部精度評価で問題となる機関が一致しているかどうかについては解析できなかったが、両者が共通している可能性はあるので、外部精度評価の結果が悪い機関については、内部精度管理の状況を詳しく調べる意義もあると考えられた(栗原)。

胸部CT検診では大量の受診者のCT画像読影が対象となり、異常所見の有無判定に基づく要精検または精検不要の判定が読影医の主な仕事である。連続確信度で回答する習慣はない。CT検診像の読影診断に不慣れた若手医師が、日ごろ見慣れていなかった画質のCT画像を読影するにあたり、異常所見の拾い上げと鑑別診断の2段階の過程が要求される。異常所見の拾い上げについてはほぼ

100%の感度が要求され、質的診断については高い感度と、ほぼ100%の特異度が望ましい。このような目標に向かって如何に効率的に学習成果をあげるかが重要であるため、その評価法を検討した。今回の検討では、さらに、CT画像をモニター診断時の読影所要時間も測定して作業能率や診断確信度の関係もみたが、さらなる検討のうえの結論づけが必要と思われた(曾根)。

Matched Pairによる経時的評価モデルの適用に関しては、ROC曲線から判断すると、sBP, dBP, TCの有効性が高く、AUCの値とケース群とコントロール群の値の変動から判断すると、sBP, dBP, TC, TG, FBS, UAの有効性が高い。既に、我々は健診データの時系列的な解析により、ホメオスタティック(恒常的)な項目は時系列的な解析に適しているという結果を得ている。一方、上記の項目の中でTG及びUAはホメオスタティックな項目とは異なり、通常はデータの変動が大きく時系列的な解析には適さない項目であると考えていた。しかし、それらの動態は病的で高値のレベルを含めた全体のデータ解析の結果であり、今回のようにデータが基準範囲に近いレベルを推移している場合、それらの項目に関しても、ホメオスターシス機能が維持され、ホメオスタティックな項目と同様に経時的評価を適用することが可能となったと考えられた。ま

た、本来、FBSはホメオスタティックな項目であると考えられるが、ROC曲線の年次の順序性が不規則であることから、経時的評価に適さないという結果が得られた。これは、今回使用した健診データの精度の信頼性に問題があったのではないかと思われた。次に、既に得ている知見をもとに、時系列的なデータ解析に適している項目のsBP, dBP, FBS, TCを健康変動batteryとし、データの変動性が高く、ホメオスタティックな項目ではないUA, GPT(肝炎なし), TG, γ -GTP, GOTを病態変動batteryとしてグループ化した。健康変動batteryに属する項目は、時系列的なデータ解析により、異常値を予見できる可能性が高い。一方、病態変動batteryの項目に関しては、異常値を予見することは難しく、発症前の介入時期を特定することが難しいと考えられた。判定結果の言語的な表現については、従来の一時点のデータから判定する方法では、悪化しても改善しても同じ判定分類で説明されることが多いのに対し、時系列的なデータ解析結果を利用すれば、「データの状態を正確に評価できる」、「今後のアクションを的確に示せる」という2つの要素が得られる可能性が指摘された(吉田)。

定期健康診断等の結果、脳・心臓疾患を発症する危険性が高いと判断された労働者に対して、脳血管及び心臓の状態を把握し、二次健康診断

及び脳・心臓疾患の発症の予防を図るための労働者災害補償保険による二次健康診断等給付制度があるが、一次健康診断結果が給付要件を満たさない場合は検査が行われない。虚血性心疾患の診断には運動負荷心電図検査が有用であるが、職域健診の場で全例に運動負荷心電図検査を施行することは困難であり、かつ効率が悪い。そこで、循環器疾患スクリーニングを目的として作成した問診票を虚血性心疾患が疑われ、冠動脈造影検査を行った症例に施行し、検査結果と問診表の回答とを照らし合わせvalidation studyを行い循環器疾患の効率的なスクリーニングとための職域健診のあり方について基礎的検討を行った。症例数の制限から今回は虚血性心疾患に関する検討にとどまった。偽陰性率が高かったが、問診票による各質問の回答に対する虚血性心疾患検出の感度は60~80%、特異度は12.5~100%、正確度は37.5~62.8%、陽性反応的中度は81.5~86.7%であった。今後は、失神、動悸に関する問診回答と心臓電気生理学的検査、Head up tilt試験などの循環器疾患精密検査との対比を含め、症例数を増やした検討が必要であり、さらに問診票の精度の向上(感度を保ち、偽陰性率を低くする)のために問診結果のスコア化を試みるなどの課題が残ると考えられた(久代)。

E. 結論

全国労働衛生団体連合会（全衛連）臨床検査精度管理調査に過去 5 年間連続して参加した全機関を対象として全衛連参考値検討委員会において真値とみなし得る目標値を設定している総コレステロール、中性脂肪、AST、ALT、 γ -GT、血糖の 6 項目について、60 点未満である項目が 1 回でも存在する機関のデータを抽出し、誤差を生じる原因について調査した結果では、5 年間に 60 点未満の成績が 1 項目以上存在した機関数は、自機関で総機関数の 9.1%、外部機関で 7.6%であったが、外部機関の 7.6%は複数の健診機関に影響を与えている例があるので、その影響は重大であることがわかった。最も頻度の高い比例系統誤差は、容易に校正可能な誤差であることから、年 1、2 回はトレーサビリティの確認と校正を行うべきであり、以上を実践することにより臨床検査の信頼性向上が果たせるとともに個人の検査データについて信頼性の高い経年的管理が可能になると考えられた（櫻井、中）。

全衛連臨床検査精度管理調査に参加した健診機関 310 機関のうち健診実績が年間 20 万件以上の大規模検査機関 66 機関の大半は、ほぼ適正な人的資源を有し、かつ臨床検査の精度（正確さ）の水準も高く、相当高度な予防医学活動を展開している機関であると考えられた。46 都道府県で健診実績が最も多い各 1 機関の検

査項目別精度管理の成績は、全衛連調査の全機関の傾向とほぼ同じ水準にあり、自機関測定機関は外部依託機関に比して 70 点未満の低評価の割合がやや高率であった。健康診断及び臨床外部検査精度管理のあり方について、産業医及び臨床検査の研究者と討議して健康診断の目的・役割、法規則の課題、企業責任と産業医責任、労働衛生活動の変容と課題、臨床検査の精度管理の課題等に関して有意義な意見と提言が得られた（徳永）。

全衛連による労働衛生検査項目についての外部精度評価に参加している全検査機関の精度管理担当者に対して、郵送法によるアンケート調査を行った結果、現行年 1 回の評価回数、4 段階の評価指標、8 項目の評価項目、6 検体の検体数については、それぞれ現行の制度を支持する機関が圧倒的に多く見られたが、アンケートにおいて、外部精度評価の回数増加と簡略化という提言を示し、意見を求めたところ、明確に反対とする意見も 15%ある一方で、全体の約 4 割からその基本的な方針に一定の支持が得られた。結論として、外部精度評価に関して次のような案が提言された。①現在の全衛連の外部精度評価を、情報交換がより起こりにくい形で年 1 回行う。②8 項目、共通 3 サンプルについて、適、不適のみの判定の精度評価を 1 に示した年 1 回の調査の 4 ヶ月後および 8 ヶ月

後に年2回、計3回行う。③1、2の精度評価は同一評価団体、あるいは連合体で、すべてワンセットとして行う。④評価については、A、Bの評価よりも、「適」をどれだけ続けてとるかを重要視するように宣伝する。⑤「不適」となった機関には、通常の原因究明および指導に加えて、内部精度管理の方法、検体の採取・保管方法についてもより徹底的に指導する(栗原)。

CT装置を用いる肺がん検診での医師の診断能力の評価法を検討し、それを基に学習効果を調査した結果では、肺がん、肺がん以外の病変、正常のCT像に対する正答率は、研修開始時にそれぞれ、45.8%、24.3%、95.7%であったが、研修後には、75.6%、60%、91.3%と上昇した。存在診断については学習効果が得られやすかったが、質的診断については学習効果に個人差があり、低X線曝射CT像による質的診断は困難であることがわかった。CT画像読影による存在診断と質的診断の学習および研修の効果判定面での有用性が示唆された(曾根)。

収縮期血圧(sBP)、拡張期血圧(dBP)、総コレステロール(TC)、中性脂肪(TG)、GOT(AST)、GPT(ALT)、 γ -GTP、FBS、UAの9項目を対象として、matched case-control studyを行い、項目毎の評価モデルの有用性を検討した結果、健診データの経年的評価モデルにより、デー

タの動態が的確に把握できるため、悪化傾向に向かうようなグループに対しては、異常性(悪化)の予見が可能となり、早期発見、早期介入が可能となることがわかった。また、MA(移動平均)、SD(標準偏差)、Slope(傾き)の指標を用いた判定結果について個人に分かりやすい言語的表現を提案した。これにより、今後、時系列的なデータ解析アプローチを発展させることにより、高精度なEBMに基づく保健指導の介入指標を開発できる可能性が示唆された(吉田)。

循環器疾患スクリーニングを目的とした問診票を用いた問診結果と冠動脈造影検査結果とを照合することにより循環器疾患スクリーニングにおける問診票のvalidation studyを行ったところ、職域健診受診者のリスク層別化と高リスク受診者の抽出が効率的に行える可能性が示唆された(久代)。

F. 健康危険情報

特になし。

G. 研究発表

特になし。

H. 知的財産権の出願・登録状況

特になし。

資料 4

職域健康診断及び地域医療における臨床検査のあり方と精度管理

中 甫 日本福祉総合医療研究所 所長

研究要旨

臨床検査における定量検査の測定値には必ず誤差が含まれる。この誤差は測定の精密さに関連する偶然誤差と正確さに関連する系統誤差に大別されるが、この両者を誤差許容範囲内に長期にわたって管理しなければ測定値の信頼性は保証できない。測定値の精密さの保証と管理は比較的容易で、長年にわたる施設内における内部精度管理法の普及と実施により、近年著しく向上している。一方、測定値の正確さの保証と管理は容易ではない。近年、臨床化学検査の正確さを基本とした測定体系が確立されて正確さの概念が明確になったが、各施設における測定値の正確さが十分保証されているとはいえない現状にある。全国の臨床検査室における測定値の客観的な正確さの指標となる外部精度評価法の一つである精度管理調査の結果をみても同様のことが言える。そこで、わが国で標準化されている偶然誤差（精密さ）と系統誤差（正確さ）の確認法および職域における健康診断を実施している施設が参加している主な精度管理調査の現状と問題点について調査を行い系統的、総括的にまとめた。

A. はじめに

工業生産部門における品質管理法 (quality control) が臨床検査に導入された 1950 年¹⁾ 以来、臨床検査の精度管理は検査結果の信頼性を管理し保証する手法として欠かすことのできないものとなっている。測定結果自体の信頼性を中心に考えられた精度管理も、現在ではユーザーが満足できる検査情報を提供するための

組織全体の取り組みとしての品質マネジメント (quality management) へと発展している。そのような視点に立って職場健康診断と地域医療に関連する臨床検査と精度管理の現状とあり方についてわが国における活動を中心に述べる。

B. 臨床検査の精度管理

1. 内部精度管理とその有用性

内部精度管理(Internal Quality Control, IQC)は同一検査室内のデータを用いて精度管理を行う手法と定義され、代表的な手法に管理試料を被検試料と同時に測定することにより管理を行う $\bar{x}-R$ 管理図法がある。わが国では 1963 年に既にコントロール血清を用いた臨床化学検査の管理図法による成績管理の報告がある²⁾。その後 40 年を経過し、その間に管理試料を用いる種々の管理法が開発されたが、 $\bar{x}-R$ 管理図法は現在でも基本的な管理法として広く用いられている。これらの管理試料を用いる方法は測定値の精密さの管理法として有用性が高いが、管理試料は一般に正確な表示値を有しないことから測定値の正確さの管理はできない。一方、管理試料を用いないで患者データを用いる管理法が開発され、代表的な方法として Hoffmann & Waid による患者データの分布を基本としたナンバー・プラス法³⁾、正常値平均法⁴⁾がある。これらの方法は個々の被検者の測定結果の信頼性を管理することはできないが、同時測定した被検者群の測定値全体にかたよりが生じたかどうかを管理する手法で、長期のデータの安定性の管理に有用である。現在はコンピュータによるデータ処理が行われていることから、1 週間または 1 カ月単位で全被検者の分布図を作成し、それを積み重ねることによ

りデータの安定性の長期管理を行っている検査室が多い。とくに職域健康診断検査を実施している機関においては、被検者数が多いことから 1 日単位で分布図の作成が可能であり、有用性が極めて高いと考えられる。

2. 外部精度評価とその有用性

外部精度評価(External Quality Assessment, EQA)は複数の検査室のデータを用いて精度管理を行う手法と定義され、代表的な手法に同一試料を多数の検査室に送付して測定させ、統計処理を行った後に評価を行う精度管理調査がある。当初は外部精度管理(External Quality control, EQC)と呼ばれていたが、現在は国際的に EQA と呼ばれている。わが国における精度管理調査は、全国規模で実施されている精度管理調査から地域的に実施されている精度管理調査まで多数に及んでいる。精度管理調査の一般的な目的を表 1 に示した。

この目的は参加施設が調査結果から得られた現行の技術水準、送付試料に対する合意値を通して自施設の実行能力を判断できるような基本的データを調査実施母体が提供すること、ならびに実行能力向上を目指した教育を通して参加者を支援するという参加者中心の考え方である。したがって検査実施機関の認定(accreditation)に関連して実施されている技能試験(proficiency

testing)とは区別して取り扱われる。わが国では検査実施機関の認証に関連する技能試験はまだ実施されていない。精度管理調査参加施設が内部精度管理により自施設内の測定値の精密さを十分に管理しているという前提に立てば、自施設における外部精度評価の結果は測定値のかたよりの有無の指標となり、合意値と自施設の測定値の関係から実行能力をの評価ができるとともに、自施設の測定値の客観的な正確さの評価が可能となる。このように外部精度評価は内部精度管理が前提にあることから、内部精度管理で精密さの管理ができていない施設の参加は、外部精度評価の信頼性を低下させる原因に繋がる。以上の問題を含むことから、外部精度評価は測定値の厳密な正確さの評価はできない。

1) わが国における精度管理調査

わが国で継続的に全国規模で実施されている主な臨床検査精度管理調査実施母体として日本医師会（以下日医）、日本臨床衛生検査技師会（以下日臨技）、日本衛生検査所協会（以下日衛協）があげられる。日医の日本医師会臨床検査精度管理調査（以下日医調査）は1967年以来年1回実施されているが、現在第36回（平成14年度）を数えており参加施設数は2800施設となっている。日臨技の日臨技臨床検査精度管理調査（以下日臨技調査）は日本医師会と

同程度に古くから年1回実施されているが、徐々に充実拡大しており平成14年度の参加施設数は2700施設となっている。日衛協の日本衛生検査所協会精度管理調査（以下日衛協調査）は1975年から年1回実施されており、現在28回（平成14年度）を数えており参加施設数は300施設となっている。日衛協調査は参加施設数が少ないが、日衛協における検査実施総件数はわが国における検査実施総件数の約半数を占めていることから大規模調査の範疇に数えられる。

予防医学に関連する全国規模で実施されている主な臨床検査精度管理調査実施母体として全国労働衛生団体連合会（以下全衛連）、予防医学事業中央会（以下中央会）、日本総合健診医学会（以下総合健医）があげられる。全衛連では前述の日医調査に参加することを義務付けているが、全衛連独自の全衛連臨床検査精度管理調査（以下全衛連調査）は1992年以来年1回実施されているが、現在第10回（平成13年度）を数えており参加施設数は325施設となっている。全衛連調査は参加施設が労働安全衛生法に基づく職場健診を実施している施設であり、先の日衛協調査と後の中央会調査の参加施設も含まれており、各施設の実施総件数は一般医療機関に比較して極めて多いことがあげられる。中央会の予防医

学事業中央会精度管理調査（以下中央会調査）は1988年以来年1回実施されているが、現在第15回（平成14年度）を数えており参加施設数は36支部(39施設)となっている。中央会の参加施設は学校検診、職場健診を含む予防医学検査を対象としており、支部（および施設）数は少ないが、日衛協、全衛連と同様に各支部（および施設）で実施している検査実施総件数は極めて多いことがあげられる。総合健医の精度管理調査（以下総合健医調査）は総合健医に所属する人間ドック実施機関を対象とした限定された調査で年4回実施されているが、今回の報告では参考程度に止める。

その他地域的な精度管理調査として、地方自治体、都道府県医師会、都道府県技師会またはこれらの合同組織が実施している精度管理調査が多数存在する。これらの地域的な精度管理調査は、全国規模の調査では実施不可能な詳細な調査または新鮮試料を用いる調査などが行われている。

2) 精度管理調査の標準化と統合化

(1) 精度管理調査の標準化

精度管理調査はそれぞれが独立した企画、方法がとられているため、相互の精度管理調査の結果を比較し、本来の目的である誤差要因を明確にしてその誤差を取り除き測定値の施設間差を誤差許容範囲内に縮小する

ことを困難にしている。菅野らは厚生省科学研究費補助金による医療技術評価総合研究事業としての「外部精度管理調査の標準化に関する研究」を行い、2000年1月に総括研究報告書を発行している⁵⁾。ここでは1998、1999年の2年間にわたる調査、解析から精度管理調査項目を測定体系の概念を基本として4群（基幹検査項目第一類、第二類、第三類、および特殊項目）に分類し、この中で第一類、第二類は真値(true value)またはそれに近い値を目標値(target value)として設定して精度管理を共通実施することが可能であること示した。また共通のコード体系が利用できるよう互換性のあるコード体系を示した。さらに目標値の設定の仕方、共通利用可能な評価基準となる統計量を提示して各調査母体が活用するよう提案した。以上により精度管理調査の標準化が可能であることを提案した。日本臨床検査標準協議会(JCCLS)は外部精度評価(EQA)専門委員会（委員長：河野均也）を組織し、わが国で実施されている外部精度管理調査を標準化して共通の理解のもとに相互に調査データを活用できる基盤を整備することを目的として上記「外部精度管理調査の標準化に関する研究」を基本骨格として実験的根拠から修正を加えて2000年12月「外部精度評価(EQA)標準化のためのガイドライ

ン」を JCCLS 指針提案として発表した⁶⁾。以上の研究と作業により精度管理調査の標準化の基盤が確立した。

(2) 精度管理調査の統合化

現在わが国で実施されている精度管理調査は、総合健医調査を除き、各実施母体において年1回実施されているのが現状である。しかし年1回の調査では調査の域を越えず、少なくとも同一プログラムで年数回の調査を実施しなければ継続的な性能の評価と維持が困難である。そこで現在精度管理調査を実施している調査実施母体を統合できないかについて議論されてきた。

日医では1999年「外部精度管理調査の標準化に関するプロジェクト委員会(委員長:河合 忠)」を組織し、わが国における臨床検査の外部精度管理調査のあり方について立案することにより臨床検査の標準化を推進するための方策の答申を求めた。その中で検討事項として①実施母体とその組織・財源、②調査項目と実施方法、③調査試料の調製と配布、④評価基準と評価方法、⑤臨床検査の標準化に対する近未来的環境整備の必要性をあげて検討を行い、2000年3月に報告を行った⁷⁾。同委員会は2000年度も継続して具体的な検討を行い2001年3月に報告を行った⁸⁾。その中で1. 事業の名称は共通外部精度評価事業(National

External Quality Assessment Scheme: NEQAS)とすること、2. 共通外部精度評価事業の意義、3. 事業のサービス対象を最終の受益者は国民であるが、直接の対象者はヒトの生体あるいは検体を用いた臨床検査を日常業務として実施している機関(臨床検査室)である病院・診療所・療養型施設等に付設する検査室、医師会立臨床検査センター、登録衛生検査所、健診センターとすること、4. 事業内容として①外部精度管理調査、②臨床検査室の認定・認証作業、コンサルテーションとすること、⑤事業組織として臨床検査精度保証会議(仮称)を日医内に置き、この事業の企画、立案、診査を行うが、そのメンバーとして日本医師会、日本臨床検査医学会、日本臨床衛生検査技師会、日本衛生検査所協会等とすること、⑥実施機関は日本臨床検査精度評価機構とすることが具体的に盛り込まれた。さらにこのような事業を実現するために、トータル・クオリティ・マネジメント(Total Quality Management, TQM)などの手法を確立することが不可欠なことから、研究、ノウハウの確立、事業化、将来展望などが必要であることが記載された。統合化される精度管理調査は日医調査、日臨技調査、日衛協調査の3調査で2004年度からの発足が予定されている。この標準化と統合化は将来の

臨床検査室の認定を視野に置いたものであり、ここに含まれない実施母体を実施する精度管理調査は、ここで確立した標準化手法に強制されることはない。しかし、可能な限りその概念および手法を活用することにより調査結果の相互比較が容易になることはいうまでもない。

3) 外部精度評価における調査項目の現状とあり方

表2に臨床検査精度管理調査を実施している上記6組織が実施している調査項目を示した。ただし特定の実施母体のみで実施されている微生物検査、病理検査、細胞診検査は表から除いた。また調査項目名に続く()内は精度管理調査の標準化の項で述べた基幹検査項目第一類、第二類、第三類の項目であることを意味する。表から明らかなように、臨床検査実施施設の全てを対象としている日医調査、日臨技調査、日衛協調査をみると基幹検査項目第一類は日衛協調査でCKの調査を実施していないことおよび日医調査でアルブミンの調査を実施していないことを除き、これら3実施母体の全てが調査対象項目としている。基幹調査項目第一類について予防医学検査を対象としている全衛連調査、中央会調査、総合健医調査も含めてみるとブドウ糖、AST、ALT、 γ -GT、総コレステロール、中性脂肪、ヘモグロビンの7項目が表にあげた全調査で

調査対象項目として実施されていることが分かる。基幹検査項目第二類は、HDL コレステロール、HbA_{1c}が日臨技調査、総合健医調査で実施されていないこと、総蛋白が日医調査、全衛連調査で実施されていないこと、赤血球数が日臨技調査で実施されていないこと、白血球数、血小板数が中央会調査で実施されていないこと、ヘマトクリットが日臨技調査、中央会調査で実施されていないことを除き、多くの実施母体で調査対象項目としてあげられている。予防医学検査の中でとくに重要な尿蛋白、尿糖半定量検査は、日臨技調査と全衛連調査のみが調査対象項目としてあげている。また基幹検査項目第一類である尿素窒素、尿酸、クレアチニンは、表中全衛連調査のみが実施していないが、これは労働安全衛生法に基づく検査項目として指定されていないことによると考えられる。以上をまとめると、基幹検査項目第一類および第二類は、医療機関検査室および予防医学検査実施施設を含めた全ての臨床検査室に対する調査項目として実施することが望ましい調査項目といえることができる。

4) 外部精度評価における調査試料の現状とあり方

(1) 調査試料の種類

外部精度評価に用いる調査試料は、日常検査の対象となる新鮮な被検者の試料に限りなく近似した性質を有

する試料が望ましい。精度管理調査開始当初は内部精度管理調査用コントロール血清が調査試料として広く用いられた。それらのコントロール血清はとくに酵素、脂質などヒト血清とは異なる動物由来の成分が添加されているものが多く、施設内の測定値の精密さを管理する内部精度管理用試料としては十分使用できたが、ヒト血清と反応態度が異なり、測定結果が測定法により異なり日常検査の性能が直接反映されないことが多かった。そのことから測定値のかたよりの有無を主として評価する外部精度評価では不適切であることが指摘されてきた。その後酵素についてはヒト培養細胞由来酵素が添加されるようになりヒト血清に急速に近づいたが、一部の酵素でヒト血清と異なる性質を示した。1996年日本臨床化学会学術連絡委員会が“常用酵素標準物質の規格”を委員会報告として発表し⁹⁾、その中で酵素の起源としてヒト体液、ヒト組織、ヒト組織培養系、ヒト細胞培養系、およびヒト酵素をコードする遺伝子を導入した組替え体(recombinant)とするとした。この規格に基づいて1997年11月にヒト赤血球、ヒト組替え体を酵素の起源とした常用酵素標準物質(CERM)が完成しその頒布が開始された。その後CERMと同じ起源の酵素が調査用試料に添加されるようになり、わが国では酵素項目におけ

る調査用試料の問題がほぼ解消された。しかし、調査試料は調製試料であることから試料マトリックスが通常の被検者試料と異なり、一部の項目または測定法でマトリックス効果による測定値のかたよりが生じることが問題として残されている。とくにマトリックスの影響を受ける測定法群としてドライケミストリー・システムがあげられる。全衛連調査では調査対象項目が限定されていることから以上のような調製試料は使用せず、予めプールした検査済み高濃度(または高酵素活性)ヒト血清と検査済み正常ヒト血清を段階的に混和して8濃度の試料を作製して調査試料としている。

ヘモグロビン、血球計数を調査対象とする血液検査用試料は、試料の安定性の問題から大規模の外部精度評価プログラムではヒト全血を調査試料として用いることはできない。したがってヒト半固定細胞や動物由来の細胞を含む試料が用いられるが、計数装置の原理により機器間差が生じて日常検査の性能を直接反映できない問題がある。この問題を解消するために日衛協調査および地域レベルの精度管理調査のような比較的小規模な調査では、新鮮なヒト全血試料を短時間内に参加施設に送付する工夫をしてこれらの問題を解決している。

(2) 調査試料の形状

調査試料は同一試料を多数の施設に送付することと、遠距離施設も存在することから試料の安定性が重要なポイントとなる。安定性に関して通常取られる手法は調査試料を凍結乾燥品として供給することである。しかし、凍結乾燥によりリポ蛋白の変性が生じることから、とくに沈殿操作を行わず直接測定を行う HDL コレステロール測定試薬において測定値の試薬間差が生じて新鮮なヒト血清による性能を直接反映しないことがある。以上のことから、日医調査では総コレステロール、中性脂肪、HDL コレステロールなどの脂質測定用試料は液状試料（-20℃の凍結試料）を調査試料として送付している。全衛連調査では 1992 年の第 1 回調査から生化学検査項目について凍結血清を凍結状態で送付している。

(3) 調査試料の本数

調査試料は一般に 2 濃度以上の複数試料が送付されることが多い。2 濃度の場合は正常濃度および異常濃度、3 濃度以上の場合は低、中、高濃度の組み合わせ試料が送付される。外部精度評価では日常測定シリーズ中での 1 回測定が原則であり同一試料の繰り返し測定は行わないことから、基本的には測定の精密さの直接評価はできない。しかし、一般に複数濃度の試料を送付することから、試料間の測定値比または散布図の作成により間接的に施設内測定値の精

密さの評価が可能である。さらに 3 濃度以上の試料を用いることにより測定値の試料濃度に対する直線性の評価が可能になる。しかし、各施設に同一試料が送付されることから、参加施設間での測定値のカンニングが生じる可能性があり外部精度評価プログラムの問題点の一つになっている。

全衛連調査では、(1) の調査試料の種類で述べた 8 試料を最低濃度または次に低濃度の試料のいずれかおよび最高濃度または次に高濃度の試料のいずれかが必ず送付されるという条件設定を行い他の濃度試料をコンピューターによりランダム化して合計 5 試料を参加施設に送付している。したがって測定値のカンニングができないプログラムになっており、さらに直線性の評価および誤差の分類評価が容易になるという長所がある。

(4) 調査の頻度

大規模な精度管理調査では年 1 回実施されることが多い。その場合は実施時点での技術水準の提示および評価に限定され、生じている測定値の施設間差の厳密な原因究明は行いがたい。測定誤差の客観的な評価のためには少なくとも年 3~4 回は実施する必要がある。先述の外部精度評価プログラムの統合により、将来この条件を満たすことが可能になるであろう。とくに検査室の認定に関

連する技能試験による評価は継続的な性能の評価を欠かすことができない。

C. 測定誤差とその管理

測定法、測定試薬および測定機器を用いて測定された測定値には必ず誤差が含まれるが、許容できる最小限の誤差を長期間一定に維持することが測定値の信頼性を保証する上で重要である。測定誤差は測定値のばらつきに起因する偶然誤差と測定値のかたよりに起因する系統誤差に大別される。偶然誤差は突き止められない原因により生じる誤差で測定値の精密さを表し、各施設内で実施される管理図法を用いる内部精度管理法で容易に管理できる。系統誤差は何らかの原因で測定値にかたよりを与えることにより生じる誤差で測定値の正確さに関連するが、管理図を用いる内部精度管理法のみでは管理できない。測定値の正確さは、理論的には正確さの根拠が明確な基準分析法で測定した測定値または基準分析法を用いて標準値を求めた標準物質と比較することにより確認でき、また校正できる。一方、外部精度評価スキーム(EQAS)の結果は、参加各機関が施設内の測定値の精密さを十分管理している場合には、測定値のかたよりにより生じる系統誤差の客観的な評価法として有用性が高いことが知られている。しかし、測定

値のかたよりの原因となる因子は複数存在することから、関連する各因子に分類してその原因の究明と是正を行う必要がある。

1) 偶然誤差とその管理

偶然誤差は測定のばらつきとなって現れる誤差で、そのばらつきの小さい程度を精密さといい、ばらつきを標準偏差またはその指定した倍数で表した値を精密度と呼んでいる¹⁰⁾。1994年日臨技では定量検査の精密さ・正確さ評価法標準化ワーキンググループを組織して検討を開始し、1997年“定量検査の精密さ・正確さ評価法指針”を発表した¹¹⁾。その後1999年にその改訂版を日本臨床検査標準協議会(JCCLS)から発行している¹²⁾。この評価指針によると精密さの評価法として、管理試料(コントロール血清)を用いる方法と患者試料を用いる方法をあげている。基本的には日常検査における同一試料の反復測定により得られた標準偏差を評価の物差しとしての個体内生理的変動の標準偏差と比較することにより評価する方法である。得られた評価が個体内生理的変動の1/2以下であれば精密度は許容できると判断し、その精密度を内部精度管理法としての管理図法(例えば $\bar{x}-R$ 管理図法)で管理する。偶然誤差の指標である精密度の評価方法の概略を下記に示した。

(1) 管理試料による評価¹²⁾

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、日常検査の中にランダムに挿入して測定し、これを 20 日以上繰り返す。その結果から分散分析法により日間精密度(SD_A)、日内精密度(SD_E)、総合精密度(SD_S)を計算し、総平均値が基準範囲上限以下のときは SD_S を個体内生理的変動の標準偏差($SD_w/2$)と比較し、上限以上のときは変動係数(CV_S)で評価するとしている。

(2) 患者試料による評価^{1 2)}

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 回目の測定を実施し、2 回目の測定を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て同日内に測定する。それぞれの患者の 2 回測定値の平均値を横軸にとり、1 回目と 2 回目の差を縦軸にとって差のプロット図を作成する。分散分析法により群内平方和(SE)、誤差分散(VE)、標準偏差(SD_E)を求め、この SD_E を $SD_w/2$ と比較して評価するとしている。

2) 系統誤差と測定体系

系統誤差は測定結果にかたよりを与える原因によって生じる誤差で、そのかたよりの小さい程度を正確さといい、推定したかたよりの限界の値で表した値を正確度と呼んでいる¹⁰⁾。具体的にはかたよりは測定値または推定量の分布の中心である平均

値と真の値(true value)との差で、濃度または平均値の真値に対する比率(%)で表現する。ここでいう真値とは測定量の正しい値と定義されるが、先述のごとく誤差を含まない測定値は存在しないことから、観念的な値であって実際には求められないことから、真値とみなし得る値が真値として用いられる¹⁰⁾。真値とみなし得る値については化学分析における正確さを基本とした測定体系の考え方が参考となる。

(1) 化学分析における正確さを基本とした測定体系

1977 年米国 NIST (National Institute of Standards and Technology, 当時は NBS) の Uriano および Gravatt は、化学分析における正確さを基本とした測定体系を発表した¹³⁾。その後米国 NCCLS ではこの考え方を臨床検査に導入し一連のガイドラインを発表している^{14~16)}。わが国でも日本臨床化学会を中心にこれらの測定体系の考え方を導入している¹⁷⁾。図 1 に分析化学における正確さを基本とした測定体系を臨床化学検査の測定体系として示した。以下に図中の各部分について解説する。

① 基本的測定単位: 1960 年国際度量衡総会において単位の統一化を図るために国際単位系(SI 単位)を決議し、国際標準化機構(ISO), 国際純正応用化学連合

(IUPAC)などの協力を得て普及の努力が続けられている。1977年WHOは第30回世界保健会議で医学領域へのSI単位系導入を推奨した。1994年日本臨床化学会学術連絡委員会では遅ればせながらSI単位換算表を作成し¹⁸⁾、臨床検査関連学会に提示して普及の努力を始めた。SI単位は長さ、質量、時間、電流、熱力学的温度、光度、物質の量の7種類の基本的単位を設定している。この中で分析化学に直接関連する単位は質量と物質の量であるが、物質の質量を天秤で直接測定できる場合は最も誤差の少ない結果が得られる。しかし臨床化学検査における微量の成分分析は直接測定はできないことから何らかの分析法により測定せざるを得ない。SI単位で直接測定したときのかたよりは約10⁻⁸%と考えられている¹⁹⁾。

- ② (絶対)基準法：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と定義される¹⁷⁾。この方法は同位体希釈・質量分析法(ID-MS)に代表される測定法で²⁰⁾、特別な技術、費用、測定装置などを必要とすることから認定を受けた基準測定検査室(reference measurement laboratory)において実施できる

測定法である。この測定法で測定したときのかたよりは約0.1~1%と考えられている¹⁹⁾。この方法は後に述べる実用基準法の開発および評価、一次標準物質の値付けに用いられる。

- ③ 一次標準物質：標準物質の特性値が、(絶対)基準法によって決定されたものと定義される¹⁷⁾。一次標準物質は実用基準法の開発と評価、正確度の高い二次標準物質の作製、厳密な精度保証の目的で使用される。
- ④ 実用基準法：測定体系上(絶対)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した正確さと精密さをもった値が得られる測定法と定義される¹⁷⁾。国際学会または国内学会において勧告される基準分析法がこの方法に相当する。この方法は特別な装置を必要とせず、熟練した技術を有すれば実施できる測定法である。この測定法で測定したときのかたよりは約1~3%であると考えられている¹⁹⁾。この方法は後に述べる日常一般法の開発および評価、日常に使用する検量用試料の作製、内部精度管理、外部精度評価における測定の正確度の評価に使用される。
- ⑤ 二次標準物質：標準物質の測定値

が、実用基準法によって決定されたものと定義される¹⁷⁾。二次標準物質は実用基準法と同様に日常一般法の開発および評価、日常に使用する検量用試料の作製、内部精度管理、外部精度管理における測定の正確度の評価に使用される。

- ⑥ 日常一般法：多数の測定試料に応じることができ、日常的に用いる精密さを有する測定法と定義される¹⁷⁾。日常検査室で一般的にルーチン法として採用されている方法がこれに当たる。これらの測定法で測定したときのかたよりは5～10%であると考えられている¹⁹⁾。
- ⑦ 血清ベース検量用試料：日常一般法の検量用試料としては目的成分の純品を秤量して精製水などの溶媒に溶解することにより作製する溶媒ベースの検量用試料とヒト血清（あるいはアルブミン）をベースとして必要に応じて目的成分を添加した血清ベースの検量用試料がある。血清ベースの検量用試料は上記基準分析法または標準物質による値付けの必要がある。現在日常一般法では自動分析装置を利用して分析を行うことが一般的であるが、自動分析装置では被検試料と検量用試料を同一試料採取ノズルで採取することから、溶媒ベースの検

量用試料では比重および粘性が異なることから、採取量に差が生じて系統誤差の原因となる。したがって血清ベース検量用試料が主として用いられるが、一方測定により値付けを行うことから値付けにおける測定誤差がかたよりの原因となる可能性がある。

- ⑧ トレーサビリティ：測定体系の上でより高い正確さに下位から次々と合わせることができると定義される¹⁷⁾。例えば日常一般法による測定値の正確さを実用基準法を用いて確認したり、一次あるいは二次標準物質を日常一般法を用いて繰り返し測定することにより確認することをトレーサビリティの確認と呼ぶ。
- ⑨ 伝達性：測定体系の上で高位の正確さを順次下位のものに合わせしていくことと定義される¹⁷⁾。例えば一次あるいは二次標準物質の標準値を血清ベース検量用試料に伝達するときなどに用いられる。
- ⑩ 校正：標準物質あるいは基準分析法による測定値を用いてかたよりを補正することと定義される¹⁷⁾。例えばトレーサビリティの確認により標準物質の標準値と日常一般法の間にかたよりがみられたとき、そのかたよりを補正することを校正するという。

(2) 酵素活性検査の測定体系

酵素活性の測定は、基質の種類および濃度、緩衝液の種類および濃度、pH、測定温度などの条件を至適化した一定条件における生物学的活性を測定していることから（絶対）基準法は存在しない。図2に日本臨床化学学会による測定体系を示した。日本臨床化学学会では、現在AST, ALT, CK, ALP, LD, γ -GTの6種類のヒト血清中酵素活性測定の勧告法を公表しているが、それらの酵素についてこの測定体系が適用される。以下に図中の各部分について解説する。

- ① JSCC 勧告法：この方法は臨床化学検査の測定体系（図1）における実用基準法に相当する。この方法は、pH、基質濃度はでき得る限り最適とし、条件を明記する、温度も必ず記載し、でき得れば30℃が望ましいという1964年の国際生化学連合(IUB)の勧告に従って測定条件の設定を行い、測定温度は30℃とした。この測定法を日本臨床化学学会(JSCC)勧告法と呼んでいる。
- ② JSCC 常用基準法：先のJSCC 勧告法は測定温度を30℃として設定されているが、世界的に臨床検査領域における日常検査の酵素活性の測定は37℃で実施されていることから、JSCC 勧告法の条件を測定温度のみ37℃とする方法をJSCC 常用基準法と呼び、この方法で得られた活性値を日常

検査法の基準とした。したがって、これらの酵素について日常検査法の正確さを確認（トレーサビリティの確認）するためにはこの方法と比較することになる。

- ③ JCCLS 認証常用酵素標準物質 (CERM)：1996年日本臨床化学学会学術連絡委員会は、常用酵素標準物質の規格を委員会報告として発表し⁹⁾、1997年その規格に基づいて作製した酵素標準物質についてJSCC 常用基準法により値付けを行いJCCLS 認証委員会により認証された“JCCLS 認証酵素標準物質”を日本臨床検査薬協会を通して供給を開始した。現在第3ロットが供給されている。
- ④ 日常検査法(JCCLS 標準化対応法)：JSCC 実用基準法で測定した値が、CERMや検量用ERMなどを利用することにより伝達されている場合にその測定法をJSCC 標準化対応法と呼ぶ。このJSCC 標準化対応法は基本的には用手法であるが、JSCC 常用基準法に準拠した試薬を使用し自動分析装置で実測Kファクターを求め、温度を37℃に管理して測定されていれば適用される。またCERMを用いてトレーサビリティが確認されている測定法および企業から供給されるCERMの標準値が伝達されている検量

用ERMを用いて検量を行った測定法も全てJSCC標準化対応法となる。すなわちJSCC実用基準法およびCERMの値を伝達することにより測定値の標準化を図ったと考えてよい。

- ⑤ 検量用ERM：これはJSCC標準化対応法の検量目的のため、企業がCERMの標準値を伝達した検量用酵素標準物質をいう。したがってこのERMの信頼性の責任はそれを提供した企業側に存在する。

3) 測定の正確さの評価

臨床化学検査における測定の正確さの基準は、先の臨床化学検査の測定体系における(絶対)基準法、実用基準法のような基準分析法により得られる測定値および一次標準物質、二次標準物質の標準値である。そのような基準分析法による測定値および標準物質の標準値を基準とした測定の正確さ評価法の概略について以下に示した。

- (1) 1種類の濃度の血清標準物質による評価¹²⁾

正確さの明確な血清標準物質を添付されている認証書の指示に従って調製し、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。その結果から平均値、標準偏差を求め、平均値の95%信頼区間を計算する。血清標準物質の標準値(X_r)、平均値(\bar{X}_B)からかたより(B)を計算す

る($B = \bar{X}_B - X_r$)。計算された平均値の95%信頼区間に X_r が含まれていれば、平均値の付近の濃度領域における有意なかたよりはないとみなす。信頼区間から外れた場合は $100 \times |B| / X_r$ が5%以下(N_a, K は2%以下)かどうかで正確さを判断する。

- (2) 3種類以上の濃度の血清標準物質による評価¹²⁾

3種類以上の濃度の血清標準物質の使用が可能なときは、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($Y = a + bx$)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)の大きさを評価できる。それぞれ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各標準血清を5回以上繰り返し測定する。その結果から直線回帰式を計算し、切片 a および勾配 b のかたよりの有無の有意検定を行う。切片、勾配の両方に統計的にかたよりがないとみられたときは、一定系統誤差、比例系統誤差を認めないと判断する。もし一方にかたよりが認められたときは、評価対象となる濃度〔医学的意思決定濃度(x_0)〕における予測値($Y_0 = a + bx_0$)のかたより B の x_0 に対する割合($|B| / x_0 = |Y_0 - x_0| / x_0$)を求め、この値が5%以下(N_a, K は2%以下)