

満たすことを前提に、DRの活用が認められた<sup>(10)</sup>。

この研究班では5人のじん肺審査医が読影実験に参加したが、じん肺健康診断に精通していない医師においても、同様の結果が得られるかどうかを追加検証する目的で、本研究を実施した。

## B. 研究方法

### 1. 症例

平成17年-18年に実施した厚生労働科学研究で収集したじん肺症例と同じ症例を用い読影実験の対象とした。

まず、福島労災病院において、院内倫理委員会の承認を受けた後、じん肺患者10名において、フィルム法による胸部エックス線写真とDRシステム(キャノン社製)による胸部エックス線写真を撮影した。さらに、種々のじん肺エックス線分類の症例を加えるために、福井大学医学部国際社会医学講座環境保健学教室で収集した、じん肺患者でフィルムとDR(キャノン社製)写真の両者を撮影した症例を加えた。

これらの症例の中から、読影実験に参加しない医師によって、エックス線分類の0型、1型、2型、3型を数例ずつ、計30例を選択した。当初は、粒状影と不整形を含めた読影実験を計画したが、不整形陰影が2例のみであったために粒状影と同時に読影実験を行うことは不適切と考え除外した。したがって、採用した症例は28例で、その内訳は0型9例、1型11例、2型6例、3型2例である。

### 2. じん肺症例の読影実験

読影者は6名の胸部放射線科医で、胸部エックス線写真の読影経験の豊富な医師である。ただし、じん肺健康診断におけるエックス線分類判定の業務に携わった経験がほとんどない医師である。

28例のフィルムとDR写真、計56枚をシャウカステン上にランダムな順序で1枚ずつ提示し、各読影者が独立して、エックス線分類を判定した。平成17年-18年に実施した読影実験では、エックス線分類に関して(0型、1型、2型、3型)のみ判定したが、今回は、さらに通常のじん肺健康診断において記載される12階尺度ではどの程度一致するかの検討を加えるために、エックス線分類は12階尺度を用いた。また、粒状影に病変を絞っていることから粒状影の大きさについても判断を求め、読影者は(p, q, r)を追加記載した。読影前に、じん肺標準写真の0型から3型のフィルムを提示し、12階尺度を含め、エックス線分類に関する説明を読影者に対して行った。読影者に各分類の症例数の情報は与えなかったが、標準フィルム(0型、1型、2型、3型)は読影実験を通してシャウカステン上に掲示し、いつでも参照できる環境とした。

読影の一致の程度をフィルムとDR間、および読影者間で比較し、定量的指標として、一致率とともにKappa valueを用いた。また、平成17年-18年の読影実験結果との比較も行った。

## C. 研究結果

### 1. フィルムとDRの一致度の検討

6人の読影者において、粒状影の密度と大きさの判定がフィルムとDR写真でどの程度一致したかをまとめたのが表1である。

12階尺度で判定した場合の粒状影密度の一致率の平均が44.6%、4段階尺度で判定した場合の平均が75.0%であった。4段階尺度でのデータのカッパ解析では、読影者間のばらつきがみられるものの、kappa valueの平均は0.5938で、許容できる一致度を示した。一方、粒状影の大きさ

の判定においても、3段階尺度での判定の一致率の平均は80.0%、kappa valueは0.5888で、やはり許容範囲の一致度であった。

6人の判定の中で多数を占める判定（4段階尺度における）をその症例の判定としたときのフィルムとDR写真の一致度をまとめたのが表2および表3である。4段階尺度での粒状影密度の一致度は、一致率が85.7%、kappa valueが0.7918と良好な一致を示した。また、粒状影の大きさに関しても、一致率84.2%、Kappa valueが0.7047と良好な一致を示した。

さらに、粒状影密度において、個々の判定がどのように分布しているのかをみるために、6人の判定をすべて1つの表にあらわしたのが、表4および表5である。粒状影の密度では、12階尺度において一致率が44.0%と低く、4段階尺度では一致率が72.6%、kappa valueが0.6041であった。また、全体の分布をみると、DR写真での判定がフィルムでの判定に比して、より高くなる傾向がみられた。

## 2. 読影者間の判定の一致度

4段階尺度における粒状影の密度の判定の一致度をすべての読影者ペア間で、フィルムとDR写真に分けて、Kappa valueを求めたのが表6である。フィルムにおける読影者ペア間の一致度の指標であるKappa valueの平均値は、0.3967で、一致度がかなり低いことが明らかになった。DR写真においても、Kappa valueの平均値は0.4486と高くはないが、フィルムにおけるKappa valueよりも高いことが示された。

## 3. 今回データと相澤班データの比較

フィルムとDR写真の判定における一致度を見

たKappa valueを、今回と同じ症例を用いて異なった読影者で行った相澤班データと比較したのが表7である。また、読影者間におけるばらつきを比較するために、相澤班において得られたKappa valueを表8に示す。

フィルムとDR写真間の一致度をみると、Kappa valueの今回の平均値は、相澤班で得られた平均値と比較して有意に低いことが明らかになった。

また、読影者間のばらつきの指標であるKappa valueは相澤班データと比較すると、フィルムにおいてもDR写真においても有意に低いことがわかった。ただ、DR写真の一致度がフィルムより良好であることは、相澤班データでも今回データでも同様に示された。

## D. 考察

### 1. フィルムとDR写真の一致度について

じん肺エックス線分類におけるフィルムとDR写真の一致度については、適切な条件で表示されたDR写真であれば、ほぼ同等の診断能があることが、相澤班の研究データによって示されているが、今回の研究はそれを追試するものである。

フィルムとDR写真における粒状影の密度判定における一致率はKappa valueが0.5938で、相澤班データの0.6975と比較するとやや低いが、読影者間でのばらつきがより大きいことを考えると、フィルムとDR写真間の比較としては十分に許容できる一致度と考えることができる。

### 2. 12階尺度と4段階尺度について

今回の研究では、粒状影の密度を判断する基準として、相澤班で行わなかった12階尺度による判定を行い、解析は12階尺度と4段階尺度の両者で行った。これは、通常、じん肺のエックス線分類として使用されていることを考慮してであ

ったが、12階尺度を用いた場合、フィルムとDR写真の判定の一致率が平均で44.6%と非常に低く、また今回は検討しなかったが、読影者間でも一致率が同様に低いことが推測された。したがって、12階尺度はあくまでも一つの目安であり、これに基づく種々の施策の実施は避けることが望ましいと考えられる。ただ、興味深い点は、6人の読影者の判定を集積したデータ(表4)を見ると、DR写真の方がフィルムよりも、より高い判定を出す傾向がある点である。これは、DR写真の特徴を反映しているのかもしれない。

これに対して、4段階尺度は、一致率が平均で72.6%あり、DR写真がフィルムと同等の判断を下せるということを再度示しているとともに、現在の4段階尺度による管理区分が妥当であることを示唆している。

また、今回の検討では、粒状影の大きさについてもフィルムとDR写真の間で判定の一致度がどの程度かを検討したが、一致率が平均で80%、Kappa valueが0.5888と許容できる一致度を示した。P, q, rという粒状影の大きさの判断自体は、じん肺患者における管理にそれほど大きな影響を与えることはないが、このデータもDR写真がフィルムと同等にじん肺健康診断に利用できることを支持する傍証になると考えられる。

## 2. 読影者間の判定のばらつきについて

今回の検討では、6人の読影者のすべての読影ペアについて、Kappa valueを求めたが、フィルムにおいては平均が0.3967、DR写真では0.4486とやや低い値となった。これは、今回の読影者は胸部放射線診断医であるものの、じん肺のエックス線分類に関する経験がほとんどないために、1型、2型といった判断を下すしつかりとした基準

がないために、個人間の違いがより強調されたものと思われる。

しかし、DR写真ではKappa valueが中等度であるもののフィルムより有意に高い点は注目される。判断基準が弱いにもかかわらず、ばらつきがフィルムより少ないことを意味しており、これは、DR写真の方がより標準的な判断が下せる可能性を示唆しているのかもしれない。この点に関しては、異なった読影者で行った相澤班におけるデータでも同様の傾向が出ており、偶然とは考えにくい。

読影者間のKappa valueよりも、フィルムとDR写真間のKappa valueが大きかったのは、方法間のばらつきがより小さいことを表しているが、これは、同じ読影者が判断するので、画像は異なっても同じ症例に対しては同じ判断基準が働くことが大きい要因として考えられる。

## 3. 相澤班データとの相違の原因について

相澤班でのデータと比較すると、フィルムとDR写真の一致度においても、読影者間の一致度においても、今回の研究の方がKappa valueが低くなり、ばらつきが有意に大きいことが明らかになった。相澤班における研究と今回の研究の違いを考えると、症例は同じであることから、読影者の違いに原因があると考えざるを得ない。

今回の研究の読影者はいずれも胸部エックス線写真の読影に関しては豊富な経験があり、診断能力が劣るものではない。ただ、今回の判定はじん肺のエックス線分類を判定するものであり、各読影者ともに、0型、1型、2型、3型の判定基準に対する知識あまりなく、じん肺健康診断に関わった経験もないことから、直前に標準フィルムによる説明を受けても、個々の症例ごとに一定の

基準で判断することが難しかったものと推測される。一方、相澤班では読影者は中央あるいは地方のじん肺審査医としての経験を積んでいる5人であったので、判定に習熟していた点が大きな違いである。結果として、フィルムとDR写真間においても読影者間においても、今回の読影者における判定のばらつきが相澤班よりも大きくなったものと考えられる。

このことは、じん肺健康診断におけるエックス線分類の判定には、この作業に対する経験を積んだ医師があたることを望ましいということを示唆しており、トレーニングを含めた教育体制の一層の整備が望まれる。

## E. 結論

じん肺患者のエックス線分類におけるアナログ画像とDR画像の一致度を相澤班とは異なった読影者で検討した。アナログ画像とDR画像の判定は許容できる一致度を示したが、じん肺判定の経験が少ない読影者では一致度が小さくなることが示され、経験を積んだ医師がじん肺健康診断にあたることを望ましいと考えられた。

## F. 参考文献

1. McAdams HP, Samei E, Dobbins III J, et al. Recent advances in chest radiology. *Radiology* 2006; 241: 663-683.
2. じん肺健康診断へのCRの活用に関する調査研究検討会-報告書、中央労働災害防止協会、平成13年3月。
3. Uffmann M, Neitzel U, Prokop M, et al. Flat-panel-detector chest radiography: effect of tube voltage on image quality. *Radiology* 2005; 235: 642-650.

4. Metz S, Damoser P, Hollweck R, et al. Chest radiography with a digital flat-panel detector: experimental receiver operating characteristic analysis. *Radiology* 2005; 234: 776-784.
5. Kroft LJ, Veldkamp WJ, Mertens BJ, et al. Comparison of eight different digital chest radiography systems: variation in detection of simulated chest disease. *Am J Roentgenol* 2005; 185: 339-346.
6. Ono K, Yoshitake T, Akahane K, et al. Comparison of a digital flat-panel versus screen-film, photofluorography and storage-phosphor systems by detection of simulated lung adenocarcinoma lesions using hard copy images. *Brit J Radiol* 2005; 78: 922-927.
7. Harner OW, Sirlin CB, Strotzer M, et al. Chest radiography with a flat-panel detector: image quality with dose reduction after copper filtration. *Radiology* 2005; 237: 691-700.
8. Bacher K, Smeets P, Vereecken L, et al. Image quality and radiation dose on digital chest imaging: comparison of amorphous silicon and amorphous selenium flat-panel systems. *Am J Roentgenol* 2006; 187: 630-637.
9. 相澤好治. 職業性呼吸器疾患の予防及び健康管理に関する研究. 平成17-18年度総合研究報告書. 平成19年3月。
10. じん肺健康診断等へのDR(FPD)の使用に関する検討会報告書. 中央労働災害防止協会、平成19年10月。

表1 フィルムとDR写真の一致度

読影者	粒状影の密度			粒状影の大きさ	
	1 2 階尺度	4 段階尺度		3 段階尺度	
	一致率 (%)	一致率 (%)	Kappa	一致率 (%)	Kappa
1	46.4	78.6	0.6216	85.9	0.7707
2	46.4	78.6	0.6866	85.7	0.7529
3	46.4	67.9	0.5556	73.7	0.4571
4	50.0	75.0	0.6245	82.3	0.4848
5	35.7	60.7	0.4359	85.7	0.5962
6	42.9	75.0	0.6384	66.7	0.4712
平均	44.6 ± 4.9	72.6 ± 7.0	0.5938 ± 0.088	80.0 ± 8.0	0.5888 ± 0.1429

表2 フィルムと DR 写真の一致度  
 (6人の判定のうち、最多のものをその症例の判定とした場合)  
 (同じ数の場合は大きい数字)

		DR(FPD)				
		0	1	2	3	計
F/S	0	7	2	0	0	9
	1	0	10	1	0	11
	2	0	1	5	0	6
	3	0	0	0	2	2
	計	7	13	6	2	28

一致率： 85.7%

Kappa value: 0.7918      SE 0.0975      95%CI 0.6008 - 0.9828

表3 フィルムと DR 写真の一致度  
 (6人の判定のうち、最多のものをその症例の判定とした場合)  
 (同じ数の場合は大きい数字)  
 (X線分類0型を除いた19例)

		DR(FPD)			
		p	q	r	計
F/S	p	8	2	0	10
	q	0	8	0	8
	r	0	1	0	1
	計	8	11	0	19

一致率： 84.2%

Kappa value: 0.7047      SE 0.148      95%CI 0.4147 - 0.9947

表4 フィルムと DR 写真の一致度 (1 2 階尺度)

6 人の判定を合算したデータ (28 症例 x 6 読影者=168 のべ症例)

All Readers	DR(FPD)											
		0/0	0/1	1/0	1/1	1/2	2/1	2/2	2/3	3/2	3/3	計
F/S	0/0	14	6	1	1							22
	0/1	1	20	11		2						34
	1/0		5	10	5	3	3		1			27
	1/1			6	11	4	1	2				24
	1/2			2	4	4	3	1	1			15
	2/1				1	4	7	3	1		1	17
	2/2					2	4	2	4			12
	2/3							1	1	1	1	4
	3/2							1	2		3	6
	3/3								1	1	5	7
	計	15	31	30	22	19	18	10	11	2	10	168

一致率 : 44.0 %

表5 フィルムと DR 写真の一致度 (4 段階尺度)

6 人の判定を合算したデータ (28 症例 x 6 読影者=168 のべ症例)

All Readers	DR(FPD)					
		0	1	2	3	計
F/S	0	41	15	0	0	56
	1	5	49	12	0	66
	2	0	7	23	3	33
	3	0	0	4	9	13
	計	46	71	39	12	168

一致率 : 72.6 %

Kappa value: 0.6041

SE 0.05

95%CI 0.5061 - 0.7021

表 6 読影者間の一致度 (Kappa value)

読影者—読影者	F/S	DR(FPD)
1-2	0.1704	0.4574
1-3	0.2462	0.4815
1-4	0.5520	0.6064
1-5	0.4380	0.4167
1-6	0.5625	0.6032
2-3	0.3670	0.4450
2-4	0.1807	0.4286
2-5	0.3333	0.4587
2-6	0.3194	0.6336
3-4	0.1855	0.4450
3-5	0.3659	0.3647
3-6	0.3935	0.7982
4-5	0.4776	0.2886
4-6	0.5692	0.5750
5-6	0.7895	0.5035
平均	0.3967 ± 0.1734	0.4486 ± 0.1577



表7 アナログ画像とフラットパネル画像の一致度  
(相澤班データとの比較)

今回のデータ		相澤班データ	
読影者	Kappa value	読影者	Kappa value
1	0.6216	1	0.7881
2	0.6866	2	0.6335
3	0.5556	3	0.4909
4	0.6245	4	0.7863
5	0.4359	5	0.7886
6	0.6384		
平均	0.5938 ± 0.080	平均	0.6975 ± 0.1334

表8 読影者間の一致度(Kappa value) 相澤班

読影者—読影者	F/S	DR(FPD)
1-2	0.4776	0.5729
1-3	0.5328	0.7488
1-4	0.6038	0.6259
1-5	0.6246	0.7009
2-3	0.5133	0.5797
2-4	0.5167	0.7878
2-5	0.7592	0.7854
3-4	0.6438	0.6316
3-5	0.6598	0.7861
4-5	0.7404	0.7488
平均	0.6072 ± 0.0967	0.6968 ± 0.0868

平成19年度厚生労働科学研究費補助金（労働安全衛生総合研究事業）  
分担研究報告書

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)

分担研究者：日下幸則（福井大学医学部環境保健学）  
分担研究者：志田寿夫（労働者福祉機構）  
研究協力者：松本 徹（放射線医学総合研究所・福井大学医学部環境保健学）

**研究要旨：**

大規模な国際的読影実験を行って ILO2000 版 QUAD 標準写真の妥当性・信頼性を確認した Jacobsen 論文、日本の標準写真が ILO の標準とどの程度トレーサビリティがあるかを英国の読影者により検証したわが国で唯一の日下論文および世界に先駆けて X 線デジタル画像 (FCR) をじん肺患者の健康管理へ応用することを検討した中央労働災害防止協会 PXQ 委員会の膨大な研究の一部をレビューすると共に、1980 年版 ILO 標準写真選定のために行われた国際的読影実験データを再解析した。その過程で ILO 標準写真選定実験や PXQ 読影実験においても、じん肺密度分類を 12 階尺度で行う時、“Middling Tendency” というじん肺診断特有の偏りが生じる可能性を指摘し、それを同定・定量する方法を示した。これらの検討結果より、X 線デジタル画像の標準写真の選定やそれによるじん肺分類の精度を評価する方法論等本研究班が今後取り組むべき課題や研究の進め方について考察した。

**研究目的**

本研究は、これまで検討された国内外の代表的な研究のレビューおよび当該研究により得られた実験データの再解析を行うことにより、本研究班が進めるエックス線デジタル撮影画像によるじん肺診断用標準写真の再構築とそれによるじん肺分類の妥当性・信頼性を評価する方法論に資する知見を探ることを目的とする。

**研究方法**

以下の項目について検討する。

1. 文献レビュー
  - 1.1 Jacobsen の研究[15, 16]
  - 1.2 日下の研究[10]
  - 1.3 中央労働災害防止協会「じん肺 X 線診断の計量化：PXQ」委員会の研究 [2, 3, 4, 7, 9, 12, 13]
2. 実験データの再解析

## 2.1 1980年版ILO標準写真の密度分類

## 2.2 じん肺診断の偏り：“Middling Tendency”の分析

### 研究結果

#### 1. 文献レビュー

##### 1.1 Jacobsenの研究

1980年版ILO標準写真(22枚)をベースに、新たな標準写真2000年版を作成し、国際的な読影実験を通して、新標準の妥当性、信頼性を確認した研究である。

新標準は2種類の写真から成る。第一は、フルサイズの1980年版ILO標準写真22枚からp/p, r/r, s/sそれぞれの0/0, 1/1, 2/2, 3/3のじん肺陰影を含む1/4サイズ(quadrant)の部分写真を切り抜き、それをp/p, r/r, s/sごとに1枚の写真として構成した4分割組写真3枚および大陰影[A], [B], [C]を含む組写真1枚、計4枚から成る。もう一種類は10枚のフルサイズILO(1980)標準写真「0/0 2枚、1/1 q/q, 1/1 t/t, 2/2 q/q, 2/2 t/t, 3/3 q/q, 3/3 t/t およびILO胸膜肥厚(quadrant), u/u(quadrant)各1枚」である。両者を合わせた14枚の新標準写真は‘QUAD’と名付けられた。本研究の目的はQUADと現在のじん肺分類尺度である1980年版ILO標準写真を用いてじん肺分類した時とで、両者の結果に偏りが生じないかどうかを確認することであった。

実験内容は以下のごとくであった。読影対象は120枚の胸部X線写真、読影者はカナダ、中国、チェコスロバキヤ、フィンランド、フランス、ドイツ、日本、ポーランド、英国、米国の10か国から4人ずつ合計40人であった。読影実験データの分析は医師でなかった1人を除外した39人の医師の読影実験データが対象とされた。読影者

を19人(IQ)と20人(QI)の2群に分け、前者は1ラウンドの実験でILO標準を参照して120枚の胸部X線写真をじん肺分類し、2ラウンドにQUAD標準を使用してじん肺分類した。後者は1ラウンドにQUAD標準で、2ラウンドにILO標準を用いて120枚の胸部X線写真を読影した。さらに3ラウンドの実験でIQ読影者19人とQI読影者18人は120枚の写真のうちの半分の60枚を、ILO標準を用いて、他の半分の60枚を、QUAD標準を用いてじん肺分類を行った。

読影実験データを分析した結果、QUAD標準写真の使用は読影者間の変動を増加することなく小陰影密度分類の再現性をILO標準使用の場合より改善したが、少数の読影者は大陰影の指摘の頻度をやや減少させた。また、少数の読影者はQUAD標準を使うことで粒状影よりも不整形影が支配的との判断が増加しILO標準使用時より不整形影と認知した写真に対して密度を高めに分類する傾向が観察された。これは本実験で使用された画像データベースの中でcategory 1以上の症例が高濃度に(約60%程度)存在したためであり、じん肺診断の現場では読影者内・間変動と区別できるほどのものでなく、QUAD標準写真を使用してもじん肺分類に重大な偏りを招くことはない結論された。

注：後述する「考察」の章で、本論文で得られたこれらの知見が本研究班の今後の研究に資することは何か検討したので参照されたい。

##### 1.2 日下の研究

我が国ではじん肺法に基づきILO国際標準写真とは別個に日本の標準写真が存在する。但し、小粒状影のp, q, rはILO分類に準じて分類す

るが、不整形影は存在の有無のみが記載される。特徴的なのはじん肺所見の有無の境にある 0/1(p), 1/0(p) の標準写真が用意されていることである。[5, 17] 本研究では、ILO 標準を「原器」とし、それに準ずる日本標準との間にトレーサビリティがどの程度あるか、英国の読影者による読影実験を行って検証した。

non-medically qualified (NM パネル)4 人と medically qualified (MQ パネル)6 人が 1980 年版 ILO 標準写真を参照してじん肺分類を行った。その結果、NM パネル、MQ パネル共、0/1, 1/0 の日本標準の密度を ILO 標準よりも低く分類した。また、NM パネルは明らかに日本の標準より高密度に、MQ パネルは日本の標準より低密度側に分類する傾向があった。しかし、両パネルの密度分類を平均した結果は、0/1, 1/0 の境界写真を除き、ILO 標準と矛盾しなかった、すなわち、日本標準は ILO 標準とトレーサビリティがあると結論された。

注：後述する「考察」の章で、本論文で得られたこれらの知見が本研究班の今後の研究に資することは何か検討したので参照されたい。

### 1.3 中央労働災害防止協会 PXQ 委員会の研究(資料 1)

本研究は、昭和 58 年～63 年にかけて、我が国独自の発明であり、デジタル X 線技術の先駆として登場した FCR がじん肺患者の健康管理に利用できるか、FCR による日本標準写真の作成は可能か、中央労働災害防止協会 PXQ 委員会により行われたものである。

じん肺診断に用いられる写真は通常、大角サイズであるが、当時の FCR は縦、横大角の 1/2 サ

イズ、面積で 1/4 の写真であった。また、FCR で行える画像処理には様々あり、そのどれがじん肺分類に最適か不明であった。そこで本節では、じん肺分類の精度に FCR の画質がいかに関わるか、1/2 サイズにフィルムを縮小したことでどのような影響を受けるか検討した結果をレビューする。詳細は PXQ 委員会報告書および本節末に添付した英文資料を参照されたい。本節で言及する Figure 番号は英文資料中のそれと対応する。英文資料にないものは図…と表示する。

#### 1.3.1 読影実験

##### 1) 方法

1980 年版ガイドラインに基づくじん肺診断の国際的尺度 (ILO 標準写真) を参照してじん肺判定する読影実験 1, 2, 3 (Fig. 1) を行った。実験 1 と 3 で使用された ILO 標準写真は 1980 年版 ILO 標準写真である。実験 2 の標準写真は大角サイズの ILO 標準をドラムスキャナーにかけてデジタル化し、FCR サイズに縮小して再生したもの (ILO 縮小標準) である。ILO 縮小標準には Fig. 2 に示す画像処理 C が施された。一方、被じん肺分類用テスト写真として使われたのはドイツの標準写真から成る 26 枚である (Table 1)。実験 1 では大角サイズのテスト写真が使われ、実験 2 のテスト写真は ILO 縮小標準写真と同様の手順でテスト原寸写真をデジタル化し、縮小し、再生したものの (テスト縮小写真) であり、Fig. 2 に示す画像処理 A, B, C, D を施された 4 セットが用意された。読影者はじん肺読影専門医 8 人である。読影は 1980 年版 ILO 読影基準に従って行われた。

#### 1.3.2 結果および討論

##### 1) じん肺密度読影結果の画質・画像サイズ依

存性

Fig. 3 は、実験を企画した専門医 3 人の合意により決定された 26 枚のテスト写真の密度と、実験 1, 2, 3 により得られた、major category 分類結果の対応関係を示したものである。処理 A のテスト縮小写真は 0 型の写真以外は対角線より下側（低密度側）に読まれ、処理 B は対角線より上側（高密度側）に読まれたが、処理 C, D および実験 1 の通常のじん肺分類（実験 1）の結果はほぼ対角線上にあることが分かる。

## 2) 密度分類の平均（じん肺分類の偏り）とばらつき（医師間変動）の画質・画像サイズ依存性

Fig. 4 は 12 階尺度で判定された結果にスコア 1~12 を割り当て、各例に 8 人の読影者が割り当てたスコアの平均値と分散を 0, 1, 2, 3 型の症例ごとに計算し、平均値を横軸に縦軸に分散をプロットした結果である。平均はじん肺分類の偏りを分散は読影者間変動を表す。図中点線で示された曲線は、実験 1（通常）の場合である。1 型の症例に対する医師 8 人のスコアの平均は 1/0（スコア 4）に近く、そこでの分散は 0, 2, 3 型の症例よりもっとも大きかった。これは通常行われているじん肺分類において 0 型(0/1)と 1 型(1/0)の境界で医師の判断が最大にばらついたことを表す。ILO じん肺分類ガイドラインによれば胸部エックス線写真によるじん肺分類は事実を客観的に決定するものであり、社会的保証の問題とは無関係と記されているが、現実には 0 型 1 型の境界にある粒状影および不整形影の密度分類で医師はもっとも迷っていることが示された。

これに対して Fig. 4 の実線で示された平均と分散の関係は、実験 2, 3 で得られた結果である。

実験 3 は実験 2 より、実験 2, 3 は実験 1 より医師間変動が大きい。特に画像処理 B で顕著である。どちらかといえば実験 2（ILO 縮小標準対テスト縮小）の処理 C の結果が通常（実験 1）の傾向にもっとも近いことが分かる。

## 3) 密度識別能の画質・画像サイズ依存性

Fig. 5 は実験 1, 2 における 0 型対 1 型、0 型対 2 型の密度識別能を表す ROC 曲線である。0 型と 1 型の識別より 0 型と 2 型の識別が容易であり、標準と被分類写真が同じサイズであれば（どちらも原寸大か縮小であれば）同程度のじん肺密度識別能が得られた。

Fig. 6 は実験 1 と 2, 3 における 0 型対 1 型の密度識別能の画質依存性を表す ROC 曲線である。ILO 標準が原寸大で被分類写真が小さい（実験 3）場合、密度識別能は劣化した。

Fig. 7 は q 形の 0/1 対 1/0 の実験 1 と実験 2, 3 の密度識別能を表す ROC 曲線である。標準も被分類写真も共に縮小サイズ（実験 2）であれば通常の前寸大同士の場合（実験 1）と同程度の識別能が得られるが、標準が大で被分類写真が小の場合（実験 3）識別能は劣化した。Fig. 8 は q 形の 1 型と 2 型の識別能を表す。

## 4) 密度分類の的中率

3 人の専門医により 0, 1, 2, 3 型に属すると確定されたそれぞれの写真を同じ型に分類した割合を計算した。Fig. 9 の実験 1 と 2 の結果は、標準も被分類写真も同じサイズであれば処理 C または D でほぼ同じ結果がえられること、Fig. 10 は標準が大で被分類写真が小の場合、通常（実験 1）と同じ結果にならないことが示された。

### 1.3.3 まとめ

①標準写真と被分類写真が同じサイズであれば小さな写真同士でも通常の大角サイズ同士の密度分類とほぼ同じ結果が得られる。

②但し、標準と被分類のデジタル写真には適切な画質変換処理を施す必要がある。本研究の場合、処理 C (Fig. 6 参照) が至適な処理であった。

③今回の実験に参加した医師は縮小写真の読影は初めてであったという事実を考慮すると本分析の結果として示されたじん肺密度分類の医師間変動は比較的小さく、じん肺診断へデジタル縮小写真を実用する可能性はあると考えられた。

### 1.3.4 追加の検討

①PXQ 委員会は上記に引き続き、被分類用じん肺写真として FCR で撮影された 31 枚からなるデジタル画像 31 枚を医師 12 人が読影実験を行い、以下の図 1 に示す処理 F が通常と同等の密度分類結果が得られることを確認した。

②さらに、「わが国では ILO 標準よりも労働省の定める標準フィルムによる読影診断が普遍的であることを考慮するに、労働省の標準フィルムでも比較読影実験を試みる必要」と判断され、日本の標準写真 21 枚をデジタル化し、画像処理 A, B, C, D, F を施し、じん肺専門医 3 人が合意のもとに画質を評価した。その結果、以下の表 1のごとくとなり、一連の実験結果で得られた解析結果と対応する結論を得た。

注：後述する「考察」の章で、本論文で得られたこれらの知見が本研究班の今後の研究に資することは何か検討したので参照されたい。

## 2. 実験データの再解析 (資料 2)

### 2.1 1980 年版 ILO 標準写真のじん肺密度分

類 [8, 19]

現在使用されている 1980 年 ILO 標準写真は 1977 年から約 2 年かけて行った読影実験の結果、選定されたものである。本報はこの時の読影実験データを再解析し、当該標準写真自体に含まれているじん肺分類の偏りや読影者間変動を明らかにする。詳細は本節末に英文資料を添付したので参照されたい。本節で言及する Figure 番号は英文資料中のそれと対応する。

### 2.1.1 材料および方法

ここで再解析された読影データは ILO ワーキンググループにより得られた。読影実験前、欧州、米国、日本から 106 枚の標準候補写真が収集された。次にこれらの写真の 1 セットが各国代表に送付され、各国内の読影者に回覧された。画像読影は 1971 年版 ILO 標準写真を参照して行われた。読影結果は ILO 作成の画像読影シートに記録され、そのシートは読影終了後 ILO に返送された。Fig. 1 に ILO により収集されたある 1 枚の写真に対する読影データの一部を示す。読影者は 27 人である。Reader1-9 は日本、Reader9-17 は米国、Reader18-27 は欧州の読影者である。本報で再解析されたデータは (1) 画質の主観的評価データ、(2) 粒状影、不整形影それぞれの 12 階尺度での分類データである。これらのデータの症例間変動、医師間変動、国間変動を評価した。

### 2.1.2 結果および討論

1) じん肺スコア (密度分類) の平均と標準偏差

読影者から各症例に 12 階尺度で分類されたじん肺陰影密度に 1-12 のスコアを割り当て症例ごとに読影者全員 27 人、日本 8 人、米国 9 人、欧

州 10 人の医師群ごとに各症例のスコアの平均値と標準偏差が計算された。平均値はスコア 1-12 の間のじん肺分類の偏りの程度を表し、標準偏差はその症例に対するスコアの医師間変動の程度を表す。Fig. 2 に 106 例の読影者 27 人によるスコアの平均値を横軸に、標準偏差を縦軸にプロットした結果を示す。粒状影、不整形影ごとに示してある。どちらも同じ傾向が見られる。すなわち、スコアの平均値が非常に小さい：0 型か大きい 3 型の症例の標準偏差は小さいが、1 型、2 型の例では大きい例が多い。読影者にとって 0 型の 0/- および 3 型の 3/+ のじん肺分類は判断が容易に行え、それに比べて 1, 2 型の症例のじん肺分類は判断に迷う例が多いといえる。このようなスコアの平均と標準偏差をもつ症例の中から IL0 は 15 例を標準写真として選択した。p, q, r (粒状影)、s, t (不整形影) の 1/1 (スコア 5), 2/2 (スコア 8), 3/3 (スコア 11) の写真である。Fig. 3 はこれら 15 例のスコアの平均と標準偏差をプロットした結果である。3 型の r の 1 例の標準偏差は 0 であった。すなわち、27 人すべての読影者が 3/3 (スコア 11) と判定したことを意味する。1 型および 3 型の 5 例の標準写真に対する医師間変動は高いものもあり比較的低いものもあった。すなわち、5 例の標準偏差の変動は 2 型の場合よりも変動の範囲が広い。2 型の標準偏差の変動は小さいが標準偏差の平均は 1, 3 型よりやや高い傾向がある。0 型から 3 型まで全体の傾向を主観的な判断に基づき手書きでなぞると Fig. 3(b) のようになる。ここで、0 型と 1 型の間に描かれたピークには 2 種類のデータが混在している。すなわち、粒状影 p, q, r の標準写真の中に含まれる不整形影の密度スコアの平均と分散および不整形影 s, t の標準写真に含まれる粒状影のスコアの平均と標準

偏差のデータである。それらに対するスコアの平均がちょうど 0/1 と 1/0 の境界で標準偏差がピークを形成した。これは 1980 版標準写真の中に含まれる 0/1 と 1/0 の境界にある陰影の密度分類に対して読影者がもっとも判断に迷ったことを表している。Fig. 4 (a) は 15 例の標準写真と同様の傾向を示した症例群に対するスコアの平均と標準偏差のプロットである。これより Fig. 3(b) の全体の傾向を滑らかな曲線でなぞった結果は妥当であったと理解される。Fig. 4(b) は 1, 2, 3 型でスコアの標準偏差 (医師間変動) が大きかった症例群の結果である。この中からは標準写真は選ばれなかった。本実験で読影者が主観的に画質の良さをスコアで回答したデータからその平均と分散をプロットした結果を Fig. 5 に示す。これより次のようなことがいえる。画質の良さを平均と分散の傾向はじん肺分類のような傾向を示さなかった。標準写真より画質の良い写真は多数存在した。これは、標準写真は画質の良さというよりもじん肺の X 線像の特徴を重視して選ばれたことを示す。ガイドラインにもそのことは付記されている [18] が、それがここに裏付けられたと考える。

## 2) じん肺密度の識別能

Fig. 6(a) は 1 型と 2 型の p, q, r の密度識別能を ROC 曲線で評価した結果である。p の密度識別能は q, r よりも難しく、1 型と 2 型の識別より 1 型と 3 型の識別が容易なことを示唆する、当然と思われる結果が確認された。

## 3) じん肺密度分類の国内・間変動

Fig. 7 に、日本の 8 人の読影者の各症例のじん肺分類密度スコアの平均 (縦軸) と欧州の 10 人

の読影者の平均(横軸)との相関を示す。粒状影、不整形影ともやや日本のスコアが対角線より上方にあるように見える。Fig. 8(a)に、ILO 標準写真として選択された 15 例に対する日本対欧州の、Fig. 7 に対応する関係を示す。但し、データに一次回帰直線がフィットされている。日本の医師は欧州の医師より密度を高く分類する傾向が観察された。Fig. 8(b)に示されるごとく日本対米国の場合も(a)と同様であった。一方、米国と欧州の医師の相関関係は Fig. 8(c)に示すごとくほぼ対角線と一致、すなわち、同一のじん肺密度分類が行われたことが確認された。Fig. 9 は 15 例の標準写真の標準スコア 5, 8, 11 に対して各読影者が付与したスコアの関係に一次回帰直線をフィットした結果である。日本の読影者は 0 型と 1 型の辺りで高密度側に判断し医師間変動が大きい傾向が認められ、欧州の医師間変動は 0~3 型まで対角線の上下に比較的大きくばらつき、米国の医師は一人を除くとほぼ対角線上に医師間変動は小さかった。米国の医師がもっとも標準写真の密度通りに医師間変動少なく密度分類したことを示唆する結果が得られた。但し、これは各症例に対する医師間変動が日本で大きく欧州や米国で大きいことを意味しない。Fig. 10 に各症例に対する医師間変動(標準偏差)の国間の違いをプロットした。日本の医師の各症例に対する変動は欧州の医師よりも小さい例が多い。しかし、p または s の 1 型に対する医師間変動が大きい。それが Fig. 9(a)と(b)の違いに反映された可能性がある。

### 2.1.3 まとめ

1980 年版標準写真選定のための読影実験データを再解析した結果以下の事実が確認された。

①日本の読影者は欧州や米国の医師よりも 12 階尺度で 1-1.5 ポイント高めに密度分類する傾向がある。

②ILO 標準写真でも p の密度分類は q, r より難しい。

③0 型と 1 型の境界で密度分類の医師間変動が大きくなる可能性が考えられる。

## 2.2 じん肺診断の偏り：“Middling Tendency”の分析[1, 19]

1974 年 W. K. C. Morgan らによりじん肺診断に固有の偏りと指摘された“Middling Tendency”現象を 1.3 PXQ 委員会および 2.1 ILO の実験データにおいても認めた可能性があるので報告する。

“Middling Tendency”とは図 2 のごとくじん肺陰影の真の分布が一樣なじん肺写真が多数であると仮定した時、これを医師が密度分類する際、0, 1, 2, 3 型の mid-category (0/0, 1/1, 2/2, 3/3) に回答が集中する現象を指す。

### 2.2.1 1980 年版 ILO 読影実験の場合

1979 年 ILO はじん肺診断用標準写真を選定するため欧州、米国、日本などから標準候補写真 106 枚を収集、これを各国の医師 27 人が ILO 1971 年版の標準写真を参照して各写真に存在する粒状影、不整形影の密度を 12 階尺度で分類する読影実験を行った。図 3 に 27 人の医師の読影結果をプールした結果を示す。ここで最下段の図の「Max」とはある医師がある写真に対して粒状影、不整形影の密度分類を行い、どちらか大きい方の密度で、その写真のじん肺密度を代表した時の結果である。いずれにおいても明らかに mid-category に分類が集中した。図 4 は欧州



10人、米国8人、日本9人の医師群後との読影結果をプールした結果である。国が異なっても同じような”Middling Tendency”が観察された可能性はある。

## 2.2.2 中央労働災害防止協会 PXQ 委員会読影実験の場合

### 1) 読影実験の方法とじん肺分類の結果

1983年上記PXQ委員会はFCRのじん肺診断への応用の可能性を探るため資料1のFig.1に示す3種類の読影実験を行った。

実験1において8人の医師が1980年版ILO標準写真(原寸大)を参照して26枚の被分類テスト用写真(原寸大)を12階尺度で密度分類した。その結果図5のような分布(8人の読影結果をプールしたもの)が得られた。実験2は実験1で用いた原寸大のILO標準写真と被分類テスト用写真をFCRサイズ(縦、横共フルサイズの1/2、面積は1/4)に縮小したもので実験1と同様の読影実験を行った。その結果を図6に示す。ここでA, B, C, Dとは画像処理により作成した画質の異なる1セット(各26枚)の縮小テスト写真を指す。なお縮小標準写真(ILO原寸標準写真をドラムスキャナーにかけてデジタル化し、FCRサイズに縮小したもの)にはCの画像処理が施された。実験3は原寸大のILO標準写真を参照して、縮小テスト用写真(A, B, C, D)を分類したもので図7のような結果が得られた。

### 2) Middling Tendencyの検出

次にMiddling Tendencyを以下の方法で検出した。図8は読影実験終了後、PXQ委員会のメンバーで実験主催者(読影実験には不参加)のじん肺読影専門医3人が、読影者8人の医師それぞれ

の症例ごとの読影結果も参照して、合意により、12階尺度分類の仕方に忠実に従って決定した各症例写真の密度分類の頻度分布(図5~7に対応する)である。図8の密度分布にはMiddling Tendency効果の混入はないか、あってもわずかであると仮定する。これより、8人の読影者の密度分類の分布全体が100%になるように正規化し、正規化された図5~7の分布から、正規化された図8の分布を差し引いた結果、図9~11が得られた。Mid-category以外に分類されるべきもの(-%)がmid-category(+)に回答されたことが示されている。すなわち、標準写真のサイズを変化した時、被分類写真の画質を変化させるとMiddling Tendencyが微妙に異なることが観察できる。

## 2.2.3 Middling Tendencyに関する考察

W. K. C. Morganの実験では8800例のじん肺写真を25人の医師が密度分類した。major category(1, 2, 3型)分類ではじん肺症の発症率が重症度(major category)と逆比例する分布を呈した。この関係は疫学的見地から妥当と考えられる。一方、同じ読影者が行ったsub-category(12階尺度)分類では、mid-categoryにじん肺分類が集中する傾向が観察された。本文献の場合、多数の症例がprospectiveに、自然発生的に集められており、selection biasは殆んど無視できるので、このようなmid-categoryへの分類の集中は”Middling Tendency”を表していると考えられる。「数多くのじん肺写真を小陰影の密度の軽いものから重いものまで順に並べれば何処にも区分はなく連続的であるのに、人工的にはっきり4区分したところに問題が生じる」[11]というこ

とである。一方、我々が確認した ILO 実験の結果：図 3, 4 および PXQ 実験の結果：図 5-7 では、読影実験の対象としたじん肺症例の中に、元来 mid-category に分類されるべき症例がかなりの頻度あり、それが反映された (selection bias が存在した) 結果と Middling Tendency の結果が複合していると推測される。PXQ 実験では、読影者 8 人以外のじん肺読影専門医 3 人の合意で 12 回尺度分類を忠実にに行った結果：図 8 を求めたが、この分類結果には、本来 sub-category に分類すべきものを mid-category に分類するという Middling Tendency は無視できる程度という仮定が成り立てば、正規化した図 5-7 の分布から図 8 の分布をそれぞれ差し引いた図 9-11 の結果は selection bias が差し引かれた Middling Tendency の効果分のみを表すと考えられる。図 2, 3 の ILO 読影実験の場合も同様の手順を踏むことで Middling Tendency の程度が評価できると思われる。また、PXQ 実験の結果は、Middling Tendency のデジタル画像の画質依存性を表すが、ILO 実験の結果：図 4 は、日本、欧州、米国の読影者グループが同じ画像を読影した結果を表す。これより簡略な方法として、例えば、図 4 の分布の「日本 - 欧州」、「日本 - 米国」、「欧州 - 米国」を求めると "Middling" の相対的な大小が比較され Middling Tendency の国間変動が定量できる。

ところで、Middling Tendency はじん肺の密度分類にとって良い現象なのだろうか、悪い現象なのだろうか。一度に大量の写真を分類しなければならない時、読影者が疲労した時、標準写真を忠実に参照しないで読影した時、画質が劣化した画像を読影する時、デジタル X 線写真を用いてじん肺分類を行った時、12 階尺度に沿ったじん肺密

度が mid-category へ丸められるという厳密さを欠いた分類が行われた結果、"Middling Tendency" が顕著に増加し、結果的に患者へのじん肺判定のばらつきに帰結すればこれは悪いじん肺分類の偏りといえる。しかし、じん肺患者の健康管理になんの影響も生じなければ気にする必要はないかもしれない。またしかし、少なくとも新しい標準写真を選択するために行われる読影実験や教育および専門家の資格認定を行う際、読影者がそのような厳密さを欠いた密度分類を行うことは見過ごしてはならないのかもしれない。"Middling Tendency" を無制限に許容しては X 線デジタル画像がじん肺患者の健康管理へ寄与することが阻害されるかもしれない。今後の研究によりそれらの真偽を明らかにする必要がある。

## 考察

### 1. 過去の研究成果に学ぶ

#### 1.1 Jacobsen 論文から本研究班へフィードバックされること

本論文に示された研究は、実験の構想、具体的な研究計画の策定、各国の協力者への協力の取り付けおよび実験手順の合意と確認、読影実験用症例画像の収集と選別、読影者のリクルート、実験用画像の各国への送付、各国での読影実験および実験データのとりまとめ、本部への実験データ転送、データ解析、報告書作成、論文作成、投稿、論文発行まで、少なくとも見積もっても 3~5 年を要した一大プロジェクトであった。実験を計画し、実行し、データを纏めて報告書や論文になるまで多大の困難があったと想像される。当初目論んだ通りの結果が得られなければ、最後に論文が受理されなければ、折角の苦労は報われなかつただろ

う。報告書は膨大だが、学会誌に公表された論文はわずかに数ページであるが、一読して内容以外の苦勞をしのばせる、中味の濃い論文である。著者をはじめこのプロジェクトを支えた関係者の熱意と情熱に敬意を表したい。症例画像は120例、読影者は40人、各国から4人ずつ、協力した国は10カ国、読影は読影者内・間変動を評価するため2群に分けられる数(偶数)で計画され、読影者はラウンド1,2,3の読影実験に従事した。じん肺診断の世界標準作りとその根拠を得るためには、これほどのことをしなければならない、ということが示されている。費用と時間と空間の規模および研究計画の用意周到さは1980年ILO標準写真選別実験の場合と同等以上のものがあったと想像される。それでいて結局、得られたアウトカムは何か。「1980年版標準とほぼ同等なので問題ない」である。ここにじん肺診断の特性と本研究班の今後の研究の取り組み方が示唆されている。すなわち、一般的に言って、研究は新しい事実が古いものと「異なるか否か」を検定し、新発見の知見であることを目指す。しかし、じん肺診断の標準作りでは「同じ」であることの証明を目指す。「異なる」を検定する統計的検定手法に比して「同じ」を検定する方法論は整備不十分であり、「同じ」であることを実験的に証明するのは「異なる」ことの証明より多大なエネルギーを要するといわれている。しかし、新しいじん肺標準作りは敢えてその難しい問題に取り組まなければならない。なぜなら、Jacobsenが細田への私信で述べているように、「物差し(1980年版ILO標準の基準)の変更はじん肺症の発生率に人工的な変化を導入し、じん肺統計の時間的推移をモニターすることに影響を与え、じん肺症を減らしダストを抑制する責任あるものにとって有害であ

り、時や国が異なるじん肺統計を比較する時の問題を一層複雑にする」[6]からである。

ここで確認しておくべきことは、本研究班は、Jacobsenが指摘した「物差しは変えない」という原則を守りつつ、1980年版ILO標準写真の画質を現代の画像の画質に合わせることを検討しつつ、被分類写真の画質が標準写真の画質と異なった場合の問題の解決に取り組んでいるということである。1980年当時、標準と被分類写真はほぼ同じ画質であった。それから28年経過した今、X線撮影技術の進歩により、標準は当時のままなのに被分類写真の画質とそれを生み出すX線撮影技術は明らかに異なっている。その変化にいかに対応するかは日本(村田班)のみならず世界的な課題である。本レビューから学ぶべきは、Jacobsen論文にその回答はない、本研究班はJacobsen論文で検討されていない問題を解決しなければならない、その方法論の基本がこのJacobsen論文に示されているということである。

## 1.2 日下論文から本研究班へフィードバックされること

じん肺症患者に対するそれぞれの国の事情に従って、その国独自の標準を整備するのは自然なことと思われる。しかし、ILO標準はじん肺分類の世界の原器である。前節でJacobsenが指摘した「物差しを変えてはいけない」の原則に従う必要がある。そこで、ILO標準と国ごとに設定された標準との間に、その国独自の事情に基づく系統的な差異がどれ位あるか明らかにし、それを差し引けばILO標準とトレーサビリティが保たれていることを確認しておく義務が当該国にあると考える。日下論文はそれを日本のじん肺学者としてはじめて、唯一「確認」した。しかも日本人では

なく(日本のじん肺医師は労働省標準写真を熟知しているため読影者になり得ないので当然であるが)英国の読影者を起用した点がユニークであった。これより以下のような興味深い事実が明らかになった。すなわち、英国の読影者は、ILO 標準にはないが日本の標準に組み入れられている 0 型(0/1)と 1 型(1/0)写真に対して密度分類を 12 階尺度で 1-1.5 程度低く判定した。一方、今回 ILO 標準写真選定のために行われた時の実験では、2.1.2 3) Fig. 8(a)に示したごとく、日本人読影者は欧州の読影者より 0 型と 1 型の辺りで、12 階尺度で 1-1.5 程度高密度側に判断した。日本の標準写真に対して英国の読影者が密度分類した結果と ILO の標準写真に対して日本の読影者が密度分類した結果は期せずして表裏一体の関係で符合したといえる。

しかし、なぜ、そうなったのか。細田によれば日本と欧州(および米国)とで少なくとも 1 ランク(12 スケールの 1 段階)以上の違いがあるのは、ILO と日本の標準作りのそもそもの発端に原因がある。

さらに付け加えると、日下論文で示された結果と ILO 読影実験データの再解析の結果は、0 型、1 型のみならず 2, 3 型の範囲でも、特に medically qualified:MQ パネル 6 人は日本の標準を低密度側に分類した。この結果は日本の医師が ILO 標準を高密度側に分類するという事実と符合する。しかし、英国の non-medically qualified:NM パネル 4 人は逆に、日本の標準の 2, 3 型を明らかに日本の標準より高密度側に判定した。なぜそうなったのか、本論文はその理由を示していない。

以上の考察結果より、本研究班が日下論文から学ぶべきは、今後デジタル化されると思われる日本の標準と ILO 標準のトレーサビリティを、何

らかの方法でもって日本の読影者を対象にして、或いはやはり外国の読影者により再確認する必要があること、0/1, 1/0 の境界写真を日本が標準写真として加えたことの意義を証明して世界に示す必要があること、それらを通じて日下論文で解決されなかったことを、デジタル画像を対象にして本研究班が明らかにすることは今後のじん肺分類の妥当性・信頼性の確保に貢献すると思われる。

### 1.3 PXQ 委員会の研究から本研究班へフィードバックされること

23 年も前に世界に先駆けて X 線デジタル画像をじん肺患者の健康管理に応用できるかどうか検討した本 PXQ 委員会の研究は、今も色あせることのない内容を含み、CR、DR のデジタル画像が日常のものとなり、それをじん肺診断へ応用する時の問題を考えざるを得なくなった現在、本研究班のために種々な示唆を与えてくれる。

第一のポイントは、画像処理によって読影者のじん肺密度分類の結果が左右される事実を証明したことである。PXQ 委員会では実験を始める前に複数の(一人の権威者でない)じん肺専門医の合意によって、どのような画像処理がじん肺診断に至適であるか凡その検討を付けて、その前後の画像処理を選び、それらとじん肺分類の精度を比較した。その結果、至適画像処理は 1.3.1 1) Fig. 6 に示される処理 C であった。そして、アナログの 1980 年版 ILO 標準写真をデジタル化して処理 C を施し、一方、アナログの被分類写真をデジタル化して処理 C を施し、C の標準で C の被分類写真を密度分類した時の結果がもっとも、通常の方法(アナログの ILO 標準を用いてアナログの被分類写真を分類する時)の結果と酷似してい