

COMMITTEE REPORT

日本肺癌学会・日本臨床細胞学会 2学会合同委員会報告： 肺がん検診における喀痰細胞診の診断一致性と標準化

佐藤雅美¹・柴 光年¹・齋藤泰紀¹・佐藤之俊¹・渋谷 潔¹・
土田敬明¹・中山富雄¹・宝来 威¹・池田徳彦²・
河原 栄²・三浦弘之²・中嶋隆太郎²・田口明美²・
矢羽田一信²・島垣二佳子²・神尾淳子²・長尾 緑²・
三宅真司²・下川幸弘²・田中良太²・遠藤千顕¹

Report of a Joint Committee of the Japan Lung Cancer Society and the Japanese Society of Clinical Cytology: Inter-rater Agreement of Sputum Cytology for Lung Cancer Screening in Japan

Masami Sato¹; Mitsutoshi Shiba¹; Yasuki Saito¹; Yukitoshi Satoh¹; Kiyoshi Shibuya¹;
Takaaki Tsuchida¹; Tomio Nakayama¹; Takeshi Horai¹; Norihiko Ikeda²;
Ei Kawahara²; Hiroyuki Miura²; Ryutaro Nakashima²; Akemi Taguchi²;
Kazunobu Yahata²; Nikako Shimagaki²; Junko Kamio²; Midori Nagao²;
Shinji Miyake²; Yukihiro Shimokawa²; Ryota Tanaka²; Chiaki Endo¹

¹Japan Lung Cancer Society, Committee for Lung Cancer Screening, Working Group for Sputum Cytology, Japan; ²Japanese Society of Clinical Cytology, Working Group for Lung Cancer Screening, Japan.

ABSTRACT — **Background.** To compare lung cancer detection rate by sputum cytology, we need some assurance that the estimates do not vary widely if the different observers evaluate the same specimens. The aim of this study was to determine interrater agreement of sputum cytology diagnoses. **Methods.** Slides of sputum cytology from 150 subjects were selected from a pool of slides held by six of the laboratories that had participated in a population-based lung cancer screening program over the last ten years in Japan. The cytotechnologists in these laboratories had considerable experience with sputum cytology. Each case was reevaluated six times. Cases that were diagnosed as the same category by all six laboratories were selected as consensus cases to serve as standardized sputum cytology cases. Sixty-four cytotechnologists with various levels of experience in sputum cytology then reevaluated these consensus cases. Interrater agreement was calculated by kappa statistics including Fleiss' kappa. **Results.** All pairs of interlaboratory agreement for the 150 cases showed statistically significant kappa values, most pairs showing substantial agreement. Fleiss' kappa value across the six laboratories was 0.5. Twenty-six cases were identified as the consensus cases, and the agreement among observers with less experience of sputum cytology showed significantly lower than the agreement among those with considerable experience. Moreover, cytotechnologists with less experience under-diagnosed the slides significantly more often than those with considerable experience. **Conclusion.** When the observers have considerable experience with sputum cytology, interobserver agreement is good. Only cytotechnologists with considerable experience in this field make accurate diagnoses on sputum cytology.

(JLCC. 2015;55:859-865)

KEY WORDS — Sputum cytology, Quality assurance, Interrater agreement, Kappa statistics, Fleiss' kappa value

¹日本肺癌学会集団検診委員会喀痰細胞診による肺癌検診小委員会；²日本臨床細胞学会肺癌検診ワーキンググループ。

要旨——はじめに、喀痰細胞診による肺癌発見率を比較する場合の前提は、同一標本に対する判定は判定者が異なっても大きく異ならないことであるが、喀痰細胞診の検査者間判定一致性の検討はこれまで我が国では詳細に検討されてこなかった。**方法**、過去10年間に我が国の肺がん検診で用いられた喀痰細胞診症例150例を喀痰細胞診の経験豊富な6施設より集積した。各症例を6施設にて再判定し、判定一致例を標準細胞症例と定義した。喀痰細胞診に対する経験値が様々な検査者計64名によって標準細胞を判定した。判定者間一致性の検討には種々のカッパ統計量が用いられた。**結果**、全150例

に対する6施設間の判定一致率は高く、Fleissのカッパ係数は0.5であった。26症例が標準細胞症例として同定された。喀痰細胞診に対する経験の浅い判定者間の判定一致性は経験豊富な者の一致性に比較して有意に低かった。また、経験の浅い者は有意に過小評価した。**まとめ**、喀痰細胞診の経験豊富な検査者間の判定一致性は高く、経験の浅い者は過小評価する傾向があるので、肺がん検診における喀痰細胞診の判定は経験豊富な施設、判定者が行うべきである。

索引用語—— 喀痰細胞診、精度管理、検査者間一致性、カッパ統計量、Fleissのカッパ係数

はじめに

肺がん検診における喀痰細胞診は肺門部早期扁平上皮癌をターゲットとしているが、我が国における肺門部肺癌に対する系統的横断調査が日本肺癌学会、日本呼吸器内視鏡学会、日本臨床細胞学会の3学会合同委員会による肺門部早期肺癌実態調査アンケート報告として報告された。¹それによると、肺門部扁平上皮癌は全国で年間約4000例の存在が推計され、そのうち、肺門部早期扁平上皮癌は年間154~270例と推定された。即ち、早期癌比率は肺門部扁平上皮癌においては10%を下回っていると考えられた。さらに、早期癌比率には、北海道16.7%、東北14.5%、関東10.1%、東海5.5%、甲信越13.1%、北陸22.6%、近畿4.1%、中国13.8%、四国16.5%、九州沖縄5.7%と有意な地域間格差も認められた。

一般的に喀痰細胞診による肺癌発見率を比較する場合、同一検体に対する検査者間での診断一致性が担保されていることが前提となる。しかし、我が国において喀痰細胞診の検査者間診断一致性に関する大規模な検討はこれまでなされてこなかった。そこで、日本肺癌学会、日本臨床細胞学会による2学会合同委員会（日本肺癌学会集団検診委員会喀痰細胞診による肺癌検診小委員会、日本臨床細胞学会肺癌検診ワーキンググループ）では、肺がん検診における喀痰細胞診の検査者間診断一致性を検討し、委員会報告としての提言を行った。

方法

検体

全国6施設において2001年から2010年の間に肺がん集団検診において用いられたパバニコロ染色²された喀痰細胞診標本を検体として用いた。この6施設は上記2学会合同委員会の委員が所属する施設で、喀痰細胞診による肺がん検診に長年従事してきた、宮城県、福島県、

千葉県、新潟県、石川県、大阪府にそれぞれ位置する6施設である。肺がん検診における喀痰細胞診は、A、B、C、D、Eの5判定区分に分類されるが、本研究では、6施設においてC、D、E判定のいずれかに判定された検体を用いた。なお、本研究は、参加各施設における倫理委員会において倫理承認を得るとともに、研究開始前に各施設において研究実施概要の公告を行った。

方法

第1部

喀痰細胞診標本は、上記6施設の原判定がC、D、E判定の各判定50症例、計150症例を集積した。内訳は、宮城68例、千葉30例、大阪20例、石川16例、新潟10例、福島6例であった。1症例につき喀痰細胞診のスライドガラスは2枚あるいは4枚が提供された。150例を無作為に1セット25例の計6セットに分割し、1セットをおよそ2週間、計6セットを約12週間で上記6施設において再判定した。各施設では2名以上の細胞検査士、細胞診専門医が再判定に参加し、各症例の施設判定を決定した。最終的に、6施設の判定が一致した症例を各判定区分の標準化された喀痰細胞診症例（以下、標準細胞症例）とした。

第2部

2014年および2015年に仙台市で開催された肺癌検診喀痰細胞診セミナー in Miyagi (<http://www.ctjsc.com/ctjsc/cms/wp-content/uploads/edae984156a4bdadfc6d2b6c5913ae87.pdf>)に参加した細胞検査士、細胞診専門医によって上記標準細胞症例の再判定を行った。

統計学的解析

2者間の診断一致性の評価にはカッパ統計量を用いた。³3分類以上の診断に関する一致性に関しては重み付けカッパ係数³を求めた。また、3者以上の多数者の診断一致性の評価には、Fleissのカッパ係数^{4,5}を用いた。カッパ値は、0.2以下を低い一致性、0.21~0.4を並の一致

性, 0.41~0.6 を中等度の一致性, 0.61~0.8 を実質的な一致, 0.81~をほぼ完全な一致として評価した. また, 分割表における分布の差の検定にはカイ二乗検定を用いた. $p < 0.05$ をもって統計学的有意とした. なお, C 判定として算定した一部には B 判定とした回答も含まれていたが, 解析上の都合からこれらも C 判定と算定した.

結果

第 1 部

150 症例のうち, 6 施設の診断が一致した標準細胞症例

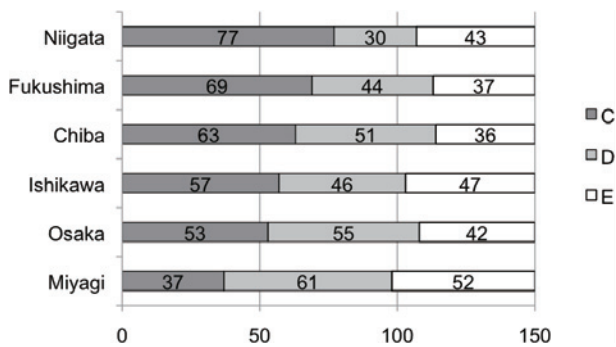


Figure 1. Distributions of sputum cytology categories by laboratory for the 150 cases. The number of cases classified as category E did not differ across laboratories. However, the number of cases classified as category C or D differed significantly (Chi square test, $p = 0.0003$).

は C 判定 5 例, D 判定 3 例, E 判定 18 例の計 26 例に達した.

6 施設による 150 症例の再判定の判定区分分布を Figure 1 に示した. 宮城の施設では C 判定が 30% 未満であったが, 新潟では 50% 以上が C 判定とされ, 判定区分の分布に有意差が認められた(カイ二乗検定, $p < 0.05$).

C, D, E 判定に対する施設間の診断一致性に関して重み付けカッパ値を Table 1 に示した. 6 施設 15 通りの組み合わせのカッパ値はすべて 0.6 以上で, 14 組で実質的な一致と評価された. また, 6 施設の診断一致性に関する Fleiss のカッパ係数は 0.5 (95% 信頼区間 0.47~0.53) であった.

肺がん検診では D および E 判定が要精査とされるので, C 判定と D+E 判定, 即ち要精査判定に対する施設間の診断一致性に関してカッパ値を Table 2 に示した. 10 の組み合わせで中等度の一致性を, 5 組にて実質的な一致と評価された. Fleiss のカッパ係数は 0.57 (95% 信頼区間 0.53~0.61) であった.

ついで, C+D 判定と E 判定, 即ち悪性診断に対する施設間の診断一致性に関してカッパ値を Table 3 に示した. 6 組で中等度の一致性を, 9 組で実質的な一致を認めた. Fleiss のカッパ係数は 0.63 (95% 信頼区間 0.59~0.67) であった.

第 2 部

2014 年に開催された第 11 回肺癌検診喀痰細胞診セミ

Table 1. Interlaboratory Agreement for the Three Categories (C vs. D vs. E) of Sputum Cytology with Linear Weighted Kappa

	Fukushima	Chiba	Niigata	Ishikawa	Osaka
Miyagi	0.67	0.67	0.61	0.68	0.74
Fukushima		0.7	0.75	0.61	0.69
Chiba			0.61	0.65	0.62
Niigata				0.62	0.67
Ishikawa					0.6

All kappa values show statistically significant ($p < 0.05$). Fleiss' kappa was 0.50 (95%CI 0.467-0.526).

Table 2. Interlaboratory Agreement for the Two Categories (C vs. D or E) of Sputum Cytology with Non-weighted Kappa

	Fukushima	Chiba	Niigata	Ishikawa	Osaka
Miyagi	0.56	0.56	0.45	0.56	0.66
Fukushima		0.68	0.68	0.52	0.65
Chiba			0.55	0.51	0.52
Niigata				0.54	0.68
Ishikawa					0.5

All kappa values show statistically significant agreement ($p < 0.05$). Fleiss' kappa was 0.57 (95% CI 0.530-0.612).

Table 3. Interlaboratory Agreement for Malignancy (C or D vs. E) of Sputum Cytology with Non-weighted Kappa

	Fukushima	Chiba	Niigata	Ishikawa	Osaka
Miyagi	0.67	0.65	0.68	0.68	0.72
Fukushima		0.58	0.69	0.57	0.64
Chiba			0.54	0.62	0.62
Niigata				0.59	0.56
Ishikawa					0.6

All kappa values show statistically significant agreement ($p < 0.05$). Fleiss' kappa was 0.63 (95% CI 0.587-0.670).

Table 4. Interobserver Agreement on the 14 Consensus Cases

Observers	Fleiss' kappa value	95%CI
Under 10 years' experience (n = 12)	0.272	0.226, 0.318
Over 10 years' experience (n = 25)	0.448	0.426, 0.470
Non-respiratory specialty (n = 22)	0.314	0.290, 0.339
Respiratory specialty (n = 15)	0.559	0.522, 0.595
Respiratory specialty with over 10 years' experience (n = 13)	0.481	0.442, 0.520
Non-respiratory specialty with over 10 years' experience (n = 12)	0.406	0.359, 0.453

Table 5. Distributions of Sputum Cytology Categories by Consensus Diagnoses for the 14 Consensus Cases

	Non-respiratory (n = 22)	Respiratory (n = 15)	P value	Under 10 yrs (n = 12)	Over 10 yrs (n = 25)	P value
Underestimate	75/308 (24.4%)	7/210 (3.3%)	<0.0001	39/168 (23.2%)	43/350 (12.3%)	0.007
Overestimate	32/308 (10.4%)	35/210 (16.7%)	0.068	18/168 (10.7%)	49/350 (14%)	0.360
Raw agreement	201/308 (65.3%)	168/210 (80%)	0.0003	111/168 (66.1%)	258/350 (73.7%)	0.072

ナー in Miyagi には、全国から 37 名の細胞検査士、細胞診専門医が参加した。細胞診の経験年数にて分類すると 10 年未満 12 名、10 年以上 25 名であった。また、専門領域で分類すると呼吸器専門 15 名、呼吸器以外専門 22 名であった。この 37 名によって標準細胞症例 14 例 (C 判定 5 例、D 判定 4 例、E 判定 5 例) を診断した。Fleiss のカッパ係数は、細胞診経験 10 年未満群 0.27、10 年以上群 0.45、呼吸器以外専門群 0.31、呼吸器専門群 0.56 であり、経験年数で 10 年以上の群が、また、専門領域で呼吸器専門の群が、有意に高いカッパ値を示した (Table 4)。経験年数 10 年以上の 25 名を呼吸器専門の 13 名と呼吸器以外専門の 12 名に分類して、同様に Fleiss のカッパ係数を算出すると、呼吸器専門群は 0.48 と、呼吸器以外専門群の 0.41 に比較して高いカッパ値を示したが、有意差は認められなかった (Table 4)。

標準細胞症例の判定に対する参加者の判定分布を検討したところ (Table 5)、経験年数 10 年未満の群は 10 年以上の群に比較して有意に、また、呼吸器以外専門の群は

呼吸器専門群に比較して有意に、標準判定より過小評価する割合が高かった。一方、過大評価に関して有意差は認められなかった。

2015 年に開催された第 12 回肺癌検診喀痰細胞診セミナー in Miyagi には、全国から 27 名の細胞検査士、細胞診専門医が参加した。27 名のうち、経験年数が 10 年未満は 5 名であった。専門領域では、呼吸器専門 17 名、呼吸器以外専門 10 名であった。この 27 名によって標準細胞症例 10 例 (C 判定 3 例、D 判定 3 例、E 判定 4 例) を診断した。Fleiss のカッパ係数は、呼吸器以外専門群の 0.15 に対して呼吸器専門群 0.42 と有意に高かった (Table 6)。また、呼吸器以外専門の群は呼吸器専門群に比較して有意に、標準判定より過小判定する割合が高かったが、過大判定に関して有意差は認められなかった (Table 7)。なお、2014 年、2015 年の両検討において、過小評価の内容を検討した結果、E 判定を D 判定や C 判定とした群と、D 判定を C 判定とした群の間には、有意な頻度の差を認めなかった。

Table 6. Interobserver Agreement on the 10 Consensus Cases

Observers	Fleiss' kappa value	95%CI
Non-respiratory specialty (n = 10)	0.150	0.092, 0.208
Respiratory specialty (n = 17)	0.419	0.382, 0.456

考 察

本研究は、本邦において初めて大規模に実施された喀痰細胞診の診断一致性に関する検討であった。第一部において、喀痰細胞診の経験豊富な施設間の判定一致性は極めて高いことが示された。また、第二部において、喀痰細胞診の経験が豊富な者は、経験が浅い者に比較して、診断一致性が有意に高いこと、また経験が浅い者は過小評価する傾向が高いことが示された。このことは、2度の独立した検討によって同様の傾向が示されたことから、信頼性の高い結論と考えられた。

喀痰細胞診に関する診断一致性を検討する際、標本の染色性を考慮する必要がある。今回の検討でも6施設のパパニコロ染色の方法は異なっており、施設ごとに細胞の染色性が異なっていた。即ち、ある施設が他施設にて作製された喀痰細胞診標本を判定する際には、染色性の相違から正しい判定がなされない可能性があり、ひいては施設間の診断一致性を正しく評価できないと考えられた。そこで、本研究では標準細胞という概念を導入した。即ち、喀痰細胞診の経験が豊富な複数施設から様々な染色性の喀痰細胞診標本を集めて、すべての施設で同一判定が下された標本は、その染色性への慣れにかかわらず、経験豊富な判定者であれば同一の判定が下され、従って標準化されると考えられる。Gold standardの設定がほとんど困難であった喀痰細胞診におけるGold standardの実践的な代替概念としてこの標準細胞を捉えることが可能と考えられ、この標準細胞を用いることによって、検査者間の診断一致性や、検査者の診断能が正確に検討可能となった。

2者間の判定一致性にはカッパ係数が用いられた。カッパ係数は粗な一致率から偶然による一致率を除いた一致率と考えられ、2者間の判定一致性の検討には最適と考えられている。³ また、重み付けカッパ値を用いることで、例えばC判定とD判定の相違より、C判定とE判定の相違をより大きな不一致として検討し得た。さらに、通常のカッパ値は2者間の一致性検討にしか用いられないが、Fleissのカッパ係数を用いて3者以上の集合における全体としての一致性の検討も行った。本研究では、このように様々なカッパ統計量を用いることで、喀

Table 7. Distributions of Sputum Cytology Categories by Consensus Diagnoses for the 10 Consensus Cases

	Non-respiratory (n = 10)	Respiratory (n = 17)	P value
Underestimate	25/100 (25%)	25/170 (14.7%)	0.036
Overestimate	13/100 (13%)	17/170 (10%)	0.449
Raw agreement	62/100 (62%)	128/170 (75.3%)	0.021

痰細胞診の診断一致性を初めて多角的に検討し得た。

Fleissのカッパ係数の検討から、C+D判定とE判定(悪性判定)の診断一致性が、CDE判定区分に対する診断一致性や、C判定とD+E判定(要精査判定)の診断一致性より高いことを示した。6施設の判定区分別の分布をみても、E判定数に大きな差はない一方、D判定数(従ってC判定数)には施設間差が認められた。このことは、喀痰細胞診においてはD判定とC判定の鑑別は容易ではないが、E判定とD判定の鑑別はそれほど困難ではないことを示唆している。そして、第二部の検討では、喀痰細胞診に対する経験値と診断一致性の検討を独立して2度行った。2度の検討ともに、喀痰細胞診に対して経験豊かなグループの診断一致性が高いこと、経験の乏しいグループでは過小評価する傾向があることを示した。いずれも統計学的有意差を認めたことから、このことは信頼性の高い結論と考えられる。我が国の肺がん検診ではD判定以上を要精査としているので、より多くの症例がD判定ではなく、過小評価されてC判定とされる場合、精査されずに肺癌発見に至らず、ひいては肺癌発見率の低下に帰着することが危惧される。

喀痰細胞診の検査者間一致性に関する既報告について述べる。検索した限りにおいては、海外の既報告は数編に限られ、いずれも2~30年以前の報告で、本研究のように様々なカッパ統計量を利用した解析は全く認められなかった。^{6,8} Holidayら⁸の報告は、比較的大規模で、喀痰細胞診105症例を11名の細胞検査士によって判定し、その一致率を検討している。彼らの用いた判定基準は6段階と多く、また粗一致率を多用しているため本研究との直接的な比較は困難であるが、彼らは粗一致率を13~60%と報告し、また、1基準違いまで、あるいは2基準違いまでを一致として取り扱った場合の粗一致率を、それぞれ47~92%、83~100%と報告した。一致率を高めるという観点からは判定基準数6段階は多いと指摘している。ちなみに、本研究第一部における粗一致率は61~74%であり、1基準違いまでを一致として扱った場合の粗一致率は95~99%であった。Browmanら⁷は、喀痰細胞診300症例を2名の熟練の細胞検査士によって、判定基準6段階で判定しカッパ値0.58と報告している。判定

者は2名に過ぎないが、カッパ値は本研究結果と類似していた。さて、本邦からは、約20年前の報告になるが、厚生省がん研究助成金研究として肺がん検診における喀痰細胞診に関するアンケート調査結果が報告されている。^{9,10} 全国27施設からの回答によって、喀痰細胞診の要精査比率および癌発見率の施設間のばらつきを指摘するとともに、呼吸器専任の細胞検査士のいる施設の癌発見率10万対146に対して、呼吸器専任細胞検査士のいない施設のそれが80と低値を示したと報告している。また、喀痰細胞診11症例の写真判定を全国25施設にて行い、原判定との判定粗一致率を検討している。細胞異型の高度な場合の判定一致性は高いが、細胞異型の軽度な早期癌症例や異型扁平上皮細胞の判定はばらつき、全体の粗一致率は65%、E判定4症例の粗一致率は83%、D判定5症例では61%、C判定2症例では42%であった。この11症例は、本研究で用いた標準細胞とは異なり、単一施設の喀痰細胞診標本であり、他施設の標本の染色性への慣れの問題は指摘されうるが、E判定に比しCD判定の一致性が低いことは本研究結果と同様であった上に、喀痰細胞診の専門性の低い施設の肺癌発見率が低値であったことは、本研究において専門性の低い判定者が過小評価する傾向を示した結果と符合する。このことは、この20年間の喀痰細胞診の精度管理が結果として一向に進歩していないのではないかと、今後もこのままの精度管理で良いのか、という問題提起であるとも受け止められる。

本研究では6施設を喀痰細胞診の経験豊富な代表的な施設と考え、6施設診断一致例を標準細胞症例としたが、これら6施設が真に代表的な施設か否かは不明である。6施設は、前述の通り、単に2学会合同委員会の委員の所属施設であるに過ぎない。本邦の肺がん検診では、スクリーニング方法や体制などの技術・体制的指標の評価指導や、都道府県単位での肺癌発見率の公開が行われ始めているが、検診実施機関ごとの肺癌発見率の評価指導は行われていない。そのため真に代表的な施設か否かの正確な評価が困難なのが現状である。しかしながら、いずれの施設も、肺がん検診の喀痰細胞診に20年以上従事し、多数の喀痰細胞診標本を診断し、肺癌発見してきたことから、代表的な施設としての資格を有していると考えられる。幸い、健康増進法においては、高危険群に対して実施した喀痰細胞診による肺がん検診の結果報告様式(地域保健・健康増進事業報告)を平成21年度より改訂し、高危険群に対する発見率を算出可能としているので、検診実施機関ごとの発見率についても比較可能と考えられ、これらの施設がどのようなレベルにあるかの検証は今後の課題としたい。

さて、細胞診経験年数と診断一致性の検討において、

カットオフ値を10年としたが、その妥当性は不明である。カットオフ値を15年や20年とすると、各群の検査者数が極めてばらつき統計解析困難となるため、10年は統計解析面から現実的な最長のカットオフ値であった。最後に本研究では、原判定としてB判定の標本は含まなかった。診断的側面からはB判定とC判定の鑑別は重要であるが、肺がん検診においてはいずれの判定でも精査不要となる。また、現状においても、B判定とD判定あるいはE判定の鑑別は容易になされていると確信し、肺がん検診においては、要精査か精査不要の分岐点であるC判定とD判定の鑑別がとりわけ重要と考え、本研究ではC判定以上の標本を扱った。

まとめ

本研究では、標準細胞症例を用いて喀痰細胞診の診断能を正確に評価する手法を構築した。その結果、喀痰細胞診の経験豊富な判定者間の診断一致性は極めて高かった。また、C判定とD判定の正確な鑑別に難渋することも判明した。さらに、喀痰細胞診の経験に乏しい場合、経験豊富な判定者に比較して診断一致性が劣り、また過小評価する傾向があることが判明した。検査者間の診断一致性は精度管理の根幹であり、約20年前の厚生省研究報告も考慮すると、現状の検診体制の継続では診断一致性の向上は期待できない。

そこで、以下を本委員会報告の提言とする。

(1) 今回構築した標準細胞症例を利用して、診断精度の評価および精度向上を目指した研修の評価といった教育面への貢献を進めるべきである。

(2) さらに、検診実施施設の診断精度が肺癌発見率にどの程度反映されているかを検証し、精度管理を進めていくために、スクリーニング方法や体制などの技術・体制的指標に加え、肺癌発見率も検診実施施設ごとに調査、公開していくべきである。

(3) 一方では、冒頭で述べたように、喀痰細胞診のターゲットである肺門部早期扁平上皮癌の発見数は、調査当時の診断精度における推定では、全国で年間270例程度であるという現状を鑑みると、肺門部早期扁平上皮癌の喀痰細胞診の経験値を上げることは容易ではない。肺門部早期扁平上皮癌を希少癌と捉えれば、その喀痰細胞診による発見率の向上は、喀痰細胞診実施施設の均てん化のみでは困難と考えられ、集約化こそが重要だと考えられる。即ち、診断一致性の向上には経験値の向上が重要である以上は、少なくとも呼吸器専門の細胞検査士、細胞診専門医が肺がん検診における喀痰細胞診の判定に携わる体制を整えるべきである。さらに、現状のように各地域に喀痰細胞診実施機関が点在するという状況を改め、実施機関の集約化を図り、実施機関ごとの喀痰細胞

診症例数を増加させ、結果として一層の経験値の向上を図ることが重要と考える。このようなことを通じて、肺門部早期肺癌の全国的な発見率の向上、地域間格差の解消が図られることが期待される。

本論文内容に関連する著者の利益相反：なし

本研究は、日本肺癌学会、日本臨床細胞学会による2学会合同委員会（日本肺癌学会集団検診委員会喀痰細胞診による肺癌検診小委員会、日本臨床細胞学会肺癌検診ワーキンググループ）によって計画、実施された。以下に委員を列記する（順不同）。日本肺癌学会集団検診委員会喀痰細胞診による肺癌検診小委員会：佐藤雅美、柴光年、齋藤泰紀、佐藤之俊、渋谷潔、土田敬明、中山富雄、宝来威、遠藤千顕。日本臨床細胞学会肺癌検診ワーキンググループ：池田徳彦、河原栄、三浦弘之、中嶋隆太郎、田口明美、下川幸弘、神尾淳子、長尾緑、島垣二佳子、田中良太、三宅真司、矢羽田一信。

なお、本研究の概要は学術誌にて公開されている。¹¹

REFERENCES

1. 佐藤雅美, 齋藤泰紀, 渋谷 潔, 中山富雄, 平野 隆, 近藤 丘, 他. 日本肺癌学会・日本呼吸器内視鏡学会・日本臨床細胞学会・3学会合同委員会報告：肺門部早期肺癌実態調査アンケート報告. *肺癌*. 2011;51:777-786.
2. Papanicolaou GN. *Atlas of exfoliative cytology*. Cambridge: Harvard University Press; 1954.
3. Mandrekar JN. Measures of interrater agreement. *J Thorac Oncol*. 2011;6:6-7.
4. Fleiss JL. Measuring nominal scale agreement among many raters. *Psychological Bulletin*. 1971;76:378-382.
5. Turner JK, Williams GT, Morgan M, Wright M, Dolwani S. Interobserver agreement in the reporting of colorectal polyp pathology among bowel cancer screening pathologists in Wales. *Histopathology*. 2013;62:916-924.
6. Evans DM, Shelley G. Respiratory cytodiagnosis: study in observer variation and its relation to quality of material. *Thorax*. 1982;37:259-263.
7. Browman GP, Arnold A, Levine MN, D'Souza T, Turner-Smith LM, Cayco R, et al. Use of screening phase data to evaluate observer variation of sputum cytodiagnosis as an outcome measure in a chemoprevention trial. *Cancer Res*. 1990;50:1216-1219.
8. Holiday DB, McLarty JW, Farley ML, Mabry LC, Cozens D, Roby T, et al. Sputum cytology within and across laboratories. A reliability study. *Acta Cytol*. 1995;39:195-206.
9. 佐藤雅美, 齋藤泰紀, 鈴木隆一郎, 関本 信, 岩波 洋, 高島 力, 他. 喀痰細胞診を用いた肺癌検診の精度管理の成績. *日本臨床細胞学会雑誌*. 1997;36:568-575.
10. 佐藤雅美, 齋藤泰紀, 鈴木隆一郎, 関本 信, 岩波 洋, 高島 力, 他. 肺癌検診喀痰細胞診判定基準に関する検討. *日本臨床細胞学会雑誌*. 1997;36:490-497.
11. Endo C, Nakashima R, Taguchi A, Yahata K, Kawahara E, Shimagaki N, et al. Inter-rater agreement of sputum cytology for lung cancer screening in Japan. *Diagn Cytopathol*. 2015;43:545-550.