

段階では基準範囲を求めることは困難である。しかし、同一レベルで評価し得る方法を確立するためには、まず活性化物質の統一化、リン脂質組成の均一化を行うことが先決である。

FIBはPTやAPTTのような複合因子の測定系ではなく単一因子の定量である。そのため、一番標準化が進み、基準範囲設定の可能性が高い検査項目と思われる。しかし、それを困難にしているのは国際標準になっているWHO89/644やWHO98/612が多数供給できないことにある。また、重量法(秤量法)により標準物質を作製するには手間と誤差要因が多く難しい。国内で標準物質を作製するにしても安定した供給は難しいが、メーカー側の協力があれば基準範囲を設定できる可能性がある。なお、FIBは加齢により増加する報告があり考慮する必要がある。

参考文献

- 1) NCCLS: Performance Goals for the Internal Quality Control of Multichannel Hematology Analyzers: Approved Standard (H26-A), 1996
- 2) 渡辺清明, 巽典之, 三輪史朗: 血球計測値の臨床的許容限界—JCCLSからの提唱—. 臨床病理, 42: 764-766, 1994
- 3) 日本医師会: 平成11年度臨床検査精度管理調査報告書, 2000
- 4) 日本臨床衛生検査技師会: 平成11年度臨床検査精度管理調査報告書, 2000
- 5) 川合陽子, 他: 血液検査の精密性・正確性—血球計数—. 臨床病理 47: 343-352, 1999
- 6) JCCLS Document: 外部精度評価(EQA)標準化のためのガイドライン(GP2-P1). 日本臨床検査標準協議会会誌, 15: 137-151, 2000
- 7) 人間ドック成績判定に関するガイドライン作成小委員会: 「人間ドック成績判定に関するガイドライン作成小委員会報告」血球算定編. 日本人間ドック学会誌 15: 140-153, 2000
- 8) 日本臨床衛生検査技師会: 全国の健康な臨床・衛生検査技師を中心とした血液正常値の現状, 1985
- 9) 金井泉, 金井正光: 各種臨床検査の基準範囲一覧. 臨床検査法提要 改訂第31版, 1998
- 10) 日本病院会予防医学委員会: 人間ドックの判定に関するガイドライン作成小委員会報告そのI. 健康医学 12: 564-577, 1997

- 11) 日本病院会予防医学委員会: 人間ドックの判定に関するガイドライン作成小委員会報告そのII. 健康医学 13: 382-403, 1998
- 12) 日本病院会予防医学委員会: 人間ドックの判定に関するガイドライン作成小委員会報告そのIII. 健康医学 14: 323-339, 1999
- 13) 安田和人: 血液検査の正常値、年齢、性別の相違、検査血液学. 臨床病理 臨時増刊 特集 97: 400-405, 1994
- 14) 米山彰子: 末梢血一般検査. 日本臨床 広範囲血液・尿化学検査, 免疫学的検査(2), 505-509, 1999
- 15) 廣田豊, 神谷忠: 付1. 正常値, 血液学(日比野進監修), 丸善, pp1935-1953, 1985
- 16) 三輪史朗(編): 正常値一覧表, 血液検査. 臨床検査技術全書3, 医学書院, 1982
- 17) NCCLS: How to Define and Determine Reference Intervals in the Clinical Laboratory: Approved Guideline-Second Edition (C28-A2), 2000
- 18) NCCLS: One-Stage Prothrombin Time (PT) Test and Activated Partial Thromboplastin Time (APTT) Test: Approved Guideline (H47-A), 1996
- 19) NCCLS: Procedure for the Determination of Fibrinogen in Plasma: Approved Guideline (H30-A), 1994

表1 血球計数検査における方法間および方法内変動 CV(%) (日本医師会)

項目	試料15		試料16		試料17	
	方法間	方法内	方法間	方法内	方法間	方法内
ヘモグロビン	0.95 (%)	0.95 (%)	1.06 (%)	1.01 (%)	18.88 (%)	52.56 (%)
赤血球数	1.48	1.20	1.86	1.27		
ヘマトクリット	5.88	1.76	6.11	1.80		
網赤血球比率						
血小板数	4.00	3.26	4.02	3.81		
白血球数	2.31	2.58	1.74	3.06		
MCV	5.45	1.43	5.71	1.39		
MCH	1.86	1.48	2.35	1.47		
MCHC	5.26	1.88	6.74	1.89		

表2 凝固検査における方法間および方法内変動 CV(%) (日本医師会)

項目	方法間				方法内			
	平成10年度		平成11年度		平成10年度		平成11年度	
	試料 N	試料 A						
PT	6.8	13.3	8.2	22.9			4.3	5.1
INR	9.9		10.0				6.4	
APTT	9.8	12.3	9.1	12.8	3.9	5.1	3.7	5.1
FIB	8.3	10.4	19.3	23.4	6.8	8.1	6.4	8.2

表3 血球計数検査における3年間の CV(%) の推移 (日本臨床衛生検査技師会)

項目	試料	平成9年度	平成10年度	平成11年度
ヘモグロビン濃度	41(正常値域)	1.2	1.2	1.2
ヘモグロビン濃度	42(低値域)	1.4	1.3	1.2
血小板数	41(正常値域)	6.5	7.6	7.0
血小板数	42(低値域)	6.3	8.1	7.4
白血球数	41(正常値域)	4.0	3.6	5.1
白血球数	42(低値域)	3.6	3.4	3.6

表 4 凝固検査における3年間のCV(%)の推移 (日本臨床衛生検査技師会)

項目	計算値	平成9年度	平成10年度	平成11年度
プロトロンビン時間	P/C比	7.6	5.9	5.5
プロトロンビン時間	INR値	12.3	9.7	9.1
複合因子A	P/C比	7.3	5.9	10.9(5.9)
複合因子A	INR値	8.2	12.7	12.1(6.6)
複合因子B	P/C比	9.0	8.9	12.0(8.3)
複合因子B	INR値	9.4	9.0	15.2(11.0)

表 5 血球計数値の臨床的許容限界 (渡辺清明,他:臨床病理 42:764-766,1994)

Test	Values proposed (%)
RBC	4
HGB	3
MCV	4
WBC	5
PLT	7

表6 日本人間ドック学会における血球算定基準範囲と基準域一覧(2000年)

男11,056 女11,141 全22,197	委員会(案) 1996年	全社連	金井	BML	MBC
WBC 男 対数 女	(3500-<9500) (3200-<8600)	3000-8900	5000-8500	3500-9700	3300-9000
WBC全部 非喫煙者	3300-<9000(3200-9500女下限、男上限をとった場合) (男3400-<8600:女3200-8300:全3200-8600女下限、男上限の場合)				
RBC 男 >60 >70 実数 女	400-<520 360-<490	430-559 380-539	410-530(470) (439) (411) 380-480(430)	438-577 376-516	430-570 380-500
Hb 男 実数 女	13.0-<16.7 11.4-<14.7	13.0-17.9 11.5-15.9	14.6-18.1	13.6-18.3 11.2-15.2	13.5-17.5 11.5-15.0
Ht 男 実数 女	37.8-<49.2 33.7-<43.5	39 --- 52 34 --- 45	36 --- 48 34 --- 43	40.4-51.9 34.3-45.2	38.7-52.4 34.6-45.0
PLT 対数 全	14---<36	12 --- 39	16 --- 43	14.0-37.9	14.0-34.0
MCH 男 女	全28-<35			28.2-34.7 26.4-34.3	28 --- 30
MCV 男 女	全78-<102			83 --- 101 -101	85 --- 102
MCHC男 女	全31<36			31.8-36.4 31.3-36.1	30.2-35.1

男11,056 女11,141 全22,197	SRL	巽ら 1994年	厚生省 1999年	新谷 (私信) CRR法1999年
WBC 男 対数 女 WBC全部 非喫煙者	3900-9800 3500-9100	3500-9300 3800-10100		3190-7950(1/X ^{1/4})
RBC 男 >60 >70 実数 女	427-570 376-500	400-540 380-490	383-565(474.3) >60(461.4) >70(443.8) 357-506(431.3)	386-522(正規) 350-464(X ^{1/4})
Hb 男 実数 女	13.5-17.6 11.3-15.2	12.0-16.2 11.4-14.7	12.2-17.4(14.8) 10.4-15.2(12.8)	12.7-16.2(正規) 11.1-14.4(X ²)
Ht 男 実数 女	39.8-51.8 33.4-44.9	36.0-48.6 34.2-44.1		37.8-48.1(正規) 32.9-42.5(正規)
PLT 対数 全	13.1-36.2	13.1-36.2 13.0-36.9		14.4-30.6(X ^{1/4})
MCH 男 女	29.0-34.6 26.3-34.3	? 36 --- 64		29.3-34.7(正規)
MCV 男 女	82.7-101.6 79.0-100.0	80 --- 100		86.7-102.0(X)
MCHC男 女	31.1-36.2 30.7-36.6	31 --- 37		33.0-34.7(正規)

規定除外処理後の2SD(白血球のみ対数変換)

喫煙者では白血球数9,100以上ではCRPを参考とすること

文献上の数値は喫煙考慮なし

新谷氏私信では赤血球系はMCH 28pgで除外後のものをCRR法で解析したもの

Table 7 Suggested Action Limits (NCCLS document H26-A, 1996)

Analyte (Units)	Action Limits	Clinical Relevance: Abnormal Results May Reflect the Following Conditions
WBC (10 ⁹ /L)	< 3	Sepsis, chemotherapy, radiotherapy, agranulocytosis marrow hypoplasia, cobalamin, folate, iron deficiency.
	> 12	Acute stress (including surgery), infection, malignancy, lymphoma, leukemia.
RBC (10 ¹² /L)	> 6.2 (M) > 5.2 (F)	Dehydration, polycythemia, shock, chronic hypoxia.
Hb (g/L)	> 180 (M) > 160 (F)	
Hct (L/L)	> 0.54 (M) > 0.48 (F)	
RBC (10 ¹² /L)	> 4.4 (M) > 3.9 (F)	Anemia from blood loss, cobalamin, folate, iron deficiency, malignancy, chronic inflammation, chronic liver disease, renal disease, marrow hypoplasia, chemotherapy, radiotherapy, hemolysis, hemoglobinopathy, thalassemia.
Hb (g/L)	> 120 (M) > 110 (F)	
Hct (L/L)	> 0.39 (M) > 0.30 (F)	
MCV (fL)	< 80	Microcytosis from iron deficiency, chronic blood loss, chronic inflammation, hemoglobinopathy, thalassemia, sideroblastic anemia.
	> 100	Macrocytosis from chronic liver disease, cobalamin or folate deficiency, sprue, smoking, hemolysis.
Plt (10 ⁹ /L)	< 50	Risk of bleeding. Idiopathic, chemotherapy, radiotherapy.
	> 800	Risk of thrombosis. Polycythemia, post splenectomy, thrombocythemia.

M: men. F: women.

生理機能検査有用性評価に関する解析

分担研究者 神辺 眞之

はじめに

臨床生理検査は、脳、神経機能、心、循環機能、肺機能などの検査から構成されている。脳波 (EEG) 検査、筋電図 (EMG) 検査、心電図 (ECG) 検査、肺機能検査が代表的な検査である。

これらの代表的な臨床生理検査の標準化の現状、可能性について、調査したので報告する。

1. 肺機能検査

1) 種類

肺機能は、呼吸性 (O₂、CO₂ ガス交換機能) と非呼吸性 (代謝機能) に分類される。臨床生理検査としての肺機能検査は、呼吸性肺機能検査である。

呼吸性肺機能は、換気機能、ガス交換機能、血液循環機能、呼吸調節の司令塔の呼吸中枢機能から成る。

それらの機能を測定する代表的な検査法は、換気機能検査として、スパイログラフィー検査、その変形であるフローボリュームテスト、He、N₂ ガス稀釈法による肺気量分画検査がある。

更に、ガス交換機能検査としては、CO 拡散能力検査、血液循環機能検査の代表的検査は動脈血ガス分析検査である。

特殊、精密検査として、ボディプレチスモグラフィ法による肺コンプライアンス検査や気道 (呼吸) 抵抗検査や心カテーテル検査などがある。

更に、呼吸中枢機能検査法として、機械的呼吸調節検査としての吸息時の口腔内圧を測定する P0.1 検査や化学的呼吸調節検査としての CO₂ 反射テストもある。

2) 検査装置間の検査精度

呼吸性肺機能検査は、自動化が進んで、多くの自動肺機能検査装置が普及している。

自動肺機能検査装置は、測定系と接続されたコンピュータによるデータ処理系から構成されている。

測定系は、換気量 (volume)、流量 (flow)、圧力 (pressure) を測定するセンサー (トランスデューサ) の測定精度によって、自動肺機能検査装置間の検査

精度は左右される。

換気量のためのセンサーは、最近、ベネヂクトロス型スパイロメータのようなボリュームメータからフローメータによる流量の積算方式に変わってきた。フローメータの種類には、リリー型、フライシュ型、熱線型、羽根車型があるが、コンピュータによる補正が出来るようになって、フローメータの種類の違いによる測定精度差は改善されている。

ボリュームメータ、フローメータの機械的測定精度を医学会レベルで先ず取り上げたのは、日本 ME 学会であった。

すなわち、流量や換気量を実測する前に、規格化されたキャリブレーションにより、流量などのキャリブレーションを義務づけ、ボリュームメータ、フローメータの測定精度の向上に貢献した。

肺機能検査装置には、CO ガスメータや He ガスメータ、N₂ ガスメータやそれに関連した測定回路があり、これらのガスメータや測定回路の死腔量が測定精度に関係する。

更に、肺機能検査が他の生理検査と違う特徴は呼吸を被検者が意思によって調節できるところから、検査装置だけでなく、被検者、検査者が総合的な測定精度に関係する。

3) 著者らの検討成績

著者らは、本邦における代表的なメーカー2社の肺機能検査装置、2機種種の測定精度について比較検討した(表1、2)。

表1に示すように、上述のキャリブレータ(校正シリンジ)を用いて、同時再現性を比較したが、10回の測定の平均値を比較すると、2機種間差は許容範囲であったが、変動係数(CV)で比較すると1リットル値が、2、3リットル値に比べ、CVが大で、バラついた。又、1リットル値で2機種間差が認められた。

又、表2で示すように、ガスメータを使用しない検査項目(VC、FEV1.0、PFRmax)に比べガスメータを使用する検査項目(FRC、DLCO、RV、TLC)の2機種間の相関係数は低値を示した。この原因は、上述したように、校正シリンジのようにガスメータの標準キャリブレーション方法がないことや、測定回路の問題(ガスメータの位置、ベル内のガスの攪拌、ガスメータのレスポンス等)があげられる。更に、コンピュータにより、同時測定時の肺活量(VC)を使わずに、別の測定で得た最高肺活量(VC)を使用し、残気量(RV)やRVを含む機能的残気量(FRC)、全肺気量(TLC)が演算されるといまちまちのデータ処理の問題もある。

4) 標準式(基準値)について

化学検査項目は、小児、成人、高齢者に分け、基準値が設定される傾向にあるが、肺機能検査項目は、成人(20才以上)について、男女別に年齢および身長、体重、体表面積などの身体因子による基準値に相当する標準式が古くから肺機能障害の判定のモノサシに使われてきた。

例えば、肺活量に関する標準式は、アメリカのBaldwinがアメリカ軍人から測定した

$$VC_{pred.} = (a - b \times \text{年齢}) \times \text{身長}$$

のように、年齢、身長の間数で表されている。

このように、肺機能検査項目に関する標準式

(基準値)は、1960年代は欧米のが用いられたが、1970年代から、日本人による日本人の標準式が作られるようになった。

日本呼吸器学会(旧、日本胸部疾患学会)も1993年2月に基準値を学会として、発表したが、高齢者数が少ない点や、スパイログラフィの表示を身長で割った値としてだしているため、

世界的な標準式(基準値)の計算方式と違っていたので、本邦のスパイロメータ各機器に採用されておらず、欧米人の基準値や各大学病院が決めた基準値が採用されている。

そこで、日本呼吸器学会は、本邦のスパイロメータ各機器に学会の決めた標準式(基準値)を採用してもらうために、2000年にスパイログラフィと動脈血液ガス分圧の基準値の見直しをした(表3)。

2. 心電図(ECG)検査

1) 種類

心、循環機能検査は、心電図検査の他、心機図検査、心音図(PCG)検査などがある。そのうち、一番利用されている心電図検査について説明する。

2) 心電図計の精度

心電図計は、一番普及している電圧心電図計の他にホルター心電図計、テレメータ心電図計、胎児心電図計、心音心電図計、ベクテル心電図計、超音波心電図計などがある。

特に電圧心電図計は、直流式、交流式、交直式があり、印刷方式により、直記式、撮影式(写真式)に分類される。

更に、マイクロコンピュータを内蔵した、診断機能付き心電図計が検診などでよく利用されている。

市販されている電圧心電図計や診断機能付き心電図計は、いずれもJIS(日本工業規格T1202-1984)、IEC(国際電気標準会議)によって、安全性、安定性、記録感度、振幅特性、総合周波数特性、チャンネル間干渉、内部雑音、弁別比、校正電圧、記録速度について詳しく規定されているので、精度は問題ない。

診断機能付き心電図計の診断ロジックは、ミネソタコードを基準にしているが、不整脈診断の性能について差があるようである。例えば、散発する上室性あるいは心室性期外収縮や心房細動に比べると、多発性心室性期外収縮や第2度、第3度房室ブロックなど、臨床であまり経験される機会のない不整脈の診断性能は劣るという報告もある。

更に、検査操作について、胸部誘導の電極の位置で多少心電図波形に違いがでてくる。標準検査手技は、金井泉著の検査法提要にも示されているが、胸部誘導の場合、肋間の取り違いによる場合が多い。

3) 基準値について

上述の検査法提要によれば、成人、小児の基準値が示されているが、日本循環病学会などの公式な基準値は報告されていない。診断機能付き心電図計は、ミネソタコードの診断基準値が心電図の各波形に使われているので、工夫によって、診断に関する標準化は可能になると推察した。

3. 脳波 (EEG) 検査

1) 歴史

脳波は、大脳機能を評価できる。最近、脳死判定に脳波検査が必須となっており、ますます脳波検査の検査精度の正確さが重要となってきた。ヒト脳の電気活動を最初に報告したのは、ドイツの Hans Berger (1929年) で、1924年に、ヒトで α 波を発見し、1931年にはてんかん例で、棘徐波複合の記録に成功したと言われている。

最初に脳波検査室が設立されたのは、アメリカのマサチューセッツ総合病院で、1937年と言われている。

2) 脳波計の精度

市販されている脳波計は、JIS規格 (T1203-1985) で規定されているから問題はない。検査についての操作方法についても、誘導法は、国際脳波学会連合標準電極配置法によっている。

脳波の誘導法は、大別すると、基準電極誘導 (referential derivation of monopolar derivation) と双極誘導 (bipolar derivation) になる。誘導部位の選び方、組合せをモンタージュと称するが、モンタージュも日本脳波筋電図学会で、標準化されている。

3) 基準値について

脳波は、周波数により、 δ 波 (4Hz 以下)、 θ 波 (4-8Hz)、 α 波 (8-13Hz)、 β 波 (13Hz 以上) に分類される。いわゆる正常脳波は、各誘導で、異常脳波を除いた脳波と定義されている。

脳波の基準値は、成人、高齢者、小児で違い、特に、小児では、新生児、学童など細かく分類されている。

脳波の成熟期は、25-45才と言われている。

しかし、専門学会レベルでの基準値の公表はない。

おわりに

臨床生理検査の肺機能検査、心電図検査、脳波検査は、特に臨床に実際に役立っている。呼吸器疾患の診断、治療、心筋梗塞の迅速診断、いろいろなてんかん、脳腫瘍の診断などである。

検査側からみると、検査装置、検査者の技術、被検者の問題があるが、先ず、検査手技の標準化が必要である。

参考文献

日本呼吸器学会肺生理専門委員会変;スパイログラフィと動脈血液ガス分圧基準値

表1 肺機能検査装置、2機種と比較
較正シリンジを用いた同時再現性

	A			B		
	1 L	2 L	3 L	1 L	2 L	3 L
mean	0.988	1.980	2.986	0.984	1.933	2.977
cv	0.427	0.244	0.282	1.285	0.339	0.161

表2 肺機能検査項目の2機種と比較(相関係数)

n=21			
VC	0.99	FEV1.0	0.97
PFRmax	0.93		
FRC	0.96	RV	0.74
TLC	0.98	DLCO	0.89

表3. 肺機能測定値と年齢・身長との回帰式より算出する基準値
スパイログラム

項目	男性(n=584)					女性(n=1227)				
	a	b	定数	R	RSD	a	b	定数	R	RSD
VC L	0.045	-0.023	-2.258	0.786	0.560	0.032	-0.018	-1.178	0.829	0.370
FVC L	0.042	-0.024	-1.785	0.778	0.573	0.031	-0.019	-1.105	0.827	0.378
FEV ₁ L	0.036	-0.028	-1.178	0.842	0.470	0.022	-0.022	-0.005	0.873	0.315
FEV ₁ %(T) %	0.000	-0.215	93.216	0.555	6.246	-0.063	-0.283	106.223	0.583	7.292
FEV ₁ %(G) %	0.028	-0.190	89.313	0.527	6.147	-0.090	-0.249	111.052	0.514	7.441
V ₅₀ L/sec	0.043	-0.046	-0.385	0.701	1.146	0.014	-0.038	3.150	0.668	0.923
V ₂₅ L/sec	0.021	-0.031	-0.073	0.710	0.710	0.003	-0.025	2.155	0.736	0.474

肺機能測定値と年齢・身長との回帰式
血液ガス分圧

項目	男性(n=164)					女性(n=769)				
	a	b	定数	R	RSD	a	b	定数	R	RSD
PaO ₂ TORR	0.084	0.014	70.489	0.070	7.461	-0.139	-0.152	114.690	0.206	9.276
Paco ₂ TORR	0.012	-0.032	41.344	0.157	3.751	0.029	0.014	35.382	0.049	3.879
HCO ₃ ⁻ mEq/L	0.015	0.000	22.340	0.062	1.890	-0.001	0.017	23.865	0.119	2.148
AaDo ₂ TORR	-0.099	0.024	29.695	0.138	7.518	0.104	0.135	-7.324	0.191	9.090

R : 重相関係数
RSD: 残査標準偏差

形態検査の有用性評価の標準化に関する研究

(組織・細胞学的形態診断・検査の有用性評価の標準化に関する研究)

分担研究者 玉井誠一
防衛医科大学校病院検査部 教授

第二次世界大戦後、特にこの30年にみられた医学・医療の発展は目覚ましいものがある。診断技術に関しては、CT、MRI、超音波などの生体内画像に関する技術、免疫学的、あるいは、分子生物学的技術を応用した生体内分子、原子の質的・量的測定法技術などの進歩は、医療の質的高度化・量的拡大をもたらしてきた。また、これに対応して治療技術の進歩、多様化も目覚ましい。こうした高度で高価な医療技術が、実際の国民の健康増進、疾患の治癒にどれほど寄与しているのかを評価する必要性が、経済成長の鈍化、これに伴う政策的経費にあてる税金などの資源の逼迫化などによって、日本でも高まってきている。

報道される平均寿命の大幅な延長、高齢者の増加などや、各種の診断機器、治療機器、試薬、薬剤などで漠然と実感できるわが国の医療技術の進歩ではなく、個々の病院や診療所の医療技術の評価することは今まで日本では一般的なものではなかった。たとえば、病院の手術成績などの開示が遅れてきており、ひとたび開示された場合には成績のあまりの幅の大きさに驚愕するという状態である(1)。

例外ともいべきもののひとつに、日本医師会などが行ってきた全国の医療機関、衛生検査所を対象とした外部精度管理アセスメントがあり、臨床検査精度という極めて限られた範囲ではあったが、臨床検査という医療技術の質を評価してきた。従来日本で実施されてきた外部精度管理アセスメントは、数値として表現されるものが多く選ばれ、統計的手法を用いデータ解析を行い、各施設の精度を評価するという方法がとられてきた。

一方で、各種の疾患の診断技術の感度、特異度の評価のgolden standardとなり、また、各種疾患の治療技術の評価の基盤ともなっている形態検査、特に、病理組織学的診断、細胞診の評価に関しては、医療技術評価の先進国である欧米においても遅れており、誤診などの裁判事例への対応が主たる関心であった。この状態の変革を迫った

のが米国のCLIA88であることはよく知られている。細胞診領域だけではあったが、形態検査の外部精度管理アセスメントが経済的な措置を伴って実施されるようになり、これが、細胞診の診断基準や異常細胞の分類の標準化を促し、これに基づいた各種診断技術や治療技術評価の標準化に結びついてきた。

上に例としてあげた日本での手術成績の病院施設間差に関しても、その原因を解析する場合先ず考慮しなければならないのは、対象とした疾患が同じものであるか、特に、病理組織学的診断が標準化された上で調査が実施されているのか、という問題があげられよう(これは、国際的な治療成績比較においても問題となるどころであり、あとで検討する重要問題でもある)。日本でも、種々の観点から、実際の医療にとって重要な役割を果たしている病理組織学的診断や、細胞診などの形態検査領域での精度管理に対する関心が現在高まってきており(2)、この医療技術の有用性を客観的に評価する必要性が生じてきているように思われる。医療のグローバル化に対応するためばかりでなく、限られた医療資源のこの方面への適切な配分を求める実際的な必要性和相まって、これから議論をすすめ、早急に結論を得なければならない問題と考えられる。

このような視点を踏まえ、ここでは、日本における病理組織学的診断、細胞診の業務の現状に関する考察を行い、日本の外部精度管理の方法などについて聞き取り調査を行って問題点を整理し、また、形態検査の有用性評価の標準化に関しては文献的に調査した。さらに、肺癌に関しては、防衛医大病院の症例を用いて生命的予後の観点から従来提唱されている分類の特性を検討した結果などについて報告する。

日本の病院での形態検査、特に、病理組織診断、細胞診の現状

病院の中央検査室などで日常の業務として実施されている形態検査には、血液形態、細胞診、病理診断など種々のものが含まれる。これらの形態検査は、臨床検査技師、臨床検査専門医、細胞診の指導医、あるいは、病理医など、形態学的判断に関する特殊な訓練を受けたものによって判断され結果が報告されている。

現在では、形態検査業務に関わる者に関しては、

その能力を試験によって審査し、合格したのものには、日本病理学会では認定病理医、日本臨床細胞学会では、細胞診指導医、細胞検査士などの認定を行うこととなっている。こうした学会の認定制度が、広く病院長や、衛生検査所の管理者、あるいは、形態検査を依頼する医師の知るところとなり、認定を受けていないものが、実際の形態検査、病理組織診断、細胞診に関与することは困難となっている。また、病理診断は医行為であるとの厚生省の認識が示されたことにより、形態検査を実施してゆくものもつべき最低限の能力に関して保証されるようになった。

資格認定後は、所定の診断数、検査数、学会出席数など更新資格の要件を満たしながら、業務にあたることになる。各病理医、細胞診指導医、細胞検査士などは、業務を実施してゆくにあたっては、各人の限られた経験の中で、正常組織の構築に関する知識、異常の場合に組織が呈する反応に関する知識、新生物の組織・細胞学的特徴に関する知識などをもとに、所与の標本を観察し、判断を下すものが、その時点でもつ知識の体系の中で適切と考えている分類にしたがって同定しているのが現状である。

そして、細胞検査士の細胞診結果報告に関しては、その報告、特に悪性の疑いがある症例に関しては指導医が再確認し、正確度、精度を維持しようと努めている。

しかし、病理組織診断に関しては、現状では多くの問題を構造的に抱えている状態であるといえよう。すなわち、認定病理医の不足、各病院の経済的制約などから病院にひとりしか病理医がおらず、ダブル・チェックなどが実施されておらず、精度管理上の問題を内包している（欧米でも *single pathologist* という語があり、精度管理上では別に取り扱わなければならないとされている）。

欧米でも従来から指摘されているところであるが、ひとり病理医の多くは日常の診断業務に精一杯であり、新たな知識を学習し、これを日常の診断業務反映する余裕が物理的にも、心理的にもなく、認定医取得時に試験勉強として獲得した知識に大きく依存している。また、身体的疲労から生じる単純ミスを防ぐことも困難な状態にあるといえよう。特に、最近の医療経済の悪化は、各病院や衛生検査所の形態検査数の増加による収入増を

強制しており、この状況は更に悪化の傾向にあると思われる。

専門職の日常業務に関する精度管理、精度保証の方法としては、外部精度管理よりも、むしろ、同僚による調査、再検討 (*peer review*) が重要とされている。後で、こうした問題点を解決する短期的に方法について触れるが、この問題の解決には、専門医の養成などを含む日本の医療全体の構造的な問題として把握されることが必要であろう。

病理組織学的診断、細胞診など形態検査の有用性評価の概念的基盤

これまで、日本の形態検査の現状における医療の構造的な問題点をみてきたが、以下には、形態検査の有用性評価のための理論的問題を考察してゆく。

正常組織・細胞について

そもそも、形態検査は、所与の標本が示す像が正常か否かをまず判別し、病的なものに関しては、病理組織・細胞学的知識を用いてその病的形態を記載するとともに、診断する者がなんらかの根拠により選択した分類体系の中で、同定し分類名をつけることを基本とする検査である。

皮膚の生検や、内視鏡検査に伴う生検などについては、生検実施者が異常と考える領域を採取し、これを標本として観察する機会が多く、すなわち、標本採取の段階で形態的異常を呈したものを診断することとなるので、正常か否かが問題となることは稀である。

一方、内科的疾患における肝生検や腎生検、あるいは、骨髄生検などでは、無作為的に採取した検体であることを前提として観察が行われるため、標本が示す像が、正常か否かが問題となる。

一方、正常の組織、細胞を対象とする学問は、解剖学、組織学である。したがって、正常か否かは、組織学の知識によると考えられるが、病理組織・細胞学的「診断」の観点から解剖学者、組織学者によってまとめられた情報は乏しく、病理学者が診断上の必要からまとめているほどである(3)。

正常値、正常範囲という言葉の不明確さを避けるために国際臨床化学連合、NCCLS(米国臨床検査標準協議会)などで採用された基準範囲 (*reference intervals*)、基準値 (*reference*

value) の基本的な考え方は、形態学的検査においても、類比的に利用できる重要な考え方であると思われる。基準組織という名称は、形態的検査ではなじまないとも考えられるが、疾患がなく、年齢や、人種などの demographic なデータ上均質な集団から採取された組織が示すヴァリエーションについてのデータが「正常組織・細胞」と判断する根拠として望まれよう。

各種病理組織学的分類の有用性評価の方法

標本が正常な組織・細胞ではなく、病的と考えられれば、それを病理組織・細胞学的に解析、解釈し、これを記載、診断名をつけることになる。この診断名は、各臓器ごとに医学・医療の進歩に対応した疾患の分類が行われ、これに応じた病理組織・細胞学的な分類が作成されており、これを利用することになる場合が多い。以下には、炎症性・変性性疾患と、腫瘍性(新生物)疾患に大きく分けて、分類に内在する問題を考え、整理してゆくことにする。

1) 炎症性・変性性疾患の病理組織・細胞学的分類の有用性評価と、評価法の標準化

特殊な病院・施設を除き、炎症性・変性性疾患の分類で、その有用性が評価できるほど多数の症例があるのは、消化管、肝、腎、肺、皮膚などがあげられるが、ここでは、前述したように、内科的な疾患で、無作為の標本抽出を想定し、その病的変化を評価する必要のある、肝の病理組織学的分類について検討してみよう。

肝の炎症性疾患生検病理組織分類

肝では、慢性肝炎の病理組織学的分類が、各種の検査技術の進歩に対応して改定されてきており、また、インターフェロンなどの治療法の有用性・有効性を評価する場合の指標として重要視されてきている。ここでは、肝生検組織標本がどのように臨床的に利用されているかについて、その現状をみ、臨床的な有用性に関して検討し、その後、病理組織学的分類法の標準化の問題点をみてゆくことにする。

臨床的な肝生検の適応は、1) B型、C型慢性ウイルス性肝炎のグレード、病期の判定、2) アルコール性肝疾患と、非アルコール性 steatohepatitis、あるいは、自己免疫性肝炎の診断、グレード、病期判定、3) 胆汁鬱滞性肝疾患、

原発性胆汁性肝硬変、原発性硬化性胆管炎の評価、4) 肝機能検査異常の評価などがある(4)。この他、患者のインフォームド・コンセントを得た上で行われる肝疾患に対する各種治療法の有用性・有効性評価のためにも肝生検が実施されている。

肝生検において、検体の固定までは、生検施行者が、これ以後に関しては病理標本作成技術者、多くは、病院の病理検査室に所属する臨床検査技師が行ってゆくものである。

生検実施者は、検体採取の部位、採取量、固定液の適切な選択などに関する配慮が要求される。肝生検に関しては、安全性の側面から年間 50 例以上(毎週 1 例程度)の生検実施実績をもつものによる実施が必要との基準を Foehlich があげているが、標本の適切な処理の面からも妥当な数値ではなかろうか(5)。

ウイルス性肝炎の診断に関しては、ウイルス感染に関する血清学的検査、ウイルス遺伝子「診断」法の開発、改良、普及、AST、ALT などのいわゆる肝機能検査法の精度の向上、普及によって大きな進歩をみせている。そして、病因に関してはこうしたウイルス感染症の各種診断法に委ね、肝生検の病理組織学的分類は、肝炎活動度の指標の数値化(グレード)と、比較的可逆性に乏しい線維化の程度の数値化(ステージ)に関して評価する方向へと移行してきている。この経過に関しては、Brunt の論文に詳しい(6)。

現在、慢性ウイルス性肝炎に関しての病理組織学的な活動度指標システムは、Knodell らが 1981 年に提唱した指標(hepatitis activity index (HAI))(7)を基礎にしているが、現在は、HAI との対応を意識して作成された Desmet らのコンセンサス版が広く利用されている(8)。

そして、肝疾患にみられる壊死、炎症反応、線維化の程度を数値化し、統計学的な分析法を用いて、各種の変化の質、程度が予後や、治療に対する反応を評価するとともに、治療法選択の場合の指標となるような努力が続けられている。

たとえば、AST や ALT などのいわゆる肝機能検査の諸項目から、肝生検で得られる情報を推定しうるか、という課題が今なお重要なテーマとして研究されている。F ter Borg らは、B 型慢性肝炎について検討を加え、AST がもっとも肝病理組織像と対応すると報告している(9)。この場合に参照する病理組織学的な肝炎の活動度(scored

hepatitis activity) は、Knodell らのスコア化のシステム (7) を一部改変、すなわち、線維化のスコアを除いた HAI スコアを用いており、また、慢性活動性肝炎の有無の基準として、Desmet らの 3 以下、以上を用いている (8)。更に、このデータを基礎として、血清 AST 値の正常上限値との比が 0.5 以下の場合には生検をせずに観察するなど、と臨床的な指針も提唱している。

一方、この論文に対し、model error rates が 40% 近くもあり、受け入れ難く、これは、F ter Borg らが採用している logistic-regression analysis のためであるとして、statistical data analysis network model を用いるよう促し、また、彼らの AST 値を用いた肝生検の適応基準に関して疑問を投げかけている (10)。

形態学的指標などの臨床的有用性を論じる場合に用いる統計学的解析法の選択の重要性を指摘したものと見えよう。

また、ウイルス性肝炎の治療法の有効性・有用性を評価するにあたって、肝の病理組織学的な反応を、こうした半定量的指標を用いて評価し、統計的に検定することが広く行われている (たとえば、peginterferon alfa-2a に関する最新の論文を参照 (11))。

各種の臨床検査法や、肝炎の治療法の有用性・有効性の評価にあたり、肝生検組織の病理組織学的分類が重要であることは、以上簡単にみてきたところからも明らかであるが、この分類の問題点も経験の中から明らかになってきた。

まず、問題となるのが、分類の再現性の問題である。すなわち、Knodell らをはじめとし、肝生検の肝炎の活動度を評価するにあたって、可能な限り客観的で、再現性のよい方法を採用しなければならないという問題意識のもとに、システムを作成してきているのであるが、同じ標本を観察しても、同一観察者でも同じグレード、ステージにならず、また、観察者が異なれば、スコアが異なってしまうという再現性が不良であることが従来から指摘されてきた (12, 13, 14)。

特に、治療法の優劣を評価するにあたっては、再現性が確保されていなければならないことはいままでもないことであろう。再現性を確保した上で、評価法の標準化が図られることが望ましいが、上にあげた臨床検査法の評価や、治療法の評価などの臨床的研究において、ほとんどが、再現性に

問題があるとされている HAI システムを利用している状況であるところに、問題解決の困難さが伺われよう。

これらの肝組織内での病変を評価するにあたって用いられる染色法などの組織標本の作製に関する標準化に関しては、必ずしも明文化されていないのが現状である。Interobserver variations の問題解決には、この点に関しても配慮が必要と思われる。

また、より根本的な問題として、肝炎の形態学的評価の臨床的有用性を何に基づいて評価すべきなのかについても、コンセンサスが得られているとはいえない状況であり、今後更に検討する必要があるであろう。

以上をまとめると、肝生検の病理組織学的診断の有用性を評価するには、各病理組織学的分類、グレード、ステージなどの臨床的意味が明らかになっている必要があるが、これを評価するのに用いる統計的解析法についても標準化が求められていること、また、標本を評価するにあたって、再現性を確保するために、標本の作製法を含めた明確な評価システムが確立している必要があるとともに、確立したシステムの標準化を推進していく必要があることが、明らかになったといえよう。

2) 腫瘍性疾患の病理組織・細胞学的分類

腫瘍の病理組織学的分類は、現在細胞診、血液細胞学的分類の基礎を構成するものとなっており、どの分類を用いるかが重要な問題となる。歴史的には、腫瘍形態学が胎生期の臓器発生過程を類比的に援用することがなされ、その発生過程における分化が基準となって、分類がなされ、また、分化度が想定されてきた。そして、形態的検査の依存する分類自身の妥当性は、その道の「権威」によって保証されてきた。

実際の診療の現場では、形態検査報告書の結果、あるいは、記載に依存して、治療の選択が行われることとなるが、症例の臨床的所見に基づく臨床医の診断と、病理・細胞診の診断に乖離がある場合には、臨床医と診断者との間で、個々の症例に関して検討し、分類が再調整されたり、国内外の「権威」がコンサルタントとして指名され、標本が送付されて、この「権威」の意見に従う、といった方法が採られるのが一般的であろう。

一方、腫瘍細胞の生物学的行動の観点から、こ

れを評価する方法として、周囲組織への浸潤の形態、リンパ管や血管などへの浸襲像、リンパ節転移などが評価される方法も採用され、また、最近では、従来から提唱されてきた形態的予後因子、あるいは、種々の理論、実験、観察に基づいて新たに想定された形態学的予後因子とされるものを、因子分析などの推計学的方法を用いて検討し、予後と有意にかつ独立して関連した因子を抽出する方法が採用するというも行われ、より客観的な妥当性に基づいて分類を作成し、これに従おうとする傾向も明確になってきた。

しかし、これらの「客観的」予後因子についても、その評価、形態学的判断においては、「権威」のもつ重要性が無視できないことは広く認められている。ところが、現在、この「権威」の間での国際的、国内的な診断のばらつきが問題とされ、正確度 (accuracy)、精度の両面が問題となってきた。以下には胃癌を例に、この事情をみてみることにする。

1) 国際的な「権威」の間での「腸型」早期胃癌診断基準の乖離

1997年のLancet誌に、消化管を専門とする欧米と日本の病理学者8名が、いわゆる「腸型」の異形成、早期癌35症例の同一の粘膜切除 (EMR) 病理組織標本を観察し、浸潤、構造異型、細胞異型、核異型などの判定基準をあげながら、診断し、その結果を評価した論文が発表された (15)。

この論文で、臨床的にも、保険などの社会的観点からも無視できない問題として明らかになったことは、同じ胃の標本を観察しても、個々の病理医によって診断が異なることだけでなく、欧米の病理学者に比べ、日本の病理学者は、病変を癌と診断することが有意に多く、核異型、構造異型を重視して、浸潤がなくとも癌と診断している場合が多かったということである。特に、平坦型や、陥凹型のもので著明であった。

日本では、そもそも歴史的に胃の腫瘍性疾患の病理学に異形成 (dysplasia) の概念がなく、粘膜内にとどまる病変であってもリンパ節転移などを生じている癌の組織・細胞学的特徴の経験から、浸潤が形態学的に確認されなくとも核異型、構造異型から癌と診断すべきであると伝統的に教授されてきた。すなわち、結果的に偽陰性となることを怖れた判断が優先されてきたといえよう。この考え方は、胃癌にとどまらず、日本の外科病理学

専門家の間に広くみられるものであって、気管支、食道、大腸などの病変においても上皮内癌の診断が病理診断能力の評価にとって重要視されてきた慣習に反映されている。

また、欧米でも、胃の異形成、腺腫の定義、診断基準に関しては混乱がみられているとの報告もある (16)。

この論文で明らかになった東西の診断乖離問題を解決する方法として、その後、Schlemperらは、noninvasive high-grade epithelial neoplasiaという概念を提唱し、これによって早期胃癌診断の国際的な比較を改善できるとしているが、病理診断の正確度の問題を回避して、診断の標準化を優先するという観点からは、実際的な解決法であると思われる (17)。しかし、形態検査の正確度 (accuracy) を正面から問題とする立場からは、解決を先送りしたものともいえよう。

腸型の胃癌という限定された病変の診断が「権威」の間でも乖離することをみてきた。次に、進行性胃癌の予後因子のひとつとして注目されている病理組織学的分類についてみてみよう。

2) 進行性胃癌の病理組織学的分類

胃癌、特に、進行性の胃癌に関しては、UICCによるTNM staging systemが広く用いられており、生命的予後や治療成績の比較に利用されてきたところである。

一方、胃癌の病理組織学的分類が、予後因子のひとつであること、あるいは、病因論との関連で重要であることが、従来から指摘されてきた。胃癌も種々の組織像を呈することが知られており、これをどのように分類するかは常に問題とされているところであるが、国際的には、Laurenの分類 (WHO分類の基礎)、Mingの分類、Gosekiの分類が用いられ、特にLaurenのものはWHO分類にも取り上げられ、広く利用されているものである。

Laurenの分類の骨子は、胃癌を腸型、瀰漫型、その他のものに大別するものであるが、最近、この診断のinterobserver variationsの検討結果が報告された (18)。これによると、分類の一致したものは、92例中71例で、kappa valueは、0.72とされて、想像以上に乖離していることが確認され、国際間の各型腫瘍発生率の比較を困難なものにしているとの結論が述べられている。

序論で述べた、日本における進行性胃癌の手術

成績施設間差を解析してゆく場合にも、こうした診断における interobserver variations の問題を考慮する必要があることを示唆する報告と思われる。

また、予後推定にどれだけ有効かを、Lauren の分類、Ming の分類、Goseki の分類に関して調査した結果が幾つか報告されている。Goseki の分類が劣り、Lauren の分類を評価するものが多いが、population-based study では、Lauren の分類は Ming の分類に劣るとの結果が報告されている (19)。さらには、Lauren の分類も、Ming の分類も予後との相関はないとの報告もある (20)。

分類のし易さでは、上にあげた3分類ともに単純であり、甲乙つけ難いが、予後因子の観点からは、統計的な評価が可能なことを示しているとともに、対象や、診断者によって、評価が分かれることも示しており、分類の有用性の評価、これの標準化に関しては多くの問題が内在していることを示しているといえよう。

以上、胃癌を例に癌の診断そのものにおける標準化の問題をみてきたが、以下には、癌の診断ではなく、その予後との相関の問題、所見報告の標準化の観点から、悪性度のスコア化が求められている前立腺癌についてみてみよう。

3) 前立腺癌のスコアにおける問題点

多くの臓器の腫瘍病理組織学的分類が、腫瘍の増殖様式を判定する組織型分類と、分化度分類によって腫瘍を分類し、混在する場合には、優勢像、あるいは、分化度の悪いものをもって判定するなどの基準を設けているが、前立腺癌の場合、Gleason が考案した Gleason's score (grade) が臨床的な予後推定、治療法選択に用いられている。

これは、前立腺生検標本を観察し、前立腺癌の分化度に相当する組織構築と周囲への浸潤様式を5段階の grade に分け、もっとも優勢な像の grade と、次に優勢な像の grade の数値を足して、その症例の score とするもので、予後とよく相関するとされている。

一方、grade の定義が曖昧で再現性が悪いことは Gleason 自身が認めているところであり、この問題の解決が求められていた。これに対する一つの回答として欧米の病理医が提案したものが、internet のホーム・ページの利用である。これに

接続し、テストに参加するものに Gleason's grading を実際に行わせ、専門家のコンセンサスを得た grade と合致しているかを評価し、その評価を参加者に知らせ、grading の標準化に寄与しようとして意図されたものである (21, 22)。

日本の病理医も、これによって標準化を図るものが多いと聞いており、これからの病理組織学的診断の標準化のあり方を示唆するものと思われる。

個別研究：肺腺癌亜分類の予後因子との関連での検討 (23)

1999年に、WHO から肺癌の分類が改定されて出版された。特に、肺腺癌の分類に関しては、第2版から大幅に改定されている。一方、個々の研究者が、各施設の症例を検討した上で、肺腺癌の分類を提唱されており、どの分類が予後とどのように関連しているかを評価する必要を感じたので、以下に概略を述べる検討を行い、結果をえた。

現在、日本で利用されている肺腺癌の病理組織学的分類としては、改定 WHO 分類 (24)、Noguchi らの分類 (25)、Silver と Askin (26)、肺癌取り扱い規約などがあるが、今回の検討では肺癌取り扱い規約に関しては、腺癌の分類に関しては予後評価を意識したものではないので検討から除いた。

対象・方法：1987年から1996年の10年間に防衛医大病院で切除された肺腺癌147例を対象とした。対象症例を観察し、上述の3分類法による再分類を行った。Silver と Askin による分類では、乳頭状腺癌と細気管支肺胞上皮癌とを優勢像で分類するよう指示されているが、容易に分類できるように、腫瘍辺縁部での進展様式に着目して優勢像を判定した。予後の検討は、Kaplan-Meier 法、Logrank 検定による単因子分析、比例ハザードモデルによる多因子分析を行い、 $p < 0.05$ を有意差判定基準とした。

結果：全例を用いた単因子分析では、3分類とも、分類ごとに予後に有意差がみられ、WHO3版分類の細気管支肺胞上皮癌、Noguchi 分類でのA、B型、Silver と Askin による分類での細気管支肺胞上皮癌に分類されるものの予後は良好で、WHO3版分類の充実型、Silver と Askin による分類での乳頭状腺癌が予後不良であった。

病期別単因子分析および病期を含む多因子分析では、Silver と Askin による分類でのみ有意差が

検出可能で、細気管支肺胞上皮癌が良好、乳頭状腺癌が不良との結果をえた。

まとめると、今回の防衛医大病院切除肺腺癌147例の検討では、WHO3版の分類や、Noguchi分類のように国内の学会で広く利用されている分類よりも、SilverとAskinの提唱した分類の方が、腺癌の予後推定、特に進行癌の予後推定には有効であるとの結論を得た。

以上、形態検査、特に、病理組織学的診断の有用性を評価する上での問題点について、例をあげて解析してきた。以下には、これらの問題点を解決し、標準化を推進してゆく上で考慮しなければならない日本の現状を述べ、解決方法の幾つかに関して言及する。

日本での病理組織学的分類の標準化と正確度、精度向上のための方法の現状

日本では、形態学的検査に関しては、診断、判断する医師の裁量に委ねられ、日本の病院における形態検査の現状の項でも触れたように、臨床的検査の有用性の評価の基礎をなす正確度、精度向上のための、系統的で、組織的な取り組みが遅れているのが現状といえよう。

現在、日本で実施されている病理組織学的分類での正確度、精度向上のための方法は、「病理診断」「病理と臨床」などの学術雑誌による啓蒙活動、(社)日本病理学会が主催する総会、秋期大会など全国レベルでの学会での教育講演(系統別病理診断講習会を含む)、国際病理学会(IAP)が主催し、(社)日本病理学会や、臨床検査医学会が協賛して行う病理組織診断のためのセミナー、地方の各支部が行う症例検討会などで行われる教育活動などがある。

病理組織分類や、細胞診での標準化を目的とした活動としては、(社)日本病理学会が外科系の各学会と共同での「癌取り扱い規約」の作成が主たるものであろう。

そして、組織学的分類の標準化の動機としては、各病院などの診療施設で、外科などの臨床医が診療を行ってゆく上での必要性から、当該の学会が用いているこの「癌取り扱い規約」などに沿った分類を、病理医に要請してくるといった病理学会外部からの要請が主たるものになっていると考えられる。

形態的正確度 (ACCURACY) 評価の標準化にいたる方法

疾患の本質的な形態学的特徴を明らかにすることが、真の意味での形態学的正確度の評価基準の設定につながるものであることはいうまでもないことであるが、上にもみてきたように、これは今後更に研究がなされなければならない領域に属し、実際的には、その領域の「権威」らの討議によるコンセンサスに基づいた診断基準設定が基本とならざるを得ないように思われる。

このコンセンサスを基準とした形態的評価の標準化には、評価者が同一の検体を評価し、その結果の一致率を kappa value として求め、0.80以上の値がでるまで評価を繰り返すことによって、評価者間の差を少なくするという方法が望ましいと考えられてきた(27)。

しかし、この評価に必要な検体は、従来は組織・細胞標本しかなかったために、多くの評価者に同時に配布することができず、多くの施設に分散して配属されている評価者の標準化は、極めて困難な状況にあった。

この隘路を打開するのに、近年、日本の病院の血液検査室、病理組織検査室など形態的検査・診断を行っている施設に普及しはじめている顕微鏡画像入力装置を接続したコンピュータを用いたファイル構築と、この画像のインターネットを介した全国の施設への大量の評価画像の配布が有効であろうと考えられる。上にも述べた、インターネットのホーム・ページを利用した前立腺癌の Gleason's grading の欧米での教育実績を参考として、標準化に取り込むことが有効であろうと予想される。今後、臨床検査医学会に所属する病理医の協力を得て、血清 PSA 値と関連する前立腺癌の Gleason 分類と、肝機能検査と関連する慢性肝炎のスコアに関して、協力の得られた施設の代表者らと研究の方法に関して検討したいと考えている。

病理学会や臨床細胞学会などが、形態学的診断におけるガイドラインをホーム・ページ上で提示するのも一法であろう。この場合には、できるだけ多数の標準的な画像を掲載するとともに、判断基準を明示し、また、国際的に用いられている診断ガイドラインなども選択肢として掲載する必要がある。

腫瘍性疾患の病理組織学的診断分類については、

日本の外科系の学会と、日本病理学会とが共同で、癌取り扱い規約を発行し、これが広く用いられている。しかし、AFIP あるいは、個々の施設で定めた分類を用いる場合も実際にはみられる。臨床的有用性の評価と、その標準化の観点からは、すくなくとも、どの分類を用いて診断しているのかを関係者に明示する必要が生じてくるものと思われる。

形態学的検査の内部精度管理（クオリティー・マネージメント）

形態学的検査の精度管理は、観察対象となる標本の精度管理が基本となる。すなわち、検体採取、採取検体の固定、脱水、包埋、薄切、染色、封入などの組織・細胞診の標本作成過程における精度管理が必要である。

上にも触れたように、検体の固定までは、生検施行者が、これ以後に関しては病理標本作成技術者、多くは、病院の病理検査室に所属する臨床検査技師が行ってゆくものである。生検実施者は、観察、診断の対象となる検体の部位、量の安全で、適切な採取から、固定液の適切な選択、固定の濃度、温度、時間の3要素に関する配慮などが要求される。

固定された検体の標本作成の精度管理に関しては、病理検査に関与する臨床検査技師などの技術者が責任をもって実施されるべきものである。多くの先人たちの努力によって、H.E.染色などの基本的な染色と、いわゆる特殊染色に関しては、種々のマニュアルが作成され、標準化に寄与してきた。また、病理技術研究会などの団体が、これの改良や、標準的技術の普及、評価を積極的に行ってきた。

しかし、最近、病理診断や、細胞診で広く利用され、診断の重要な補助的診断に利用される免疫組織学的検査に関しては、精度管理が困難な状況にある。

実際の診断者の再現性に反映されるような標本作成が実施されているか、という点にまで配慮したトータルな精度管理、質のマネージメントが望まれているといえよう。

形態学的外部精度管理（Proficiency）評価の方法

予後や、治療に対する反応などを根拠とした分類の有用性、妥当性に関する統計的研究、有用と

考えられ、標準と考えられる分類の普及、教育に関する制度の整備、実際に診断を行うものが、標準化された判断（診断）基準を用いて分類が行われるための「精度管理」調査研究などに基づいた評価基準の作成し、これを明示した上での外部精度管理の実施の必要性が以前に増して強く感じられるようになってきている。

こうした形態検査の外部精度管理アセスメントは、各地区の臨床検査技師会などによって行われている。しかし、各県などの医師会が、各地区の臨床検査技師会などに委託して実施しているので、その地区の技師会など、実施するものの、問題意識や、力量に依存しており、アセスメントの全国的な標準化が遅れていると思われる。

埼玉県では、臨床検査技師会所属の細胞診検査士の活動が活発で、県が実施する精度管理用の問題作成にあたって、検体の疾患別出現頻度なども参考にするなど、実際的な proficiency 評価のための努力を続けており、成果をあげている（28）。

病理組織学的診断に関しては、peer review を基本とする内部精度管理の実施が望まれるが、病院のリスク・マネージメントの観点から、日本病理学会などによる外部精度管理の導入も要望される時代になる可能性もあると思われる。

おわりに：形態検査の有用性評価の方法の開発に向けて

歴史的にも、疾患の認識にとって形態的な異常が重要であり、ほかに適切な方法がなかったこともあり、病理組織学的診断をはじめ形態検査が有用であることは、従来からいわば a priori に認められてきた。

実際に、疾患の診断における病理組織学的、細胞学的診断が重要性は、MRI などの画像診断装置、あるいは、種々の測定機器、試薬の開発を背景とした臨床検査の発達をもってしても代えがたい側面をもっているものと思われ、むしろ、確定診断としての重要性は増しているようにも思われる。

しかし、形態学的検査における有用性の概念は、未だ十分に整理されているとはいえない状況であり、最近では、医学、医療の発展、特に、免疫学的検査や、遺伝子検査などの診断技術の発展は、個々の診断技術の客観的な指標による有用性を評価する必要性を生じさせている。

一方、上にもみてきたように、いままでも組織

学的分類に関しては、分類しやすく、多くの人々が共通して用いることができ（すなわち、再現性に優れており）、また、臨床的な応用が可能であること（すなわち、予後を予測できるものであること、各種治療法による反応を予測できるものであること、臨床的な病像との対応が優れている）などの観点から評価されてきたといえよう。ただし、医師の「診断」であることから外部精度管理になじまないとされ、学問的、研究的な段階にとどまり、この評価が、系統的、組織的に行われてこなかったところに問題があるとも考えられる。

病理組織学的診断が、疾患の診断、分類において絶対的な位置を占めていた時代は終わろうとしている。今後は、病理組織学的検査などの形態学的検査が診断技術の gold standard であるとする根拠を、自らのデータを用いて明示しなければならない時代となるであろう。形態学的検査を実施するものは、こうした時代の変化を自らのものとして、統計学的な種々の方法を利用した有用性評価の方法を自ら考案するとともに、方法の標準化に向けて、概念の整理を行ってゆかなければならないと思われる。

臨床生化学検査などのように、測定技術に関して有効性を発揮した外部精度管理アセスメントの形態学的検査への安易な導入は、かえって混乱を招くだけであり慎まねばならないが、医療の質向上のためには、種々の臨床的有用性に関する根拠を示した病理組織学的診断や、細胞診などを含む形態検査の正確度、精度の向上が望まれる。

臨床的有用性に関する根拠に基づいた病理専門家による分類のコンセンサスの形成、コンセンサスのインターネットなどを通じた普及、病理医を含む医師たちに対する生涯教育の方法の改良など、有効と思われる種々の方法を開発、研究してゆくことを制度的に保証することが日本のこれからの医療の質向上に必要と思われる。

文献

1. Yokota T, Kunii Y, Kikuchi S, Saito, Yamauchi H: Disclosure of results of gastric cancer operations in Japan. *Lancet* 356: 1689-1690, 2000.
2. 真鍋俊明, 向井清他: ワークショップ 病理診断の精度管理。日本病理学会会誌 90:120-121, 2001.
3. Sternberg SS: *Histology for pathologists*. 2nd edition. Lippincott Williams & Wilkins. 1997.
4. Bravo AA, Sheth SG, Chopra S: Liver biopsy. *New Engl J Med* 344:495-500, 2001.
5. Froehlich F, Lamy O, Fried M, Gonvers JJ: Practice and complications of liver biopsy: results of a nationwide survey in Switzerland. *Dig Dis Sci* 38:1480-1484, 1993.
6. Brunt EM: Grading and Staging the Histopathological Lesions of Chronic Hepatitis: The Knodell Histology Activity Index and Beyond. *Hepatology* 31:241-246, 2000.
7. Knodell GK, Ishak KG, Black WC, et al.: Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* 1:431-35, 1981.
8. Desmet VJ, Gerber MA, Hoofnagle JH, Manns M, Scheuer PJ: Classification of chronic hepatitis: diagnosis, grading and staging. *Hepatology* 8:493-96, 1988.
9. ter Borg F, ten Kate FJW, Cuypers HTM, et al.: Relation between laboratory test results and histological hepatitis activity in individuals positive for hepatitis B surface antigen and antibodies to hepatitis B e antigen. *Lancet* 351: 1914-18, 1998.
10. Habersetzer F, Trepo C, Minor J, Comanor L: Predictions of liver histology from laboratory tests and demographic data. *Lancet* 352: 907 1998.
11. Heathcote EJ, Shiffman ML, Cooksley GE, et al.: Peginterferon alfa-2a in patients with chronic hepatitis C and cirrhosis. *New Engl J Med* 343:1673-80, 2000.
12. French METAVIR Cooperative Study

- Group: Intraobserver and interobserver variations in liver biopsies in patients with chronic hepatitis C. *Hepatology* 20: 15-20, 1994.
- 1 3. Westin J, Lagging LM, Wejstal R, Norkrans G, Dhillon AP: Interobserver study of liver histopathology using the Ishak score in patients with chronic hepatitis C virus infection. *Liver* 19: 183-187, 1999.
 - 1 4. Goldin RD, Goldin JG, Burt AD, et al.: Intra-observer and inter-observer variation in the histopathological assessment of chronic viral hepatitis. *J Hepatol* 25: 694-654, 1996.
 - 1 5. Schlemper RJ, Itabashi M, Kato Y, et al.: Differences in diagnostic criteria for gastric carcinoma between Japanese and Western pathologists. *Lancet* 349: 1725-1729, 1997.
 - 1 6. Goldstein NS, Lewin KJ: Gastric epithelial dysplasia and adenoma: historical review and histological criteria for grading. *Hum Pathol* 28: 127-33, 1997.
 - 1 7. Schlemper RJ, Kato Y, Stolte M: Well differentiated adenocarcinoma or dysplasia of the gastric epithelium: rational for a new classification system. *Verh Dtsch Ges Pathol* 83: 62-70, 1999.
 - 1 8. Shibata A, Longacre TA, Puligandla B, Parsonnet J, Habel LA: Histological classification of gastric adenocarcinoma for epidemiological research: Concordance between pathologists. *Cancer Epidemiology Biomarkers and Prevention* 10:75-78, 2001.
 - 1 9. Roy P, Piard F, Dusserre-Guion L, et al.: Prognostic comparison of the pathological classification of gastric cancer. A population-based study. *Histopathology* 33: 304-310, 1998.
 - 2 0. Garnier P, Vielh P, Asselain B, et al.: Prognostic value of the Lauren and Ming classifications in gastric adenocarcinoma. *Multidimensional analysis. Gastroenterol Clin Biol* 12: 553-558, 1988.
 - 2 1. Egevad L: Test your accuracy in Prostate Cancer Grading !!! <http://www.eur.nl/fgg/pathol/education/Pctest.htm>
 - 2 2. Kronz JD, silberman M, Allsbrook WC, Epstein JI: Gleason's grading of Prostate needle biopsies; Tutorials and testing Web site. <http://162.129.103.34/prostate/>
 - 2 3. 相田真介、島崎英幸、玉井誠一: 予後因子としての肺腺癌亜分類の検討 *日本病理学会会誌* 90 (1) : 217, 2001.
 - 2 4. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E: *Histological Typing of Lung and Pleural Tumours. Third edition. Springer, Berlin Heidelberg, 1999.*
 - 2 5. Noguchi M, Morikawa A, Kawasaki M, Matsuno Y, Yamada T, Hirohashi S, Kondo H, Shimosato Y: Small adenocarcinoma of the lung. *Cancer* 75:2844-52, 1995.
 - 2 6. Silver SA, Askin FB: True papillary carcinoma of the lung. A distinct clinicopathologic entity. *Amer J Surg Pathol* 211: 43-51, 1997.
 - 2 7. Brennan P, Silman A: Statistical methods for assessing observer variability in clinical measures. *BMJ* 304: 1491-4, 1992.
 - 2 8. 埼玉県、埼玉県医師会 : 平成 11 年度臨床検査精度管理事業報告書 pp. 241-252, 2000.