

臨床検査情報を活用した各種業務支援・診断支援手法の
開発とその有用性評価

(Development and validation of intelligent support systems
for management and clinical interpretation of laboratory tests)

平成 27 年 3 月

山下 哲平

山口大学大学院理工学研究科

論文要旨

臨床検査の役割は、生体に起こった変調を客観的に捉え、その情報を病気の診断、進行度・治療効果の判定、病気の予後予測に役立てることにある。一般に、臨床検査法の性能評価では、臨床検査が病態以外で変化する要因（年齢、性、体動、体位など）を明らかにし（**臨床検査の変動要因分析**）、臨床検査が病態の把握にどの程度有効かを評価すること（**臨床検査の臨床的有用性の評価**）である。

これらの性能評価は、直接的な調査や実験によって実施しうるが、十分なデータ数（症例数）が必要で、多くの時間と経費を要する。そこで、日常検査情報が大規模に蓄積される臨床検査情報システム（LIS: laboratory information system）の情報を利用して、それらの分析・評価を行う試みが、これまで広くなされている。しかし、実際にはLISのデータは、多種多様な病態に由来する雑多な検査値の集合であるため、的確な前処理を施し、目的とする分析に利用しうるデータを効率よく取り出す技術が重要となる。

私の研究では、この展望のもと、これまで実用化されていない、(1)日常検査で問題となる検体取り違いの検出法の開発、(2)男女の判別における臨床検査値の有用性の評価の2つのテーマを設定し、それぞれ、新しいアプローチによりその解決を試みた。

第1の研究テーマである**検体取り違いの検出法の開発**では、検査値の前回値との偏差を目安に検体の取り違いを検出することになるが、実際には、検査項目毎に単位や分布が異なることが問題となる。そこで、まず**検査項目毎に測定値をべき乗変換方式で正規化**し、しかも健常者の値の範囲（基準範囲）が、 ± 2 の範囲となるように標準化することで、全ての検査値を統一スケールで表示可能とした。その後、前回値との偏差（D）の絶対値を個体内変動幅の逆数を重みとして平均し、それを取り違いの指数（wCDI: weighted cumulative delta check index）とした。ここで、個体内変動幅の大きさは、Dの分布から反復切断補正法により求めた。

wCDIによる取り違い検出力の評価は、LIS上でランダムに取り違い例を発生させた場合（混同群）のwCDI値と、実際に前回値の存在する場合（正常群）のwCDI値の比較から、感度・正診率・ROC分析を用いて行った。その結果、判定用検査項目数6以上で利用でき、10項目以上あれば検出感度が95%以上となり、十分実用に耐える性能が得られた。

第2の研究テーマである、**日常検査情報からの男女判別**の試みは、検体取り違いシステムの補助手段としての利用を目的に着手した。まず、多重ロジスティック分析（MLRA: Multiple Logistic Regression Analysis）により、最適な判別が得られる検査の組み合わせを求めたところ、男女差の大きい主要4項目の検査値が存在すれば、ほぼ的確に男女判別ができることが明らかとなった。しかし、MLRAでは欠損値への対応に問題があるため、より柔軟な方法として、市原による加重対数尤度診断法（WALL: Weighted Average Log Likelihood）の応用を試みた。これは、男性の検査値の分布を確率密度関数とみなし、検査値 x が分布の中心に近いほど高い尤度 $L(x)$ を与え、任意の症例の対数尤度 $\log(L(x))$ を、分散分析で求まる男女差指数を重みとして平均する。男性では平均値は大きな値となるが、女性では明瞭に小さな値となる。MLRAと、WALLの男女判別性能を比較したところ、ほぼ同等の正診率となった。

研究1で開発した、検体の取り違い検出システムは、すでに実用になっているが、その副産物として、異常値を多く含む LIS の検査値でも、分布両端の極端値の除外によりどの検査項目も正規分布に変換できたという事実は非常に重要である。なぜなら、正規変換後、基準範囲によって標準化することで、全ての検査項目を計測単位によらず統一表示できるためである。現在この技術の臨床応用を検討している。さらにその研究から臨床検査の誤差許容限界の目安となる検査値の生理的変動幅を、同じ患者の2回分の検査値の偏差Dの分布から算出できた点も重要で、今後各病院で独自にその幅を算出して精度管理業務に利用できることになる。

研究2では、男女の判別を基本的な検査の組み合わせで行えることを示したが、臨床検査の診断的有用性の評価では、これまで、ROC (Receiver operating characteristic) 分析や MLRA が一般的で、特に複数検査の組合せ診断では、MLRA が必須とされてきた。しかし今回の研究で WALL を用いた男女判別から、MLRA のそれと同等の性能が得られたことの意義は大きい。なぜなら WALL は、分布型に依存せず尤度を計算でき、任意の検査を柔軟かつ動的に組合せて、その対数尤度平均から疾患群との類似度を算定できるからである。さらに、複数の疾患群に対して、それぞれに WALL を計算することで、どの疾患に最も類似しているかを判定できるため、病気の自動診断システムにそのまま応用できることを示している。

本研究では、日常検査情報を対象として様々な情報技術、統計手法を駆使して、そこから検査業務支援、診療支援につながる情報・知見の獲得を目指した。その結果、一定の成果が見られたという事実は、今後の臨床検査情報の利用を推進する上で重要と考えている。

目次

第 1 章 緒論	1
1-1 本論文の背景.....	1
1-2 研究の経緯.....	1
1-3 本論文の新規性.....	2
1-4 本論文の構成と概要.....	2
第 2 章 導入	4
第 3 章 検体取り違いを高率に検出する加重累積デルタチェック法の考案とその性能評価 ... 17	17
3-1 はじめに.....	17
3-2 手法.....	18
3-2-1 加重累積デルタチェック法(wCDC)の数理.....	18
3-2-2 検証手法.....	22
3-3 結果.....	25
3-3-1 調整 Box-Cox 変換の検証.....	25
3-3-2 個体内変動 SD.....	26
3-3-3 wCDC による取り違い検出性能.....	28
3-4 考察.....	30
第 4 章 基本検査による男女判別の正確度：多重ロジスティック回帰分析法と加重対数尤度診断法の比較検討	36
4-1 はじめに.....	36
4-2 方法.....	36
4-2-1 解析対象.....	36
4-2-2 男女差指数.....	37
4-2-3 多重ロジスティック回帰分析法.....	37
4-2-4 加重対数尤度診断法.....	37
4-2-5 判別性能の比較評価法.....	40
4-3 結果.....	40
4-3-1 男女差指数.....	40
4-3-2 多重ロジスティック回帰分析法による判別性能評価.....	40
4-3-3 加重対数尤度診断法による判別性能評価.....	42
4-4 考察.....	43
4-5 結語.....	45
第 5 章 結論と今後の展開	46
謝辞.....	50

第1章 緒論

1-1 本論文の背景

医療では、臨床検査などの体の変調をとらえる情報に基づき、投薬や輸血などの治療の必要性を判断する。このとき他の患者の検査結果が間違っ報告されてしまうと、治療すべき患者を見過ごす、必要のない投薬を行うなどの重大な医療事故が発生してしまう。日本医療機能評価機構の平成18年度報告では、2003年から2006年までの間に、臨床検査に関連した取り違いとして、患者取り違いが91件、検体取り違いが101件報告されている¹⁾。

このような患者（検体）取り違いを防止するために、様々な方法が用いられ、実際に取り違いの件数は減少してきている。例えば、採血などの検体取得時に起こる患者取り違いを防ぐ方法としては、患者のリストバンドの確認や、患者自身に名乗ってもらい確認するといった従来の方法に加えて、リストバンドにつけられた患者固有のバーコードと採血管のバーコードを、採血室に設置した読み取り機や、PDA(Personal Digital Assistant)を使って確認し、一致しているか認証するシステムなどがある²⁾。しかしながら、導入コストの問題があるため全ての病院でこのようなシステムを導入できるわけではない。

一方、採血後に取り違いを検出する方法として、前回の検査値と今回の検査値を比較し、個体内変動幅を超えた変化をしたものを検体取り違いとして検出する方法が用いられてきた^{3, 4, 5)}。これらの従来の手法では単純な前回値との差を用いているため、判定に利用可能な検査項目が限定されたり、複数項目の組み合わせが固定されたりするといった問題がある。

1-2 研究の経緯

本論文では、このような現状から、日常の臨床検査データを用いて検体取り違いを検出することを目的として2つのアプローチを試みた。

まず、複数の検査項目の前回と今回の検査値の差を用いて、検体取り違いを検出する手法の開発を行った(第3章)。これを実現するために、前回値との偏差が分布型に影響されること、検査項目によって単位や変動域が異なるため、偏差を組み合わせた判定が難しいという課題があった。そこで全ての検査値の分布型を正規分布に変換し、各項目の基準範囲をもちいて標準化することで、分布型や単位の違いのない標準化値として統一して取り扱う方法を考案した。また、検査項目を組み合わせる取り違いを検出する際の、各検査項目の重要度の目安として、個体内変動幅をデータマイニング技術により取り出し利用することとした。これらの新しく考案した一連のデータ解析方法により、日常検査情報であっても高い精度での検体取り違い検出が可能となり、また様々な検査項目の組み合わせに柔軟に対応可能であることが明らかとなった。これらの手法については、山口大学および株式会社エイアンドティーと共同で特許出願を行っている。(特開2012-256196)

続いて、前回値を必要としない取り違い検出手法として、検査結果から性別を判定する方法を検討した(第4章)。臨床検査分野では、複数の検査値から疾患などの有無を判定する手法として多重ロジスティック回帰分析が一般に用いられる。そのため、まず多重ロジスティック回帰分析

を用いて検査値からの男女判別精度を確認した。しかしながら、多重ロジスティック回帰分析では、男女判別に有効な検査項目でも、互いの相関が強い場合には調整されてしまうため、同時に利用できないという問題が明らかとなった。そのため、異なる手法として、市原の開発した対数尤度診断法を応用し、男女差指数を組み合わせた男女判別の性能評価を行った。その結果として、多重ロジスティック回帰分析法と比較して同程度の男女判別性能があることが明らかとなった。

1-3 本論文の新規性

本論文では、検体取り違いの検出を目的として、前回検査値との比較と男女判別という 2 つの研究を行い、それぞれ効果的な手法が開発できたことについて述べている。

第 3 章 加重累積デルタチェック法

- ・検体取り違い検出に、すべての検査項目の結果を利用しうることを示した
- また、その副産物として以下の 2 点がある。
- ・分布型や極端値の問題により実現できていなかった日常検査値の統一標準化表示を、測定値分布の正規化と標準化の技術を組み合わせて成功した
 - ・日常検査情報から個体内変動幅が算出可能であることを示した

第 4 章 男女判別

- ・検体取り違い検出の補助手段として高い精度で検査値から男女の判別が可能であることを示した
- ・「加重対数尤度診断法：WALL 法」が、現在、臨床検査分野で診断に広く用いられる多重ロジスティック回帰分析法に匹敵する判別力を有することが分かった

1-4 本論文の構成と概要

「第 1 章 緒論」では、本論文の背景と目的について述べた。

「第 2 章 導入」では、日常の臨床検査データを取り扱うために、次章以降で用いる技術について、その概説と今回の研究で行った変更について述べる。

「第 3 章 検体取り違いを高率に検出しうる加重累積デルタチェック法の考案とその性能評価」では、日常検査情報から検体取り違いを行う手法の開発と、実際の施設のデータを用いて人工的に作成した検体取り違いの検出精度の評価について述べる。

「第 4 章 基本検査による男女判別の正確度：多重ロジスティック回帰分析法と加重対数尤度診断法の比較検討」では、前回値を必要としない検体取り違い検出方法として、検査値から男女を判別する手法に関する検討について述べる。

「第 5 章 結論と今後の展開」では、本論文の研究にて得られた結果についての説明と、今後その結果がどのように展開可能であるのか述べる。

本章の参考文献

- 1) 日本医療機能評価機構 医療事故情報収集等事業 平成 18 年 年報
http://www.med-safe.jp/pdf/year_report_2006.pdf
- 2) 鈴木 友宏. 採血時の患者誤認防止採血認証システム. 日本医療マネジメント学会雑誌 2010;
10: 589-92
- 3) Nosanchuk JS, Gottmann AW. CUMS and delta checks. A systematic approach to quality control. Am J Clin Path 1974; 62: 707-12.
- 4) Sheiner LB, Wheeler LA, Moore JK. The performance of delta check methods. Clin Chem 1979; 25: 2034-7.
- 5) Iizuka Y, Kume H, Kitamura M. Multivariate delta check method for detecting specimen mix-up. Clin Chem 1982; 28: 2244-8.

第2章 導入

臨床医学において、臨床検査の役割は、生体に起こった変調（病態）を客観的に捉え、その情報を病気の診断、進行度の判定、治療の必要性の判定、治療後のモニター、将来の病状の予測（予後推定）に役立てることにある。臨床検査は、大きく、体の変調に関わるシグナルを直接生体からとらえる生理機能検査（心電図、脳波、筋電図など）と、生体から採取した成分を検体として、その中の生体物質を測定し、生体に起こった変調を捉える検体検査に分かれる。

一般に、臨床検査法の測定意義や有用性を評価する重要な観点は、大きく臨床検査が病態以外で変化する要因（年齢、性、体動、体位など）を明らかにすること（**臨床検査の変動要因分析**）と、臨床検査が病態の診断（把握）にどの程度有用であるかを評価すること（**臨床検査の臨床的有用性の評価**）に分かれる。

これらの観点に基づく臨床検査の変動要因の分析や診断特性の評価は、直接的な調査や実験によって実施されるが、十分なデータ数（症例数）を取得するためには多くの労力と経費を要する。そこで、日常の臨床検査値が大規模に蓄積される臨床検査情報システム（LIS: laboratory information system）の情報を利用して、それらの分析や評価を行う試みが、これまで広くなされてきた。しかし、実際の LIS のデータは、多種多様な病態に由来する雑多な検査値の集合であるため、分析を行うためのソースデータを取得するためには、適切な前処理が必要となる。

このような展望のもと、これまで実用化されていない、(1) 検体の取り違いの検出法の開発、(2) 男女の判別における臨床検査値の有用性の評価、の2つのテーマを設定し、いずれも新しいアプローチによる研究を行い、その成果を本学位論文にまとめた。

各研究内容の具体的な方法論とその成果は、3章、4章で記載するが、この章では、これらの研究で用いた、日常検査情報の活用に必要な**分析対象データの前処理法**と、目的に応じた情報の分析に必要な**統計解析法**について概説する。

1) 分析対象データの前処理

- (1) 分布型の確認
 - ・度数分布図、箱ひげ図、正規確率紙
 - ・尖度、歪度
- (2) 極端値の除外
 - ・ 3σ 法、Tukeyの方法
- (3) 分布型の調整
 - ・補正 Box-Cox べき乗変換
- (4) ノイズ除去、
 - ・反復切断補正法
 - ・潜在基準値法

2) 統計解析技術

(1) 変動要因分析

- ・分散分析法
- ・重回帰分析

(2) 診断特性分析

- ・ROC 分析
- ・多重ロジスティック分析法
- ・加重対数尤度推定法

以下にそれぞれの方法論の概要を述べる。

1) 分析対象データの前処理

日常の臨床検査情報を分析するうえで問題となるのが、その検査値の分布である。日常検査情報には患者などの病態特有の検査値や、手術前後・透析前後などの極端値が含まれ、また検査項目によって歪んだ分布となるものも多い。しかし、多変量解析などの多くの統計手法では、分布の歪みが、その結果に大きく作用する。このため、臨床検査情報を分析する前に、グラフや統計量による分布の状態把握や、極端値の除外、歪んだ分布を正規分布へと補正する手法などが必要である。ここでは、①分布型の確認、②極端値の除外、③分布型の調整、④ノイズ除去、の4つの主要な観点に分けて、前処理手法について述べる。

(1) 分布型の確認

測定値の分布型を把握する方法は、度数分布図や箱ひげ図、正規確率紙といった**作図による視覚化手法**と、歪度・尖度、 χ^2 適合度検定などの**統計量から判定を行う方法**に大別される。

a) 作図法（度数分布図、箱ひげ図、正規確率紙）

視覚的に分布型を把握する手法として、**度数分布図**や**箱ひげ図**がある（図 2-1）。**度数分布図**では、一定の階級値の間に含まれる測定値の数（度数）を集計し、棒の高さで表現する。また、箱ひげ図は、第 1 四分位数（Q1）、第 2 四分位数（Q2、中央値）、第 3 四分位数（Q3）により箱の下端・中心・上端を示す。ひげの上端と下端は、2.5 パーセンタイル値と 97.5 パーセンタイル値を表したもののや、極端値を除外した場合の最小値と最大値（Tukey の方法）を表したものなどがある。

また、分布型が正規分布に適合しているか、より明確に把握する作図法として**正規確率紙**がある。これを作成するには、まず度数分布図を作成する（図 2-2）。次に下側の度数を含んだ累積度数分布図を作成し、それぞれの棒の高さの位置に点を打つ。最後に、正規分布であれば直線となるように縦軸を調整したものが正規確率紙となる。一般に、度数分布図が右裾広がりの測定値では、直線に対して上に凸なグラフとなり、左裾広がりの測定値では下に凸なグラフとなる。

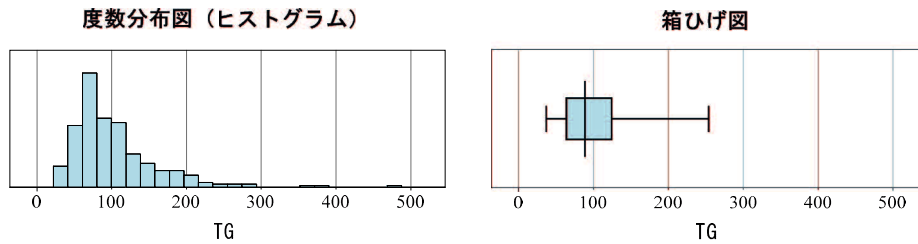


図 2-1 度数分布図および箱ひげ図による分布型の確認

測定値の分布型を視覚的に確認する方法として「度数分布図」や「箱ひげ図」などがある。

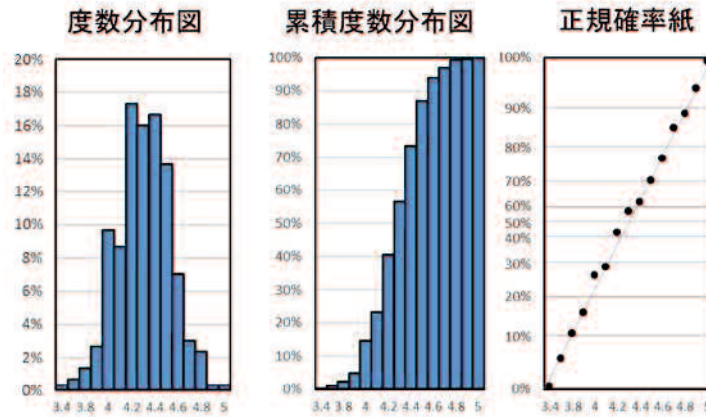


図 2-2 度数分布図と正規確率紙の関係

正規確率紙は、測定値が正規分布に適合しているか、視覚的に評価する方法である。正規確率紙は、度数分布図から累積度数比を求めて累積度数分布図を作成し、その縦軸（累積度数比）を調整することで、正規分布であれば直線となるようにしたものである¹⁾。

b) 歪度、尖度

歪度(skewness: Sk)、尖度(kurtosis: Kt)は、それぞれ分布の左右非対称度と、分布の扁平さを示す統計量であり、以下の式で表される¹⁾。

$$\text{歪度 Sk} = \sqrt{n} \frac{\sum(x_i - \bar{x})^3}{\{\sum(x_i - \bar{x})^2\}^{3/2}} \quad \text{式(2-1)}$$

$$\text{尖度 Kt} = n \frac{\sum(x_i - \bar{x})^4}{\{\sum(x_i - \bar{x})^2\}^2} - 3 \quad \text{式(2-2)}$$

正規分布であれば、歪度と尖度はともに 0.0 となる。歪度は、右裾が長い場合には 0 よりも大きな値をとり、左裾が長い場合には 0 よりも小さな値をとる。また、尖度は、分布が正規分布に比べて扁平な形状である場合には 0 よりも大きな値をとり、分布が正規分布に比べて凸である場合には 0 よりも小さな値をとる。また、どちらの指標も絶対値がおよそ 0.3 を越えると、実質的に明瞭な正規分布から乖離していると、本研究では判断した。

(2) 極端値の除外 (3σ法、Tukeyの方法)

分布における極端値の除外方法として、3σ法や、箱ひげ図などで使用される **Tukeyの手法**²⁾ などがある。3σ法では、平均値 μ と標準偏差 σ を計算し、 $\mu \pm 3 \times \sigma$ の範囲を外れたものを除外するが、分布に偏りがある場合には適切ではない。一方 **Tukey** による除外法は、第1・第3四分位偏差(Q1, Q3)と、四分位範囲(Inter Quartile Range : IQR = Q3 - Q1)を用いて、 $Q1 - 1.5 \times IQR$ から $Q3 + 1.5 \times IQR$ までの範囲を外れたものを除外する方法であり、分布に偏りがある場合でも適切な除外を行える。

(3) 分布型の調整 (調整 Box-Cox べき乗変換式)

多くの統計解析法では、分布が正規分布であることを前提としているが、臨床検査値の分布では、右裾、または左裾の長い歪んだ分布型が多くある。このような分布型を補正する手法として、**Box-Cox べき乗変換**³⁾がある。さらに、臨床検査分野で利用するため市原によって変換原点の項が追加された**調整 Box-Cox べき乗変換**⁴⁾があり、次の変換式が用いられる。

$$x^T = \frac{(x - a)^p - 1}{p} \quad (p \neq 0) \quad \text{式(2-3)}$$

$$x^T = \log(x - a)^p \quad (p = 0) \quad \text{式(2-4)}$$

ここに p はべき乗値、 a は変換原点である。図 2-3 に示すように、 $p = 1$ の場合に無変換であり、右裾が長い場合は p に 1 を下回る値を設定し、左裾が長い場合は p に 1 を上回る値を設定する。また、 p が 1 から遠ざかるほど変換は強く作用するため、 p により変換の強弱を調整できる。また、臨床検査値では、分布の最小値が 0 でないものも多く、この様な分布を補正する場合には変換原点 a に分布の最小値周辺の値を設定することにより、適切に変換を行うことができる。

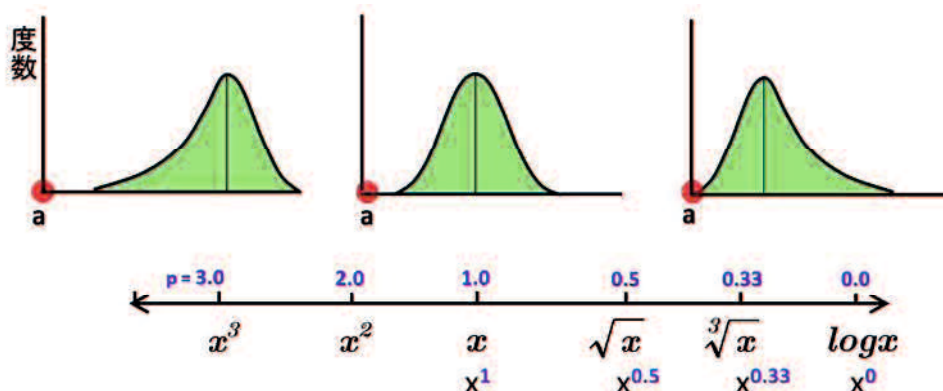


図 2-3 分布型とべき乗値の関係

臨床検査値の分布の多くは正規分布とはならず、それ前提とした統計処理を行うためには分布型の補正が必要である。調整 Box-Cox 法は、最も確かな正規分布への変換法で、右裾広がりであればべき乗値 p に 0 や 0.33 などの 1 より小さい値を設定することで、分布を正規分布に補正することができる。

べき乗値 p と変換原点 a は、以下の手順 1～5 で推定する

1. a の初期値 a_0 として、分布の最小値から、中央値と最小値の差の 1/10 低い値を用いる。

$$a_0 = x_{min} - (Me - x_{min})/10$$

2. p の初期値 p_0 は、0 から 4 まで 0.1 間隔で p を段階的に変化させ、正規分布への適合度が一番良くなる値とする。このとき、適合度の評価には χ^2 適合度検定を用いる。
3. a_n と p_n と、変換済みの分布の平均値 M と標準偏差 SD 用いて、 a_{n+1} を計算する。

$$a_{n+1} = p_n \times (M - 4SD)^{1/p_n} + a_n$$

4. Newton-Raphson 法により a_{n+1} と p_n から p_{n+1} を計算する。
5. p_n と p_{n+1} の違いがなくなるまで 3 と 4 の過程を繰り返す。

調整 Box-Cox べき乗変換で導入された変換原点 a は、臨床検査の検査値分布を変換するために必要な項であるが、分布の最小値より小さな値しか設定できないという制約がある。本研究では日常検査値の分布を対象としているが、このような分布の下方には極端値を多く含まれるため、適切な変換原点を設定しえず、正規分布への的確な変換が困難となる。そのため、分布両端、特に低値域の極端値を除外することが、適切な変換に重要となる。本論文の第 3 章では、日常検査データの分布を的確に正規分布に変換するため、データ分布の低値側 1% と、高値側 1% を除外し、その後に変換係数 p 、 a の推定を行ったが、除外の有無と変換の成否の関係についても詳しく検討を行った。

(4) ノイズ除去

a) 反復切断補正法

臨床検査では、病院・検査機関内にて精度を保つため行われる内部精度管理に加え、異なる病院間で検査値を一致させるための外部精度管理が行われている。この外部精度管理では、同一の試料を各病院に送付し検査値を収集、検査値が一致するか調査を行う。同じ試料の反復計測値であるため、その分布は正規分布になると仮定されるが、実際には測定法の誤登録などにより、本来の分布から外れた一群の結果が含まれることが多い。このような多くの外れ値を含みながら、その分布の中央部には明らかに正規分布が存在すると仮定できる場合、その中央部の正規分布の形状(平均値、標準偏差)を抽出する方法として、市原は反復切断補正法(iterative truncation and correction method : ITC 法)を開発した⁵⁾。

平均値と標準偏差を計算する場合、極端値を除くために、まず平均値 M と標準偏差 SD を計算し、 $M \pm k \times SD$ の範囲外を除外した上で、再度、平均値 M' と標準偏差 SD' の計算を行う。このときの切断係数 k は一般に $k = 3.0$ 程度を設定し、約 0.3% のデータを除外する。しかしながら多くの極端値が含まれる場合には、最初の SD が大きな値となり、適切な除外が行えないため、再度計算される平均値 M' と標準偏差 SD' も、本来の分布の形状をとらえることができない。

このため ITC 法では、切断係数として $k = 1.645$ 程度の値を設定し、より大きな範囲を除外対象とする。しかしながら、 $k = 1.645$ の場合、除外されるデータは約 10% となり SD は小さめに計

算される。このため、除外後に計算された SD は本来の大きさの 0.7896 倍となるため、切断係数を $k = 1.645$ とする場合には、標準偏差を 1.2664 倍に補正する必要がある。つまり、ITC 法の特徴は、小さめの切断係数による除外と、切断係数に応じた補正処理を交互に行うことにより、大量の極端値を含む分布に対応できることである。

図 2-4 では、平均値 $M = 0$ 、標準偏差 $SD = 1$ となる正規分布に対して、高値側に極端値の集団がある例を示している。除外を行わずに計算された平均値は、最下段に示したように平均値が $M = 0.531$ と極端値側に偏り、標準偏差も $SD = 2.131$ と大きめに計算されている。このとき、切断係数を $k = 1.645$ とすると、図中では赤色で示した範囲が除外される。この範囲を除外した上で、再度平均値と標準偏差を計算し、補正を行った結果が $M = 0.148$ 、 $SD = 1.525$ となる。この手続きを繰り返すことで最終的には最上段に示すように $M = -0.015$ 、 $SD = 1.000$ と偏りのない平均値と標準偏差が計算できる。

外れた位置にある集団の割合が、分布全体のデータ数に対して大きい場合には、切断係数 k を小さめに設定することで対応が可能である。しかしながら、小さめの切断係数で除外を行い計算された SD は、本来の SD よりもかなり小さい値となる。このため、切断係数 k に応じた補正係数を、計算された SD に掛け合わせることで、正しい SD を得ることができる。本研究では、第 3 章にて個体内変動幅の算出にてこの手法を使用した。

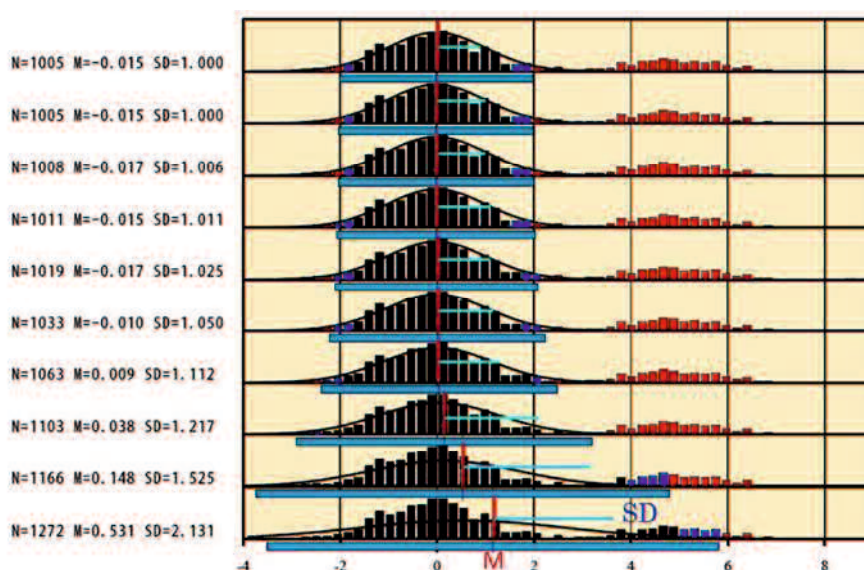


図 2-4 反復切断補正法の概要

正規分布の一方または両側に 1 群の極端値が含まれる場合に、反復切断補正法を用いることでその影響を除いた平均値 M と標準偏差 SD を計算できる。例えば正規分布の右側に極端値の集団が含まれていると、最下段のごとく M は極端値側に偏り、 SD は広めに計算されている。しかし $M \pm k \times SD$ から外れたものを除外して再度 M 、 SD を計算すると下から 2 段目のように偏りが補正される。これを繰り返すことにより、最上段のような偏りのない M 、 SD を計算できる。

b) 潜在異常値抽出法

臨床検査において、検査項目の基準範囲を計算するためには、健常者からなる**基準個体**の集団が必要となる。基準個体の選択では、問診による生活習慣の確認や、医師による診察などがおこなわれるが、実際にはこれらの方法では潜在病態を有するものを完全には除外できない。そこで、日常検査から基準範囲を設定するために市原によって開発されたのが**潜在異常値除外法**である⁶⁾。その原理は、

図 2-5 に示すように「基本検査項目に異常値がなければより健常相当者である」ということであり、より適切な基準個体であるという仮説に基づいている。

潜在異常値除外法では、この仮定に従い、ある検査項目の基準範囲を計算するときには、他の検査値が基準範囲を外れていない個体を基準個体とする。

図 2-6 に潜在異常値除外法の手順を示す。初回は、検査項目毎に、個別に基準範囲を $L_0 \sim U_0$ として設定するが、潜在的な異常値の混入により、基準範囲はやや広めに求まる。2回目からの計算では、他の検査項目の値に基準範囲を外れるものがあればデータから除外し、それぞれの検査項目の基準範囲を更新する。更新の前後で基準範囲の値の変化が無ければ計算を終了する。

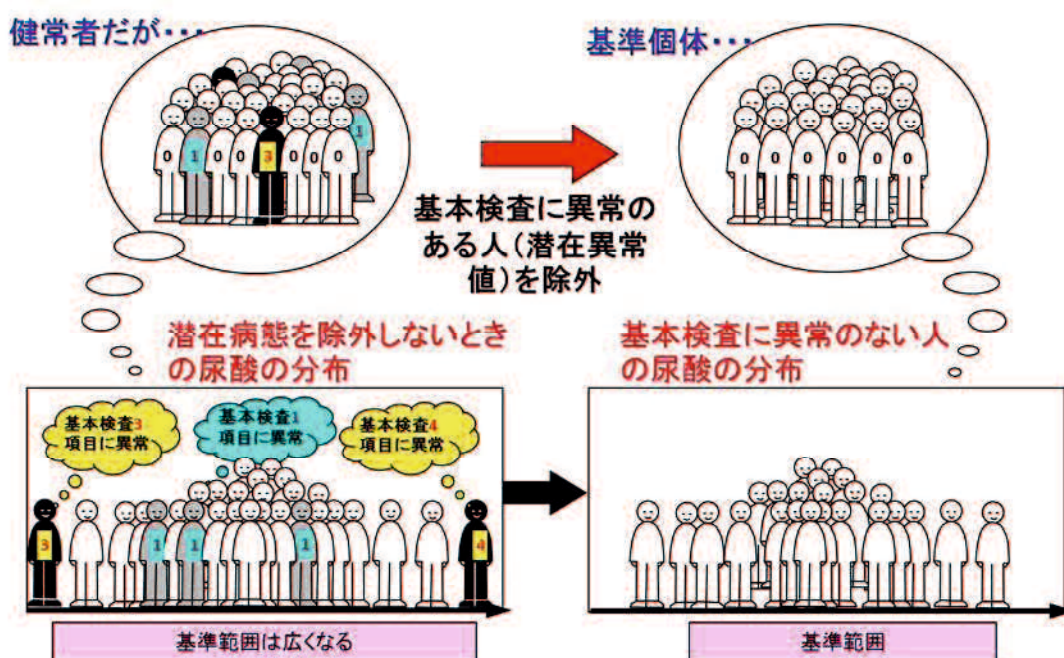


図 2-5 潜在異常値除外法の原理

問診や診察で健常者を選別し、その測定値から基準範囲(RI)を求めても、メタボリック症候群などの潜在病態を持った人が含まれるため、RIは広めに計算されてしまう。潜在異常値除外法では、他の検査に異常値のないデータに絞り込むことで、最適なRIを計算することができる。

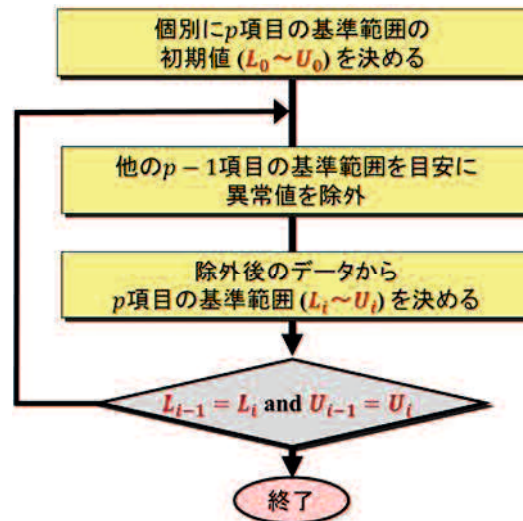


図 2-6 潜在異常値除外法による基準範囲の計算フロー

p 個の検査項目について基準範囲(RI)を同時に設定する場合、初回は個別に RI を設定するため、潜在的な異常値の存在で、RI は少し広めに求まる。しかし 2 回目の設定では、初回に決めた RI を外れる値をもつ個体を除くため、異常値の混入率が減少し、RI は初回より狭く計算される。これを繰り返すことでより適切な基準個体に絞こめるため、RI の計算値が最適化される。

2) 統計解析技術

臨床検査情報に対する統計解析手法としては、検査値と変動要因の関連を調べるもの（**変動要因分析**）や、検査の診断能を評価するもの（**診断特性分析**）、複数の検査項目の共通性を調べたり統合する方法（**主成分分析・因子分析**）などがある。

この研究に関連する分析手法として、(1) **変動要因分析に関わる手法**である「分散分析」と「重回帰分析」、(2) **診断特性分析に関わる手法**である「ROC 分析」、「多重ロジスティック回帰分析」と「対数尤度診断法」について、それぞれの手法と、臨床分野での利用方法について述べる。

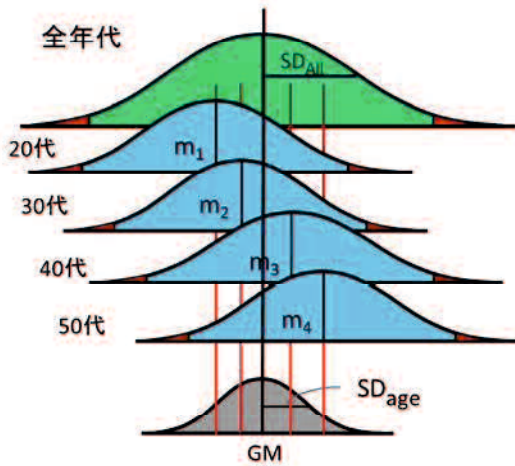
(1) 変動要因分析

a) 分散分析法と群間差指数

特定の要因で分けられた群間の測定値に、差があるかどうかを分析する手法として**分散分析法** (ANOVA)がある²⁾。一元配置分散分析では、群間分散と群内分散の比から群間の差を検定することができるが、群間 SD と個体間 SD から、群間の離れ具合を示す値となる**群間差指数 SDR**を計算することができる⁵⁾。図 2-7 に、年代によって分けた群の群間差指数 SDR の計算の例を示す。SDR の計算では、まず群間 SD を計算する。群間 SD は、全群の測定値から得られる総平均 GM に対する、各群の平均値のばらつきを示す値であり、以下の式により定義される。

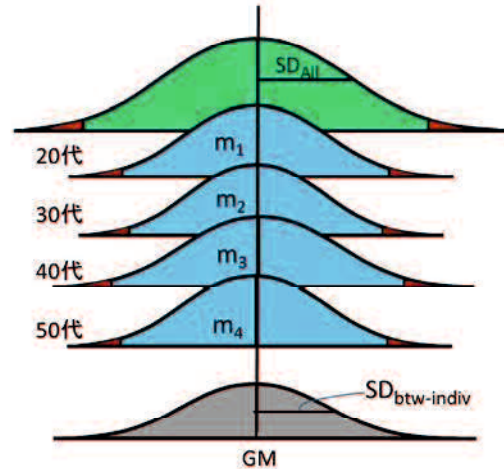
$$\text{群間SD} = \sqrt{\frac{\sum_{i=1}^k (m_i - GM)^2}{k - 1}} \quad \text{式(2-5)}$$

1 群間SDの計算



SD_{age} : 総平均に対する
各年代の平均値のばらつき

2 個体間SDの計算



$SD_{bt看-indiv}$: 年代の調節済みの個体間SD

$$\text{年代間差指数: } SDR_{age} = \frac{SD_{age}}{SD_{bt看-indiv}} \geq 0.3$$

図 2-7 群間差指数の計算手順

一元配置分散分析を用いると、群間差を示す単位のない指標として群間差指数(SDR)を計算することができる。SDR は、総平均からの各群の平均値のばらつきを示す群間 SD を、個体間 SD によって割った値である。一般に臨床検査分野では、測定値の誤差許容限界の理論から SDR が 0.3 を越えると、意味のある群間差があると見なされる。

ここに k は群の数であり、図 2-7 の例では $k = 4$ となる。また、 $m_i (i = 1, \dots, k)$ は各群の平均値を示す。次に、個体間 SD を計算する。個体間 SD は、群による測定値の違いを除いた純粋な個体間のばらつきを表す値であり、各群の平均値を GM に揃え、そのうえで各群の SD の平均的な値を計算したものである。このようにして群間 SD と個体間 SD を計算し、群間 SD を個体間 SD で割ったものが群間差指数 SDR である。この群間差指数 SDR は単位をもたないため、異なる検査項目間で比較が可能である。また、臨床検査値を対象とする場合、一般に 0.3 以上であれば群間に意味のある群間差があるといえる⁷⁾。

b) 重回帰分析

目的変数と多項目からなる説明変数との関係式の推定および、関連性の確認を行う一般的な手法として重回帰分析がある^{8,9)}。重回帰分析の一般式(回帰式)は、

$$(\text{目的変数}) = b_0 + b_1(\text{説明変数 1}) + b_2(\text{説明変数 2}) + \dots + b_p(\text{説明変数 } p)$$

の形で表される。ここで p は説明変数の個数を示し、 $b_i (i = 0, \dots, p)$ を回帰係数と呼ぶ。回帰係数 b_i にはそれぞれの標準誤差 $SE(b_i) (i = 0, \dots, p)$ が与えられるため、各回帰係数が 0 とみなせるか検定できる。説明変数にかかる回帰係数 b_i が 0 であるとき、その説明変数は目的変数に対して相関

がないことを示すため、検定は説明変数と目的変数の関係の有無を判定することになる。また、ある説明変数の回帰係数は、それ以外の説明変数の影響（交絡作用）を除いた、純粋な目的変数との関係を示す値である。この原理のため、収縮期血圧と拡張期血圧のように相関の強い項目を、同時に説明変数に設定すると、どちらか関係の強い項目が残るか、両方とも関係が弱いと判断される。そのため、相関の強い回帰係数がある場合、どちらかを回帰式から除くことになる。

(2) 診断特性分析法

a) ROC 分析

検査による診断（判別）の能力を評価する手法として **ROC (receiver operating characteristic) 分析**がある¹⁰⁾。ROC 分析では、各カットオフ値に対応した感度と特異度から ROC 曲線を描画し、その曲線下面積（AUC: area under curve）により判別性能を評価する。

ROC 曲線の描画手順を図 2-8 に示す。図 2-8 は、検査値 x を用いて疾患群と非疾患群を判別する例である。ROC 曲線を描画するためには、まず疾患群と非疾患群に対象を分け、カットオフ値に対応する感度と特異度を計算する必要がある。感度は、疾患群の中でカットオフ値を超えて陽性となる割合であり、特異度は、非疾患群の中でカットオフ値を下回り陰性と判定できる割合である。図 2-8 の例では、4 以上を「疾患あり」とするカットオフ値を設定した場合に、感度が 0.35、特異度が 0.92 となることを示している。このようにして全てのカットオフ値に対応する感度と特異度を計算し、感度と $1 - \text{特異度}$ をプロットしてカットオフ値の順に繋いだ曲線が ROC 曲線である。このように描画された曲線の下側の面積が AUC であり、AUC が高いほど検査による判別能が高いことを示す。AUC の最大値は 1 であり、この場合は、検査によって完全に疾患の有無を判別できることを示す。

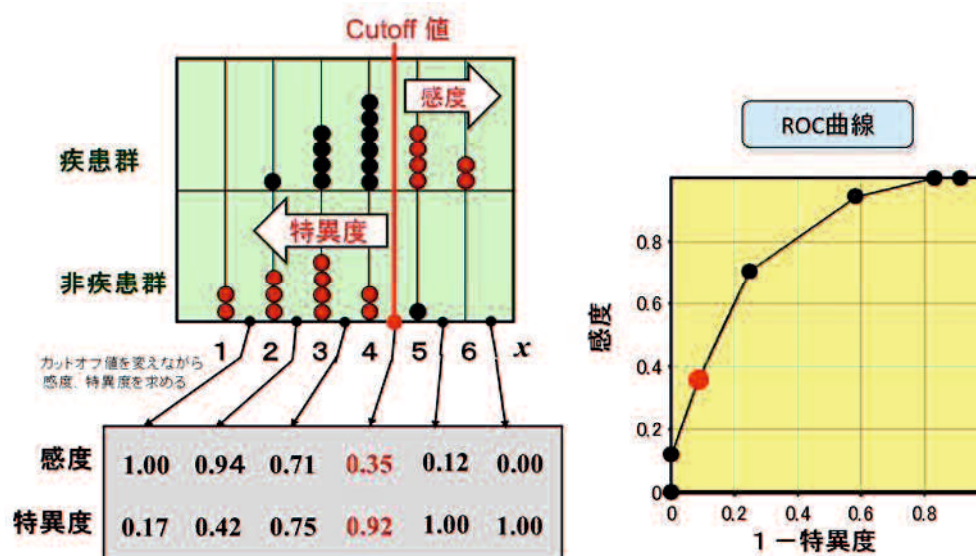


図 2-8 ROC 曲線の描画方法

検査による特定疾患の判別能を示す指標として ROC 曲線の曲線下面積（AUC）がある。ROC 曲線は、各カットオフ値における感度と特異度を計算し、感度と $1 - \text{特異度}$ を順番につなぐことで描画できる。このときの曲線の下側の面積（AUC）が大きいほど検査の判別性能が高いことを示す。

b) 多重ロジスティック回帰分析

複数の臨床所見や検査所見を組み合わせた回帰式を作成し、病態の有無、予後の良好・不良などの診断（判定）を行う手法として多重ロジスティック回帰分析がある^{11, 12)}。多重ロジスティック回帰分析の一般式は、

$$p = \frac{1}{1 + e^{-X}} \quad \text{式(2-6)}$$

$$X = b_0 + b_1x_1 + \dots + b_px_p \quad \text{式(2-7)}$$

の形で表される。ここで p は説明変数の個数を示し、 $b_i (i = 0, \dots, p)$ を回帰係数と呼ぶ。回帰係数 b_i にはそれぞれの標準誤差 $SE(b_i) (i = 0, \dots, p)$ が与えられるため、各説明変数が診断に有効であるか検定を行うことができる。また、ある説明変数の回帰係数は、他の説明変数による交絡作用を除いて計算されたものであるため、重回帰分析と同様に、相関の強い回帰係数がある場合にはどちらかの説明変数を回帰式から除く必要がある。

c) 対数尤度診断法

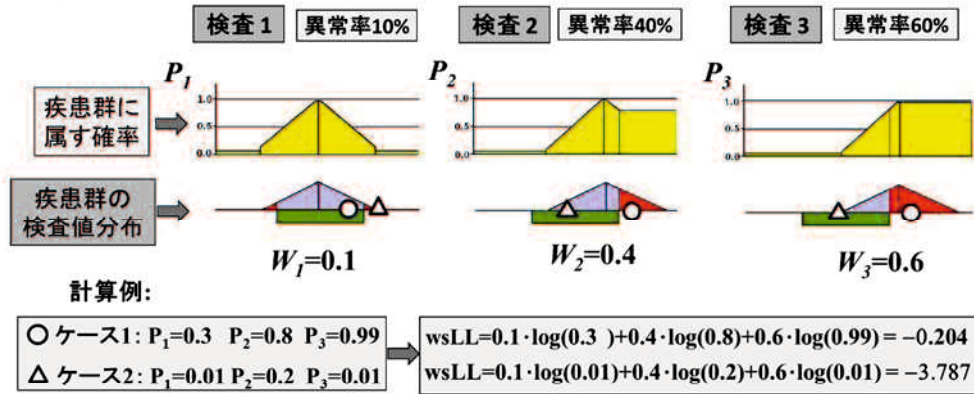
対数尤度診断法は、複数の臨床所見や検査所見を組み合わせて、ある個体が特定の疾患と類似しているかを判定するために市原の開発した手法である¹³⁾。この手法では、検査所見による判定では**疾患群の検査値分布を尤度関数とし、臨床所見による判定ではその陽性率を尤度**として、疾患判別における各所見の有効度を考慮することで、疾患への類似度($wsLL$)を算出する。検査所見と臨床所見のそれぞれの場合について、尤度の計算と $wsLL$ の算出方法を

図 2-9 に示す。

図 2-9 の定量検査セットでは、検査 1(P1)～検査 3(P3)について、それぞれの尤度関数を黄色、疾患群の検査値分布を青色、検査値の基準範囲を緑色で示している。ある検査所見の尤度関数 $L(x)$ は、疾患の検査値分布と照らし合わせ、検査値分布の中心に近い値であれば高い尤度となるよう設定する。検査 1 では、○で示した位置の尤度は、0.3 となり、△で示した位置の尤度は 0.01 となる。また、検査 2 や検査 3 のように、基準範囲に対して疾患の検査値分布が高値側にある場合には、基準範囲を超えた検査値の尤度は一定値となるよう尤度関数を設定する。また、基準範囲に対して疾患群の検査値分布が大きく外れる検査項目は、疾患を判別する場合に有用な検査項目と考えられる。このため、疾患群の検査値分布のうち基準範囲を外れるものの割合を異常率として、検査の有用度と考える。よって、対数尤度診断法では、検査に対する尤度と異常率をかけた値の総和を $wsLL$ 値とし、これが高い値であるほど疾患である可能性が高いとして、疾患の判別を行っている。

一方、臨床所見セットでは、各所見の陽性率と陰性率を元に尤度を求める。すなわち、ある所見が見られる場合には陽性率、無い場合には陰性率が臨床所見に対する尤度となり、陽性率の高い所見ほど、疾患の診断に有用であるため、対数尤度と陽性率をかけたものの総和を $wsLL$ として、疾患の判別を行っている。

■検査所見セットの場合



■臨床所見セットの場合

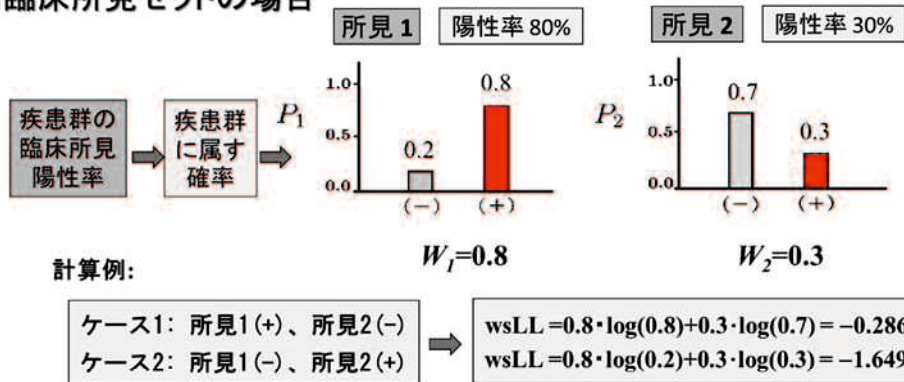


図 2-9 対数尤度診断法の計算手順

対数尤度診断法における、検査所見と臨床所見についてその尤度の計算と、類似度(wsLL)の計算方法について示す。検査所見では、尤度関数は黄色、疾患の検査値分布は青色、基準範囲を緑色とした。また、臨床所見では、所見の陽性率を赤色、陰性率を灰色のバーで示した。

本章の参考文献

- 1) 市原 清志. バイオサイエンスの統計学. 南江堂(1990)
- 2) 市原清志, 佐藤正一. 新版 カラーイメージで学ぶ統計学の基礎. 日本教育研究センター (2014)
- 3) E.P.Box, D.R.Cox. An Analysis of Transformations. *Journal of the Royal Statistical Society* 1964; 26, 2: 211-52
- 4) Determination of Reference Intervals for 13 Plasma Proteins Based on IFCC International Reference Preparation (CRM470) and NCCLS Proposed Guideline (C28-P, 1992): Trial to Select Reference Individuals by Results of Screening Tests and Application of Maximal Likelihood Method. *Journal of Clinical Laboratory Analysis* 1996; 10:110-17
- 5) Ichihara K, Kawai T. An iterative method for improved estimation of the mean of peer-group distributions in proficiency testing. **Clin Chem Lab Med** 2005; 43:412–21.
- 6) 市原清志：潜在基準値法による日常検査情報の活用. *臨床検査 増刊号* 2005; 49: 1471–85
- 7) Ichihara K. Statistical considerations for harmonization of the global multicenter study on reference values. **Clin Chim Acta** 2014; 432: 108–118.
- 8) Kleinbaum DG, et al. *Applied Regression Analysis and Other Multivariable Methods*. 3rd ed., Duxbury Press, 1997
- 9) 市原清志. 重回帰分析による臨床検査の変動要因の解析法. **臨床検査 増刊号** 2005; 49: 1297–306
- 10) 久繁哲徳. 検査の有用性の指標と ROC 分析. **臨床検査 増刊号** 2005; 49: 1335–40
- 11) Kleinbaum DG. *Logistic Regression : A Self-Learning Text*. Springer-Verlag, New York, 1998
- 12) 市原清志. 多重ロジスティック分析による診断的有用性の評価法. **臨床検査 増刊号** 2005; 49: 1341–53
- 13) Ichihara K, Sato K. Evidence-based laboratory interpretation system built on a large collection of case records with well-defined diagnosis. **Clin Chem Lab Med** 2001; 39: 1033–44

第3章 検体取り違いを高率に検出する加重累積デルタチェック法の考案とその性能評価

3-1 はじめに

近年、オーダリングシステムや自動化の普及により、大量の検体処理が可能となってきた。しかしながら、臨床検査に関連した患者取り違いについて、公益財団法人 日本医療機能評価機構の平成 18 年度報告¹⁾には、2003 年から 2006 年までの間に、患者取り違いが 91 件、検体取り違いが 101 件報告されているなど、多くの検体取り違いが発生している。また、浜松医科大学医学部附属病院では、2006 年から 2008 年の間に輸血のための血液型の検査において、6 件のインシデントが報告されており、その事例を契機として 2010 年にバーコードによる検体照合システムの導入が行われた^{2, 3)}。このように患者と検体を認証する仕組みを導入により、取り違いを防止する動きが広がっているが、2013 年に報告された、天理よろず相談所病院の例では、外来患者の採血では全ての工程を通してバーコードによる管理が行われているが、病棟での採血では一元的なバーコード管理が導入されておらず、2008 年から 2012 年までの間に採血時の患者誤認が 5 件、バーコードラベル貼り付け間違いによる患者誤認が 4 件確認されている⁴⁾。

このように、検体処理には、いまだに人の手によって行われる作業が含まれるため、ラベルの張り間違いや、患者の取り違いなどによる検体の取り違いが発生する可能性を含んでいる^{5, 6)}。事実、無視できないような取り違いが発生しているが^{7, 8)}、このような検体の取り違いは、臨床検査技師などによって事前に発見されている。しかしながら、そのような検体取り違いを見逃してしまうと、検査室への不信感へとつながるため、検査結果報告の前に、必ず検出する必要がある。そのため検査室では、検査値が一定の閾値を越えるか、関連のある検査項目同士を比較し一致がみられなければ自動的に再検査を行うという単純な手法を用いている。ただし、極端な検査値が多く存在するため、このような手法では、多くの費用が必要となっている。

当然のことながら、検体の取り違いを検出するための唯一の方法は、その日測定された今回の検査値と、前回の検査値を比較することである。検査情報システム(laboratory information system : LIS)に含まれる基本的な機能としては、自動的に前回値との比較を行い、その差が非常に大きい場合に、取り違いの可能性があると警告するものがある。しかしながら、最終的な判断は臨床検査技師によって行われている。

過去には、自動的に取り違いを検出する方法として、いくつかの手法が論文にて提案されている。これらの基本的な原理としては、同時に測定された複数の検査項目のうち、全てもしくは必要な検査項目について前回値と今回値の差(デルタ)を単純に加算する方法⁵⁾や、判別方程式を作成する方法^{6, 7)}が用いられている。しかしながら、検査値の分布は多種多様であり⁸⁾、これらの手法は分布型を考慮してないため、実用では上手く動作しないことが予想される。正規分布でなく歪んだ分布である場合、そのデルタをそのまま用いることは取り違いの検出に大きな影響力をもつ。そのうえ、同一個体間での変動が検査項目によって大きく異なるという問題がある。例えば、グルコースや中性脂肪などは、採血時の状況より大きく変動する。それゆえ、デルタによって取り違い検出を行うためには、項目によって異なる分布の形状や、個体内での変動の大きさに

ついて十分に考慮しなければならない。

そこで本研究では、べき乗変換による検査値分布の正規化手法や、膨大な過去データから検査項目ごとの個体内での変動の大きさを、反復切断除外法⁹⁾により数値化する手法を使用するなど、分布や変動の影響を考慮した新しいデルタチェック法を開発した。この手法は計算により得られた個体内変動の大きさを加重として、各検査項目のデルタの加重平均を行うため、加重累積デルタチェック法(**weighted Cumulative Delta Check: wCDC**)と命名した。この章では、加重累積デルタチェック法による計算手順と、モデルデータを用いた取り違い検出の精度の確認を記述した。

3-2 手法

3-2-1 加重累積デルタチェック法(wCDC)の数理

3-2-1-1 分布型の正規化

臨床検査値を対象とする場合、検査値の分布が正規分布となる項目だけでなく、多くの項目で対数正規分布のような右裾の長い分布形状をとる。図 3-1 は、検査値分布が正規分布となる総蛋白と、右裾広がり分布となる中性脂肪についての同一個体の前後 2 回の関連を示す相関図である。グラフ内に示した直線は、前後の値が同じ点を結んだものであり、各点の直線からの距離は前後差と比例する。総蛋白ではこの直線からの距離は検査値の高低に関わらずほぼ一定であるが、中性脂肪では低値側では前後差が小さく、高値側での前後差が大きくなり、値の範囲によって前後差が異なることが分かる。

このような右裾広がり分布となる検査項目では、分布を正規分布に変換することで高値側と低値側の前後差を一定にすることができる。このため、以下の調整 **Box-Cox** べき乗変換式により、検査値分布を正規分布に変換する。^{12, 14)}

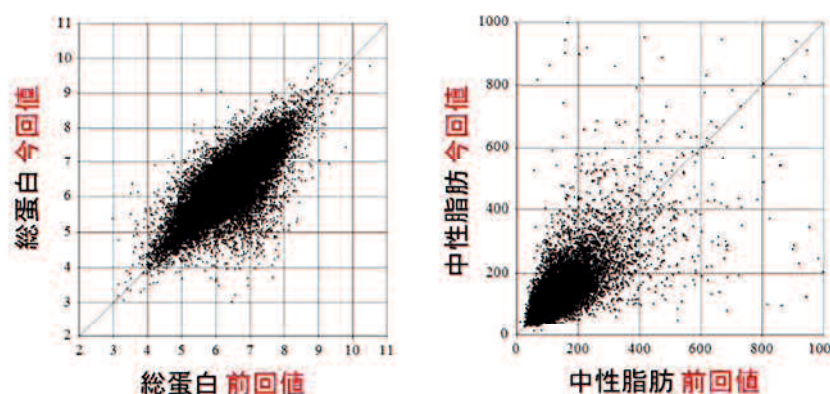


図 3-1 総蛋白と中性脂肪の前回と今回との値の相関

正規分布となる総蛋白と、対数正規分布となる中性脂肪について、同一個体の前回値と今回値の相関図を示す。直線は前回値と今回値が同じ場合を示しており、総蛋白では低値側でも高値側でも直線からの距離に大きな違いはない。一方、中性脂肪では高値側では直線からの距離が大きく低値側に比べて変動が大きいことを示している。

$$X^T = \frac{(x - a)^p - 1}{p} \quad (p \neq 0) \quad \text{式(3-1)}$$

$$X^T = \log(x - a) \quad (p = 0) \quad \text{式(3-2)}$$

ここに x と X^T は、変換前の検査値と、変換後の検査値を示している。また係数 p 、 a は、それぞれべき乗値と変換原点を示す。適切にこれらの係数を設定することで、変換式による正規分布への変換が可能であるが、日常の検査値分布には大量の極端値が含まれているため、係数 p 、 a の推定に影響が出る。このため、前処理として検査値分布の1パーセント値以下、および99パーセント値以上の計2%を除外し、係数 p 、 a の推定を行うこととした。

係数 p と a は互いに連携するため、最尤推定法による推定をそれぞれ交互に行うことで、最適な p と a を推定している。まず、検査値分布の最小値 X_{min} と中央値 Me を用いて a の初期値を $a_0 = X_{min} - (Me - X_{min}) / 10$ と設定し、 p の初期値 p_0 の推定を行う。次の a_1 は、 a_0 、 p_0 によってべき乗変換した検査値の平均値 M と標準偏差 SD を用いて $a_1 = p \times (M - 4SD)^{1/p_0} + a_0$ とし、この a_1 を用いて p_1 の推定を行う。この手順を、両方の係数が安定するまで繰り返し、係数 a 、 p を決定する。また、男女で分布型が若干異なるため、係数 p 、 a の推定は男女別に行った。

3-2-1-2 検査値の標準化

検査項目間の単位の違いを除くために、検査値ごとに設定された基準範囲を用いて検査値分布の標準化を行った。まず、検査項目に対応した変換係数 p 、 a を用いて、その項目の基準範囲の下限(Lower Limit : LL)と上限(Upper Limit : UL)を補正 Box-Cox べき乗変換したものを LL^T 、 UL^T とする。基準範囲は95%の基準個体が含まれる範囲であるため、基準個体の分布の平均値 M^T と、標準偏差 SD^T に対して以下の関係式が成り立つ。

$$LL^T = M^T - 1.96 \times SD^T \quad \text{式(3-3)}$$

$$UL^T = M^T + 1.96 \times SD^T \quad \text{式(3-4)}$$

この式より、基準個体の M^T と SD^T は以下の式で定義できるため、基準範囲の LL^T と UL^T から仮想的に基準個体の M^T と SD^T を計算することができる。

$$M^T = \frac{UL^T + LL^T}{2} \quad \text{式(3-5)}$$

$$M^T = \frac{UL^T - LL^T}{3.92} \quad \text{式(3-6)}$$

この基準個体の M^T と SD^T を用いることで、べき乗変換済みの個々の検査値 X^T を以下の式により標準化値 z_x に変換できる。

$$z_x = \frac{X^T - M^T}{SD^T} \quad \text{式(3-7)}$$

この標準化も、正規化と同様に男女それぞれの基準範囲を用いて行う。前項の標準化と正規化の手順を図 3-2 に示す。図には、べき乗変換による正規化 (1→2) と、基準範囲による標準化 (2→3) と、次項で説明する前後差から反復切断補正法を用いて個体内変動を算出するもの (4) が含まれている。図中の青色の領域は患者の検査値分布を示し、緑色の領域は基準範囲とその範囲から推定される健常者の検査値分布である。健常者の分布に歪み (偏り) のあるものは、基準範囲にも歪みがあり、患者の分布にも歪みを生じる。また、患者由来の検査値分布には健常者の分布には見られないような極端な値が両端に含まれるために、さらに大きな歪みを生じる。この歪みは、極端値の除外処理と、調整 Box-Cox べき乗変換により解消することが可能であるが、この変換係数は健常者分布の変換にも適用可能であるため、基準範囲の変換にも利用できる。よって、次の手順では、基準範囲の下限値 (LL) と上限値 (UL) から、基準範囲の平均値 M と標準偏差 SD を仮想的に算出し、個々の測定値の標準化に用いる。

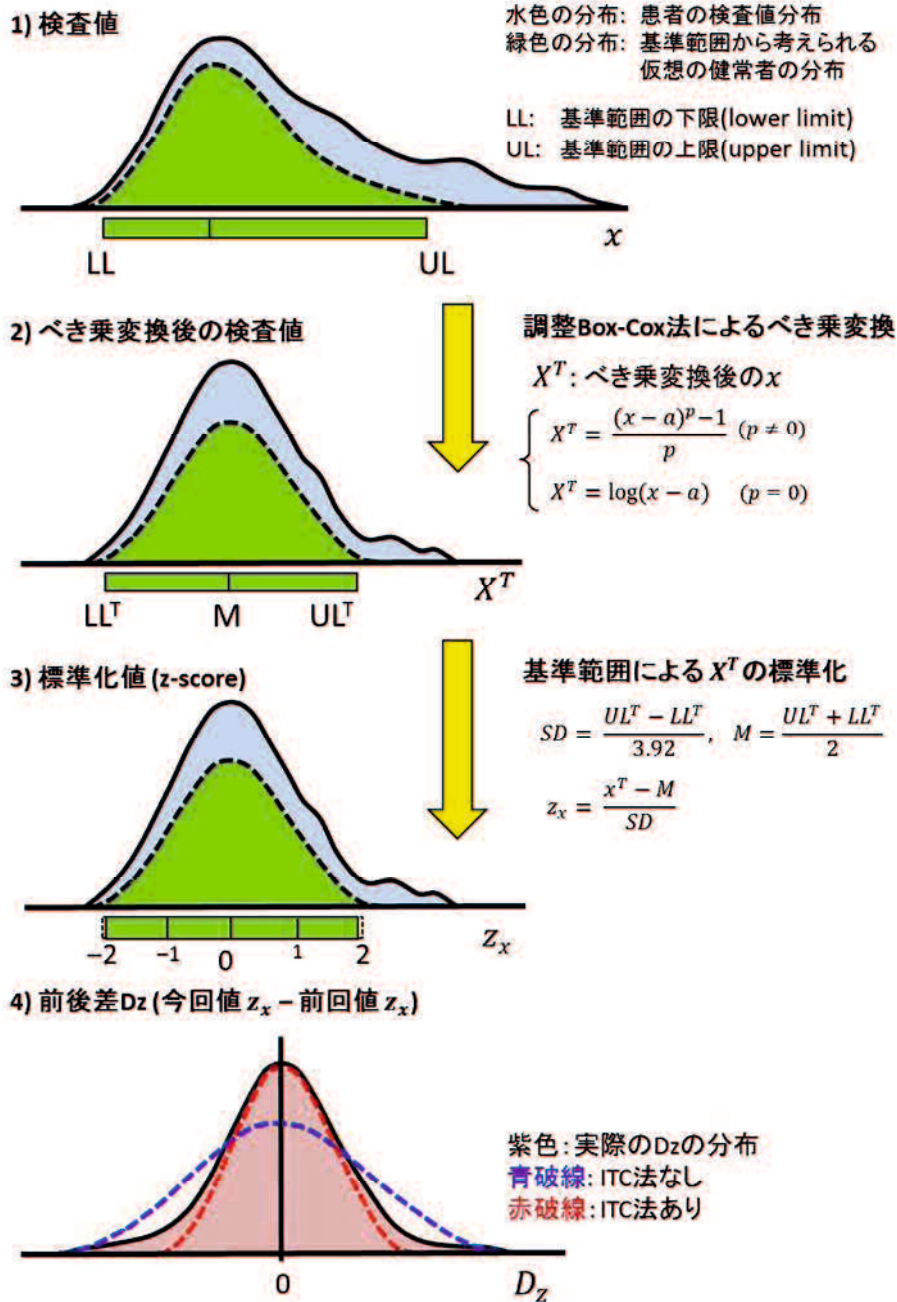


図 3-2 調整 Box-Cox べき乗変換により

wCDC 法にて検査値 x を標準化値 z に変換する過程と、前後差 D_z の分布から反復切断法による調整済みの SD を計算する流れを示した。1)~3)の青色の分布は、LIS から得られる患者の検査値分布を表し、軸の下側に示す緑色の長方形は基準範囲であり、緑色の分布は基準範囲から仮定した健常者の検査値分布を表している。4)の紫色の分布は LIS から抽出した前後の標準化済み検査値の差 D_z の分布を表している。

3-2-1-3 個体内変動 SD の算出

次の手順では、検査項目ごとの個体内変動を調べるために、個人の検査結果について現在から過去に向かって直近の過去の検査値を検索し、今回値と前回値の差を調べていく。この前回値 x_{prev} と今回値 x_{curr} の差は、そのまま検査値の差分をとるのではなく、調整 Box-Cox べき乗変換と、基準範囲による標準化をした標準化値を用いる。変換済みのそれぞれの値を z_{prev} 、 z_{curr} とす

ると、その差 D_z は $D_z = z_{curr} - z_{prev}$ と計算できる。これを全ての前後のペアの値について計算し、それぞれの検査項目の個体内変動を示す D_z の分布として集積する。このとき D_z の分布のSDが検査項目の個体内変動の大きさを示す指標となる。しかしながら、図3-2の4)に示すように、 D_z の分布は左右対称ではあるが、極端な変動が存在により左右の裾の長い分布を形成し、SDは見かけより大きめに計算される。このため、分布周辺の極端値の影響を除き、偏りのない平均値と適切な標準偏差を計算するために、本研究では反復切断補正法(Iterative Truncation and Correction: ITC)を用いた⁹⁾。元々ITC法は、分布の周辺に極端値の集合がよく見られる外部精度評価において、偏りのない平均値と標準偏差を計算するために開発された手法である。

ITC法では、分布の両端に存在するデータの集合を、除外係数 k に沿って、繰り返し処理によって除外する($M \pm k \times SD$)。この除外処理が問題なく動くためには、 $M \pm k \times SD$ によって残された中心部が正規分布していることが前提である。よって、我々はこの手法を用いて D_z の分布から調整されたSD(aSD)を求め、各検査項目の個体内変動幅とした。

3-2-1-4 加重累積デルタチェック指数の計算手順

前項の手順によって得られた個体内変動幅 aSD は、値が小さい検査項目ほど検体取り違いに有効な項目であることを示している。そのため、取り違い検出のために必要となる各検査項目の重要度 w としては、下記の式に示すように aSD²の逆数を用いた。

$$w = \frac{1}{aSD^2} \quad \text{式(3-8)}$$

この重要度 w を用いて、検体取り違いを示す指標は下記の式で表され、これを加重累積デルタ指数(weighed cumulative delta index : wCDI)とし、この指標により取り違いを検出する方法を加重累積デルタチェック法(weighed cumulative delta check method: wCDC法)とした。

$$wCDI = \frac{\sum_{i=1}^k w_i |D_{z_i}|}{\sum_i w_i} \quad \text{式(3-9)}$$

ここに k は取り違い検出に用いる検査項目数を示し、 D_{z_i} は i 番目の検査項目($i = 1, 2, \dots, k$)の前回値と今回値との差を示す。このように wCDI は各検査項目の差の絶対値を、各検査項目の w で加重平均したものとなる。wCDI は、前回値との差 D_z が大きいほど大きな値となる。取り違いが発生した場合には、別人の前回値との差をとることになるため D_z は大きく、wCDI も大きな値となる。よって、得られた wCDI が、正常な前後の検査値から算出される wCDI 値に対して十分に大きな値であるとき、その検体が取り違いであると判定することができる。

3-2-2 検証手法

3-2-2-1 検証用データセット

検証には、1998年に検査情報の教育のために抽出された、匿名化済みのデータセットを用い

た。このデータセットには、22,677名の患者からなる約1年間の171,547件の検査依頼(入院79,307件、外来92,240件)が含まれている。検査項目の種類としては総タンパク(total protein: TP)、アルブミン(albumin: ALB)、ナトリウム(sodium: Na)、カリウム(potassium: K)、クロール(chloride: Cl)、カルシウム(calcium: Ca)、無機リン(inorganic phosphate: IP)、HDL-コレステロール(high-density lipoprotein cholesterol: HDL-C)、LDL-コレステロール(low-density lipoprotein cholesterol: LDL-C)、総コレステロール(total cholesterol: TC)、中性脂肪(triglyceride: TG)、コリンエステラーゼ(choline esterase: CHE)、グルコース(glucose: GLU)、尿酸(uric acid: UA)、アスパラギン酸アミノトランスフェラーゼ(aspartate aminotransferase: AST)、アラニンアミノトランスフェラーゼ(alanine aminotransferase: ALT)、アルカリホスファターゼ(alkaline phosphatase: ALP)、アミラーゼ(amylase: AMY)、乳酸脱水素酵素(lactate dehydrogenase: LDH)、 γ -グルタミルトランスぺプチターゼ(gamma-glutamyltransferase: GGT)、血中尿素窒素(blood urea nitrogen: BUN)、クレアチニン(creatinine: CRE)、赤血球数(red blood cell: RBC)、白血球数(white blood cell: WBC)、血小板数(platelet: PLT)、ヘモグロビン(hemoglobin: Hgb)、ヘマトクリット(hematocrit: Hct)、平均赤血球容積(mean corpuscular volume: MCV)、平均赤血球ヘモグロビン量(mean corpuscular hemoglobin: MCH)、平均赤血球ヘモグロビン濃度(mean corpuscular hemoglobin concentration: MCHC)、活性化部分トロンボプラスチン時間(activated partial thromboplastin time: APTT)、プロトロンビン時間(prothrombin time: PT)の、32項目が含まれている。また、1件あたりの検査項目数は平均17.4項目である。5項目以上の検査値があればwCDIを計算できるとした場合、対象となる依頼件数は137,134件であった。

3-2-2-2 正規性と標準性の検証

1) 調整 Box-Cox ベキ乗変換により、検査値の分布を正規分布に変換できているか評価するために、歪度と尖度を用いた。歪度(Skewness: Sk)は分布の偏りを示す指標であり、対象の分布が完全な正規分布である場合にSk = 0となり、下側に裾が広い場合にSk < 0、上側に裾が広い場合にSk > 0となる。尖度(Kurtosis: Kt)は分布の扁平さを示す指標であり、正規分布であればKt = 0であり、より尖っている場合にKt < 0、より平らである場合にKt > 0となる。また、どちらの指標も絶対値が0.3を超えると正規分布でないといえる。

このSkとKtは、分布の極端値による影響を受けやすいため、下記の式で表される下限(eLL)と上限(eUL)を用いたノンパラメトリック的な除外手法により、極端値を除外した上で計算している。

$$eLL = Q_1 - 3.0 \times (Me - Q_1) \quad \text{式(3-10)}$$

$$eUL = Q_3 + 3.0 \times (Q_3 - Me) \quad \text{式(3-11)}$$

ここにMeとQ₁およびQ₃は、中央値と第1四分位数、第3四分位数を表す。分布の偏りを考慮する除外方法としてはTukeyの方法があるが、Tukeyの方法ではQ₁からの下限値、およびQ₃からの上限値までの幅は、どちらも四分位範囲IQRの1.5倍である。本研究では、より偏りの強い分布に対応するため上記の式を用いた。

2) 個体内変動の計算に、ITC法を用いたが、その除外により的確に正規分布の範囲を抽出できているかを確認するため、 χ^2 適合度検定を用いた。 χ^2 検定では、正規分布との適合性を検定するため、平均値Mと標準偏差SDによるM - 1.6 × SDからM + 1.6 × SDまでの7つの境界値を用い、8つの区分に分布を分割しそれぞれの区分に含まれる観察度数 O_i ($i = 1, 2, \dots, 8$)を数える。正規分布であるときこれらの8つの区分に含まれるデータ数、つまり期待度数 E_i ($i = 1, 2, \dots, 8$)は予測できるため、この観察度数と期待度数の違いに対して下記の式により χ^2 を求めることにより、正規分布との適合性を検定できる¹⁶⁾。

$$\chi^2 = \sum_{i=1}^8 \frac{(O_i - E_i)^2}{E_i} \quad \text{式(3-12)}$$

しかしながら、統計的検定はデータ数が極端に多い場合に鋭敏に反応しやすい。このため、本研究では、対象の分布からn = 200のサンプリングを100回行い、その平均の χ^2 値を用いて検定を行うこととした。

3-2-2-3 wCDC法による判別性能の検証

wCDC法の性能評価では、検証用データセットを用いたシミュレーション手法を用いた。検証の考え方を図3-3に示す。まず、過去の全てのデータについて、前回値と今回値の組み合わせに対する自然なwCDIを計算し、これを自然群(natural group)とする。一方、検体取り違いを模して、ランダムに選択した今回値に対して、同日に測定された別人の前回値との組み合わせを作成し、wCDIを計算したものを人工群(artificial group)とする。wCDIを用いて、自然群と人工群の判別がどの程度可能であるかをROC分析¹³⁾の曲線下面積(AUC: Area Under Curve)により評価し、これをwCDC法の性能とした。

また、検査項目の組み合わせでは、まず一般的に測定される、3種類の検査項目セット(6~9項目)を対象として、それぞれ組み合わせの場合のAUCを計算した。続いて、検査項目数と判別性能の関係を確認するため、検査項目数を変化させた場合のAUCを計算した。

1) 自然な変化の場合(自然群)

$$A_{prev} : x_1, x_2, \dots, x_i$$

$$A_{curr} : x_1, x_2, \dots, x_i$$

自然群の
wCDIの分布

2) 取り違いの場合(人工群)

$$A_{prev} : x_1, x_2, \dots, x_i$$

$$B_{curr} : x_1, x_2, \dots, x_i$$

人工群の
wCDIの分布

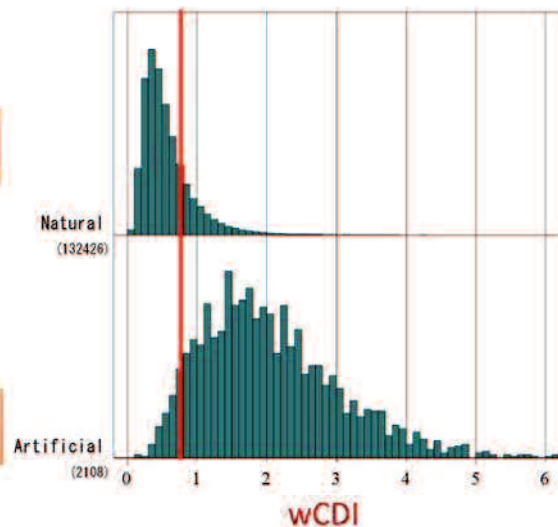


図 3-3 wCDC 法の判別性能検証

wCDC 法による判別性能の評価方法を示した図である。1)は、正常な組み合わせとして同一の個体の前後の検査値から wCDI を計算したものであり、対照群となる。2)は、取り違いとして、別々の人物の組み合わせから wCDI を計算したものであり、判別群となる。

3-3 結果

3-3-1 調整 Box-Cox 変換の検証

調整 Box-Cox べき乗変換による検査値分布の補正について、その性能評価結果を表 3-1 に示す。表中の n^a は各検査項目のデータ数を示し、 p および a はべき乗変換の係数を示す。Skは歪度、Ktは尖度を表し、絶対値が 0.3 を越えるものに記号(*、†)を付与している。変換前後と、分布の両端 1%のデータの除外の有無による効果を確認するために、3 つのケースについて、それぞれSkとKtの計算を行った。Case 1 は変換を行わない元の検査値の分布である。また、Case 2 と Case 3 はそれぞれ、変換前の除外なしの場合と、除外ありの場合のべき乗変換の係数と、SkとKtである。また、Case 3 については、除外後のデータ数 n^b も示した。変換が必要な項目についての効果の確認であるため、無変換で歪度の絶対値が 0.3 を越える項目のみを表した。また、この結果は男性の検査値分布についてのものであるが、女性の検査値分布についても同様の結果であったため女性の結果は省略している。

無変換の場合に歪みのある分布のみを選択しているため、Case 1 では、全ての検査項目が正規分布から外れている。また、Case 2 と Case 3 のように調整 Box-Cox べき乗変換を行った場合には、かなりの割合で正規分布への変換が成功していることがわかる。Case 2 と Case 3 の比較では、Case 3 でより多くの検査項目の変換に成功していることが明らかである。

Case 2 と 3 の変換係数について比較すると、変換の強さを示す係数 p に大きな違いは見られないが、係数 a の値は大きく異なる。係数 a には分布の最小値以上の値を設定できない。除外を行わない Case 2 では、分布から低値側に外れた極端値により、適切な係数 a を設定できていないことを示している。極端値除外処理を行った Case 3 では、最適な係数 a を設定できるため、係数 p が Case 2 と同様でも、良い変換結果を得られている。

表 3-1 調整 Box-Cox 法による正規分布への変換結果

Name	Without truncation							With truncation				
	1) Original			2) Box-Cox transformation				3) Box-Cox transformation				
	n ^a	Sk	Kt	p	a	Sk	Kt	n ^b	p	a	Sk	Kt
WBC	77568	0.48 *	0.13	0.230	0.0	0.08	-0.07	76135	0.251	0.8	0.03	-0.05
TP	46085	-0.42 *	-0.06	2.773	0.0	0.01	-0.30	45247	6.969	1.9	1.12 *	0.81 †
ALB	36826	-0.42 *	-0.46	1.776	0.0	-0.13	-0.72	36198	0.932	1.1	-0.47 *	-0.39
ALP	40974	1.03 *	0.93 †	0.025	0.0	0.50 *	0.13	40164	0.005	101.5	0.10	-0.04
ALT	50371	1.31 *	1.33 †	0.031	0.0	0.48 *	-0.08	49434	0.000	5.8	0.13	-0.19
AMY	19161	0.69 *	0.21	0.036	0.0	-0.01	-0.10	18786	0.180	21.8	-0.09	-0.08
AST	50396	1.23 *	1.28 †	0.042	0.0	0.65 *	0.22	49468	0.012	9.9	0.16	0.01
BUN	74366	0.91 *	0.70 †	0.032	0.0	0.26	0.11	73082	0.000	5.8	0.06	-0.08
CRE	75057	1.11 *	1.42 †	-0.685	0.0	0.48 *	0.27	73638	0.000	0.3	0.48 *	0.36 †
GGT	58351	1.59 *	2.09 †	0.027	0.0	0.67 *	-0.04	57190	0.000	10.0	0.17	-0.31
GLU	21088	1.19 *	1.00 †	0.029	0.0	0.86 *	0.34 †	20682	0.000	67.7	0.24	-0.22
HDL-C	13152	0.49 *	-0.02	0.194	0.0	0.13	-0.24	12900	0.333	19.5	0.00	-0.23
LDH	41535	0.72 *	0.34 †	0.022	0.0	0.34 *	0.00	40706	0.123	100.6	0.00	-0.04
TG	23449	0.99 *	0.56 †	0.026	12.3	0.16	-0.23	22999	0.060	39.0	-0.13	-0.16
Na	49105	-0.97 *	0.99 †	4.435	0.0	-0.90 *	0.67 †	48331	2.025	124.5	-0.42 *	-0.39
Ca	22693	-0.49 *	-0.10	2.000	0.0	-0.38 *	-0.19	22294	4.175	0.0	-0.17	-0.35
APTT	17378	1.09 *	0.81 †	1.224	0.0	1.15 *	0.94 †	17033	0.080	21.4	0.06	-0.11

^a Number of original dataset

Sk: Skewness (* |Sk|>0.3)

^b Number of truncated dataset

Kt: Kurtosis († |Kt|>0.3)

3-3-2 個体内変動 SD

各検査項目の Dz の分布について個体内変動の大きさを示す SD を計算した。ITC 法を用いなかった場合と、用いた場合の 32 項目の平均値と SD と χ^2 値を表 3-2 に示す。表 2 では、ITC 法を使用した場合の SD が、使用しなかった場合の SD に対して明確に小さい値であることが示されている。また χ^2 適合度の検定結果でも、Dz の分布の中心を ITC 法によりの確に捉え、除外を行った結果が大きく正規分布から外れていないことを示している。結果は男性の分布に対するものであるが、女性でも同じ傾向を示している。

6 つの代表的な検査項目(ALB, ALP, Ca, Cl, GGT, GLU)の Dz の分布を図 3-4 に示す。Dz の分布は 0 を中心とした左右対称な曲線となるが、正規分布と比較して分布の裾野が長い形状となっている。図中では、ITC 法を未使用の場合の平均値と標準偏差を用いた理論正規分布の曲線を黒色、ITC 法を使用した場合の理論正規分布を赤色で示している。後者の理論曲線は Dz の分布に限りなく一致しており、ITC 法により極端値の影響を除外することにより分布の正確な SD(adjusted SD: aSD)を計算できていることが分かる。つまり Box-Cox べき乗変換と標準化により、各検査項目の個体内変動の大きさを aSD として比較が可能となる。表 3-2 より aSD の小さな検査項目としては順に、TP、MCV、ALP、MCH、ChE となる。

表 3-2 ITC 法の有無による個体内変動 SD の比較

Name	Without ITC method			With ITC method		
	Mean	SD	χ^2	Mean	aSD	χ^2
RBC	0.05	1.09	23.99 *	0.03	0.71	9.78
WBC	0.03	1.46	36.31 *	0.07	0.77	9.95
PLT	-0.02	0.92	42.14 *	0.00	0.46	10.74
Hct	0.05	1.09	21.90 *	0.02	0.73	9.80
Hgb	0.07	1.28	25.86 *	0.04	0.81	10.26
MCH	0.01	0.37	20.89 *	0.01	0.25	10.08
MCHC	0.01	0.38	12.05 *	0.01	0.33	10.37
MCV	0.00	0.51	60.47 *	-0.01	0.21	9.86
Na	-0.02	0.95	16.73 *	-0.01	0.73	10.08
K	-0.03	1.36	15.28 *	-0.04	1.03	10.80
Cl	0.00	1.04	14.76 *	0.00	0.83	9.95
Ca	0.02	1.33	64.08 *	-0.01	0.66	10.08
IP	0.06	1.88	73.92 *	0.00	0.82	9.82
TP	0.02	1.06	107.13 *	-0.02	0.20	12.41 *
ALB	0.05	1.10	21.90 *	-0.02	0.75	9.50
TC	0.01	0.65	14.97 *	0.00	0.51	10.92
HDL-C	-0.01	0.60	15.43 *	0.00	0.47	10.92
LDL-C	0.02	0.78	16.74 *	0.01	0.59	10.48
ALP	0.00	0.50	62.23 *	0.02	0.23	9.58
ALT	0.00	0.82	36.37 *	0.05	0.45	10.10
AMY	-0.01	1.22	74.66 *	0.00	0.46	10.14
AST	0.02	1.08	37.83 *	0.04	0.57	9.93
BUN	0.00	1.17	40.24 *	0.00	0.61	10.52
CHE	0.01	0.42	25.89 *	0.00	0.26	10.07
CRE	0.01	1.04	54.67 *	0.01	0.51	9.97
GGT	0.00	0.78	59.82 *	0.05	0.34	11.32 *
GLU	0.04	1.76	60.80 *	0.02	0.74	10.42
LDH	0.04	0.91	44.28 *	0.04	0.46	9.42
TG	0.01	0.79	19.01 *	0.01	0.55	10.12
UA	0.03	1.11	72.94 *	0.00	0.51	10.32
APTT	-0.04	0.75	47.60 *	0.00	0.35	9.52
PT(%)	0.03	1.46	30.06 *	-0.01	0.86	10.54

*P<0.05, $\chi^2_{(df=5, P=0.05)} = 11.07$

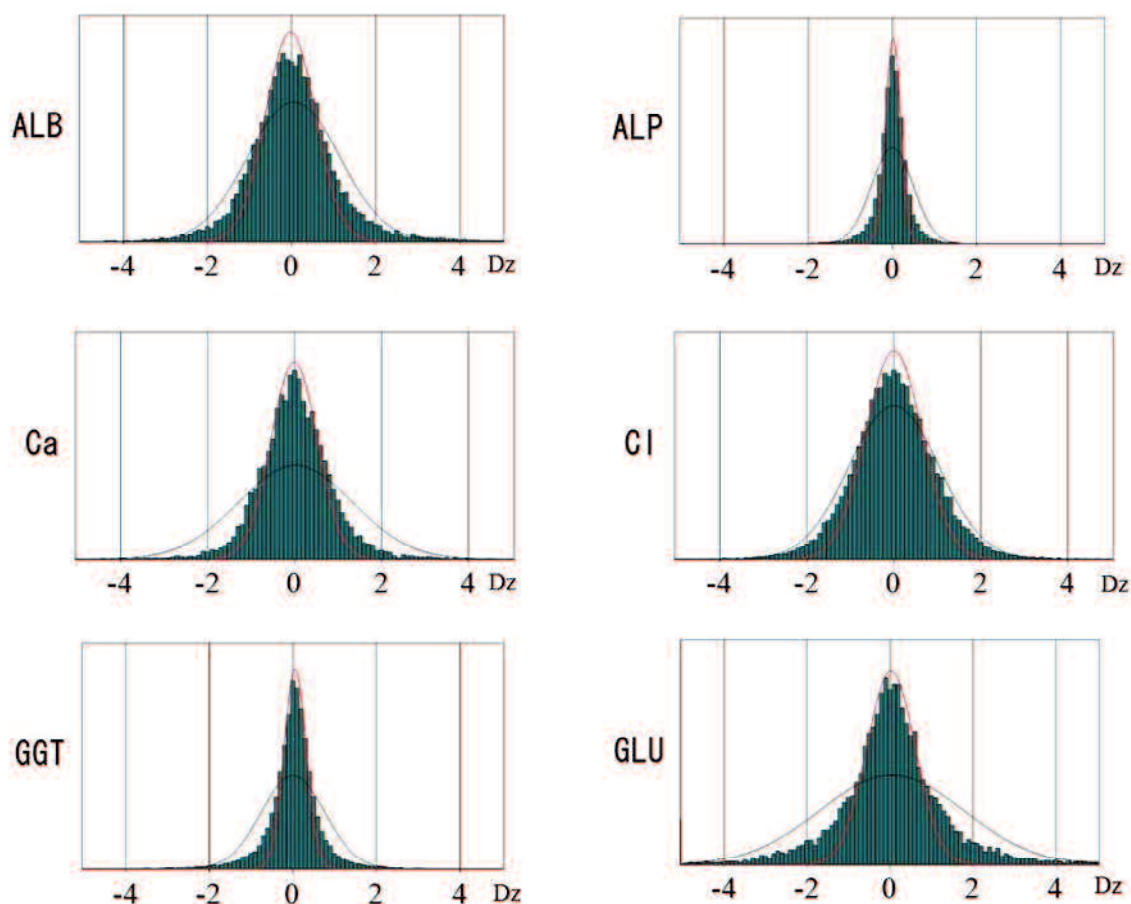


図 3-4 前回値との差 Dz の分布

6つの検査項目について、同一個体のデルタ Dz（標準化済みの前回値と今回値の差）の分布を示す。黒色の曲線は、ITC 法を用いずに計算された平均値と標準偏差によって描画した正規分布の理論曲線であり、極端値により標準偏差が広めに計算されていることが分かる。赤色の曲線は ITC 法により調整済みの平均値と標準偏差で描画した理論曲線であり、調整後の SD が妥当であること示されている。

3-3-3 wCDC による取り違い検出性能

3-3-3-1 固定検査項目セットでの検証

wCDI による取り違い検出の性能評価を行うために、人工的に取り違い例を作成し取り違いをシミュレートする手法を用いた。まず、使用する検査項目を固定し、セットとして依頼される一般的な 3つの組み合わせについて wCDI の性能評価を行った。使用した検査項目は Set 1 (WBC, RBC, Hb, Ht, PLT, MCV, MCH, MCHC)、Set 2 (TP, Alb, BUN, CRE, UA, Na, K, Cl, Ca)、Set 3 (ALT, AST, LDH, ALP, GGT, TP)とした(表 3-3 A)。

3つの組み合わせについての wCDI による、人工群と自然群の判別性能である AUC は、0.937～0.967 と高値を示している。自然群の偽陰性率(false-positive rate :FPR)を 5.0、7.5、10%と設定した場合の人工群の感度はそれぞれの組み合わせで 63.2–84.8%、74.0–89.2%、80.7–91.6%となっている。また、wCDI の計算において aSD による加重をしなかった場合の判別性能を表 3-3 B に示す。加重なしの場合では十分な精度が得られていない。

表 3-3 wCDC 法による性能評価（検査項目固定）

A) Fixed test item (with weight)

	Data size		AUC	Sensitivity % (cut-off value)		
	Artificial	Natural		FP=5%	FP=7.5%	FP=10%
Set1 (WBC, RBC, Hb, Ht, PLT, MCV, MCH, MCHC)	126211	105691	0.967	84.8% (0.82)	89.2% (0.73)	91.6% (0.67)
Set2 (TP, Alb, BUN, CRE, UA, Na, K, Cl, Ca)	4087	10205	0.953	74.8% (1.13)	81.7% (1.02)	86.2% (0.94)
Set3 (ALT, AST, LDH, ALP, GGT, TP)	36327	39255	0.937	63.2% (1.07)	74.0% (0.92)	80.7% (0.82)

B) Fixed test item (without weight)

	Data size		AUC	Sensitivity % (cut-off value)		
	Artificial	Natural		FP=5%	FP=7.5%	FP=10%
Set1 (WBC, RBC, Hb, Ht, PLT, MCV, MCH, MCHC)	126211	105691	0.952	76.2% (1.21)	82.5% (1.07)	86.2% (0.98)
Set2 (TP, Alb, BUN, CRE, UA, Na, K, Cl, Ca)	4087	10205	0.944	70.0% (1.17)	77.8% (1.06)	83.5% (0.98)
Set3 (ALT, AST, LDH, ALP, GGT, TP)	36327	39255	0.933	61.2% (1.37)	71.8% (1.19)	78.8% (1.07)

FP: False positive

3-3-3-2 検査項目の動的組み合わせの検証

我々は、任意の検査項目の組み合わせについて wCDC 法の性能評価を行った。検査項目固定の検討と同様に、人工的に作成した取り違い群を用いたシミュレーション手法を用いた。性能は、組み合わせに使用する検査項目数を変え、それぞれの場合に wCDI による自然群と人工群の層別化がどれだけ可能であるか評価した。図 3-5A には、加重を使用した場合の検査項目数ごとの AUC(実線)と、加重を使用しない場合の AUC(破線)を示した。図 3-5B には、偽陰性率を 5, 7.5, 10%と変えた場合の項目数ごとの wCDI のカットオフ値を示し、図 3-5C にはそのときの感度の値を示した。図 3-5A の AUC と図 3-5C の感度をみると、項目数 k が 10 未満の範囲では、項目数に従った取り違い検出性能の上昇がみられるが、項目数 k が 10 以上では大きな変化はみられない。また図 3-5B のカットオフ値も、 k が 10 となるところまでは減少しているが、 k が 10 以上では大きな変化は見られない。それゆえ、 k が 10 以上である場合には、検査項目数を問わず偽陰性率 5, 7.5, 10%に対するカットオフ値は 0.9, 0.83, 0.75 とほぼ一定といえる。これらの結果より、十分な検査項目数がある場合には、検査項目の組み合わせによるカットオフ値の設定は必要でないため、あらかじめ全ての検査項目の組み合わせに対するカットオフ値を計算する必要が無いことが分かった。

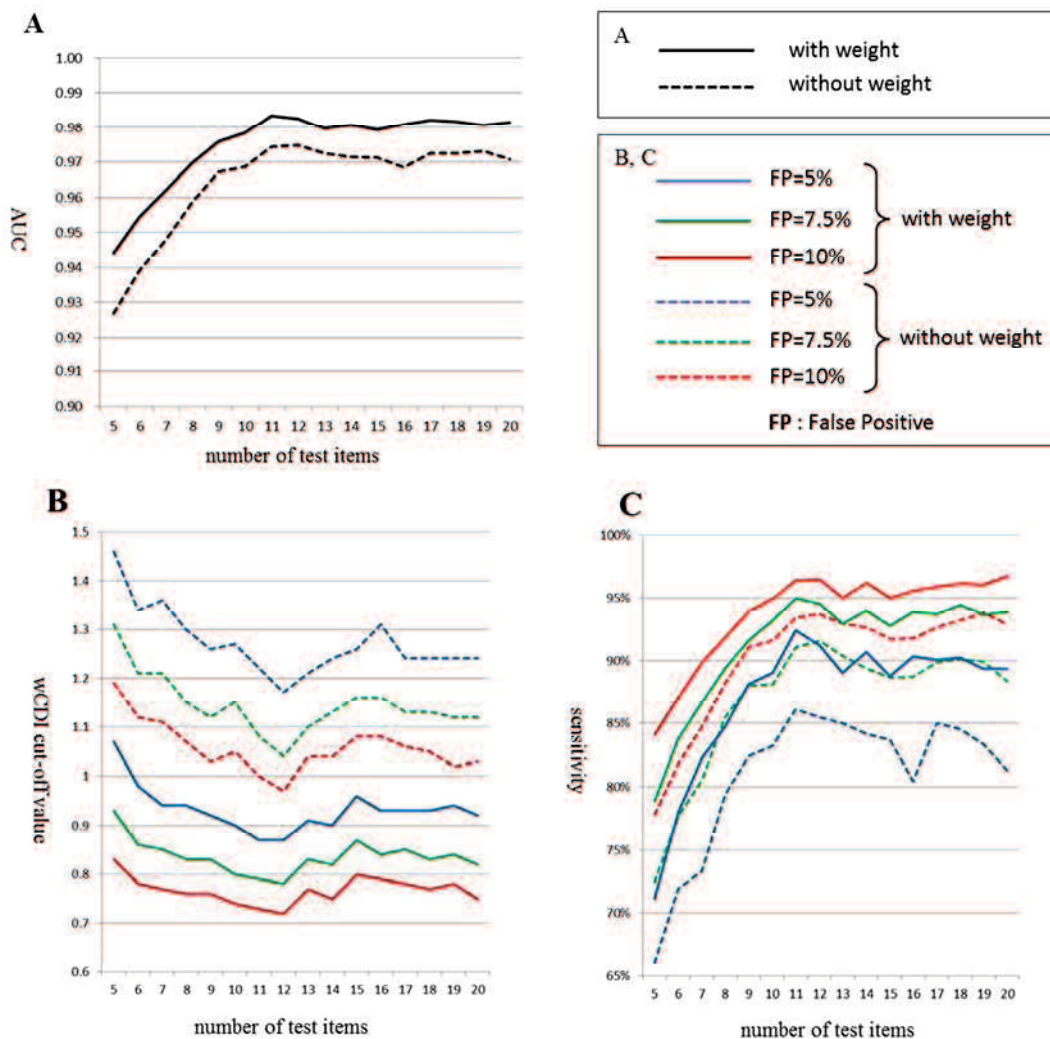


図 3-5 wCDC 法による性能評価（検査項目可変）

検査項目数が変化した場合の wCDC 法の性能評価を示す。(A)は、ウエイトの有無による判別性能の違いを示したグラフであり、実線がウエイト有り、点線がウエイト無しである。(B)および(C)は、それぞれ偽陽性率を 5%(赤色)、7.5%(緑色)、10%(青色)とした場合の、カットオフ値と、感度(sensitivity)である。A と同様に実線がウエイト有り、点線がウエイト無しの場合となっている。

3-4 考察

過去には、情報技術を用いて、日常臨床検査データに含まれる検体取り違いを検出するための様々な取り組みが行われてきた。しかしながら、現実の臨床検査データでは多くの極端値が含まれているため、単純な統計的手法による検体取り違いの検出には様々な問題があった。我々は様々なデータ分析手法を複合的に用いることにより、これらの問題に対応した。

まず、検査値の分布の形状が項目によって異なるため、分布の両端 1%ずつを除外し、調整 Box-Cox べき乗変換^{12,14)}を用いて単一の形状である正規分布への変換を行った。この除外処理を追加したことにより、的確な変換係数を推定できたため、ほぼ全ての検査項目を正規分布に変換することに成功した。患者由来の検査値分布は、その極端値の影響によりそのままでは分布の変換が行えないと推測していたが、単純な極端値除外の有無により歴然と変換の成功率が変化したこと

は非常に興味深い事実である。また、過去にも臨床検査値の標準化を試みた例があったが、いずれも実用化には至らなかった。これは、分布の形状を考慮していないため、低値側と高値側での値の違いを適切に表現できないためであった。しかしながら今回、全ての検査項目を正規分布に変換でき、かつ基準範囲を標準化の目安としたことで、どの検査項目も臨床的な意味を含む新たな尺度に変換することが可能となり、真の統一化表示に成功したことは大きな成果といえる。

また、分布の正規性の検証では、歪度と尖度を評価指標として使用したが、これらの計算式には偏差を三乗または四乗する項が存在する。これにより極端値が計算結果に大きく影響するため、極端値を除外することが必要とされている。本研究では、この極端値処理に Tukey の方法に改良を加えたものを利用した。Tukey の方法では第 1 四分位偏差(Q1)と第 3 四分位数(Q3)と、四分位範囲(IQR = Q3 - Q1)を用いて、下限を $Q1 - 1.5 \times IQR$ 、上限を $Q3 + 1.5 \times IQR$ としてこれを外れたものを除外する。この手法の場合、除外範囲の起点を、下限は Q1、上限は Q3 とすることで分布の非対称性に対応しているが、検査値特有の極端に歪んだ分布では十分とはいえない。本研究では、分布の非対称性への対応を強化するため、中央値(Me)を用いて、下限を $Q1 - 3.0 \times (Me - Q1)$ 、上限を $Q3 + 3.0 \times (Q3 - Me)$ とする除外範囲を用いた。この範囲は、非対称度の強い分布では、Tukey の方法よりその分布に適合するが、正規分布の場合には Tukey の方法や 3σ 法と同じ範囲を除外する。臨床検査値を対象とした除外では、このような考え方は重要であると考えている。

次に、検査値を標準化値 z に変換することについて述べる。本研究では、検査値の単位や、取り得る範囲の違いを除くために、基準範囲を用いて標準化を行う手法を用いた。この手法により、標準化値 z は、基準範囲の幅を -2 から 2 とした場合の位置を示す値となるため、臨床的に解釈しやすくなる。つまり、どの検査であっても z が 0 であれば基準範囲のちょうど真ん中に位置しており、 2 を越えると基準範囲を高値側に外れていることがわかる。図 3-6 に、標準化値で表したときに可能となるグラフの例を示す。図の A では、前回の検査値を黒丸、今回の検査値を白丸で表している。例えば ALP の値が基準範囲内から基準範囲外に変化しているが、分布から考えると変化量は非常に小さいことが分かる。図の B では、前回値からの変化量を矢印の向きと長さで示し、検査値ごとの個体内変動幅を青色の棒で示している。このように見ると、ALT 以外は個体内の変動の範囲で変化しているが、ALT は個体内変動幅で考えられるより大きな変化となってい

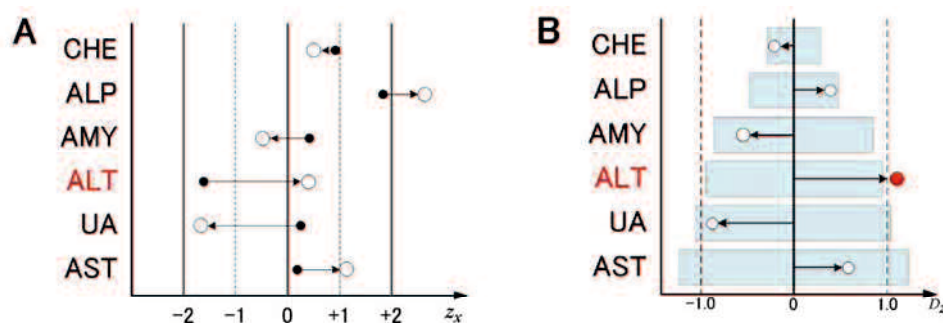


図 3-6 標準化値による検査値の表示例

wCDC 法により猥体取り違いが検出された場合に、標準化値によりどの検査に以上が含まれるのかを確認する方法として、2つの表示方法を提案する。(A)では、前回の検査の標準化値を黒丸、今回の標準化値を白丸で表し、基準範囲を -2 から 2 として表現したものである。(B)では、各検査値の前後の変化量を矢印の長さで表し、検査項目ごとの個体内変動幅を青色の棒で表している。

る。特に図の B の表示では、取り違いとして検出された場合に、どの検査項目が変化しているが把握することに有効と考えられる。

wCDC 法を実現するために、最も重要なことは、各検査項目の個体内変動幅を推定することである。患者データを対象とした場合、前後の検査値の間には、手術や透析など、自然にはない極端な変動が含まれる。そのため、単純に前後の差 Dz の分布を作成すると、連続的で左右対称な形状を示すが、その両裾は非常に長い。このため、単純に SD を計算しても、自然な個体内変動を表す値とはならない。しかしながら、ITC 法¹³⁾によって分布の両端を除外することで、 χ^2 適合度検定にてほぼ正規分布といえる範囲を抽出できることが分かった。ITC 法は元々、外部精度管理において、極端値による検査値分布の偏りの影響を除外した平均値（中心）を抽出するために開発された手法である。

Dz の中心部分の正規性の確認では、 χ^2 適合度検定を用いた。しかしながら、統計的検定手法はデータ数が多い場合に極端に感度が高くなる傾向が知られている¹⁸⁾。実際に、検定に使用するデータ数が m 倍となる場合を考えると、自由度および有意となる境界の χ^2 値は変化しないが、計算される検定統計量 χ^2 値は m 倍となる。

$$\chi^2 = \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad \text{式(3-13)}$$

$$\chi^2 = \sum_{i=1}^k \frac{(m \times O_i - m \times E_i)^2}{m \times E_i} = m \times \sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i} \quad \text{式(3-14)}$$

この統計的検定の性質は、今回使用するような膨大なデータに対しては問題となる。このため我々は、分布から繰り返し小規模のサンプルを抽出して χ^2 値を計算し、それらの平均値を用いて判定を行う方法を採用した。この方法により、ITC 法を用いた場合と、そうでない場合の分布のパターンの客観的な判断を行った。この変更は、ITC の有無で分布中心部の正規分布へのフィッティングに違いが大きく改善することを示すことを目的としているため、適切な範囲内であると考えている。

新しいデルタチェック法の一般への適用性に確認するために、外来患者に限定しないものと限定したものについて、それぞれ人工的に取り違い群を作成し性能を評価した。正常でない検査値を有する割合は異なるが、判別性能に大きな違いは見られなかった。これは、検査値分布の正規分布への変換により、疾患特有の極端な変化の影響が押さえられているからであることを示している。

もう一つ重要な点としては、新しいデルタチェック法で比較する今回値と前回値の測定間隔の許容性が挙げられる。前回の測定から今回の測定までの期間が大きい場合、年齢による変化の大きい検査項目、特に小児や高齢者の集団などでは、性能に影響することが考えられる。それゆえ、今回のシステムでは、最長の間隔を 1 年と限定した。さらに、ユーザーが測定の間隔について判断を行えるように、システムでは自動的に前回値との間隔に関する情報を提供する。また、システムでは検査項目毎の SD を自動的に更新するように動作できるため、検査システムや患者の変

化にも対応できる。

加重累積デルタチェック法では、実用化のために一施設で行った 2 ヶ月間の予備検討において、検体の取り違い例が 2 件検出されている。また研究段階で予測していたとおり、取り違い以外の例として、手術前後、透析前後、大きな外傷前後、輸血前後といった、生体の機能に急激な変化がおこったケースも取り違いの疑いとして検出されている。今後の課題としては、この検出された極端な変動のパターンを学習し、自動的に分類することである。これらの手術、透析、外傷、輸血などは、変化の生じる検査項目に限られるため、比較的容易に自動分類が可能であると考えられる。このような取り違い以外の変動を自動的に除外することで、検体取り違い疑いとして検査技師がチェックする対象を減らすことができるため、さらなる業務支援となると考えている。

加重累積デルタチェック法のさらなる改良点としては、wCDI 値の計算では、検査項目間の相関関係について含まれていないため、これらを考慮した計算方法が行うことで、より精度を向上できる可能性があることである。これについては、既存研究にあるマハラノビス距離による取り違い検出手法について、本研究において使用した検査値分布の正規化および標準化の手法を導入することにより、相関を考慮した判定が行えると考えている。しかしながら、wCDI の計算では、検査項目の組み合わせを柔軟に変更可能である点と、組み合わせの変更時の検査項目数が変化しても一定のカットオフ値を使用できる点は本研究の長所といえるだろう。

本章の参考文献

- 1) 日本医療機能評価機構 医療事故情報収集等事業 平成 18 年 年報
http://www.med-safe.jp/pdf/year_report_2006.pdf
- 2) 山田 千亜希, 藤原 晴美, 小川 奈緒子. 当院における採血患者誤認によるインシデント 6 件の解析と今後の対策. 臨床病理 2011; 59: 42-5
- 3) 永井 聖也, 山田 千亜希, 藤原 晴美. 採血患者誤認を契機とした検体照合システムの導入と技師の病棟採血への参加 -臨床側との連携をめざして-. 臨床病理 2014; 62: 749-54
- 4) 大峠 和彦. 採血業務および検体管理に於ける患者誤認対策. 臨床病理 2013; 62: 739-44
- 5) Quillen K, Murphy K. Quality improvement to decrease specimen mislabeling in transfusion medicine. Arch Pathol Lab Med 2006;130:1196-8.
- 6) Morrison AP, Tanasijevic MJ, Goonan EM, Lobo MM, Bates MM, Lipsitz SR, et al. Reduction in specimen labeling errors after implementation of a positive patient identification system in phlebotomy. Am J Clin Pathol 2010;133:870-7.
- 7) Carraro P, Zago T, Plebani M. Exploring the initial steps of the testing process: frequency and nature of pre-preanalytic errors. Clin Chem 2012;58:638-42.
- 8) Dunn EJ, Moga PJ. Patient misidentification in laboratory medicine: a qualitative analysis of 227 root cause analysis reports in the Veterans Health Administration. Arch Pathol Lab Med 2010;134:244-55.
- 9) Nosanchuk JS, Gottmann AW. CUMS and delta checks. A systematic approach to quality control. Am J Clin Path 1974;62:707-12.
- 10) Sheiner LB, Wheeler LA, Moore JK. The performance of delta check methods. Clin Chem 1979; 25:2034-7.
- 11) Iizuka Y, Kume H, Kitamura M. Multivariate delta check method for detecting specimen mix-up. Clin Chem 1982;28:2244-8.
- 12) Ichihara K, Boyd JC; IFCC Committee on Reference Intervals and Decision Limits (C-RIDL). An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537-51.
- 13) Ichihara K, Kawai T. An iterative method for improved estimation of the mean of peer-group distributions in proficiency testing. Clin Chem Lab Med 2005;43:412-21.
- 14) Ichihara K, Kawai T. Determination of reference intervals for 13 plasma proteins based on IFCC international reference preparation (CRM470) and NCCLS proposed guideline (C28-P,1992): trial to select reference individuals by results of screening tests and application of maximal likelihood method. J Clin Lab Anal 1996;10:110-7.
- 15) Sokal RR, Rohlf FJ. The normal probability distribution. In: Biometry. 2nd Ed. New York: WH Freeman & Company 1981:98-127.
- 16) Bland M. The chi-squared goodness of fit test. In: An introduction to medical statistics. 2nd Ed. New York: Oxford University Press Inc. 1995:244-5.

- 17) Zweig MH, Campbell G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 1993;39:561–77.
- 18) Boero G, Smith J, Wallis KF. The sensitivity of chi-squared goodness-of-fit tests to the partitioning of data. *Econometric Reviews* 2004;23:341–70.

第4章 基本検査による男女判別の正確度：多重ロジスティック回帰分析法と加重対数尤度診断法の比較検討

4-1 はじめに

我々は、複数の検査項目について前回値との差を組み合わせることで検体の取り違いを検出する手法の開発を行ってきた。しかしながら、前回値がない場合には使用できないなど実用での限界があるため、検体取り違い検出を行える補助診断的な手法の開発を目指してきた。臨床検査値の多くには性差があることが知られている。このため、測定した検査値と、性別との食い違いを検出することで、取り違いの可能性が高くなり取り違い検出の補助手段として利用できるのではないかと考えた。

医療分野におけるコンピュータを用いた診断手法では、複数の疾患に対応可能という利点から、人工ニューラルネットワーク (ANN) を用いた試みが行われていた。しかしながら、判断論理の明確な表現が難しくブラックボックス化してしまう、他の研究者が学習結果を再利用するためには同じ構造の ANN を構築する必要がある、中間層のユニット数や各種パラメータの初期値などの調整が必要であり取り扱いが難しい、といった課題があった。そのため臨床検査分野では、判別方程式、近年では多重ロジスティック回帰分析法²⁾が広く用いられている。

しかしながら、臨床検査項目の分布のいくつかは歪んでおり、パラメトリックな手法である多重ロジスティック回帰分析では分布型の補正や極端値の処理が必須である。また、実用化を目指す上では極端値の影響なども問題となることが考えられるため、ノンパラメトリックな方法が有効となるのではないかと考えた。そのため、本研究ではノンパラメトリックな手法である市原の開発した対数尤度診断法³⁾を応用した手法と、パラメトリックな手法である多重ロジスティック回帰分析法と比較し検討を行うこととした。

加重対数尤度診断法は、特定の診断すべき群（本課題では、男性または女性の集団）の検査値の分布そのものを尤度関数（確率密度関数）と見なし、その群の検査値パターンとの合致度を対数尤度和として求める方式である。もしも、任意の個体の検査値パターンが、特定の診断すべき群のパターンと適合している場合には、個々の検査の尤度が高くなるため、対数尤度和も高値となる。逆に、適合していない個体では、対数尤度和は低値となる。この原理を利用して、今回は、健常男性の多項目検査セットを尤度関数群として、任意の個体について、各検査値分布に対する尤度を求め、それを各検査の“男女差指数”を重みとして累和して、加重平均対数尤度として算出する方法を考案した。

4-2 方法

4-2-1 解析対象

本研究では、共有基準範囲設定のために 2009 年に東・東南アジア地域で行われた国際調査で収集された健常者のデータセットのうち、日本国内から得られた 1,944 件（年齢 20～64 歳、男性 856 例、女性 1088 例）のデータを用いた^{4,5)}。これは、協力者の同意を得て収集され、連結化不可能形で匿名化されたデータである。分析対象は、日常最も広く利用される基本検査のうち、以下に示す生化学 21 項目に限定した：総蛋白(TP), アルブミン(ALB), Na, K, Cl, 総 Ca, 無機リ

ン IP, 尿素窒素(BUN), creatinine (CRE), 尿酸 (UA), 総コレステロール(TC), HDL コレステロール (HDL-C), LDL コレステロール (LDL-C), 中性脂肪(TG), ブドウ糖(GLU), AST, ALT, LDH, ALP, GGT, アミラーゼ(AMY)

4-2-2 男女差指数

基準個体のデータを、男女で層別化し、次式により男女差指数（男女差を表す SD 比）を計算した^{6,7)}。

$$\text{男女差指数 : SDR} = \frac{SD_{\text{性別}}}{SD_{RI}} \quad \text{式(4-1)}$$

ここに分子の $SD_{\text{性別}}$ は、男女差を表す標準偏差 SD を表し、分母の SD_{RI} は性差調整後の基準範囲(RI)の変動域を表す SD に相当する。それぞれの SD は一元配置分散分析法により求めた⁸⁾。また、加齢とともに性差が変化する項目が多いため⁴⁾、45 歳を境界として 45 歳以下と、46 歳以上に分けた場合の男女差指数も計算した。

4-2-3 多重ロジスティック回帰分析法

性別（男性 0、女性 1）を目的変数、年齢および検査値を説明変数として回帰係数の推定を行った。検査項目の組み合わせについては、臨界とする有意確率を $P=0.05$ とし、ステップワイズ法により判別に有効な検査項目を選択した。年代により男女判別可能な検査項目が異なることが考えられるため、全年代に加え、45 歳以下のみ、46 歳以上のみの 3 つの条件に対して回帰係数の計算と判別性能の評価を行った。なお、測定値の分布に歪みのある検査項目 TG, GLU, AST, ALT, GGT では、あらかじめ測定値を対数変換してから分析を行った。

また、回帰分析ではデータ数が少ない場合に過剰適合(over fitting)が起こることが知られている。そこで本研究では、データセットの 3 分の 2 を判別式作成用の教師群データに、残りを判別式の性能評価用の検証群データとなるよう、全データをランダムに 2 群に振り分けた。

4-2-4 加重対数尤度診断法

加重対数尤度診断法は、複数のパラメータ（検査所見・臨床所見）を組み合わせ、臨床検査診断を任意の条件で行うために市原らによって考案された動的診断技法である³⁾。これには特定の診断群(diagnostic category)毎に、複数のパラメータを登録しておき、各検査項目の測定値の分布を尤度関数と考えて、任意の症例の特定のパラメータの測定値を与えたときに、その値が分布とどの程度マッチングしているかを尤度 L（下側累積確率または上側累積確率）として表す。本研究の男女判別では、まず検査項目毎の男性の分布を作成し、この男性の分布を用いて尤度を計算する。すなわち、男性に比べて女性の分布が下側に位置する場合には、男性の分布の下側累積確率を尤度とする。これを任意の p 個のパラメータに対して行い、各々の尤度 L_i ($i=1\sim p$) の対数を、各々の男女差指数を重み W_i として加重平均したものを加重平均対数尤度 (weighted average log-likelihood: WALL) 指数とした。

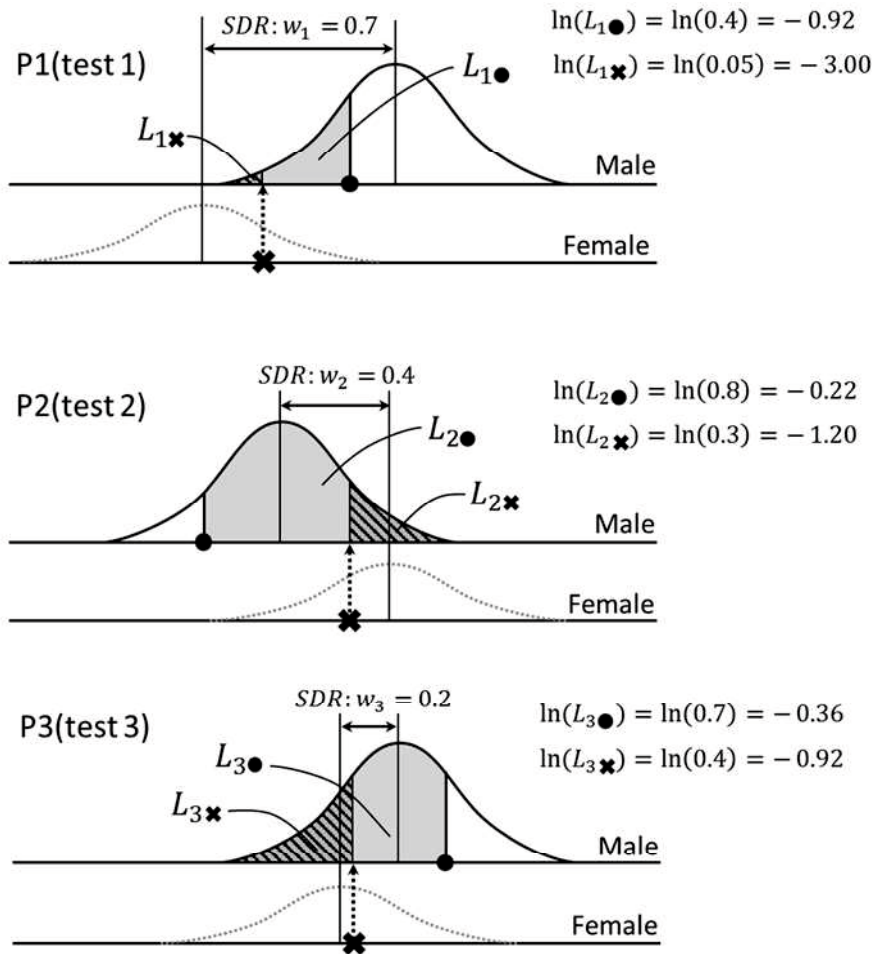
$$WALL = \frac{\sum_{i=1}^p w_i \ln(L_i)}{\sum_{i=1}^p w_i} \quad \text{式(4-2)}$$

図 4-1 に、WALL の計算例として、男性 (●)、女性 (×) の 2 症例について、P1、P2、P3 の 3 つのパラメータによる計算の手順を示す。各パラメータは性別によって層別化され、男性の分布を実線、女性の分布を破線で表している。ここで WALL の計算に必要な各パラメータの重み w_1 、 w_2 、 w_3 として、男女差指数を設定した。この例では、 $w_1 > w_2 > w_3$ であるため、P1 が最も男女判別に有用な検査項目となる。各パラメータに対する男性例●の尤度は、分布全体の面積に対する着色部の面積であり、女性例×の尤度は、分布全体の面積に対する斜線部の面積である。ここで、男性の分布を基準にそれとの適合度を示す尤度（男性尤度）を計算しているため、計測値が女性 < 男性となる場合（P1、P3）では、男性群分布の下側累積確率を尤度とした。逆に、計測値が男性 < 女性となる場合（P2）では、男性群分布の上側累積確率を尤度とすることで、男女差が WALL 値によりの確に反映されるようにした。

各パラメータに対する尤度の対数を男女差指数によって加重平均したものが WALL である。どのパラメータの尤度も $p \leq 1.0$ であり、 $\ln(p) \leq 0$ となるため、WALL 値は常に 0 以下の値をとる。例示した男性の WALL 値は -0.669、女性例の WALL 値は -2.403 となっており、概して男性では WALL 値はより 0 に近い負の値を、女性ではより 0 から離れた負の値となりやすいことが分かる。すなわち、WALL 値が一定のカットオフ値以上であれば男性、カットオフ値未満であれば女性として、男女の判別を行うことができる。

WALL 値の計算では男女差指数での重み付けを行うことにより、判別に有効でない検査項目の貢献度を小さくできる。しかし、男女差指数の大きな項目に限定して WALL 値を計算することで、より高い判別性能を得る可能性がある。そこで、この点を検討するために、全ての検査項目を用いた場合と、男女差指数 ≥ 0.3 、 ≥ 0.5 の検査項目に限定した場合の 3 通りについてそれぞれ男女の判別度を計算し、その性能の比較も行った。また、多重ロジスティック回帰分析と同様に、年代について 3 群による比較も行った。

なお、対数尤度診断法においても、多重ロジスティック回帰分析法と同様に教師群データと検証群データにランダムに分けて判別の性能評価を行った。



● : a male case

$$\begin{aligned} \text{WALL}(\bullet) &= \frac{\sum_{i=1}^p w_i \ln(L_{i\bullet})}{\sum_{i=1}^p w_i} = \frac{w_1 \ln(L_{1\bullet}) + w_2 \ln(L_{2\bullet}) + w_3 \ln(L_{3\bullet})}{w_1 + w_2 + w_3} \\ &= -\frac{0.7 \times 0.92 + 0.4 \times 0.22 + 0.2 \times 0.36}{0.7 + 0.4 + 0.2} = -0.669 \end{aligned}$$

✕ : a female case

$$\begin{aligned} \text{WALL}(\times) &= \frac{\sum_{i=1}^p w_i \ln(L_{i\times})}{\sum_{i=1}^p w_i} = \frac{w_1 \ln(L_{1\times}) + w_2 \ln(L_{2\times}) + w_3 \ln(L_{3\times})}{w_1 + w_2 + w_3} \\ &= -\frac{0.7 \times 3.00 + 0.4 \times 1.20 + 0.2 \times 0.92}{0.7 + 0.4 + 0.2} = -2.403 \end{aligned}$$

図 4-1 WALL 係数の計算例

WALL 値は、任意の検査値の組み合わせが、診断群にどれだけ適合しているかを表す値である。本研究は、男性を診断群としているため、男女差を示す各検査項目の SDR を加重として、男性の各検査値分布への適合度を平均した値を WALL 値としている。

この図は、性差のある 3 つの検査例を用いて、男性の検査値セット(●)と、女性の検査値セット(✕)についての WALL 値の計算の例を示した。

女性に比べて男性の分布が高値側にある場合は、男性の検査値分布での下側累積確率を尤度(L)とする。逆に女性の分布が高値側にある場合には、男性の検査値分布の上側累積確率が L となる。

4-2-5 判別性能の比較評価法

男女判別の性能評価には、ROC 分析による曲線下面積（area under curve: AUC）と、感度特異度曲線より得られる正診率を用いた。正診率は、感度＝特異度となる位置にカットオフ値を設定し、その場合の偽陽性、偽陰性の例数を割合として計算した。すなわち、正診率＝（総数－偽陽性例数－偽陰性例数）／総数とした。

4-3 結果

4-3-1 男女差指数

一元配置分散分析法により、全年齢と、45 歳以下、46 歳以上の 3 群に対する男女差指数を計算し、表 4-1 に示した。また表では 45 歳以下と 46 歳以上の 2 群の SDR の差を計算し右端の列に示した。また、表では SDR の値の高さを、棒グラフで表している。また、臨床検査分野では、許容誤差限界の考え方から SDR が 0.3 を超えるものに意味のある差があると考え^{8,9)}。

全年齢で最も強い男女差を示す検査項目は CRE であり、次いで BUN、GGT、ALT、HDL-C、TG となる。また、45 歳以下と 46 歳以上を比較したものでは、かなりの項目にわたって 46 歳以上では SDR の値が小さくなっている。これは、高い年齢では性差が小さくなる検査項目が多いことを示している。最も年齢による性差の変化が大きいものは ALP であり、続いて Ca、Cl、Na となる。また、IP では高い年齢でより性差が現れる結果となった。

4-3-2 多重ロジスティック回帰分析法による判別性能評価

表 4-2 に多重ロジスティック回帰分析の結果を示す。回帰式の計算では、全年齢、45 歳以下のみ、46 歳以上のみの 3 群について、それぞれ最適な検査項目の組み合わせをステップワイズ法により選択した。この計算では、データ全体の 2/3 を教師群として回帰係数および検査項目の選択を行い、残りの 1/3 を検証群として回帰式による判別性能の確認に用いている。

表中の数値 z は、検査項目別の男女判定への貢献度を表す。空欄は判別への有意な貢献度を認

表 4-1 検査項目毎の男女差指数

Name	All	≤45	46≤	Diff
TP	0.024	0.088	0.025	-0.062
ALB	0.404	0.497	0.273	-0.224
UN	0.311	0.398	0.118	-0.280
CRE	1.679	1.703	1.634	-0.070
BUN	1.312	1.391	1.150	-0.241
Na	0.473	0.583	0.231	-0.352
K	0.251	0.270	0.183	-0.087
Cl	0.240	0.346	0.000	-0.346
Ca	0.408	0.552	0.106	-0.446
IP	0.323	0.192	0.618	0.426
GLU	0.307	0.288	0.339	0.051
TG	0.612	0.688	0.474	-0.214
TC	0.000	0.059	0.138	0.079
HDL-C	0.623	0.647	0.582	-0.066
LDL-C	0.235	0.309	0.030	-0.279
AST	0.425	0.517	0.221	-0.296
ALT	0.656	0.759	0.444	-0.315
LD	0.169	0.279	0.062	-0.217
ALP	0.399	0.619	0.000	-0.619
GGT	0.662	0.689	0.648	-0.041
AMY	0.169	0.220	0.036	-0.184

表 4-2 多重ロジスティック回帰分析の結果

		Age group		
		All	≤45	46≤
z-value	Age	0.386	0.173	0.940
	TP	3.112	2.665	
	ALB	4.429	3.997	
	BUN		3.124	
	CRE	11.282	8.090	6.326
	UA	7.011	5.337	3.705
	Na	4.187	3.060	2.459
	K			2.144
	Cl	5.191	4.682	
	Ca			
	IP	4.981	3.296	4.063
	GLU	3.589	3.293	2.769
	TG			
	TC	4.606		4.040
	HDL-C	4.306	5.445	
	LDL-C		3.192	
	AST			
	ALT		2.037	
	LD			
	ALP		2.660	
GGT	3.653		2.139	
AMY				
Training data				
N	1249	864	393	
AUC	0.992	0.995	0.990	
Accuracy	95.1%	96.5%	94.2%	
Validation data				
N	672	460	217	
AUC	0.986	0.990	0.980	
Accuracy	94.2%	95.1%	92.6%	

めなかった検査項目である。表の下端には教師群データ (training data) と検証群データ (validation data) のそれぞれについて、用いたデータ数と判別性能である AUC、正診率 (Accuracy) を示した。

この結果から、45 歳未満を対象とした場合に AUC・正診率がともに高く男女判別の性能が最も良かった。これに対し、46 歳以上に限定するとその性能の低下がみられた。一方、年齢を区別しないで作成すると、判別度はその中間的な値となった。

注目されるのは、年代により判別に貢献する検査項目セットが変化することで、CRE、UA はどの群でも常に高い貢献度を示すが、LDL-C、ALT、ALP などは 45 歳以下の場合にのみ貢献し、TC、GGT などは 46 歳以上でのみ貢献するという結果となった。この結果から、もう一つ注目される点は、判別式に組み込まれた検査項目で、男女差指数があまり大きくないもの (Na、Cl、IP、TC など) 含まれていること、また、逆に男女差指数が大きくとも判別式には組み込まれなかった項目 (AST、ALT、TG) があることである。

一方、過剰適合の可能性については、3 つの年代別分析のいずれにおいても、検証群データでの AUC・正診率は、教師群データと比べてやや低下する傾向を認めた。しかし、実際には、その程度はわずかで、実用上問題になる程度ではないと考えられる。

4-3-3 加重対数尤度診断法による判別性能評価

表 4-3 に加重対数尤度診断法の結果を示す。WALL 法の計算では、全ての検査項目を組み合わせて使用した場合と、男女差指数を基準として 0.3 以上の項目、0.5 以上の項目の組み合わせ、検査項目の組み合わせを選択し、全年齢、45 歳以下のみ、46 歳以上のみの 3 群について分析を行っている。表では、各条件で使用した検査項目を○で示し、表の下部には多重ロジスティック回帰分析法と同様に、教師群データと検証群データについて、データ数、AUC、正診率を表記した。

年齢による 3 群の比較では、45 歳以下での AUC および正診率が高く、46 歳以上では若干の低下が見られた。検査項目の組み合わせについて、全 21 項目使用した場合と、 $SDR \geq 0.5$ の項目に限定した場合の比較では、使用する検査項目数が半数以下であるにもかかわらず、AUC 及び正診率に大きな違いは見られなかった。また、教師群データと検証群データとの比較では、検証群データで AUC および正診率について、多重ロジスティック回帰分析法と同程度の低下がみられた。図 4-2 に、45 歳以下で 0.3 以上の検査項目の組み合わせの場合に、検証群データを対象として計算された WALL 指数の分布を示した。

表 4-3 加重対数尤度診断法(WALL 法)による結果

	All				≤45				46≤			
	SDR	All	0.3≤	0.5≤	SDR	All	0.3≤	0.5≤	SDR	All	0.3≤	0.5≤
TP	0.024	○			0.088	○			0.025	○		
ALB	0.404	○	○		0.497	○	○		0.273	○		
BUN	0.311	○	○		0.398	○	○		0.118	○		
CRE	1.679	○	○	○	1.703	○	○	○	1.634	○	○	○
UA	1.312	○	○	○	1.391	○	○	○	1.150	○	○	○
Na	0.473	○	○		0.583	○	○	○	0.231	○		
K	0.251	○			0.270	○			0.183	○		
Cl	0.240	○			0.346	○	○		0.000	○		
Ca	0.408	○	○		0.552	○	○	○	0.106	○		
IP	0.323	○	○		0.192	○			0.618	○	○	○
GLU	0.307	○	○		0.288	○			0.339	○	○	
TG	0.612	○	○	○	0.688	○	○	○	0.474	○	○	
TC	0.000	○			0.059	○			0.138	○		
HDL-C	0.623	○	○	○	0.647	○	○	○	0.582	○	○	○
LDL-C	0.235	○			0.309	○	○		0.030	○		
AST	0.425	○	○		0.517	○	○	○	0.221	○		
ALT	0.656	○	○	○	0.759	○	○	○	0.444	○	○	
LD	0.169	○			0.279	○			0.062	○		
ALP	0.399	○	○		0.619	○	○	○	0.000	○		
GGT	0.662	○	○	○	0.689	○	○	○	0.648	○	○	○
AMY	0.169	○			0.220	○			0.036	○		
Training data												
N	-	1249	1249	1249	-	864	864	864	-	393	393	393
AUC	-	0.977	0.979	0.978	-	0.986	0.986	0.986	-	0.973	0.973	0.975
Accuracy	-	92.2%	92.7%	93.1%	-	94.7%	94.7%	94.7%	-	93.0%	93.0%	91.4%
Validation data												
N	-	672	672	672	-	460	460	460	-	217	217	217
AUC	-	0.973	0.975	0.976	-	0.981	0.981	0.982	-	0.973	0.975	0.974
Accuracy	-	91.1%	91.0%	91.1%	-	93.2%	92.8%	92.8%	-	90.5%	91.6%	91.0%

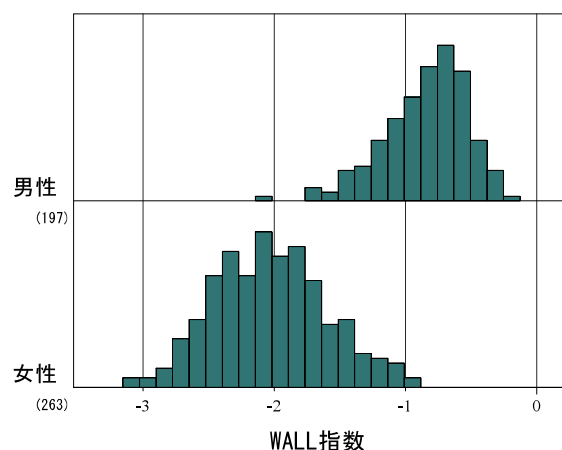


図 4-2 45 歳以下の男女の WALL 指数分布

4-4 考察

基準範囲の設定において、性や年齢など様々な要因で、基準個体を層別化するニーズがあり、そのため定量的な判断基準として、市原は分散分析法の理論に基づく群間差指数（SDR）を考案した^{7,8)}。ある変動要因に対する群間差指数とは、その変動要因による層別化後の群間差 SD を分散分析法で算出し、それを個体間 SD（その要因で調整後に設定した基準範囲の 1/4 の幅に相当）で割った値である。SDR の臨界値は、Fraser の提唱した臨床検査の偏りに関する許容誤差の考え方に基づき、 $SDR \geq 0.3$ が実質的に有意な群間差であるとして、現在、世界規模の基準範囲設定プロジェクトで実際に使われている^{8,9)}。

今回分析に用いたデータは、2009 年のアジア地域共有基準範囲プロジェクト⁵⁾から得られたデータであるが、対象を日本人に限定し、かつ年代の影響の強い検査項目が多くあるため、年代別の男女差指数も新たに計算した。年代の層別化として 45 歳以下と 46 歳以上で分けたのは、女性で 45 歳前後での変化を多くの項目で認めたため、男女の判別で使用するためにはこの区別が必要であると考えたからである。ほぼ予想された結果であるが、最も男女差指数が大きいのは CRE と UA であり、46 歳以上では若干の低下が見られるが、どちらの年代でも高い数値を示した。どちらの年代でも男女差指数が 0.3 以下となったのは TC, TP, AMY, LD, K であった。年代別の男女差指数の違いに目を向けると、IP のように 46 歳以上で男女差が大きくなる項目もあるが、項目の全体の傾向としては、男女差指数が 0.3 以上となるのは、45 歳以下では 14 項目であるのに対して、46 歳以上では 8 項目と減少している。特に ALP、Ca、Na などは、45 歳以下で大きな男女差があるにもかかわらず、46 歳以上では全く差を示さなかった。そのため 45 歳以下の男女差判別と比較すると、46 歳以上では方法に依らず男女差判別は難しくなると考えられる。

本研究では従来の多重ロジスティック回帰分析法と、加重対数尤度診断法の 2 つの手法による男女の判別を行ったが、判別の正確性として AUC と正診率を計算した結果から、どちらの手法

を用いても基本的な臨床検査の値による男女の判別は高い精度で行えることが分かった。どちらの手法においても、45歳未満を対象とした場合に AUC・正診率がともに高く男女判別の性能が最も良かった。これに対し、46歳以上に限定するとその性能の低下がみられた。これは男女差指数による考察と一致しており、高齢での男女判別が難しいことを示している。

また、2つの手法はその原理が大きく異なり、多重ロジスティック回帰分析法では、判別に対する各検査項目の貢献度(z-score)は、互いの相関関係も考慮して決定される。この場合に UA、ALT、GGT、TG、HDL-C など互いに関連性の強い項目がある場合、重複情報として扱われ、最も貢献度の高いパラメータだけが残るか、打ち消し合ってどのパラメータも残らない(多重共線性)という状況が起こりやすくなる。實際上、TG、HDL-C、LDL-C、ALT、GGT は、どちらの年代であっても高い男女差指数を示す項目であるが、45歳以下では HDL-C、LDL-C、ALT、46歳以上では GGT のみが貢献し、TG はどちらの年代でも貢献しないという結果となった。一方、対数尤度診断法では、関連性の強い項目同士であっても、同じ傾向の変化が複数存在することは、より積極的に診断に貢献するとして、組み合わせる用いることが可能で、これにより判別精度を高めることができると考えられる。

また、多重ロジスティック回帰分析法では判別式に使用する検査項目によって判別式の回帰係数が異なるため、様々な組み合わせに対応するためには、あらかじめその組み合わせに対応した判別式を作成しておく必要がある。一方、加重対数尤度診断法では、男女差指数さえ大きければ、任意の検査の組み合わせで WALL 値を計算できるため、その利用上の自由度が高く、利用可能な情報に応じた動的診断への応用が可能であることから実用性が高いと考えられる。今回使用したのは健常者のみのデータであったが、病院などの日常検査情報に含まれる患者データを対象とする場合、極端値の存在により分布型の変換は容易ではない。臨床検査項目の多くは分布型が歪んでおり、多重ロジスティック回帰分析法ではこの歪みが実用のための課題となる。一方、加重対数尤度診断法で尤度として使用している下側・上側累積確率は、分布型や極端値の影響を非常に受けない、ノンパラメトリック的な数値であるため、患者データへの対応でも極めて有効に働くことが期待される。また、今回のデータには欠損値はほとんどみられないが、病院のデータであれば検査値の組み合わせがかなり限定される事が考えられる。そのため、使用できる検査項目の組み合わせを、病院検査で一般的なものに限定した場合についての検討が必要である。

本研究で使用した加重対数尤度診断法は、市原が報告したような診断の確定した症例群毎への適用での有用性が確認され、多重ロジスティック回帰分析とは異なり対照群を必要としない点は大きなメリットとされていたが、対照群を設定してその有用性を検証する形の検討はなされていなかった。今回は、その方法論を男女差判別に応用したことになるが、これは男女どちらか一方を被判別群にすると、自動的に他方が対照群になるという意味で、対数尤度診断法のより明確な検証になっていると考えられる。今回の比較検討で、その診断能(正診率)がロジスティック分析のそれとほとんど変わらなかったことは、対数尤度診断法の分別能の高さを証明したことにもなる。さらに、教師群データと検証群データの判別度に明瞭な差違を認めなかったことは、この方式の信頼性の高さを示していると考えられる。

4-5 結語

多重ロジスティック回帰分析法と加重対数尤度診断法による男女判別では、両手法とも基本検査項目から高い精度で男女の判別ができることが示された。これらの手法が取り違い診断の補助的手段として有効であることが示唆された。

本章の参考文献

- 1) Yamashita T, Ichihara K, Miyamoto A. A novel weighted cumulative delta-check method for highly sensitive detection of specimen mix-up in the clinical laboratory. *Clin Chem Lab Med*. 2013; 51: 781–9
- 2) 市原清志：多重ロジスティック分析による診断的有用性の評価法. *臨床検査 増刊号* 2005; 49: 1341–53
- 3) Ichihara K, Sato K. Evidence-based laboratory interpretation system built on a large collection of case records with well-defined diagnosis. *Clin Chem Lab Med* 2001; 39: 1033–44
- 4) Ichihara K, Ceriotti F, Tam TH, et al. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. *Clin Chem Lab Med* 2013; 51:1429–42
- 5) Ichihara K, Ceriotti F, Mori K, et al. The Asian project for collaborative derivation of reference intervals: (2) results of non-standardized analytes and transference of reference intervals to the participating laboratories on the basis of cross-comparison of test results. *Clin Chem Lab Med* 2013; 51:1443–57
- 6) Ichihara K., Itoh Y, Lam CWK, Poon PMK, Kim JH, Kyono H, Chandrawening N, Muliaty D, the Science Committee for the Asian-Pacific Fed of Clin Biochem. Sources of variation of commonly measured serum analytes among 6 Asian cities and consideration of common reference intervals. *Clin Chem* 2008; 54: 356–65
- 7) Ichihara K, Boyd J. An appraisal of statistical procedures used in derivation of reference intervals. *Clin Chem Lab Med* 2010; 48:1537–51
- 8) Ichihara K. Statistical considerations for harmonization of the global multicenter study on reference values. *Clin Chim Acta* 2014; 432: 108–18.
- 9) Ozarda Y, Ichihara K, Barth J, Klee G. Protocol and standard operating procedures for common use in the worldwide multicenter study on reference values. *Clin Chem Lab Med* 2013; 51:1027–40.

第5章 結論と今後の展開

「第2章 導入」では、臨床検査の意義と、臨床検査値を用いた分析で必要となる技術について説明した。3章と4章では、これらの技術を用いて、臨床検査値による検体の取り違い検出と、男女判別についての研究結果を述べた。この章では、3章と4章で記した研究内容の特に重要な結果について整理し、その知見の意義と、今後それをどのような方向に利用可能であるのかを述べる。

「第3章 検体取り違いを高率に検出しうる加重累積デルタチェック法の考案とその性能評価」では、日常の臨床検査情報の中から検体の取り違いを検出するために、**加重累積デルタチェック法**を開発し、過去の検査結果データベースにある自然な前回値と今回値の組み合わせの中から、人工的にランダムにデータを組み合わせることで、擬似的な取り違い群を作成し、その性能評価を行った。

その中で、複数の検査項目の前回値と今回値の差を同時に扱うための検査値特有の課題として(1) 検査値が正規分布でない項目がある、(2) 検査値の取り得る値の範囲と単位が異なる、(3) 検査値毎の個体内変動幅を考慮する必要がある、という克服すべき問題点があった。

(1) については、日常の臨床検査値分布であっても、極端値の除外を適切に行うことで調整 Box-Cox べき乗変換によって正規分布へと変換することが、調べたほぼ全ての検査項目について可能であることが明らかとなった。

(2) については、分布の正規化に加えて検査値に変換を加えその基準範囲の中央値が0、両端が-2~2となるように標準化する方式を採用した。この分布の正規化と標準化処理によって、単位や大きさの異なる検査値を、全て共通の値で統一表示することが実現した。

これにより全ての検査値は、基準範囲の幅に対して、その中心からの外れ具合を示す値に変換される。臨床検査値を取り扱う場合、医師や検査技師は検査の値そのものだけでなく基準範囲に対する外れ具合や、そこに対する変化量を考えて判断を行っている。この標準化値 z では、例えば値が2を超えれば、検査項目によらず基準範囲上限を超えていることを示すなど、検査値の相対的な解釈が容易になる。

(3) については、検査結果データベースの中から、同一個体の前後の検査値のペアを取り出し、その標準化値の差の分布 Dz の標準偏差を、検査項目の個体内変動とすることとした。この Dz の分布は、同一個体の偏差の分布であるので恒常状態では正規分布になると想定できるが、重篤な病態では著明な偏差をとり、それらの極端値の影響で左右の分布の裾が非常に長い形状となるため、反復切断補正法により補正した標準偏差 aSD (adjusted SD) を計算した。

このような検査値を標準化する手法と、各検査項目の個体内変動幅を算出し、それを重みとし

て検査毎の偏差を累和することにより、複数の検査項目を利用したデルタチェックでの取り違いの検出が可能となった。また、取り違い検出の性能評価では、一般的にオーダーされることの多い検査項目セットを用いた取り違い検出の性能評価と、検査の種類は問わずその項目数のみで検出性能がどのように変化するかを比較検討した。やはり特定の項目セットのほうが少し高い判別性能を示したが、特に項目数と性能の関係では、項目数が 10 以上であれば検出性能はほぼ一定であり、検査項目の組み合わせに関係なく同一のカットオフ値を利用でき、かつ十分な性能が得られることが分かった。この結果は、組合せて判定する検査項目数を指定すれば、取り違い判定のカットオフ値を容易に設定できることを示している。

以上の結果・知見で、特筆すべきことは、調整 Box-Cox べき乗変換による変換結果である。今回の研究で適用した対象は、主に患者で構成された日常検査データであり、極端値除外を行わない場合は、当然ながら良い変換結果が得られていない。しかし両側の各 1% ずつを除外することで、ほぼ全ての検査項目の変換に成功している。このことは、患者データであっても分布の正規化が可能で、それをどの検査に対しても一樣に行えば、極めて雑多な臨床検査情報を、統一的に取り扱え、その統計処理も容易になることを意味する。

また 2 つめの特筆すべき点は、基準範囲に揃える形での標準化処理である。基準範囲に対して標準化された値は、分布が正規化されているので、基準範囲の幅に対する相対的な位置を示しているので、その値を見るだけで基準範囲をどの程度超えているか判断できる。このため、互いに測定意義やスケールの異なる複数の臨床検査値を同次元で取り扱え、その数値の臨床的な解釈も容易になる点である。

もう一つ特筆点は、日常検査情報から個体内変動幅を算出できたことである。臨床検査の領域では、測定値の精度保証のために同一試料の反復測定を行い、あらかじめ決められた許容誤差限界に照らして、検査の精度管理 (QC) を行っている。その許容誤差限界の目安として、世界的に広く利用されているのが、個体内変動の CV である。従来は、複数の個体について、日を変えて繰り返し試料を採取して、その測定値の自然な揺らぎを計算することで算出されていたが、手間と時間がかかりしかも対象個体数も少なかった。今回、私の研究で、異常値を多く含む LIS 中の日常検査情報からでも、間接的に個体内変動の CV を設定できたことは、きわめて重要な成果といえる。この技術により、どの病院においても、その施設の測定系にマッチした個体内変動 CV を自在に算定できるようになり、それを利用した誤差許容限界を QC に利用できることになる。

今後の展開としては、複数の施設で同様に高い取り違い検出精度が得られるか検証を行うことが必要である。このため、現在、他の施設の検査結果での検証作業を行っている。この施設で行った 2 ヶ月間の予備検討において、検体の取り違い例が 2 件検出されている。

一方、手術前後など外的要因による急激な変化が起こったケースも検出されたが、この検出された極端な変動をさらに小分類に分別することが必要である。手術、透析、外傷、輸血などで変化の生じる検査項目は限られるため、変動のパターンを学習し、それを自動的に分類してコメン

トすることができれば、検査技師のチェックが必要となる対象を減らすことができるため、さらなる業務支援となると考えている。

「第4章 基本検査による男女判別の正確度：多重ロジスティック回帰分析法と加重対数尤度診断法の比較検討」では、日常の臨床検査値から男女の判別をどの程度行えるのか、従来の多重ロジスティック回帰分析と新しい対数尤度診断法を対比させて検討を行った。多重ロジスティック回帰分析はパラメトリックな手法であるのに対して、対数尤度診断法はノンパラメトリックな手法であるので、2つの手法の原理は大きく異なっているが、判定結果はどちらの手法でも同様の高い精度での判定が可能であることが明らかとなった。

多重ロジスティック回帰分析では、あらかじめ判別すべき2群（通常は疾患群と非疾患群、今回の研究では男性群と女性群）を用意しておき、その区別に最もふさわしいパラメータ（臨床所見や検査所見）の組み合わせから回帰式を推定し、それを判別式とする。このため、特定の集団を判別するためには、対照となる集団を設定する必要がある。また、判別式を利用するには、式に含まれる説明変数が全て揃っている必要があるため、その検査項目が測定されていないと判別を行えないという制約がある。実際上、診療の現場では、疾患の特定や病状確認に必要な検査項目しか測定されないことも多く、男女判別に有効な検査が全て測定されていないため判別式を利用できない可能性がある。これに対処するためには、測定されうる p 個の検査項目に対して、 2^p 通りの組み合わせの判別式を事前に用意しておく必要がある。さらに、各検査値の分布型の影響を受けるので、大きく歪んだ分布ではその調整が必須となる。また、類似した検査項目は同時に判別式に取り込めないため、2群の違いを表す類似の検査結果が多重に存在しても、それを極端な場合には1つの検査情報としてしか取り扱えない。これらの点が、多重ロジスティック回帰分析を様々な診断的ニーズに合わせて、柔軟に利用する上での大きな制約となっている。

一方、**加重対数尤度診断法**では、男性群の検査値分布そのものを確率密度関数（尤度関数）として、各検査値の男性群としての適合度（尤度）を、対数尤度の形で計算し、これを各検査の男女差指数を重みとして、平均した値（加重対数尤度平均）を判別の指標として用いる。このため、判別に必要となるのは各検査値分布に基づく尤度関数と、各検査の男女差指数のみであり、任意の検査項目を組み合わせ、男性群との類似度を計算できる。このため一将来的に新しい項目が増加した場合にも対応が容易である。また、検査値分布からノンパラメトリックな方法（分布型に依存しない形）で尤度を計算するため、分布型の影響をうけない。特に重要な特性は、男性群の尤度関数により得られる判別の指標は、検査値の男性群への類似度を示す値であり、対照群となる女性群の分布を必要としないということである。

第4章で重要な結果は、多重ロジスティック回帰分析法と同等の診断性能を、加重対数尤度診断法により得られたことである。臨床検査の診断的有用性の評価では、これまで、ROC分析や多重ロジスティック回帰分析法が一般的で、特に複数検査を組合せて診断を行う場合には多

重ロジスティック回帰分析法が必須とされてきた。しかし、今回の研究で、加重対数尤度診断法で十分な判別性能を得られたことは、診断手法として広く利用可能であることを示している。また、今回の研究で用いたのは健常者のみのデータであったが、対数尤度診断法の分布型に影響を受けないという頑強性は、分布の歪みが起こりやすい患者データを対象とした分析では、特に重要なアドバンテージとなるであろうと考えている。

第4章で用いた臨床検査値は、基準範囲作成のための健常な集団のものである。今後の展開としては、病院での男女判別性能を確認するために、日常検査情報での判別性能を検討する必要がある。また、加重対数尤度診断法の有用性が示されたが、特に重要なのが対照群の設定を必要としないという特性である。今後、たくさんの疾患群を含む、疾患別症例データベースを作成すれば、各疾患との類似度を照合することが可能となり、対数尤度診断法を用いた診断支援システムを構築し、それを日常診療で利用できるようにすることも可能となる。実際に、このような構想のもと、自動診断システムの構築を目指し、疾患別の専門家グループの協力を求めることで、診断の明確な症例を大規模に集積する全国レベルのプロジェクトが、市原らによって計画されている。

今回の研究では、実際の日常検査情報を対象として様々な手法を用いてきた。その結果、一定の成果が見られたという事実は、今後の臨床検査情報の利用を推進する上で重要と考えている。今後、これらの技術により、臨床検査情報を用いた医療のさらなる発展に期待したい。

謝辞

学位論文をまとめるに当たり、多くのご支援とご指導を賜りました山口大学大学院理工学研究科 教授 宮本文穂先生に深く感謝しております。

本研究を遂行するに当たり、日々の研究だけでなく多くのご指導をしていただき、辛抱強く見守って下さいました山口大学大学院医学系研究科 教授 市原清志先生には深く感謝しております。時に応じて、厳しくご指導いただいたこと、またやさしく励ましてくださったことを通して、私自身の至らなさを実感することができたことは今後の努力の糧となるものであります。

本研究は当時、株式会社エイアンドティーに所属されていた鈴木賢治氏に患者データの分析を持ちかけられたことがきっかけとなり、そのスタートを切りました。また、現在では株式会社エイアンドティーの渡邊達久氏、長尾麻由氏の協力のもと、他の施設での検証を行っています。このような研究に必要なデータや環境をご提供いただいたことに感謝しております。

第 3 章の検体の取り違い検出手法は、山口大学 大学研究推進機構 知的財産センターの田崎泰孝氏、株式会社エイアンドティー 石田博文氏の協力により「臨床検査装置、臨床検査方法および臨床検査プログラム（特開 2012-256196）」として特許申請を行うことができました。本当に有り難うございました。

現在の職場である滋慶医療科学大学院大学の教員の皆様の精神的な支えなくしては論文の執筆は成しえませんでした。ここに感謝いたします。

そして、ここに名前を挙げさせていただいた方々だけではなく、私を支えてくださったすべての方に、心よりの感謝の意を込めて謝辞とします。