

- Within 1 year, establish guidelines and an approval process for facilities, equipment, and image quality. Seek consistency for the quality assurance processes with oversight and documentation. Define acquisition protocols with guidelines (scatter reduction, beam quality, rating of generators, exposure index, exposure monitoring over time, etc.). Define file format and QC guidelines. Conduct beta testing of the approach.
- In the medium term (1–3 years), create a more representative and complete phantom with automated analysis. Determine exposure requirements. Develop integrated QC with the phantom.
- In the long term (3–6 years), develop automated disease classification. Develop automated image quality assessment based on image data.

Discussion

Dr. Clunie wondered about the possibility of having influence over the sites' acquisition of equipment. We need to define values such as exposure index and influence vendors, perhaps through a NIOSH request for proposals. Dr. Samei cited the goal of classification in the requirements proposed. Dr. Ravin emphasized the basic point that digital images are better than film images. Dr. Henry noted a paradoxical trend in which radiographers produce poorer-quality images as a result of their overconfidence in a better technique (digital). Dr. Flynn suggested that exposure indicator/range may be a key value on which to focus. Dr. Weissman noted that NIOSH will have to engage Federal lawmakers to specify digital regulations or standards. A main goal, reminded Dr. Peterson, is to protect the continuity of the program to protect the health of coal miners—a public health function. The workshop participants cited a need to establish an accreditation system or update the current NIOSH accreditation system.

2 – Image Presentation, Including Processing and Display

Dr. Flynn summarized the results of the second subgroup's discussion of image processing, focusing on three areas and making the following recommendations:

- Develop a reference case library for digital radiography to elaborate the effect of processing on classification, to provide a basis for a digital reference set, and to provide training material. Images would be in “for processing” form. Recent and ongoing studies could provide data for the library, which would be modest in size—perhaps a few thousand cases. Various institutions (e.g., National Library of Medicine, National Cancer Imaging Archive) have experience in developing such a collection. The NIOSH effort could include digital standard films, documentation on disease stages, and perhaps correlated film-screen and CT images.
- Develop consensus on display hardware specifications, for example, a 2-monitor, 24-inch, 3-megapixel system with grayscale cards, calibration, and navigation. This should be a discrete flat-panel technology, such as LCD. The system should include embedded test images, and there should be a document specifying aspects of the reading environment (ambient light, etc.).
- Develop display software for reading images and comparing them to standards. It should support a NIOSH/ILO viewer tailored to present reference comparisons, and it should be incorporated on a CD. One problem that must be avoided is the presence of an image from a prior reading left on a screen, which is read mistakenly as a current image. Cross-sectional cases present difficulties. A report template should be included in the electronic viewer presentation (it could then be easily transmitted electronically).

Discussion

Dr. Audrey Banyini noted that South Africa has a large database of images from medical surveillance, some of which might be shared with the reference case library. The workshop participants proposed standards and testing for reader vision. Dr. Henry wondered about incorporating the American Association of Physicists in Medicine test pattern in discs. Dr. Samei suggested that in addition NIOSH develop a reference image or pattern. The workshop participants agreed that there should be procedures for measuring and controlling, perhaps with

the aid of a light meter, the screen brightness and ambient lighting. The second subgroup proposed that, for image presentation in general, the first efforts involve tools that resemble film-screen tools (this could change over time).

3 – File Interchange, Including Formats and Interoperability

Dr. Clunie presented the results of the third subgroup's discussion of file interchange, listing assumptions, caveats, and requirements for image filing issues. The group assumed that the approach will be reusable in similar settings, will use NIOSH-supplied equipment, and will depend on a limited NIOSH support staff. There are questions about training and needed changes in regulations. The acquisition site must have pre-qualification of system and transfer processes. A central site will require pre-qualification processes and tools, QC processes and tools, and archival and disaster-recovery capabilities. Readers must receive images, read them, dispose of them, and send results.

Dr. Clunie listed the following recommendations for a CD-based solution in the short term (3-month):

- The acquisition site burns CR/DX to CD from a modality or PACS. Features include one patient per CD, identity in the header, a "for presentation" image (or, optimally, with "for processing" as well), the DICOM GP-CDR and IHE PDI profile, and no lossy compression. The site submits an initial pre-qualification CD.
- The central site receives and checks the CDs. Required tools include standalone PCs, displays, viewers, an automated CD format checker, an automated DICOM file checker, a CD duplicator, and a DICOM header editor. The central site duplicates CDs for archives and disaster recovery, sends CDs and ID documents to B readers, and receives the completed document from the B readers.

- The B readers' equipment is installed and calibrated by NIOSH and supplied in a configured state. The B readers complete reading and evaluation forms, then destroy the CDs.
- Viewer features include the following: custom or off-the-shelf, supporting calibrated 3 megapixel grayscale displays, read from a DICOM CD, a PA CXR display on left monitor, a single reference image on right monitor, scrolling through a reference set, supporting all grayscale variants, supporting window level and width with a sigmoid function, supporting pan and zoom, with identification and technique annotation and linear distance measurements.

Dr. Clunie listed the following recommendations for the longer term:

- The central site builds or uses an off-the-shelf PACS/RIS
- Acquisition sites submit CDs or send the information via the Internet
- Central site PACS match identifiers and archive images
- B readers view images on PACS remotely and securely over the Internet, complete forms, and make use of hardware as in the short-term solution

Dr. Clunie listed requirements for a reference set. The short-term recommended solution should assume the ILO 2008 standard, with highest-fidelity digitized data. DICOM encoding should feature DX "for presentation," contrast adjustment for P value grayscale output, window values either sigmoid or linear, black borders to reduce glare, identifying attributes in the header, spacing attributes that allow measurement of nodule size, orientation attributes that allow correct hanging, and validation as DICOM standard. The DICOM header identification should follow the current rules. The modality operator should be allowed to change social security number to a patient ID, to change date to DICOM study date, and to change the ALOSH approval number.

Facility name and address will be fixed by the field engineer. The central site (NIOSH) might have to perform some cleanup of headers during CD copying.

Discussion

It was noted that NIOSH will install B reader stations only for participants in the NIOSH program. As for a broader program, Dr. Clunie suggested that the paradigm for filing could be scaled-up. He noted that a header attribute could indicate no lossy compression. Responding to a question about their long term stability, he concluded that CDs might not be suitable for long-term archiving of data. Of course, the longer-term solution possibly will not use CDs at all. We should consider eventually a QC program for archiving.

THE ILO REVISION

Dr. Petsonk asked the workshop participants to offer ideas for revision of the ILO standards for reference images. How should the revision proceed? How should it reach out to partners? Should the ILO manufacture an image or accumulate and validate existing images?

Dr. Fedotov suggested that a next meeting focus more on international points of view as they relate to revising the guidelines. Tests of placing ILO standards on CDs have identified some problems. Perhaps NIOSH could work with ILO to prepare a CD with calibration and a test pattern and to conduct a study using various monitors. The result might lead to an international standard.

Dr. Balmes wondered whether the ILO was willing to develop generated standard images. In fact, noted Dr. Wagner, this is a research question. Dr. Samei stated that his research group has been performing simulations and validation and has learned that it takes about one year to develop a validated simulated image/pathology. Dr. Henry noted that actual candidate radiographs offer the benefit of cross-sectional comparison. Boundary radiographs of normal and abnormal cases would be helpful. Yet, noted Dr. Petsonk, how would we select from current

standard images? Dr. John Parker suggested considering only validated standard films, and he stated that we cannot simulate what dust does to the lung. Dr. Michael Jacobson stressed the need to maintain continuity with the past in advancing classification. Joseph Burkhardt proposed a two-step approach featuring research followed by the acquisition of films.

COMPARISON OF DIGITAL RADIOGRAPHS WITH FILM-SCREEN RADIOGRAPHS FOR CLASSIFICATION OF PNEUMOCONIOSIS

Alfred Franzblau, MD*, Ella A. Kazerooni, MD, Ananda Sen, PhD, Mitchell M. Goodsitt, PhD, Shih-Yuan Lee, MPH, Kenneth D. Rosenman, MD, MPH, James E. Lockey, MD, Cristopher A. Meyer, MD, Brenda W. Gillespie, PhD, E. Lee Petsonk, MD, Mei Lin Wang, MD

***Correspondence to Dr. Franzblau at:**

University of Michigan School of Public Health

109 South Observatory Street, Ann Arbor, Michigan 48109-2029, USA.

Email: af Franzblau@umich.edu, Phone: (734) 936-0758, Fax: (734) 763-8095

Support for this research was provided by grant #S2200-22/23 from the Association of Schools of Public Health (ASPH) and the Centers for Disease Control and Prevention (CDC)/National Institute for Occupational Safety and Health (NIOSH).

Abstract

Purpose: This investigation assessed the equivalency of digital radiographic images (hard copy (HC) and soft copy (SC)) with traditional film-screen radiographs (FSR) for recognition and quantification of parenchymal and pleural abnormalities due to pneumoconiosis and other forms of fibrotic lung disease using the International Labour Organization (ILO) classification system.

Materials and Methods: This report is based on analyses of readings of FSR, HC and SC images from 107 subjects by 6 certified B readers.

Results: Readings of FSR and SC for small parenchymal opacities did not differ significantly; in contrast HC readings demonstrated a significantly greater prevalence of small opacities compared to FSR and SC. The prevalence of large opacities differed significantly among all three image formats, with HC>FSR>SC, however, no difference was seen between FSR and SC readings when images with 'ax' were grouped with large opacities. The prevalence of pleural abnormalities differed significantly among all three image formats, with FSR>HC>SC.

Conclusion: ILO classification outcomes are similar using either film-screen radiographs or soft copy digital images (but not hard copy images) for small parenchymal opacities and for large opacities (with 'ax'). Both film-screen and soft copy digital images can therefore be recommended for the recognition and classification of dust-related parenchymal abnormalities. The role of digital radiography in reading for pleural abnormalities requires additional clarification.

じん肺分類密度の医師間変動を評価する新しい方法 ：CE-IRT モデル分析法の提案

松本 徹、日下幸則、志田寿夫、菅沼成文、田村太朗

研究要旨：

Item Response Theory : IRT (項目反応理論) として一般に知られる項目 (読影者) 特性モデルに、我々が考案した Conflict-equilibrium : CE (葛藤-平衡) モデルを組み合わせたじん肺分類密度分析法(CE-IRT モデル分析法と命名)を提案した。本法の基本は 1980 年 ILO 実験データに示されている医師全体の読影傾向を ILO 標準写真と同じく国際的なじん肺分類傾向の標準と見做すことにある。これより統計変動の中に存在する、当該医師のじん肺分類密度と 1980ILO 標準の対応関係を、理論モデルに基づいて滑らかな曲線として顕在化し、医師個人が 0 型か 1 型以上か等 2 つのカテゴリ間の境界を分ける密度を特定する。次いで同一対象例のじん肺分類に参加した任意の二人の医師間のじん肺分類密度の変動および各医師のじん肺分類密度と 1980ILO 標準の偏りを定量する。ここでは提案手法の原理を示し、その妥当性を 1980 年版 ILO 標準写真選定実験データを対象に検証した結果を報告した。

A. 研究目的

アナログ画像である 1980ILO 標準写真¹⁾をデジタル X 線画像 (CR, DR) の標準へ切り替えるための研究の一環として現在村田班は粉塵曝露歴が明らかなじん肺画像を班研究協力施設から収集している。筆者が想定するその後の手順は以下の通りである²⁾。1) 複数じん肺専門家が各施設から収集されたじん肺画像を鑑定し、当該画像のじん肺分類 (密度、サイズ、形状他) を決定し、じん肺画像データベースを構築する。2) 当該画像データベースからじん肺分類実験用画像を選択し、それを、我が国を代表する複数のじん肺診断医がじん肺分類する実験を行う。3) 実験データを分析

し、その結果に基づきデジタル標準画像候補を決定する。4) 以下略。本研究の目的は上記 3) のための分析法を提案し、その妥当性を検証することである。

B. 研究方法

1. 本法が対象とするデータ

複数の画像を複数の医師が 1980ILO 標準画像を参照しつつ観察して各画像のじん肺密度を 12 階尺度分類する読影実験を想定する。その時、前節 3) の分析法に関与するデータを Fig. 1 に示す。

ILO 標準画像には標準密度-1②と表記したじん肺陰影が含まれるが、それらは、1980ILO

標準写真選定実験^{3,4,5)}で27人の医師が標準写真候補106枚をじん肺分類した密度「標準密度-0」①の一部であることに注目する。すなわち、ILO標準写真の密度は、じん肺密度分類用物差しの1型、2型、3型のmid-category(1/1, 2/2, 3/3)を表す目盛りに相当する。本法はこの大きな目盛りの間に、これまで誰も指摘しなかったので表立つことはなかったが細かい目盛り「標準密度-0①」が存在すると考え、これを1980ILO標準との偏りを定量分析するための物差しとして活用する。じん肺分類される画像は前節1)で複数専門家によりじん肺分類結果が確定されたものでこれを「確定密度」③と表記する。1980ILO標準写真とは別に、それぞれの国情によって定められた各国の標準写真の密度もILO標準ではないというレベルにおいて「確定密度」に属する。それらの確定密度写真を対象に各医師がじん肺分類した密度を「分類密度」④と表記する。①、②対③/④、③対④の関係(Fig.1右図)から医師二人の分類密度④-1、④-2間二次的関係(Fig.1左図)が得られる。

2. 本法が分析目標とするデータ間の関係

Fig.1①、②、③対二人の医師のじん肺分類密度④-1、④-2の対応の基本形態とじん肺分類結果の期待値、非期待値の関係をFig.2に示す。医師二人のじん肺分類密度の対応(3)からその基になった(1)、(2)の一意的関係を復元することは原理的に不可能である。すなわち、(3)の関係のみからでは二人の医師が標準に照らして、確定密度または標準密度からどの程度偏ってじん肺分類を行ったか確認することはできない。同様にFig.2(1)の結果か

ら(3)は得られるが(2)は得られない。(2)から(1)、(3)の関係は必然の結果として明らかになる。本分析法はFig.1(右図下辺)の細かい目盛り:「標準密度-0①」の物差しを使って、標準密度対分類密度/確定密度および二人の医師の

分類密度間に存在する滑らかな関係Fig.2(2)を後述する理論モデルに基づいて推測する。

3. 分析の流れ

本分析法全体の流れをFig.3に示す。以下これに沿って方法論を述べ適用結果を示す。

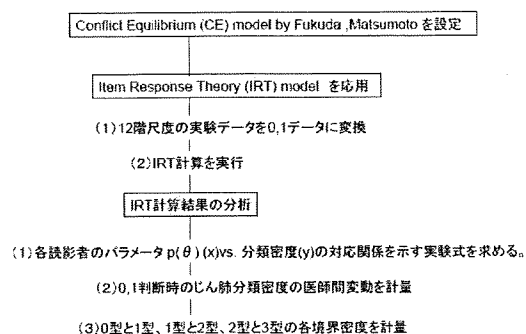


Fig.3 分析全体の流れ

4. 本分析法のじん肺密度分類の原理

4.1 葛藤—平衡 (Conflict-Equilibrium: CE) モデル

読影者がじん肺X線画像を読影してじん肺陰影の有無、ここでは0型か1型以上か(以下これを0又は1と表現)判断する場面を想定する。その時Fig.4に示すような葛藤が心の中に生じ、葛藤の程度に応じた一定時間経過後、「1」という考えが心の中に占める割合(確信度: f_1)と「0」という考えの割合(確信度: f_0)が平衡状態に到達した時、読影者はじん肺陰影の有無(1, 0)を決断すると仮定

する。但し、 $f_+ + f_- = 1$

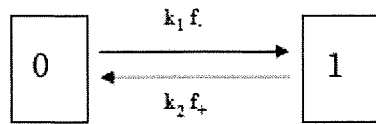


Fig. 4 葛藤—平衡モデル

k_1 は「0」→「1」への遷移確率（「1」かもしれないと心が揺れる確率）、 k_2 は「1」→「0」への遷移確率（「0」かもしれないと心が揺れる確率）を表す。

葛藤が平衡に達した時、

$$k_1 f_- = k_2 f_+$$

$$k_1 (1-f_+) = k_2 f_+$$

これより、

$$f_+ = k_1 / (k_1 + k_2)$$

ここで k_1 、 k_2 に対して以下を仮定する。

① $k_2 = \text{定数}$ （読影が慎重なほど大きくなる。）

② $k_1 = k_0 \text{EXP}[\alpha (S/N)]$

S は検出が期待されるターゲット（じん肺陰影）の信号成分、 N はターゲットに関係した雑音成分、 k_0 は比例定数、 α は読影者（個人差：読影の癖、経験、学習）、画像（胸部 X 線写真、CT：撮影条件、表示条件）等により異なる係数

k_1 は S/N が大きくなると指数的に増大する。以上のことより、 $t = S/N$ とおくと、人が X 線画像を読影して、じん肺陰影があるという考えが心の中を占める割合（確信度） f_+ は次式で表せる。

$$\begin{aligned} f_+(t) &= k_0 \text{EXP}(\alpha t) / [k_2 + k_0 \text{EXP}(\alpha t)] \\ &= 1 / [1 + (k_2/k_0) \text{EXP}(-\alpha t)] \\ &= 1 / [1 + \text{EXP}(-(\alpha t - a^*))] \end{aligned} \quad (1)$$

但し、 $a^* = 1/n (k_2 / k_0)$

(1)式で表される確信度 f_+ は単調増加関数であり、じん肺陰影密度 t が大きいほど高くなり、「1」と判断される確率が增大することを表す。

4.2 項目反応理論 (Item Response Theory : IRT) モデル⁶⁾

じん肺密度分類を 0 か 1 かで行った時の IRT モデルとして (2) 式の 2 パラメータロジスティックモデルを適用した。

$$P_j(\theta) = 1 / [1 + \exp\{-Da_j(\theta - b_j)\}] \quad -\infty < \theta < \infty$$

(2)

ここで j は IRT における項目番号を表す。本報では読影者を項目 j 、症例(画像)を被じん肺分類潜在特性値 θ （じん肺と判定されやすさの指数）と見立てた分析を行う。 a_j 、 b_j は項目 j （読影者）の特徴を表すパラメータで、 a は (2) 式を観測値の 0, 1 データに当てはめた時の項目特性関数の傾斜を表す。 b はその関数全体が特性値 θ 軸上、より右側にあるか、左側にあるかを表すパラメータである。 $P_j(\theta)$ は潜在特性値 θ における、読影者（項目） j の「1」判断率を表す。(2) で表される関数は θ の単調増加関数であり、特性値 θ が大きいほど項目（読影者）が「1」判断する確率 $P_j(\theta)$ が增大することを表す。

4.3 CE モデルと IRT モデルの結合

CE モデルと IRT モデルは (1) 式および (2) 式において、 $\alpha t - a^* = Da(\theta - b)$ と置くことにより対応付けられる。すなわち、

定数； $D=1.7$ 、 $a^* = -Dab$ 、 $\alpha = Da$ 、 $t = \beta \theta : \beta$ は比例定数

(3)

$$f_+(\theta) = 1 / [1 + \text{EXP}(-Da(\theta - b))] = P(\theta) \quad (4)$$

$$f_-(\theta) = 1 - f_+(\theta) = 1 - P(\theta)$$

後述する実験データの実際分析では、じん肺陰影の有無判断 (1, 0) の結果に IRT モデルの (2) 式を当てはめて a, b, θ を求め、 $P(\theta)$ を計算する。CE モデルはじん肺陰影密度分類の問題において (2) 式を用いることの理由やその意義を読影者の心理学的側面から説明し、変数 t を (3)、(4) により IRT モデルと結合し、 θ 、 $P(\theta)$ と関係付ける。

ここで改めて以上を整理すると、2 つの理論では以下を仮定している。

- ① CE モデルにおいてじん肺陰影の存在確信度 $f_+(\theta)$ または $f_-(\theta)$ の根拠となる、定量的、客観的パラメータ $t = S/N = \text{肺内ダスト量} \propto \text{読影者により判定されるじん肺密度} (1-12) \propto 0, 1$ 判定結果の存在および IRT モデルにおいてそれに対応する潜在特性 θ の存在
- ② CE モデルにおいて確信度特性曲線 $f_+(\theta)$ または $f_-(\theta)$ の存在および IRT モデルにおいてそれに対応する「1」判断率 $P(\theta)$ の存在

5. 実験材料

CE-IRT 法の有用性を検証するため以下の ILO 読影実験データを使用した。ILO は 1977 年から約 2 年かけて標準写真選定のため国際的な読影実験を行った [3]。日本、米国、欧州から 106 枚の標準候補写真が収集され、これらの写真のコピー各 1 セットが各国代表に送付され、各国の読影者に回覧された。読影者は 27 人、内訳は、日本 9 人、米国 8 人、欧州 10 人であった。じん肺密度分類は 1971 年版

ILO 標準写真を参照して行われた。読影結果は ILO 作成のレポートシートに記録され、読影後シートは ILO 本部に返送された。これより、ILO は各写真に対する 27 人の読影者のじん肺診断結果を Fig. 5 のようにまとめた。本報で対象としたデータは、読影者が 12 階尺度で回答した 3 種のじん肺密度分類結果：(1) 円形陰影の密度、(2) 不整形陰影の密度、(3) 円形陰影および不整形陰影のどちらか支配的な密度 ((1)、(2) の密度のうち大きい方を採択：以下 Max 陰影と略) である。

6. CE-IRT モデル投入場面

モデルは以下の場面で生じる 1, 0 データに適用される。すなわち、読影者個人がじん肺分類を 12 階尺度上のあるカテゴリの境界において、これより上か下か判断する場合を想定し、境界より下側の密度に分類された時 0 を、上側の密度に分類された時 1 を割り振る。Major category 4 段階の時のカテゴリ境界は 3 個、12 階尺度では 11 個ある。原理的にはそのすべての境界の 1, 0 判断に対して IRT モデルが適用できるが、12 階尺度の両端で 0 又は 1 のデータのどちらかが極端に少なくなった場合、計算不能となるか計算誤差が増大することに注意する必要がある。

7. 入力データの作成

Fig. 5 に示される元データを以下のごとく 0, 1 データに変換する。各読影者が各症例に対して 12 階尺度で回答したじん肺密度：1, 2, 3 (0 型)、4, 5, 6 (1 型)、7, 8, 9 (2 型)、10, 11, 12 (3 型) に対して 0 型か 1 型以上の場合に対して IRT 分析する時は 1, 2, 3 (0 型)

に回答されたものに 0 の値を、4 以上に回答されたものに 1 の値を割り当てる。1 型以下か 2 型以上かの場合は、1, 2, 3(0 型)、4, 5, 6(1 型)に回答されたものに 0 を、それ以外に 1 を割り当てる。以下同様にして設定した境界より下側に分類された密度に対して 0 を、境界より上側の密度を 1 に置き換える。

8. 1, 0 データ配列

1, 0 データの配列を如何にすれば所定の分析結果が精度よく得られるか、本モデルを考案した現時点ではデータ配列最適化の方法は定まっていない。本報告では Table 1 左に示すごとく列方向に読影者 No. 1-27 の項目を置き、それぞれに対して、行方向に(1)症例 No. 1-106 の円形陰影に対する 0、1 データ、(2) 症例 No. 1-106 の不整形陰影に対する 0、1 データ、(3) 症例計 318 行のデータを配置した。

9. IRT 計算の実行

BILOG-MG3⁷⁾および BILOG を Excel から利用するためのマクロパッケージを使用した。

10. 出力データ

IRT モデル計算により Table 1 右に示すごとく読影者 27 人に対応する項目特性曲線のパラメータ a, b と症例 106 例の円形陰影密度、不整形陰影密度、Max 陰影密度それぞれに対する θ が得られる。27 個の a, b と延べ 318 症例の θ を(2)式に代入すると読影者 27 人それぞれの $P(\theta)$ データが得られる。

11. CE-IRT モデルパラメータ間の対応

本分析法では、CE モデルから導かれる肺内

ダスト量を表すパラメータ $t=S/N$ が以下の IRT モデルパラメータおよび観測と対応することが仮定されている。

$t=S/N$ (ダスト量：未知)

\propto 被じん肺分類画像の潜在特性値 θ_k : $k = 1-318$

\propto 項目 (読影者) 特性曲線の $P_j(\theta)$: $j = 1 - 27$

\propto 日本、米国、欧州読影者の平均項目特性曲線の $P_j(\theta)$ 、 $P_U(\theta)$ 、 $P_E(\theta)$

\propto 27 人の読影者の平均項目特性曲線の $P_m(\theta)$

\propto 各症例に各読影者が 12 階尺度で判定したじん肺密度の 27 人分の平均 : t_{m1-12}

\propto 各症例に各読影者がじん肺陰影の有無を 0, 1 判断した 27 人分の平均 : $t_{m0,1}$

12. 標準密度-0①の設定と分類密度④(Fig. 1 参照)の推定

読影者 27 人に対する観測値 : t_{m1-12} を従属変数、IRT 計算結果である $P_m(\theta)$ を独立変数として $y_m(t_{m1-12})$ を推測する実験式を求める。0 型か 1 型以上かじん肺分類した場合を(5)式に示す。本実験式は医師個人が 0 型か 1 型以上か判断した時の分類密度が 1980ILO 標準とどれだけ偏っているか定量するための ILO 標準(物差し)を表す。(このような実験式は Major category 4 段階で 3 個、12 階尺度では 11 個求められるが略。)各読影者および国別の $P(\theta)$ を実験式に代入し、それぞれの推測分類密度④(Fig. 1) :

$y_m(t_{m1-12})$ を推定する。

$$y_m(t_{m1-12}) = 1.5242 + 9.1316(P(\theta)) - 19.635(P(\theta))^2 + 17.2041(P(\theta))^3$$

(5)

13. じん肺密度分類の医師間（内）変動の定量とカテゴリ境界密度の特定

①読影者 27 人のそれぞれの項目（読影者）特性曲線および読影者全体 27 人、日本 9 人、米国 8 人、欧州 10 人それぞれの平均項目特性曲線を求めて比較した。

②2 者間の変動を以下のごとく定義した。

読影者 27 人の任意の二人および日本対米国、日本対欧州、米国対欧州の特性曲線から同一 θ （被じん肺密度分類潜在特性）の $P_1(\theta)$ 対 $P_2(\theta)$ およびそれらを (5) 式により密度変換した y_1 対 y_2 より、

$$\Delta P(\theta) = P_1(\theta) - P_2(\theta) \quad (6)$$

$$\Delta y_m(t_{m1-12}) = y_1(P_1(\theta)) - y_2(P_2(\theta)) \quad (7)$$

$\Delta P(\theta)$ 、 $\Delta y_m(t_{m1-12})$ を標準密度 -0 ① (Fig. 1) : $y_m(t_{m1-12})$ の関数として評価する。

③複数読影者の平均の項目特性曲線上で医師間変動が最大になる点および個人の特性曲線上で医師内変動が最大になる点は $P_j(\theta) = 0.5$ の時、すなわち、読影者が 0 か 1 か決断する境界で生じる。

その時の θ は $dP_j(\theta)/d\theta = Da_j P_j(\theta)(1 - P_j(\theta)) = 0$ を満たす $\theta = b_j$ である。 $P_j(\theta) = 0.5$ に対応する密度 $y_m(t_{m1-12})$ をカテゴリ境界密度とする。

C 研究結果

1. 本法が基本とする関係—0 型か 1 型以上か判断の場合

Fig. 6 上段に日本 9 人、米国 8 人、欧州 10 人の各人の項目特性曲線 $P(\theta)$ を示した。日本の読影者は曲線が θ 軸上比較的左方に位置する 4 人と右方の 5 人に二分したことが観察される。下段は各国の平均特性曲線と各症例に各国読影者が 0, 1 判断した時の平均値 $t_{m0,1}$ を θ の関数としてプロットしたものである。CE モデルを反映した観測値の 0, 1 判断結果の平均の間に国別平均 $P(\theta)$ 曲線があることが示されている。

Fig. 7 左は読影者 27 人の平均 $P_m(\theta)$ と 106 の各例に 27 人が 12 階尺度分類した平均値 t_{m1-12} を対応させた結果である。平均 $P_m(\theta)$ を独立変数、平均値 t_{m1-12} を従属変数として 3 次多項式を当てはめた時の推測曲線 y_m と 2.7 で示した実験式 (5) が表示されている。本分析法ではこの関係（実験式）を医師が 0 型か 1 型以上か判断する時の 1980 ILO 標準と見なす。その使い方については後述する。Fig. 7 右は横軸を θ にとった場合である。Fig. 6 上段に日本 9 人、米国 8 人、欧州 10 人の各人の項目特性曲線 $P(\theta)$ を示した。日本の読影者は曲線が θ 軸上比較的左方に位置する 4 人と右方の 5 人に二分したことが観察される。下段は各国の平均特性曲線と各症例に各国読影者が 0, 1 判断した時の平均値 $t_{m0,1}$ を θ の関数としてプロットしたものである。CE モデルを反映した観測値の 0, 1 判断結果の平均の間に国別平均 $P(\theta)$ 曲線があることが示されている。

Fig. 7 左は読影者 27 人の平均 $P_m(\theta)$ と 106 の各例に 27 人が 12 階尺度分類した平均値 t_{m1-12} を対応させた結果である。

読影者 27 人の平均 $P_m(\theta)$ を独立変数、平均値 $t_{m-12}3$ を従属変数として 3 次多項式を当てはめた時の推測曲線 y_m と 2.7 で示した実験式 (5) が表示されている。本分析法ではこの関係 (実験式) を医師が 0 型か 1 型以上か判断する時の 1980ILO 標準と見なす。その使い方については後述する。Fig. 7 右は横軸を θ にとった場合である。

2. 0 型か 1 型以上か判断の医師間変動

二人の読影者の医師間変動を本法により求める一連の工程とそれに対応する観測値の関係を Fig. 8 に示す。左図は IRT モデル計算で得られた日本医師 No. 3 と米国医師 No. 1 の $P(\theta)$ 曲線である。Fig. 8 右図 A は日本医師 No. 3 が 0 型か 1 型以上か判断した時の分類密度 $y_m(P(\theta))$ を縦軸にとった場合を示す。 $P(\theta)$ から y_m への変換は、Fig. 7 に示す標準曲線において $P(\theta)$ と同一値の $P_m(\theta)$ に対応する密度 y_m として求めた。同様に B は米国医師 No. 1 の場合である。各医師の推定分類密度曲線の周囲に散在する点は二人の医師が 106 例に対して与えた 12 階尺度の観測値である。Fig. 8 右下段に、A と B を対応させた場合および二人の分類密度の差： $\Delta y_m = A - B$ (0 型か 1 型以上か密度分類の医師間変動) を Fig. 7 の縦軸 y_m (1980ILO 標準) の関数とした場合を示す。ばらつきの大きい観測値の間に分析モデルに基づいて推測された滑らかな関係が示されている。

以下同様にして読影者 27 人の任意の 2 者間のじん肺密度分類の変動を評価することができる。Fig. 9 左上段に一部の例として日本医師

No. 3 対米国医師 No. 1-8 の医師間変動： Δy_m を示す。同一の θ (肺内ダスト量に比例するパラメータ) に対して日本医師 No. 3 は米国の 3 人の医師よりは明らかに高く、他の 3 人よりは明らかに低く、他の 2 人とは同程度にじん肺分類したことが y_m (1980ILO 標準) の関数として示されている。読影者 27 人の任意のすべての組み合わせ (351 対) における密度分類の変動を一度に見渡せるようにするため Fig. 9 左下のごとく 2 者一対の Δy_m 曲線の平均を計算した。一人の医師対他の 26 人の間で 26 個の Δy_m の平均が得られる。Fig. 9 右上段に、27 人ごと、国ごとのその結果を箱ひげ図で比較した。

3. 境界密度の特定と 1980ILO 標準との偏りの定量

0 型か 1 型以上かの判断を医師がどの密度で行ったかを 2.8③に示す方法で特定する。すなわち、Fig. 10のごとく各読影者の $P(\theta) - y_m$ の関係を求め $P(\theta) = 0.5$ に対応する y_m を読影者が 0 型と 1 型を分けた密度とする。次に 1 型と 2 型を分ける境界密度を求める。すなわち、1 型以下の 12 階尺度分類密度を 0 に、2 型以上のそれを 1 に置換し、IRT 計算を実行する。1 型以下か 2 型以上か判断する時の $P_m(\theta) - y_m$ の関係を求めこれに対して 2.7(5) 式に相当する ILO 標準を表す実験式を立てる。Fig. 10のごとく $P(\theta) - y_m$ の関係を求め $P(\theta) = 0.5$ の時の y_m を 1 型と 2 型の境界密度とする。2 型と 3 型の境界密度も同様の手順で求める。Fig. 11 に Major category の境界密度の期待値 (縦軸)：3.50、6.50、9.50 対日本、米国、欧州の平均境界密度を横軸にブ

ロットした結果を示す。折れ線は境界密度間隔が等間隔でないこと、対角線との隔たりは読影者の分類密度と 1980ILO 標準の偏りを表す。

D 考察

1. 理論モデルのパラメータの意味

本報で提案した CE モデルは、読影者が 0 か 1 か決断する時の葛藤が確信度 0~1 の間で平衡に達した時、確信度 f_c がじん肺画像に撮像されたじん肺陰影密度 t を変数とするロジスティック関数(1)で表されるとしたものである。一方、IRT モデルは、読影者が「1」判断する確率、すなわち、項目（読影者）特性曲線の縦軸： $P(\theta)$ が被じん肺分類画像の潜在特性値 θ ：横軸の関数である 2 パラメータ・ロジスティックモデル(2)式で表されるとしたものである。特性曲線は読影者の数だけ計算される。両モデルの対応から θ は読影者がじん肺分類の対象とした各症例のじん肺密度 t ：S/N(ダスト量)に比例するパラメータと推察される。すなわち、 θ が小さいほどじん肺密度は 12 階尺度で「1」の側に、 θ が大きいほど「12」の側に分類される特性をもつ。 θ と対応する $P(\theta)$ は何を意味するか。複数読影者の平均の特性曲線の場合と 1 読影者の特性曲線の場合とで解釈が異なる。前者は潜在特性値 θ をもつじん肺症例に対して「1」と判断した読影者の割合を、後者は読影者が「1」かもしれないと確信する度合いを表すと考えられる。Fig. 6 下段に示されるごとく日本、米国、欧州医師の平均項目特性曲線は観測値である各症例の 1, 0 判断の平均 $t_{m,1}$ と比較的よ

く適合した。この事実は、 $P(\theta)$ が特性値 θ の被じん肺分類画像に対する「1」判断率：複数読影者中何人が「1」と判定したか表すことを裏付けたものである。Fig. 7 左：27 人の平均項目曲線 $P_m(\theta)-y_m$ に示される ILO1 型標準写真（赤丸）は 0 型か 1 型以上かの判断において 27 人中約 80%が「1:1 型以上」、20%が「0:0 型」とし、残りの標準写真（赤丸）はほぼ全員が「1:1 型以上」と判断したことを表している。

2. 本法の特長

最近の従来法は Fig. 2(3) に示す 2 者の分類密度の一致の評価結果でもってアナログからデジタルの標準に切り替えることを目指しているように見える。⁹⁾ それに対して本分析法は Fig. 2(1)、(2) の関係を対象とし、2 者の分類密度の差を ILO 標準(物差し)に照らして定量評価することを基本とする。Fig. 2(3) の結果は付随的に必要ならいつでも分析し提出できる。

観測値をベースにした分析結果は対象症例の偏りに依存する傾向があるが、Fig. 8 に示したごとく本法は大きな統計変動を含む観測値の中にある医師全体および医師個人の系統的な「偏りの傾向」を理論モデルに基づいて 12 階尺度の全体にわたって推測する。Fig. 10、11 に示したごとく医師の分類密度および国情に基づいて設定された各国標準写真の確定密度や複数専門家による読影実験用画像の確定密度を国際的原器である 1980ILO 標準密度の関数として求める。これより Fig. 9 のごとく 2 者の分類密度の差が算出され比較される。さらに Fig. 11 に示した通り、分類カテゴリの

境界をどの密度で分けているかを医師ごと、国別に特定する。その結果、日本の医師は欧米医師より 0 型と 1 型の境界をやや低めの密度で分けていること、従って結果的に 12 階尺度全体で平均 0.5-1.0 程度高めに分類していることが明らかにされた。これらの知見は以前の報告^{3,4,5)}を裏付けており、本法の妥当性を示すものである。本分析法はじん肺診断の基本に則った方法であると考えられる。

3. 今後の課題

方法論をさらに拡充する必要がある。最も基礎的で具体的な課題は Table 1 で示した 1, 0 データ配列の最適化を理論的、実験的に検討することである。また、ILO 標準と確定密度とのトレーサビリティを明らかにするため確定密度を仮の標準に据えた時と ILO 標準を用いたときの分類密度評価結果の異同を評価する方法論を確立する必要がある。さらに、本報で使われたデータは医師 27 人が 106 枚の胸部 X 線写真を読影するという大規模なものであった。研究や教育のため通常行われる比較的小規模な実験データに対しても応用可能かどうか一般性を検証する必要がある。

E. 結論

1980 年 ILO 標準写真選定実験データが示す読影傾向全体を ILO 標準写真と同じようにじん肺密度分類評価用の「標準」と見做し、医師が 0 型か 1 型以上か等じん肺分類カテゴリ境界で 0, 1 判断した時の ILO 標準との偏りを検出する新しい方法：CE-IRT モデル分析法を提案した。本法を 1980 ILO 実験で得られたじん肺分類データに応用し、観測値の統計的ば

らつきの中に埋没して肉眼では観察が容易でないか認められても定量が困難なじん肺密度分類の医師間変動を滑らかな傾向として顕在化できることを示した。また、じん肺密度カテゴリ間の医師内変動を最大にする医師個人のカテゴリ境界密度を特定し 1980 ILO 標準とのじん肺分類密度の偏りが定量できることを確認した。

謝辞

CE モデルの考案およびその応用に関して長年にわたり共同研究を賜った故福田信男先生（放医研臨床研究部）、じん肺診断についてご指導ご教示賜った故野辺地篤郎先生（聖路加国際病院）、本報の基となるじん肺分類研究の機会にご支援をいただいた中央労働災害防止協会 PXQ 委員会、全衛連、厚労省がん研究助成金長谷川班、安曇プロジェクト、新潟大―富士通共同研究ほか、関連の多くの施設・皆様に感謝する。

F. 参考文献

- 1) The ILO international classification of radiographs of pneumoconioses (Revised Edition 2000), Occupational safety and health series NO. 22 (Rev. 2000), ILO Publications: 2000 年版じん肺 X 線写真国際分類使用のガイドライン、監訳者：日下幸則、相澤好治、志田寿夫他、保健文化社、2003. 11. 10
- 2) 日下幸則、松本 徹、志田寿夫他：じん肺におけるエックス線デジタル撮影画像活用のための基礎的検討（標準写真関連研究のレビ

- ューと標準作り・読影結果評価法の考察)、厚生労働科学研究費補助金労働安全衛生総合研究事業「じん肺健康診断におけるエックス線デジタル撮影画像の活用に関する研究」平成19年度総括・分担研究報告書(主任研究者 村田喜代史)、平成20年3月、28-65, 2008
- 3) Y. Kusaka, T. Matsumoto, T. Shida et al.: Diagnostic Variation in ILO standard films (rev. 80) for the diagnosis of pneumoconiosis, 厚生労働科学研究費補助金労働安全衛生総合研究事業「じん肺健康診断におけるエックス線デジタル撮影画像の活用に関する研究」平成19年度総括・分担研究報告書
(主任研究者 村田喜代史)、平成20年3月、58-65, 2008
- 4) 細田 裕、松本 徹、野辺地篤郎: じん肺標準写真の歴史、臨床放射線、36(2):195-203, 1991
- 5) Y. Hosoda: 57-70, 2003, (Eds.) Y. Kusaka, K. G. Hering, J. E. Parker: International classification of HRCT for Occupational and environmental respiratory diseases, Springer-Verlag Tokyo
- 6) 芝 祐順: 項目反応理論—基礎と応用, 初版、東京: 東大出版会; 1991
- 7) M. F. Zimowski, E. Muraki, R. J. Mislevy et al.: BILOG-MG3: Scientific Software, 2007
- 8) 菊地賢一: VBAを用いたBILOGMG 3.0を利用するためのExcelワークブックの開発, 日本テスト学会第2回大会発表論文抄録集, 2004. 8. 28
- 9) 厚生労働科学研究費補助金労働安全衛生総合研究事業「じん肺健康診断におけるエックス線デジタル撮影画像の活用に関する研究」平成19年度総括・分担研究報告書(主任研究者 村田喜代史)、平成20年3月

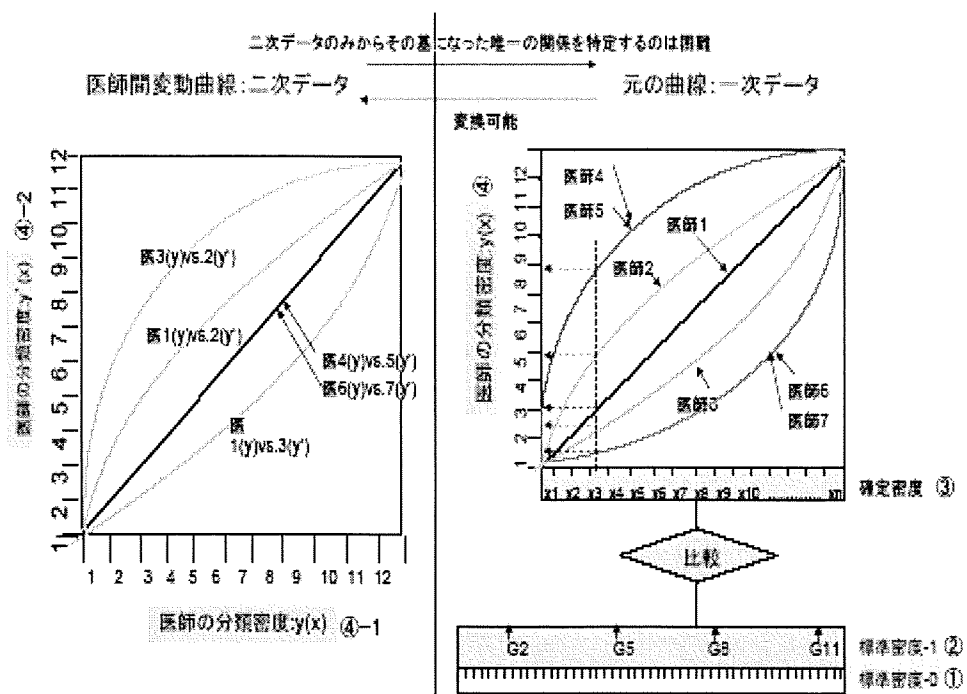


Fig. 1 分析に關与するデータ一覧

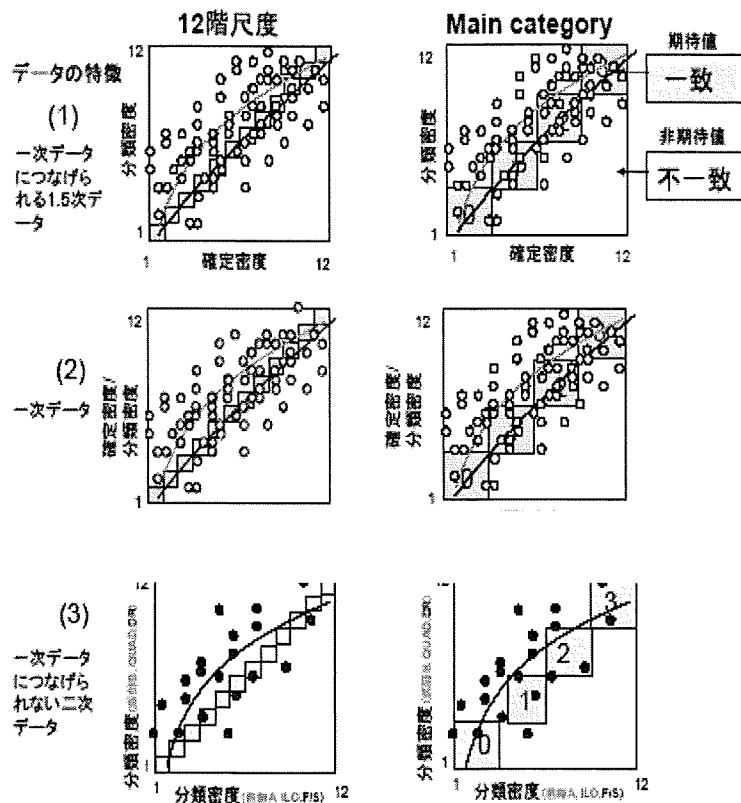


Fig. 2 分析対象となる二人の医師のじん肺分類密度対応の基本形態

FILM # 1																											ALL JAPAN USA EUROPE																														
READER #																											MEAN	MEAN	MEAN	MEAN																											
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	SO	SO	SO	SO																											
FILM QUALITY																											2	1		1	2	1	1	2	1	1	1	2	1	2	2	2	2	2	2	2	1	2	1	2	1	1					
ROUNDED OPAC: TYPE																											R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R				
PROP																											11	11	11	10	11	10	11	11	10	11	10	11	12	9	11	11	10	10	0	0	12	9	11	11	11	11	10.37	10.67	10.43	9.55	
ZONES: UR																											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	1.11	0.50	0.92	1.52
ZONES: UL																											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
ZONES: RD																											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
ZONES: RL																											X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X				
ZONES: LR																											X																														
ZONES: LL																											X																														
IRREG. OPAC: TYPE																																																									
PROP																												2	2				2	2	2	2	2	2	2	2	2	2	2	2							2	2	2	2.00	2.00	2.00	2.00
ZONES: GR																																																						0.0	0.0	0.0	0.0
ZONES: UL																																																									
ZONES: RD																																																									
ZONES: RL																																																									
ZONES: LR																																																									
ZONES: LL																																																									
COMBINED PROP																											1			2						2			11					0			9						5.50	1.50	0.50	4.50	
MAX (ROUNDED, IRREG) PROP																											11	11	11	10	11	10	11	11	10	11	10	11	12	9	11	11	10	10	0	0	12	9	11	0	11	11	10.37	10.67	10.43	9.55	
																																																						1.11	0.50	0.92	1.52

Fig.5 1980ILO 標準選定実験データ:候補写真 No.1 に対する 27 人の医師のじん肺分類結果
 [Reader No.1-9(Japanese), No.10-17(American), No.18-27(European)]

Table 1 各読影者の IRT 入力データと出力データ
 (1)入力 (2) 出力

Shape	Assessment	Image	Reader 1	Reader 2	...	Reader 27	θ	$p1(\theta)$	$p2(\theta)$...	$p27(\theta)$
Rounded	1/2point scale	Image 1	0	0	1	0	$\theta 1$	$p1(\theta 1)$	$p2(\theta 1)$...	$p27(\theta 1)$
		Image 2	1	1	1	1	$\theta 2$	$p1(\theta 2)$	$p2(\theta 2)$...	$p27(\theta 2)$
		Image106	0	1	1	1	$\theta 106$	$P1(\theta 106)$	$p2(\theta 106)$...	$p27(\theta 106)$
Irregular	1/2point scale	As above	0	0	1	0	As above	As above	As above		As above
		As above	0	0	1	0	As above	As above	As above		As above
		As above	1	0	0	0	As above	As above	As above		As above
Max	1/2point scale	As above	0	0	0	0	As above	As above	As above		As above
		As above	0	0	1	1	As above	As above	As above		As above
		As above	0	0	1	1	As above	As above	As above		As above

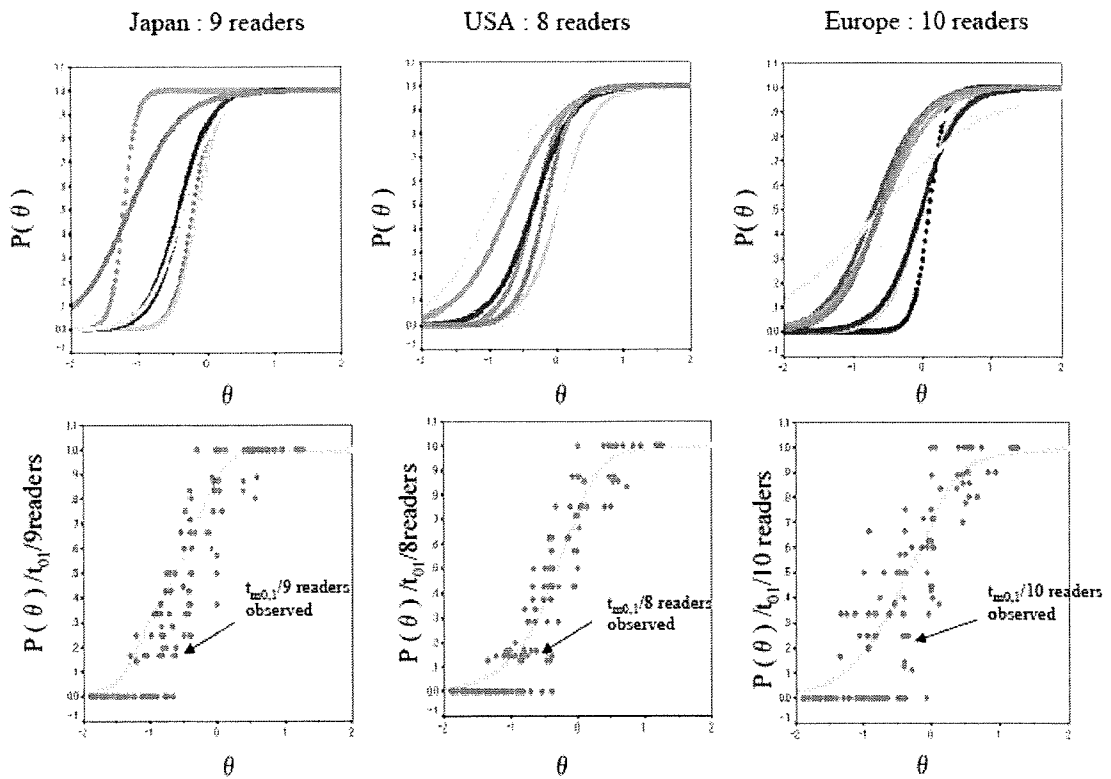


Fig.6 IRT モデル計算結果、上段: 医師個人の $P(\theta)$ 曲線、下段: 日本、米国、欧州医師の平均 $P(\theta)$ 曲線と 1,0 データの平均との対応

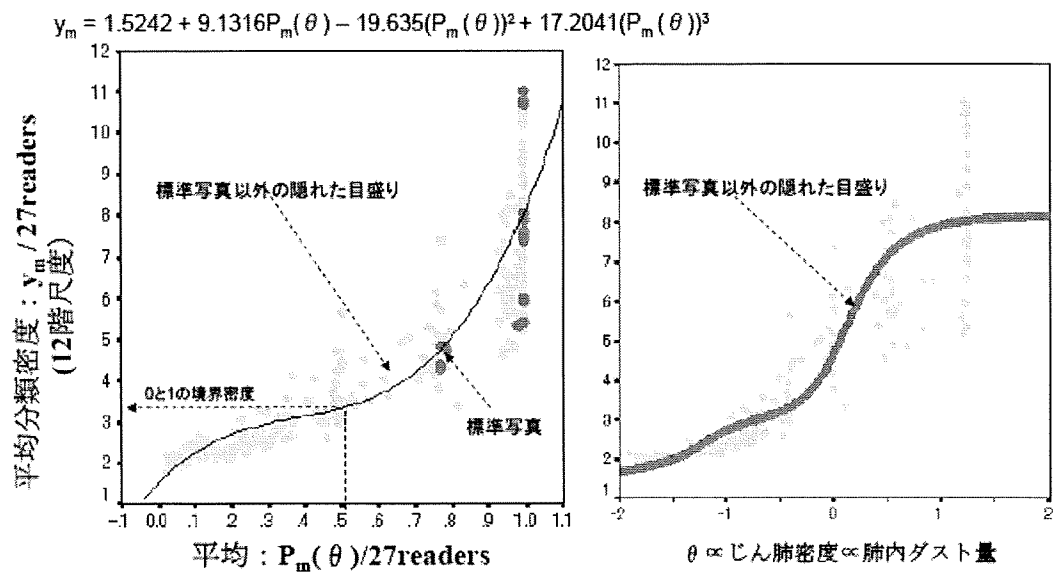


Fig.7 0 型か 1 型以上か密度分類する時の 1980ILO 標準: $P_m(\theta) \rightarrow y_m/\theta$