

フィルムオートチェンジャーの改善	- / 3
読影医師の画質評価方法の理解向上	
画質改善の重要性の理解	
その他	0 / 11

評価が低迷した機関を下表に示す。画質向上などによる精度管理の必要性、重要性の認識が機関に浸透してないことが最大の問題とみられた。パターン 5 を示す 10 機関中の 7 機関では全衛連からのコメントが活用されていなかった。10 機関中の 6 機関では、そこでの担当医師の画質評価が全衛連のそれと乖離していた。

項目	機関数(平成 14/15 年度)
全衛連審査にもとづくコメントの活用せず	6 / 7
感光材料、増感紙 X 線フィルム組み合わせの 是正	5 / 5
グリッドの格子比の改善	2 / 4
X 線管電圧の改善	2 / 1
自現機、現像条件の改善	
付加フィルターの改善	2 / 1
フィルムオートチェンジャーの改善	1 / 1
読影医師の画質評価方法の理解不足	0 / 2
画質改善の重要性の理解	0 / 6
その他	0 / 3

## F. まとめ

1. 継年的にみて評価が上昇傾向にある機関では全衛連から提示した評価コメントをよく活用していた。このことから、全衛連からのコメントは概ね適切な内容であることがわかった。
2. 評価が低くて変動の少ない機関についてはこれが活用されていなかった。機関毎の特種事情、財政的な問題、機関の担当医師の全衛連とは異なった考え方、担当技術者による無関心、軽視などが原因として考えられるが、解決可能な事情があり得よう。コメントを活用しない機関についてはよく連携をとる必要がある。
3. 経年的に評価の低い機関に対しては、改善がみられるまで積極的な指導を継続して行うことが重要と思われる。
4. 精度管理が行き届いていて評価の高い優良な機関を公表するなど、機関の精度管理に対するインセンティブを高める努力が全衛連サイドでは必要であろう。

## 資料 20

### 胸部 CT 画像読影の能率と精度および学習効果の評価法に関する研究

安曇総合病院院長 曾根脩輔

#### はじめに

医療技術革新の代表ともいえるべき CT 装置が胸部疾患診断の能率や精度の向上に寄与していることは疑問の余地がないように思われる。しかし、せっかくの技術もこれを活用する医師の読影技術が未熟であれば、患者や医療経営者に対するその恩恵は X 線被曝のリスクに比し不十分なものとなる。最近の研修医制度の導入とも関連して、CT 検査の安全と質を一層高めるため、研修医が医療現場で日常業務をこなしつつ、CT 画像読影技術をいかに習得し、その効果のほどをいかに確認するか、CT 画像診断の能率と精度を的確に評価できる方法論の確立が重要課題となっている。

このような現状に鑑み本報では、安曇総合病院に研修目的で来院した一外国人医師を対象に、CT 画像読影の能率と精度を評価する方法を検討、それを胸部 CT 画像読影の修練を目的にした読影実験に応用し学習効果を評価した。その結果、興味ある知見が得られたので報告する。

#### 1. 材料および方法

胸部 CT 画像読影の学習効果を評価するために考案した方法論とそれを用いて行った読影実験の手順および材料を以下に示す。

##### 1) 読影実験—1

研修目的で中国（上海）から安曇総合病院に来院した医師 1 名が勤務に入る直前に確定診断付き胸部 CT 画像データベースを読影、存在診断を行う読影実験—1 を行った。読影対象となった CT は 1 症例中の 1 スライスを延べ 270 例分集めたものからなる。胸部 CT 画像データベースの異常所見の有無および疾患の内訳を表 1 に示す。異常所見が有る場合は 1 画像中、重要な所見 1 個が定義された。研修医にも読影前、その旨が知らされた。

研修医は上海大学医学部卒、31 歳、女性、呼吸器内科医として 7～8 年の経験を有する。検診目的の CT 画像読影は今回が初めてであった。

読影対象の CT 画像は、実験担当者が操作する PC 画面(12 インチ)上に表示されたものをコネクタ接続された別の CRT 装置画面(19 インチ)上に表示した。

読影医は、その画像を読影、異常所見の有無判定を行い、その結果を、所見有りの場合は気管支分岐部辺の CT 画像シェーマ上に局在位置を示す印をつけ、かつ、その存在確信度をシェーマ右端に記された 0 から 100%表示のメジャー上にチェックした。その際、大別して 0~50%のチェックは所見無し判断、50~100%のチェックは所見有り判断であること、従ってそのどちらにも属さない 50%の確信度はチェックしないこと、所見無しまたは有りの確信の差がわずかな場合でも 50%を境にそのどちらかに決断することを申し合わせた。異常所見が存在しないと判断された場合は、0~50%の確信度のみがチェックされた。

## 2) 読影実験—2

実験—1 に引き続き、研修医は存在診断を再度行った後、質的診断も行う実験—2 を行った。CT 画像シェーマ上に、所見の局在および存在確信度を実験—1 と同様に記録するとともに、肺がん所見：1、肺がん以外の病変：2、正常：3 の鑑別結果を記録した。症例の疾患分布は知らせなかった。

## 3) 存在診断の学習—1

実験—2 に引き続き、No.137-270CT 画像（画像データベースの後半）に対する実験—2 の読影結果（存在診断、質的診断の結果）が研修医に開示され、自身の読影結果と照合するという存在診断の学習が行われた。

## 4) 実験—3

その後、研修医は No.1-136（画像データベースの前半）に対して再度、存在診断と質的診断を行った。

なお、以上の実験—1, 2, 3 は同日中の午後 1:00~4:00 の間に、随時休憩を挟んで実施した。読影は時間制限せず行ったが、1 スライス画像当たりの所要時間は、所見位置、確信度入力、質的診断結果入力を含めて、最大でも 30 秒以内、全体では休憩時間も含めて約 3 時間を要した。

## 5) CT 画像読影診断全般に関する学習—2

来日直後の実験—1, 2, 3 および学習—1 に引き続き、研修医は研修担当医師（曾根）と共に約 3 週間にわたり画像データベースの各例に対して詳細な症例検討を行った。その一方で、研修医は当院における日常業務の一部に参加、胸部 CT 読影に関する技能の研鑽に励んだ。なお、当該画像データベースに関する学習は滞在期限一週間前に終了した。

## 6) 実験—4

研修一ヶ月後の最終日、研修医は画像データベースの No.1-136 例（画像データベース前半）に対して実験—2 と同様の方法で存在診断と質的診断を行った。

次に、画像データベース後半；No.137-270の読影に対して、CRTモニター上で、CT画像表示、検出所見位置入力（ROI表示）、確信度入力、次画面表示、以下繰り返す一連の作業を全てPC画面上で行うことができるビューアシステムを用いて読影を行った。

実験前、読影済みのCT画像データベース前半の数症例を対象に、所見検出時の所見位置入力法、確信度入力法を練習してから本番の読影に臨んだ。

270スライス画像全体の読影所要時間は、実験—1または2と同程度（約1時間）であった。ビューアシシステムを用いて読影した画像データベース後半の134例では全体の読影時間と各例の読影所要時間がビューアシシステムにより自動記録された。

#### 7) 読影結果の評価法

上記実験において、研修医が決断した存在診断および質的診断の結果を確定診断の結果と比較した。これより読影者が下した存在診断および質的診断の正診、誤診および読影診断の能率を以下の方法により評価した。

##### ① 存在診断の精度

画像データベースの異常所見の有無は、確定診断の結果、1スライス画像中に異常所見は1個またはなしとされている。それぞれに対して研修医が下した判断が確定診断と一致した場合はTPまたはTNとなる。確定診断と不一致の場合は次の3種類が有りうる。

確定異常所見見逃し；FN,正常例に異常所見の存在を指摘；FP,確定診断所見を見逃したがそれ以外の別な箇所異常所見を指摘した場合；FN&FPである。

このようなデータを基に通常のROC解析を適用するため確定診断；異常所見あり例を異常所見有り領域となし領域からなると考えた。確定診断；正常例は異常所見なし領域である。各領域に対する研修医の判断の正誤（TP,TN,FN,FP）を、回答された確信度が50%より上か下で分類する。確定診断；異常領域の個数は異常所見有りの例数と同じ（s）である。確定診断；異常所見なし領域は正常例の領域nと異常例の異常所見なし領域sの合計（n+s）である。

確定診断；正常例（異常所見なし領域）に対して読影者が回答した確信度pが50%以上の時、0か1かの判定結果（0,1判定）はFP、50%以下の時はTNに分類される。各領域のFP,TNに対する確信度はpである。確定診断；異常例の異常あり領域に当てられた確信度（p）が50%以上の時、異常所見あり領域に対する0,1判定はTP、確信度はp、異常所見無し領域の判定はTN、確信度は100%-p（50%以下となる）であるとした。異常所見あり領域に対する確信度pが50%

以下の時、読影者の 0,1 判定は FN、異常所見無し領域に対する 0,1 判定は TN、確信度はともに p を当てた。確定診断：異常所見あり例で、異常所見なし領域に確信度 50%以上の異常所見が判定された時は FP と分類され、その時の確信度は p、異常所見あり領域の判定は FN に分類され、確信度は 100% - p を当てた。

以上の異常所見あり領域群 (s 個) に下された確信度 (0~100%) の頻度分布と異常所見なし領域群 (s+n 個) に下された確信度 (0~100%) の頻度分布を基に、 $TPR = \text{異常所見り領域の TP 数} \div s$ 、 $FPR = \text{異常所見なし領域の FP 数} \div (s+n)$  を ROC 解析の通常の方法に従って算出、ROC 曲線を描き、その曲線下面積 Az を求めた。ROC 曲線の描出および Az の計算は SPSS で行った。

## ② 質的診断の精度

読影結果と確定診断 (肺がん、肺がん以外の病変、正常) のクロス集計表を求めた。

## ③ 読影診断の能率

CT 画像読影の能率を評価するため CT 画像ビューアシステムに、実験開始時間、画像表示から所見有りを判断して所見位置を記録する直前までの時間、所見位置を記録し存在確信度を入力するまでの時間および次画像表示までの時間 (所見無し判断の場合は画像表示一次画面表示までの時間)、実験終了時間を区別して記録できる機能を付加した。実験—4 に本ビューアシステムを応用し上記作業時間を測定・集計した。

## 2. 結果

### 1) 存在診断

画像データベースの前半 No.1-136 画像に対する実験—1, 2, 3, 4 の結果を図 1 に示す。実験—1 (存在診断のみ実施) に比較して、その後、存在診断のほか質的診断を行った実験—2 の結果は、ROC 曲線がやや下方に位置し、Az もやや低下した。さらにその後、実験—2 の No.137-270 に対して真の答えが公開され、自身の読影結果と照合する学習が行われた後の実験—3 の ROC 曲線と Az は実験—1, 2 に比しやや上昇した。

一ヵ月後 (研修終了間際) 再度読影した No.1-136 画像に対する実験—4 の成績は一ヶ月前の実験—1, 2, 3 に比し、ROC 曲線と Az は明らかに上昇した。

異常所見の存在確信度分布の形状は、実験—1 では 0%~100%まで連続分布したが、実験—2 以降は 0%か 100%のどちらかで回答されたことが明らかであ

る。

## 2) 質的診断

質的診断を行った実験—2 と 4 の結果と質的確定診断結果を比較したクロス集計を表 2 に示す。実験—2 (研修前) の肺がん、肺がん以外の病変、正常の正答率はそれぞれ、45.8%、24.3%、95.7%、確定診断との一致率  $k=0.284$  であったが、実験—4 (研修後) では、75.6%、60%、91.3%、 $k=0.614$  と上昇した。

## 3) 読影所要時間の定量

### ①存在診断

実験—4 で画像データベース後半 137-270 の画像に対してビューアシステムで自動記録した存在診断の正誤と判断所要時間の結果を図 2 に示す。画像が CRT 画面に表示されてから異常所見ありと判断して所見位置を記録する直前までの時間 (TP; 99 例) は約 5 秒であった。正常と判断され、次画面を表示するまでの判断時間 (TN; 30 例) は約 12 秒を要した。結果的に FP となった (3 例)、異常所見ありと判断してその局在位置を記録する直前までの時間は約 20 秒であった。異常所見無しと判断し次画面を表示するまで、結果的に FN となった判断 (2 例) に要した時間は約 25 秒であった。

### ②質的診断

肺がん病変を他病変に、またはその逆に鑑別するのに要した平均時間は  $28.5 \pm 9.7$  秒、肺がんを肺がん、それ以外の病変をそれ以外と鑑別するのに要した平均時間は  $22.7 \pm 8.9$  秒であり、両者の間に有意の差 ( $p < .005$ ) が検出された。

## 3. 考察

### 1) 存在診断の評価

実験—2 および 3 により行われた研修開始直後の学習—1 は、読影者が判定したばかりの自分の結果 (画像データベースの後半画像 No.137-270 で) と真の答えを照合するという単純なものであった。その後、真の答えが未公開であった画像データベース前半 No.1-136 の CT 画像をもう一度読影して ROC 曲線 ( $A_z$ ) で学習前後を比較した。結果は図 1(a),(b) に示されるごとく ROC 曲線 ( $A_z$ ) の向上によりその効果が示された。図 1(c)~(f)に示すごとく、初めて CT 画像を読影した時の確信度の回答(c),(d)は 0~100%の間に連続分布した。一方、2 回目以降の読影実験-2,3,4(e),(f)では、確信度はすべて 0 か 100%のどちらかで回答された。これは学習効果というよりは学習の影響というべきかもしれない。

異常所見の存在を確信度で回答するのを読影者に求めたのは診断精度を ROC

曲線で定量評価するのに必要なデータを取得するためであった。この種の読影実験では通常、実験開始前、読影者にその旨要請するのが常であり、今回も実験—1の直前そのように指示した。しかし、実験—2以降は新たにそのような指示はせず読影者の自然の行動に任せた。

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。ROC解析できるようにするため、異常所見の有無を確信度で回答させた今回の実験は、臨床現場と異なる読影の仕方を読影医に強要した可能性がある。

研修医は、日ごろ見慣れない検診目的の条件で撮影された画質のCT画像を来院早々読影するよう指示された。従って、初回の読影実験—1では、自信の動揺や実験者の指示に忠実になろうと努力した結果が確信度に反映された。しかし、同時に確信度の回答という不慣れた行為が精神的負担を生じ、それが2度目以降の実験で0,1判断につながったと推測される。さらに根本の原因は、課せられた仕事が「存在診断」だった可能性もある。自分が異常と思って見えているものや異常所見は見えていないと思った判断に、敢えて100%または0%以外の確信度を与える必要はないと結論した研修医の態度は、実験—1の経験を通して検診目的の読影行動の基本を自己学習した当然の結果とも考えられる。(注；質的診断では、存在診断の結果、異常所見が歴然と見えている場合であっても癌かそれ以外の病変かの確信が50%前後になることは有り得る。また、今回の存在診断の実験で禁止した50%の回答も禁止できない。)

問題は、このような医師の行動を是認した場合、その結果を定量的、客観的に評価する方法はあるかということである。現在、世界に流布するMetzのROC解析プログラムは実験—1図1(a),(b)のデータは解析できるが、0,1判断を行った実験—2以降のデータでは結果が出力されない。(台形公式によるSPSSの方法はAzを算出できるが、過小評価された結果となる。)かといってROC解析プログラムによる定量評価のため、現実と異なる確信度の回答を求めた実験によるROC解析の結果は現状を反映していないかもしれず、そもそも確信度を読影医が自ら申告することができるのかという意見もある。その解決法を暗示する結果が図2(実験—4)に示されている。所見の有無を判断し、結果的にTN,TPであった判断所要時間に比し、結果的にFP,FNであった判断所要時間は明らかに延長した。このような事実は他の画像診断においても多く観察されており、一般性のある現象である。これより、異常所見の有無判断に要する時間が確信

度の強さと比例することを根拠に読影者の無意識的確信度を推定し、ROC 曲線を描く方法が検討されている<sup>1)</sup>。本法によれば、0,1 判断（胸部 CT 検診による要精検か精検不要かの判断）の結果からでも、読影医の判断所要時間から確信度が推定され、存在診断の精度を定量評価する従来の ROC 解析（Metz のプログラム）が適用できる。但し、その方法はアイデアの提案の段階にある。0,1 判断に対しても ROC 解析が適用可能な方法の開発は今後の重要な課題である。

図 1(a)に示されるごとく、実験—2 以降の ROC 曲線は 0,1 判断によるものであり、FPR、TPR の原点(0,0)と到達点(100,100)との間に動作点は 1 点しか存在しない。従って、台形公式から算出された  $A_z$  は真の（類推するしかない ROC 曲線の）値より過小評価されている。それにもかかわらず実験—1 の連続確信度に基づく ROC 曲線に比較して、実験—2 以降の結果は学習後のそれぞれの特徴を反映していると思われる。すなわち、実験—2 は実験—1 の直後、続けて存在診断のほか、質的診断も行うよう指示されたものであり、読影者のモチベーションの低下がそのまま  $A_z$  の若干の低下となって現れたと推測される。しかし、存在診断の再現性は表 2 に示されるごとく十分確保されている。

実験—3 は画像データベースの後半(No.137-270)の自身の読影結果を真の答えと比較する学習を行った直後、画像データベース前半(No.1-136)を再読影したものであり、 $A_z$  の若干の上昇と確信度の回答の仕方に大きな変化が生じたことは前述した通りである。このような傾向は、将来、CT 画像読影スクリーナーを務める可能性を秘めた医療技術系大学生を対象に今回と同じ画像データベースを用いて行った別の読影実験でも再現された。図 3 にその結果を示す。CT 画像読影が初体験の学生 42 名に対して今回と同じく、自身の読影結果と真の答えを照合するという学習を行い、再度同症例画像を読影した結果である。研修医と同様の効果が観測されている。

学生および研修医で示されたこれらの事実は、実験—2 の後試みたような単純な学習でも効果が期待できることを示したと考えられる。但し、学習効果には個人差がある。研修医は当院で初めての画質に接したとはいえ、呼吸器内科医として CT 画像読影の現場の経験が 7~8 年ある。そのようなベースの基で、今回行ったような単純な学習後の(実験—3)の  $A_z$  の上昇率は CT 画像読影がまったく未経験の学生の平均より相対的に小さかったように思われる。

研修医の存在診断能の絶対値を日本の医師と比較した場合はどうか。研修医とほぼ同様の方法で同画像データベースを CRT 読影した日本の医師 5 名の読影結果を図 4 に示す。このときの  $A_z$  は 0.864~0.945 の範囲に分布した。研修医



の学習前の Az (実験—1, 2) は図 1 に示すごとく 0.880, 0.843、学習後 (実験—3) は 0.896 であったから、日本人医師と遜色ない。特に、研修 1 ヶ月後 (実験—4) の Az は 0.972 であり効果は歴然である。但し、実験—1, 2, 3 の後、読影医は同画像データベースを用いて研修担当医と一対一の詳細な症例検討を行った。従って、帰国前 1 週間の冷却期間をおいた後の実験—4 の時点で真の回答が研修医の記憶に残っていた可能性が無視できない。実験—4 の Az は、実験—1, 2, 3 の経験、日本の CT 肺癌検診で使われた画質への順応、当院における CT 画像読影に関する学習および現場での研修、自己の研鑽等が総合的に盛り込まれた結果と思われる。

いずれにしても、今回の一連の実験で得られた研修医の Az 向上の推移は、CT 画像診断の基本である存在診断に対して行われた当院での学習および研修が医師にとって意義あるものであったことを示していると考えられる。

## 2) 質的診断の評価

最近改定された肺癌取り扱い規約では肺癌疑い (E 判定) に対して確信度に近い概念が導入されている。しかし、今回の読影実験では、研修医は肺癌疑いの確信度は回答せず、所見なしの正常か、異常所見を検出した場合は肺癌か肺癌以外の病変かを分類した。その結果 (表 2) 研修前の実験—2 より研修一ヶ月後の実験—4 の質的診断能は明らかに向上した。しかし、向上後の肺癌 75.6%、肺癌以外の病変 60% の正診率は研修担当医と約 3 週間にわたり詳細に症例検討を行った割には決して高いとはいえない。ちなみに図 4 の存在診断能を示した日本人医師の肺癌正診率も 80% 前後であり、研修医と同等であった。以上のことより CT 画像診断のみから肺疾患を的確に鑑別診断できるようになるには、存在診断の場合より一層の研鑽が必要であることが示唆された。

## 3) 今後の課題

CT 画像読影の能率を表す指標のひとつに読影所要時間がある。モニター画像診断が次第に主流になりつつある昨今、CRT 読影に伴う医師の行動や CAD 参照効果を評価するための研究上、VDT 作業による医療従事者の健康管理上、HIS/PACS の管理上、CT 画像読影にどれだけ時間を要するか知ることは意味があると考えられる。しかし、いざ定量する段になり採られる通常の方法は、当事者の自己申告か手作業による測定・集計である。前者のデータは主観に基づく回答の集積であり妥当性に難があり、後者は莫大な手間と費用を要するため継続が困難という欠点がある。

今回、CT 画像をモニター診断する時、読影の開始から終了の間、CT 画像読

影に伴う一連の PC 操作の節目となる事象の発生と時刻を逐一記録することができるビューアを開発し、本ビューアを研修一ヵ月後行った実験—4の一部に应用することができた。図 2 は医師の CT 画像読影行動時間の 1 例を示したに過ぎない。現在、本ビューアを用いて CT 画像読影診断が学習や研修により能率が如何に向上するか評価する実験を計画している。また、今回の学習・研修評価実験に使用された画像データベースは通常の CT 画像（ボリューム）の中から選別された 1 スライス画像を 270 例分集めたものであった。現場に即した結果を得るため、本来の CT 画像からなる画像データベースを作成し、これを対象に本ビューアを使用した読影実験を行う予定である。

## 結論

研修目的で来院した外国人医師を対象に CT 画像読影の能率と精度を定量的、客観的に評価する方法論を検討、それを用いた読影実験を行い、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を確認することができた。

## 謝辞

本研究は松本 徹（放射線医学総合研究所）、和田真一（新潟大学）、富士通バイオ IT 事業開発本部の協力により行われた。

## 文献

1) T. Matsumoto, N. Fukuda, M. Tsuchikawa et al: Observer performance study for CT-image reading of one slice or multi-slice by the cine display mode of CRT system- An application of the Diagnosis-Dynamic Characteristic(DDC) model, CARS 2001 Computer Assisted Radiology and Surgery,1190, Elsevier Science, Amsterdam,2001

存在確定と質的確定のクロス表

度数		質的確定			合計
		1	2	3	
存在確定	0			36	36
	1	66	34		100
合計		66	34	36	136

度数		質的確定			合計
		1	2	3	
存在確定	0			69	69
	1	131	70		201
合計		131	70	69	270

表1 疾患分布 (画像データベース前半 136 例、全体 270 例)、質的確定 1; 肺がん、2; 肺がん以外の病変、3; 正常、存在確定 0; 異常所見なし、1; 異常所見あり

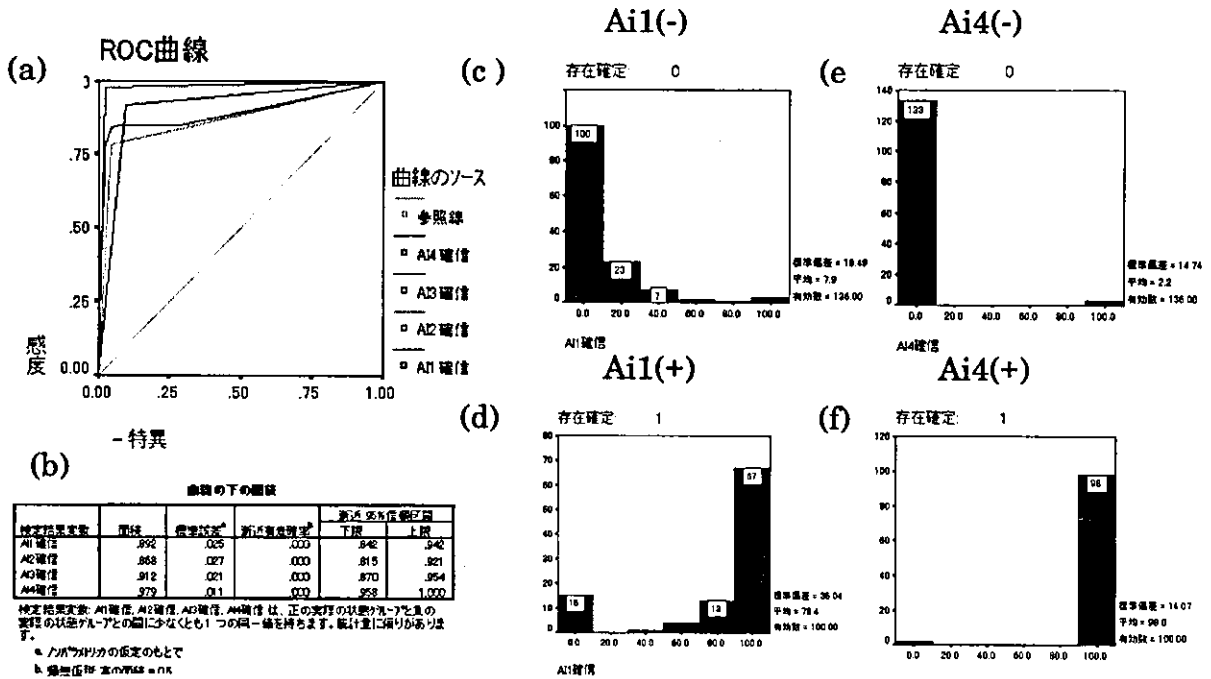


図1 存在診断の学習効果

(a) 実験-1,2,3,4ROC 曲線、(b)Az,(c) 確信度分布;実験—1 所見無し領域,(d) 同左;実験—1 所見あり領域,(e)同左;実験—4 所見無し領域,(f)同左; 実験—4 所見あり領域

加減表

		AI質的			合計
		1	2	3	
質的 確定	1 患者数	60	38	33	131
	質的確定の%	45.8%	29.0%	25.2%	100.0%
	AI質的の%	59.4%	65.5%	29.7%	48.5%
	総和の%	22.2%	14.1%	12.2%	48.5%
2	患者数	41	17	12	70
	質的確定の%	58.6%	24.3%	17.1%	100.0%
	AI質的の%	40.5%	29.3%	10.8%	25.9%
	総和の%	15.2%	6.3%	4.4%	25.9%
3	患者数		3	66	69
	質的確定の%		4.3%	95.7%	100.0%
	AI質的の%		5.2%	59.5%	25.6%
	総和の%		1.1%	24.4%	25.6%
合計	患者数	101	58	111	270
	質的確定の%	37.4%	21.5%	41.1%	100.0%
	AI質的の%	100.0%	100.0%	100.0%	100.0%
	総和の%	37.4%	21.5%	41.1%	100.0%

加減表

		AI質的			合計
		1	2	3	
質的 確定	1 患者数	99	28	4	131
	質的確定の%	75.6%	21.4%	3.1%	100.0%
	AI質的の%	76.7%	38.4%	5.9%	48.5%
	総和の%	36.7%	10.4%	1.5%	48.5%
2	患者数	27	42	1	70
	質的確定の%	38.6%	60.0%	1.4%	100.0%
	AI質的の%	20.9%	57.5%	1.5%	25.9%
	総和の%	10.0%	15.6%	.4%	25.9%
3	患者数	3	3	63	69
	質的確定の%	4.3%	4.3%	91.3%	100.0%
	AI質的の%	2.3%	4.1%	92.6%	25.6%
	総和の%	1.1%	1.1%	23.3%	25.6%
合計	患者数	129	73	68	270
	質的確定の%	47.8%	27.0%	25.2%	100.0%
	AI質的の%	100.0%	100.0%	100.0%	100.0%
	総和の%	47.8%	27.0%	25.2%	100.0%

封存性による類似度

	値	近似標準誤差 <sup>a</sup>	近似T値 <sup>b</sup>	近似有意確率
一致の判定方法 あり	285	.044	6.768	.000
有効なケースの数	270			

- a. 帰無仮説を仮定しません。
- b. 帰無仮説を仮定して近似標準誤差を使用します。

封存性による類似度

	値	近似標準誤差 <sup>a</sup>	近似T値 <sup>b</sup>	近似有意確率
一致の判定方法 あり	614	.042	14.046	.000
有効なケースの数	270			

- a. 帰無仮説を仮定しません。
- b. 帰無仮説を仮定して近似標準誤差を使用します。

表2 質的診断の学習効果：

研修前；Ai2、研修後；Ai4、確定診断と読影診断の一致率；k

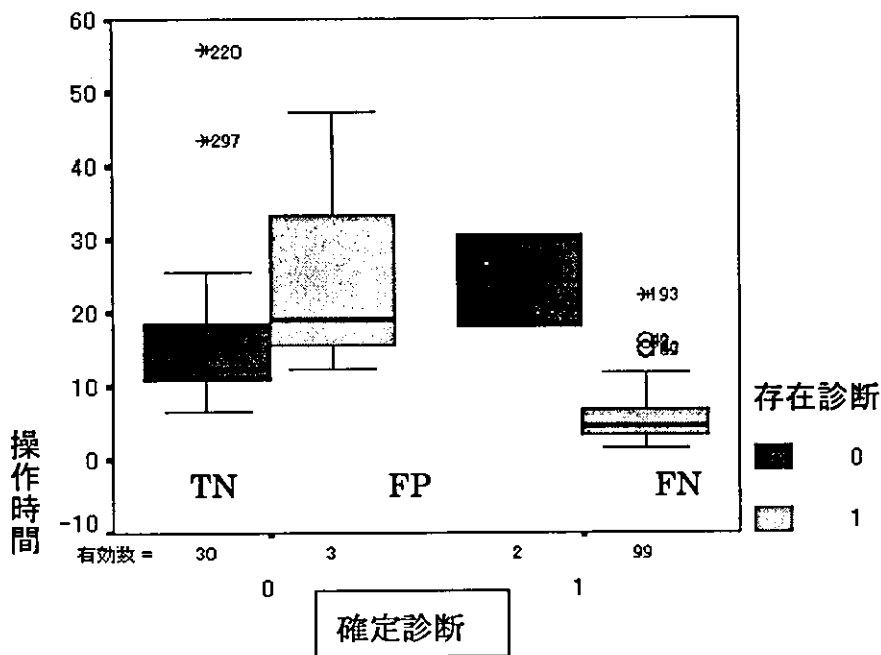


図2 存在診断の正誤と判断所要時間

AI1TFFP と AI2TFFP のクロス表

			AI2TFFP		合計
			0	1	
AI1 TFFP	0	度数	286	9	295
		AI1TFFP の %	96.9%	3.1%	100.0%
		AI2TFFP の %	91.7%	5.7%	62.6%
		総和の %	60.7%	1.9%	62.6%
	1	度数	26	150	176
		AI1TFFP の %	14.8%	85.2%	100.0%
		AI2TFFP の %	8.3%	94.3%	37.4%
総和の %	5.5%	31.8%	37.4%		
合計	度数	312	159	471	
	AI1TFFP の %	66.2%	33.8%	100.0%	
	AI2TFFP の %	100.0%	100.0%	100.0%	
	総和の %	66.2%	33.8%	100.0%	

対称性による類似度

名義と名義		値	漸近標準誤差 <sup>a</sup>	近似 t 値 <sup>b</sup>	近似有意確率
アイ	Cramer の V	.841			.000
一致の測定方法	カパ <sup>c</sup>	.838	.026	18.245	.000
有効なケースの数		471			

- a. 帰無仮説を仮定しません。
- b. 帰無仮説を仮定して漸近標準誤差を使用します。

表 3 実験—1 と 2 の存在診断の再現性

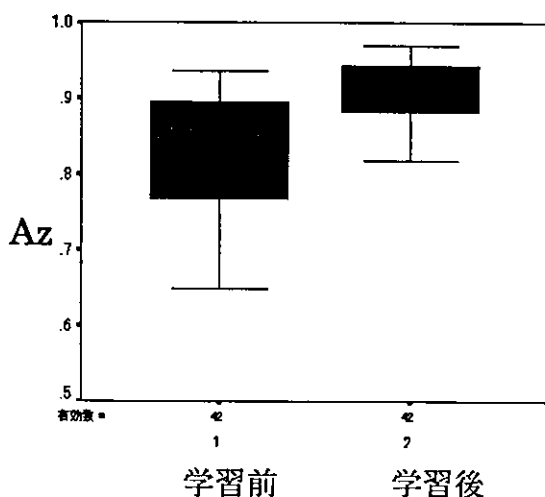


図 3 医療技術系大学生の学習結果

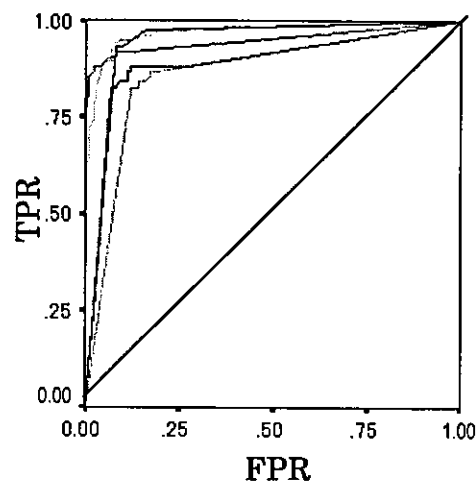


図 4 日本人医師 5 人の ROC 曲線 (Az:0.864~0.945)

## 時系列健康診断データの有効利用と精度管理

分担研究者 吉田 勝美 聖マリアンナ医科大学 予防医学教室教授

### 研究要旨

健診実施目的を考えた場合、その判定基準値が乖離しているため、新しい基準値が必要であると考えられる。健診データを時系列的に判定する手法研究においては、個人値の変動を用いることの有益性が示唆された。

### 研究協力者

佐々木 敏雄 バイオコミュニケーションズ株式会社  
企画室長  
小林 祐一 HOYA 株式会社グローバル本社 産業医  
座間 聡子 HOYA 株式会社グローバル本社 産業医  
柴山 順子 バイオコミュニケーションズ株式会社  
企画室主任

異性の高い検査方法を実施しなければならない。検査精度においても、異常値域も“正確に”、“精密に”測定するということが求められる。また早期発見という課題に対しては、実施間隔を短縮し、健診を実施しなければならないということになる。

当班は、健診の意義や目的を労働者の健康保持・増進、疾病予防、有所見者の状態把握に役立てることとし、健診検査精度を研究することとする。

これからの健診検査精度は、個人データが時系列的に評価される時代を迎えていることを考えれば、健診検査データは経年的データ活用に耐え得る精度であることや、いわゆる基準値域が正確に精密に測定できることが要件となる。健診判定においては、時系列的な判定手法を導入し、個人毎に具体的な事後指導対策を立

### A. 研究目的

健診の有効性について多くの議論がなされてきたが、その大部分は健診の目的を疾病診断や疾病の早期発見に置いて評価したものであった。もし健診の目的をそのように捉えるならば、臨床診断と同様に対症疾病に該当する検査項目を選び、疾病特

てられるような手法の開発が求められる。例えば従来の『所見あり、なし』という判定方式から脱し、『基準範囲・上昇傾向』というような判定分類をつけることにより、今後予想されるデータの上昇を抑制できるような保健指導を行うことができる。

一方、臨床診断においては学会等で診断基準が示されているが、健診の場合、判定は健診医の裁量に委ねられているため、判定内容に格差が生じ、保健活動や職場改善のための資源として役立てることが困難である。この点については判定基準の検討が必要である。また、少子高齢化社会において高齢就労者の健康を考えると、有所見者が多く含まれ、疾病者も仕事に従事する状況が生じる。有所見者や疾病者であってもそれなりの状態評価とその評価に沿った健康支援活動ができるような判定方式も研究したい。

## B. 研究方法

本研究作業範囲は『時系列的データ評価』、『多項目評価』、『生活習慣介入評価』、『個人に分かりやすい表示方法』であるが、本年度は準備作業として「解析データの入手」、「コンバート／リファイン作業」を行い、『時系列的データ評価』については一部のみ実施した。また、現状の基準値調査についても一部実施した。解析対象データは、某製造業種企業

の健診データ(退職者を含む)1982年～2001年まで、1人最大13回分の健診データである。個人情報を保護する観点から、連結不可能匿名化処理を行ったデータを使用している。整理した情報は、BMI、sBP、dBP、TC、TG、HDL-C、GOT、GPT、 $\gamma$ -GTP、BS、HbA1c、RBC、Hb、Ht、UAである。

データのリファインとしては、不要になったデータが削除されていないものもあったためそれらを精査し、同一人で同一受診日のデータになるようにした。

尚、某製造業種企業からのデータの入手は、分担研究者および研究協力者が提供者の了解を得て使用しているものである。個人情報を保護する上で、上述の連結不可能匿名化処理を行って使用している。

## C. 研究結果

### 1. データ件数

全レコード件数は表1の通り、49,522件である。

個人毎の受診回数の件数は表2の通り11～12回受診している人が約1,300名である。

2000年度受診者における性別年代構成は表3の通り2000年度時点においてはどちらかと言えば若年層も多く含む若い会社と言える。しかし、現在、海外へ生産を移しているため国内での新規採用が減っており、今後、

この集団の年齢は毎年上昇をし続けるものと思われる。表4のように、受診時の平均年齢を男性で見ると、1995年は平均36.35歳、2000年では平均38.92歳と急激に上昇している。

この年齢上昇は今後さらなる労働衛生の充実策を進める必要があるということを意味している。

リファインした結果の項目件数は、表5の通りである。身体計測や血压検査の件数は多いが、血液検査は若年層が受診していないため、件数が少ない。

## 2. 基準値等の調査

現在、某製造業種企業が運用上使用している基準値は表6の通りである。

健診データ(1999年度)の基本統計量は表7の通りである。表7の結果は反復切断法によるものである。本来除外すべき疾病者を含めていること、また一次健診時に要健康管理健診も同時に実施しているので、一般的な健診データよりも異常値が多少多く含まれているという母集団になっている。

## 3. 個人値の時系列解析(個人値の算出結果)

個人のデータを時系列的に処理するには同一項目における件数が課題となるが、今回は4件以上のデータが登録されている受診者を対象とした。

本年度はsBP、dBP、BSを対象とし、個人毎の平均値(AV)、標準偏差(SD)、一次回帰直線による傾き(Slope)を算出した。sBP、BSの個人毎の平均値(AV)、標準偏差(SD)の算出結果のヒストグラムは図2、3の通りである。

## D. 考察

### 1. 基準値と反復切断法(臼井法)との比較

母集団に対し、臨床診断上でよく使用されているといわれる反復切断法(臼井法)によりデータを収斂させる処理を行った。疾病者や有所見者が含まれているため、反復切断法による基準値範囲( $\pm 1.96SD$  95%信頼区間)は広がっている。疾病者を除いた比較や、年代格差の評価方法を検討する必要がある。

### 2. 基準値の施設間比較

図1は『総合精度管理事業参加機関の臨床検査の基準値調査結果』(平成12年度全衛連)の項目毎Reference Intervalsに基づいて作成したものであるが、施設によって変動がみられる。

### 3. NCCLS と反復切断法(臼井法)との比較

NCCLSでは表8にあるように『除外されるべき基準例』(Table1)、『基準値として区分すべき要素例』(Table2)、『批評要素の概要』(Table3)、



『分析前に考慮すべき要素』(Table4)等のガイドラインがあり、Reference Intervals が必然的に狭いものになると考えられる。

この NCCLS の手法は労働衛生機関にも少なからず影響を及ぼし、その結果、上述のように施設による Reference Intervals の格差となって現れているものと考えられる。

#### 4. その他の基準

一方、TC は日本動脈硬化学会のガイドラインでは管理目標値は220mg/dl とされている。

検査値の基本統計量とは大きくかけ離れている。平成13年3月、労災保険二次健康診断等給付制度では、脳・心疾患予防の観点からは、単項目ではなく複数検査項目の組み合わせにより判定を行うというようになってきている。更に、労働衛生機関の一部には生活習慣病予防に主眼を置いて基準値を狭く定めているところも現れ、判定の方法や基準が多様化しつつある。

#### E. 結論

時系列的評価方法の検討；個人値の AV と SD による評価

個人値の AV や SD について基準算定のために基本統計量を算出した。sBP の個人値の SD は  $1SD=10.41$ 、 $1.96SD=12.80$  となった。時系列的評価基準値の検討は次年度以降に行う予定である。本年度は個人 AV、個人

SD による散布図を用いて健康状態が把握できるかどうか検討した。その結果、本方法でもある程度の健康状態が把握ができることが示唆された。(図4)

件数は個人毎に異なるため、個人 SD だけでは評価できず、個人 AV と組み合わせて評価することが現実的と考えられた。短期間でデータ件数が少なければ最近の健康状態を示すことになり、長期間で多数のデータが含まれていれば当人の全体的健康状態を示すことになると考えられる。また、最近の例えば4件に対するデータの Slope を見ることにより、上昇や下降の傾向が把握できると考えている。(図5)

#### 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。

表1 年度別件数(元データ)

年度	男(件)	女(件)	計
1982	3		3
1983	10		10
1984	11		11
1985	17		17
1986	21		21
1987	44		44
1988	73	4	77
1989	197	24	221
1990	933	124	1057
1991	1441	211	1652
1992	1783	311	2094
1993	1966	356	2322
1994	2286	430	2716
1995	4259	621	4880
1996	4903	761	5664
1997	5193	909	6102
1998	5395	1053	6448
1999	5539	1210	6749
2000	5497	1234	6731
2001	2177	526	2703
計(1982~2001)	41748	7774	49522
性別不明者			14
生年月日不一致	1	1	2
合計			49538

①性別不明14件のうち詳細調査を行い、9件について個人を特定できたので性別を補完し、正規データへ参入した。5件は補完不可能であったため、データを削除した。  
 ②生年月日不一致者2名について詳細調査を行った結果、生年月日を修正し、男女共に11コト→12コトへ参入されることとなった。

表2 個人毎のレコード件数(整理後) N=49,533

重複回数	男(人)	女(人)	計(人)	性別不明
性別不明				5
生年月日不一致	0	0	0	
1	223	238	461	
2	168	170	338	
3	118	146	264	
4	145	115	260	
5	147	104	251	
6	140	80	220	
7	133	57	190	
8	149	43	192	
9	200	47	247	
10	424	77	501	
11	1,200	176	1,376	
12	1,213	105	1,318	
13	213	13	226	
計(員数)	4,473	1,371	5,844	
計(レコード件数)	41,756	7,777	49,533	5

表3 2000年度受診者における性別、年代別員数、件数構成

年代	男(人)	女(人)	計(人)	男(件)	女(件)	計(件)
~19	4	10	14	11	15	26
20~29	737	519	1256	1144	563	1707
30~39	1412	252	1664	1961	296	2257
40~49	1112	142	1254	1509	167	1676
50~59	618	136	754	797	160	957
60~69	49	22	71	56	27	83
70~	1		1	1		1
計	3933	1081	5014	5479	1228	6707

年齢は受診時における年齢

表4 受診年度毎の年齢比較

性別	1995			2000		
	N	AV	SD	N	AV	SD
男(人)	3310	36.35	9.12	3933	38.92	9.82
女(人)	529	33.32	10.99	1081	33.90	11.19
男女(人)	3839	35.93	9.46	5014	37.84	10.34

表5 項目毎累積件数(1982年～2001年、外れ値除外はしていない)

項目名	N	AV	SD	項目名	N	AV	SD
身長	49457	167.68	7.48	ハマトリット	15016	44.64	4.00
体重	49455	63.24	10.56	T-P	5261	7.37	0.49
BMI	49454	22.43	3.09	ZTT	4729	6.79	4.39
視力 遠点裸眼右	28645	0.99	0.48	T-BII	4691	0.81	0.38
視力 遠点裸眼左	28633	0.99	0.48	GOT	24188	23.57	15.70
視力 遠点矯正右	22377	1.10	0.36	GPT	24272	26.37	23.85
視力 遠点矯正左	22404	1.11	0.34	γ-GPT	24194	43.24	53.36
聴力所見1000Hz右	19559	1.03	0.17	ALP	4781	141.31	75.06
聴力所見1000Hz左	19571	1.03	0.17	Ch-E	1194	4590.57	3921.07
聴力所見4000Hz右	19495	1.08	0.27	LDH	3382	297.97	101.85
聴力所見4000Hz左	19504	1.08	0.27	T-CH	24525	199.77	35.80
聴力 その他	23826	1.01	0.15	T-G	24500	128.90	100.63
sBP	49442	120.67	14.11	HDL-C	20583	55.23	14.92
dBP	49442	73.43	10.89	LDL-C	20067	122.13	32.60
尿蛋白	49032	1.03	0.20	UA	18912	5.76	1.39
尿糖	49030	1.03	0.21	BUN	15963	14.13	3.97
白血球数	14460	6072.69	1723.15	Crea	16005	0.88	0.29
赤血球数	23118	479.94	41.57	FBS	16359	96.07	22.26
ヘモグロビン	23114	14.94	1.40	HbA1c	2667	5.79	1.57

表6 運用上使用している基準値

検査項目	
BMI	19.8～24.1
収縮期血圧	140未満
拡張期血圧	90未満
白血球数	4000～9000
赤血球数(万)	M:410～550 F:380～500
ヘモグロビン量(g/dl)	M:13.5～18.(F:12.0～15.0
ヘマトクリット値(%)	M:40～52 F:34.5～48
GOT	8～40
GPT	5～35
γ-GTP	59未満
総コレステロール(mg/dl)	140～219
中性脂肪(mg/dl)	35～149
HDLコレステロール(mg/dl)	40～80
尿酸(mg/dl)	M:3.5～7.5 F:2.5～6.0
空腹時血糖(mg/dl)	70～110
HbA1c(%)	2～6