

<p><b>Lung Cancer, 48: 331-337</b> Cytological characteristics of pulmonary large cell neuroendocrine carcinoma.</p>	<p>2005</p>	<p>Hiroshi K, Abe S, Ebihara Y, Ogura S, Kikui M, Kodama T, Komatsu H, Saito Y, Sagawa M, <u>Sato M</u>, Tagawa Y, Nakamura S, Nakayama T, Baba M, Hanzawa S, Nirano T, Horai T</p>
<p><b>胸部外科 58(5):371-377</b> EBM に基づく高齢者肺癌外科治療ガイドライン作成の試み</p>	<p>2005</p>	<p><u>佐藤雅美</u>, 庄司文裕, 一瀬幸人, 山下眞一, 芳賀克夫,</p>
<p><b>成人病と生活習慣病,35(12):1337-1339</b> EBM の手法による肺癌診療のガイドライン 2003 年度版</p>	<p>2005</p>	<p><u>佐藤雅美</u>, 藤村重文</p>

# 肺癌検診の現状と 将来展望

大阪府立成人病センター研究所 鈴木隆一郎

## KEY WORDS

- 肺癌
- 検診
- 有効性
- らせんCT

Lung cancer screening  
and the future  
Takaichiro Suzuki (特別研究員)

## はじめに

わが国における肺癌検診の現状を知るために、1972年に厚生省がん研究助成金「肺門部早期がんの診断体系の確立と診断法の開発に関する研究」班(主任研究者 池田茂人)、通称「肺癌集団検診研究班(池田班)」が組織され、先駆的な肺癌検診を開始して以来、現時点に至るまでの肺癌検診研究の経過と、それに対する厚生行政の対応とを概観する。

また、現時点におけるわが国の肺癌検診の数量的な側面は、厚生労働省統計情報部が毎年度公表している「地域保健・老人保健事業報告」により観察する。

なお、老人保健事業としての肺癌検診の有効性評価を確立した「肺癌集検研究班(第二次藤村班)」が終了した時点(2000年)に立ち戻って、その後現時点までの肺癌検診研究の展開を眺め、その彼方に将来を展望してみたい。

## I. わが国における 肺癌検診研究と 行政の対応

わが国における肺癌検診研究と行政の対応を年表風に表1に示した。この表では、「肺癌検診研究と日本肺癌学会・胸部CT検診研究会の対応」と「厚生行政の対応」の2つの流れを示した。前項に述べたように、わが国における肺癌検診研究の支柱をなしてきた研究班は長い正式名称を有しているが、ここでは最も短い略称を使用し、「池田班」、「成毛班」、「金子班」、「藤村班」、「久道班」、「鈴木班」と表記している(各主任研究者は、池田茂人、成毛詔夫、金子昌弘、藤村重文、久道茂、鈴木隆一郎)。

1972年に始まった第一次池田班から1986年に終了した第四次池田班までに、肺癌検診は、胸部X線検査で早期発見が期待される肺野末梢部早期肺癌と、高危険群に喀痰細胞診を行うことで早期発見が期待される肺門部早期肺癌の

2種類があることを基本に早期発見をめざすべきことを確立し<sup>1)</sup>、かつ、各地で市町村に働きかけて、結核検診を活用して肺癌検診を実施する実践的な活動が行われてきた。この状況を受けて、日本肺癌学会は「肺癌取扱い規約」のなかに「肺癌集団検診の手びき」<sup>2)</sup>を新設し、また、厚生省(当時)は1987年度からの第二次保健事業のなかに「胸部X線検査と高危険群に対する喀痰細胞診の併用」方法による肺癌検診を取り入れた。

時を同じくして研究を開始した第二次成毛班は、研究課題を肺癌検診の有効性評価に切り換えた。ここで有効性とは、肺癌検診を実施することによって肺癌死亡が減少するか否かという問題である。いまだ実施されていない新しい検診の場合は実験的な研究方法である無作為化比較試験によることが可能であるのに対して、わが国における肺癌検診のように、(その前駆的な結核検診まで含めると)すでに長年にわたって全国的に広く実施されてきた検診の有効性評価は、観察的な研究方法である症例対照研究を複数実施して、いずれにおいても有効であることを確かめねばならない。そこで表1に示したように、1992年に論文が掲載された「成毛班の症例対照研究」から、2002年に論文掲載の「藤村班の『群馬』症例対照研究」まで、合計6個の症例対照研究<sup>3)~8)</sup>が実施され、すべてにおいて肺癌検診による肺癌死亡減少効果が示された。

この研究成果を受けて、2001年の「新たながん検診の有効性評価研究班」報告書<sup>9)</sup>では、「わが国における現行の肺癌検診(胸部X線検査と高危険群に対する喀痰細胞診の併用)」について、

表1. わが国における肺癌検診の研究と行政の対応

肺癌検診研究と日本肺癌学会・胸部CT検診研究会の対応	
1972年	がん研究助成金「肺癌集団検診研究班」(池田班)
1975年	がん研究助成金 第二次池田班
1978年	がん研究助成金 成毛班
1981年	がん研究助成金 第三次池田班
1984年	がん研究助成金 第四次池田班
1987年	がん研究助成金 第二次成毛班 日本肺癌学会「肺癌集団検診の手びき」 <sup>2)</sup> 公表
1990年	がん研究助成金 第三次成毛班
1992年	成毛班の症例対照研究 <sup>3)</sup> 論文掲載
1993年	がん研究助成金 金子班
1995年	がん研究助成金 第二次金子班 日本肺癌学会「肺癌集団検診の手びき」改訂
1997年	がん研究助成金 藤村班
1998年	「がん検診の有効性評価研究班」(久道班)報告書
1999年	がん研究助成金 第二次藤村班 「肺がんCT検診の有効性評価研究計画班」(鈴木班) 金子班の症例対照研究 <sup>4)</sup> 論文掲載 日本肺癌学会「肺癌集団検診の手びき」改訂
2001年	「肺がんCT検診の有効性評価研究班」(鈴木班) 藤村班の「宮城」症例対照研究 <sup>5)</sup> 論文掲載 藤村班の「新潟」症例対照研究 <sup>6)</sup> 論文掲載 藤村班の「岡山」症例対照研究 <sup>7)</sup> 論文掲載 「新たながん検診の有効性評価研究班」(第三次久道班)報告書 <sup>9)</sup>
2002年	藤村班の「群馬」症例対照研究 <sup>8)</sup> 論文掲載
2003年	日本肺癌学会「肺癌集団検診の手びき」改訂
2004年	日本肺癌学会「低線量CTによる肺癌検診の手引き」出版 胸部CT検診研究会「胸部検診用CT撮影マニュアル」公表 胸部CT検診研究会「CT検診精度管理ガイドライン」公表

「肺がんの生存率は一般に極めて低い。しかし、わが国における現行の肺癌検診は、適切に行うならば、死亡率減少に寄与する可能性が高く、継続して実施する相応の根拠がある」と結論し、有効性を認めている。

## II. 老人保健法による肺癌検診

老人保健事業の各癌検診受診者総数のみは、毎年度の「国民衛生の動向」で、年次推移として、簡単にみることで

きる。第二次保健事業初年度の1987年度の検診受診者数でみると、胃癌検診の363万人に対して肺癌検診は266万人であったが、翌1988年度には388万人が肺癌検診を受診して、胃癌検診の373万人を凌駕した<sup>10)</sup>。このように肺癌検診受診者数が急増した理由は、第二次計画の開始にあたり、厚生省が結核難病感染症課長と老人保健課長の連名で「保健事業の肺癌検診に利用する胸部エックス線写真について」という通知<sup>11)</sup>を出して、結核検診で撮影された胸部X線写真を利用して肺癌検診を行

厚生行政の対応

- 1980年 厚生省「がん予防対策打合せ会」報告書
- 1982年 老人保健法施行
- 1983年 第一次保健事業開始(胃がん検診, 子宮がん検診 開始)  
「保健事業実施要領」制定
- 1987年 第二次保健事業開始(肺がん検診, 乳がん検診 開始)  
「保健事業実施要領」改訂  
「老人保健法による健康診査マニュアル」出版
- 1991年 老人保健法とがん検診に関するシンポジウム
- 1992年 第三次保健事業開始(大腸がん検診 開始)  
「保健事業実施要領」改訂  
「老人保健法による肺がん検診マニュアル」出版
- 1994年 「老人保健法による健康診査マニュアル」改訂1版 出版
- 1996年 「保健事業実施要領」改訂
  
- 1997年 「老人保健法による健康診査マニュアル」改訂2版 出版
- 1998年 がん検診費を国庫負担金から外し, 地方交付税で措置  
「保健事業実施要領」改訂

うようにとの方針を示し、結核検診を土台に肺癌検診を展開する方向を打ち出したことによると思われる。1992年度からは、性別・年齢5歳階級別・検診結果区分別・都道府県別の実績が、「厚生省統計情報部編：老人保健事業報告・厚生統計協会」として、毎年度市販されるようになった。官庁再編成後も、若干名称を変えて継続されている。その資料から、肺癌検診受診者総数と検診結果区分別数について、「男女計・全年齢・全国値」を抜粋し、表2にまとめた。この表は、わが国にお

ける肺癌検診の現状をよく表している。

まず受診者数の下に記した「対前年」増減をみると、毎年3%程度着実に増加していたが、1998年度にわずかではあるが減少したことが読み取れる。これは表1をみれば、「1998年がん検診費を国庫負担金から外し、地方交付税で措置」(一般財源化)に対応して発生した現象であることがわかる。すなわち、国は「補助金行政」を批判する地方分権推進会議の勧告に従って、がん検診費を、実施件数に対して市町村へ補助する方法から、原則的には市町村が

自由に執行できる地方交付税の積算基礎へ算入する方法に移したのであったが、該当金額が直接みえないために市町村側に検診継続についてある程度の混乱が生じたものと推測される。しかしながら、すぐその翌年から受診者数が再び着実に増加していることをみると、市町村は肺癌死亡の急増に対する住民の要望に対応していると考えられる。

要精検率は2%台に留まっており、精検システムにかかる負担は不当に大きなものではない。

一般財源化以降、少々気になることは「要精密検査者」中の「未把握・未受診」割合の漸増である。この補数が精検受診率の最小見積もりであり、2002年度では74.7%と計算される。「癌であった者」の割合、すなわち肺癌発見率の最小見積もりは0.047%であるが、これを「要精密検査者」中に占める割合、すなわち陽性反応の中率としてみると1.7%となっており、「要精密検査」といわれるほどの高危険要因は他にない。精検受診率を可能な限り100%へ近づけることが精度管理の第一歩である。一般財源化によって、市町村が精検受診を奨励し、その結果を把握する力に弛みが生じているとすれば、わが国の肺癌二次予防策の中核をなす保健事業の肺癌検診にとって重大な問題である。

### Ⅲ. 肺癌検診の将来展望 -むすびにかえて-

表1に戻ってみると、行政の対応は、1998年の一般財源化で留まっているのに対して、肺癌検診研究は、なお成長していることがわかるであろう。すなわち、保健事業型の肺癌検診の有効性

表2. 老人保健法に基づく肺がん検診の成績

年度	受診者数 [対前年]	要精密 検査者	異常 認めず	がんで あった者	がんの 疑いの ある者	がん以外 の疾患で あった者	未把握・ 未受診*3
1992年度*1	5,869,993 (100.0) [ - ]	124,460 (2.1) (100.0)	60,618 (48.7) (100.0)	2,575 (0.044) (2.1)	1,487 (1.2) (1.2)	33,195 (26.7) (26.7)	26,585 (21.4) (21.4)
1993年度*1	6,298,527 (100.0) [7.3%]	143,228 (2.3) (100.0)	67,724 (47.3) (100.0)	2,726 (0.043) (1.9)	1,418 (1.0) (1.0)	40,039 (28.0) (28.0)	31,326 (21.9) (21.9)
1994年度*1	6,503,415 (100.0) [3.3%]	159,998 (2.5) (100.0)	76,840 (48.0) (100.0)	2,869 (0.044) (1.8)	1,640 (1.0) (1.0)	42,326 (26.5) (26.5)	36,323 (22.7) (22.7)
1995年度*1	6,703,359 (100.0) [3.1%]	167,255 (2.5) (100.0)	79,916 (47.8) (100.0)	3,144 (0.047) (1.9)	2,082 (1.2) (1.2)	45,778 (27.4) (27.4)	36,335 (21.7) (21.7)
1996年度*1	6,926,451 (100.0) [3.3%]	179,671 (2.6) (100.0)	78,006 (43.4) (100.0)	3,134 (0.045) (1.7)	2,325 (1.3) (1.3)	57,244 (31.9) (31.9)	38,962 (21.7) (21.7)
1997年度*1	7,061,535 (100.0) [2.0%]	172,561 (2.4) (100.0)	76,854 (44.5) (100.0)	3,436 (0.049) (2.0)	2,375 (1.4) (1.4)	51,967 (30.1) (30.1)	37,929 (22.0) (22.0)
1998年度*1	7,030,639 (100.0) [-0.4%]	180,527 (2.7) (100.0)	77,512 (42.9) (100.0)	3,464 (0.049) (1.9)	2,653 (1.5) (1.5)	58,684 (32.5) (32.5)	38,214 (21.2) (21.2)
1999年度*2	7,127,240 (100.0) [1.4%]	189,302 (2.7) (100.0)	79,387 (41.9) (100.0)	3,582 (0.050) (1.9)	2,744 (1.4) (1.4)	63,475 (33.5) (33.5)	40,114 (21.2) (21.2)
2000年度*2	7,257,718 (100.0) [2.0%]	191,200 (2.6) (100.0)	76,723 (40.1) (100.0)	3,611 (0.050) (1.9)	3,151 (1.6) (1.6)	64,692 (33.8) (33.8)	43,023 (22.5) (22.5)
2001年度*2	7,412,212 (100.0) [2.0%]	198,725 (2.7) (100.0)	76,972 (38.7) (100.0)	3,483 (0.047) (1.8)	3,424 (1.7) (1.7)	68,018 (34.2) (34.2)	46,828 (23.6) (23.6)
2002年度*2	7,490,412 (100.0) [1.1%]	207,830 (2.8) (100.0)	76,137 (36.6) (100.0)	3,557 (0.047) (1.7)	3,928 (1.9) (1.9)	71,544 (34.4) (34.4)	52,664 (25.3) (25.3)

[男女計, 単位は人, ( )内は%]

\*1厚生省統計情報部編:平成4年度老人保健事業報告.厚生統計協会より,  
同:平成10年度老人保健事業報告 までの各年度の「肺がん検診」より抜粋引用。

\*2厚生労働省統計情報部編:平成11年度地域保健・老人保健事業報告(老人保健編).厚生統計協会より,同:平成14年度地域保健・老人保健事業報告(老人保健編) までの各年度の「肺がん検診」より抜粋引用。

\*31998年度まで「未把握」と表章,1999年度から「未把握」,「未受診」に分けて表章されている。  
ここでは「要精密検査者」から「異常を認めず」~「がん以外の疾患であった者」を減じた残差を計上した。

が確立し、その論文を国際学術誌へ掲載する努力中に、一方では、肺癌CT検診の有効性評価研究が開始された。これは1987年に発売された「らせんCT装置」が高速低線量化し、1995年にはこれを搭載した「低線量らせんCT検診車」が出現するなど、肺癌検診への導入が技術的には可能になったためである。

前述の久道班報告書<sup>9)</sup>では「らせんCTによって全肺スキャンを行えば、より小さな肺がんを発見できることが知られている。しかしながら、この装置が検診の分野に普及し始めたのは、わが国においても海外においても、ごく最近のことであり、肺がん死亡減少効果を測定する研究は、現在、着手されようとするところである。したがって、その結果が判明するまで評価を保留する」と結論しているが、その記述の通り、この報告書と同年に研究を開始した「鈴木班」では、すでに実施済みのCT検診受診者全員を振り返って登録・追跡する「コホート研究」の手法により肺癌CT検診の有効性評価研究を実施中である。この件に加え、これまでに観察されたCT検診の効果については、近著<sup>12)</sup>に解説した。

なお、1998年度の一般財源化は、一面では、検診方法が自由化されたとも考えられるため、CT検診への転換を考える市町村が少なからず出現する状況になっている。しかしながら前述のように、CT検診の有効性自体が現在

研究中であるため、有力な研究グループの協力を得て、研究的な評価に耐えられるように実施できる態勢が不可欠である。

また、最近の「らせんCT装置」の進歩は急速である。すなわち、前述の有効性評価研究の検診はすべてsingle detector CT (SDCT) によるものであるが、1999年8月に4列multi detector CT (MDCT) が発売され、続いて2002年2月に8列MDCT、2002年12月に16列MDCT、2004年12月には32列MDCTが発売された。SDCTによる検診の有効性さえ研究中の現時点で、被曝量の正確な実測すら間に合わないほどの機材の急激な発展に直面している有様である。どのような機材をどのように使用すれば、最大の有効性、最高の安全性、最適な経済性が得られるのか、現在、CT検診を実施中の組織、CT検診を開始しようとしている組織の関係者は、これらの疑問への回答に対して、なんらかの貢献を志して取り組んでいただきたい。

## 文 献

- 1) 成毛昭夫：早期肺癌の分析。池田茂人編、肺癌集検の実際。東京、医学書院、49-56, 1986
- 2) 日本肺癌学会集団検診委員会：肺癌集団検診の手びき。日本肺癌学会編、肺癌取扱い規約 改訂第3版。東京、金原出版、139-156, 1987

- 3) Sobue T, Suzuki T, Naruke T, et al : A case control study for evaluating lung-cancer screening in Japan. *Int J Cancer* 50 : 230-237, 1992
- 4) Okamoto N, Suzuki T, Hasegawa H, et al : Evaluation of a clinic-based screening program for lung cancer with a case-control design in Kanagawa, Japan. *Lung Cancer* 25 : 77-85, 1999
- 5) Sagawa M, Tsubono Y, Saito Y, et al : A case-control study for the efficacy of mass screening program for lung cancer in Miyagi Prefecture, Japan. *Cancer* 92 : 588-594, 2001
- 6) Tsukada H, Kurita Y, Yokoyama A, et al : An evaluation of screening for lung cancer in Niigata Prefecture, Japan : a population-based case-control study. *Br J Cancer* 85 : 1326-1331, 2001
- 7) Nishii K, Ueoka H, Kiura K, et al : A case-control study of lung cancer screening in Okayama Prefecture, Japan. *Lung Cancer* 34 : 325-332, 2001
- 8) Nakayama T, Baba T, Suzuki T, et al : An evaluation of chest X-ray screening for lung cancer in Gunma prefecture, Japan : a population-based case-control study. *Eur J Cancer* 38 : 1380-1387, 2002
- 9) 鈴木隆一郎：肺がん検診。久道 茂編、新たながん検診手法の有効性の評価報告書。東京、日本公衆衛生協会、277-304, 2001
- 10) 厚生統計協会編：国民衛生の動向。厚生指標 41 : 122, 1994
- 11) 厚生省保健医療局老人保健部老人保健課：老人保健法による健康診査マニュアル。東京、日本公衆衛生協会、187, 1987
- 12) 鈴木隆一郎：CT検診の意義と現状。呼吸器科 6 : 265-269, 2004

各種がん検診から学ぶ精度管理—肺癌

---

中山富雄<sup>1</sup>・楠 洋子<sup>1</sup>・鈴木隆一郎<sup>1</sup>

<sup>1</sup>大阪府立成人病センター調査部疫学課

*Japanese Journal of Lung Cancer*

肺癌 第45巻 第2号 2005年4月

## 各種がん検診から学ぶ精度管理—肺癌

中山富雄<sup>1</sup>・楠 洋子<sup>1</sup>・鈴木隆一郎<sup>1</sup>

**要旨** — 我が国の市町村で行われている肺癌検診は、年間 700 万人以上が受診しているが、肺癌発見数は 3600 人弱 (発見率 51/10 万人) で、受診者の増加に伴っていない。都道府県別にみると発見率は最大 111.5、最小 18.5 でその差は最大 6.2 倍とバラツキは大きかった。市町村別に見ると、精検受診率や精検完了率が 30% を下回る市町村も認められ、受診者数 5000 人以上の市町村の 6.3% で過去 2 年間肺癌は 1 例も発見されていなかった。肺癌検診は外的精度管理の枠組みがないまま、急速に普及していった。市町村事業と位置づけられてからは、実施数や費用のみが問題とされており、精度については検討されておらず、受診者にも精度に関する情報は提供されていない。今後は対象者や精度管理指標の基準値を盛り込んだ運用指針の見直しと、第 3 者機関による検診成績の情報公開が必要である。(肺癌, 2005;45:183-187)

**索引用語** — 精度管理, 肺癌, 検診

## Quality Control in Lung Cancer Screening in Japan

Tomio Nakayama<sup>1</sup>; Yoko Kusunoki<sup>1</sup>; Takaichiro Suzuki<sup>1</sup>

**ABSTRACT** — The number of people screened for lung cancer has been increasing and now exceeds 7000000 per year in Japan, but the number of detected lung cancer is only about 3600 cases (detection rate; 0.051%) and has not been increasing. When detection rates were compared according to prefectures, in the detection rate varied by a factor of as much of 6.2 as large. In some municipalities, the rate of compliance and completion of diagnostic work-up fell below 30%. None of the lung cancer cases had been detected in the past two years in 6.3% of municipalities with 5000 or more screened people. Lung cancer screening has spread rapidly, though there is no framework for external quality control in Japan. Therefore, only the number of screening participants and the cost of screening are managed, the level of accuracy of the screening is not controlled in each municipalities offering the screening program. Thus screening participants are not offered a uniformly accurate of screening program. It is necessary to innovate guidelines including a quality control index, and to disclose information of the screening result by a central coordinative organization. (JLCC. 2005;45:183-187)

**KEY WORDS** — Quality control, Lung cancer, Screening

我が国では、住民を対象とした肺癌検診は、すでに普及していた結核検診で撮影されたフィルムを利用し、高危険群に対して喀痰細胞診を併用する形で、昭和 58 年よ

り開始された。喀痰細胞診については、日本臨床細胞学会からの資格認定を受けた細胞検査士と指導医が従事することが義務づけられた。しかし X 線撮影と読影に関し

<sup>1</sup>大阪府立成人病センター調査部疫学課。

別刷請求先：中山富雄，大阪府立成人病センター調査部疫学課，〒537-8511 大阪府大阪市東成区中道 1-3-3 (e-mail: nakayama-to@mc.pref.osaka.jp)。

<sup>1</sup>Division of Epidemiology, Department of Cancer Control and Statistics, Osaka Medical Center for Cancer and Cardiovascular Diseases, Japan.

Reprints: Tomio Nakayama, Division of Epidemiology, Department of Cancer Control and Statistics, Osaka Medical Center for Cancer and Cardiovascular Diseases, 3-3 Nakamichi 1-chome, Higashinari, Osaka 537-8511, Japan (e-mail: nakayama-to@mc.pref.osaka.jp).

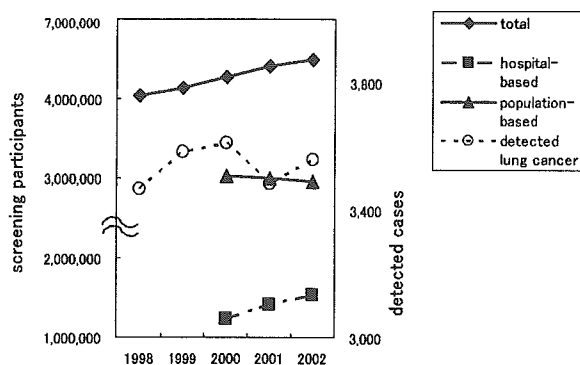
© 2005 The Japan Lung Cancer Society



**Table 1.** Quality Control Index of Lung Cancer Screening by a Screening System

	Total	Hospital-based	Population-based
Positive rate	2.7%	3.2%	2.6%
The rate of compliance of diagnostic work-up	87.0%	81.9%	88.5%
The rate of completion of diagnostic work-up	76.4%	61.5%	80.7%
Detection rates *	48.0	51.6	47.2

\*; per 100000 population.

**Figure 1.** Changes in the numbers of participants in lung cancer screening and detected lung cancer.

ては、資格認定がなされないまま、全国的に普及し、現在にいたっている。

現状での肺癌検診の精度管理は、診療放射線技師や細胞検査士による内的精度管理の色彩が濃く、またその努力が対外的には評価されていない。一方外的精度管理の枠組みはなく、実施する市町村や、受診者に相当する一般の人たちにも、自らの身を守るための検診の精度を容易に知るすべはない。

今回、全国の市町村の検診実施成績が集計されている地域保健・老人保健事業報告を用いて、肺癌検診の精度の評価を試みた。

## 方法

平成10～14年の地域保健・老人保健事業報告を用い、全国・都道府県別・市町村別に集計を行った。老人保健事業報告は、厚生労働省統計情報部の統計表データベースシステム ([http://www.dbtk.mhlw.go.jp/IPPAN/ipcart/sko\\_K\\_Nyuuryoku](http://www.dbtk.mhlw.go.jp/IPPAN/ipcart/sko_K_Nyuuryoku)) よりダウンロードして、使用した。集団方式と個別方式の分離集計は、平成12年度から掲載されているため、過去3年分を利用した。都道府県別の発見率の比較には5年間の平均値を用いた。市町村別の集計は13年度から掲載されているので、過去2年分を使用した。市町村合併がこの2年間の間に行われているものは、合併後の市町村単位で集計した。

## 結果

### 1. 全国集計での解析

肺癌検診の全受診者数は約700万人で、平成10～14年の5年間にかけても、約50万人の増加を認めているが、発見肺癌数は変動はあるものの、受診者の増加に伴っていない (Figure 1)。検診方式別に平成12年からの3年間の受診者数を比較すると集団方式はほぼ一定であったが、個別検診の増加が著しい。この3年間の平均の精度管理指標を見ると、要精検率2.7%、精検受診率87.0%、完了率76.4%、肺癌発見率は10万人あたり48.0人であった (Table 1)。個別方式では、発見率が51.6と若干集団方式よりも高いが、精検受診率が81.9%と低く、精検完了率は61.5%にすぎなかった。

性・年齢階級別に見ると、40歳代の女性受診者が約70万人存在するが、発見数は50人に満たなかった (Figure 2)。50歳代では受診者の4分の3が女性であった。60歳代から男性受診者数が増加し、発見数も男性が女性を上回った。要精検率は、胸部X線では男女とも年齢とともに増加し、発見率も増加した。一方喀痰細胞診の要精検率は年齢に全く関係なく、男性では肺癌の発見率が65歳以上で急増したが、女性では全年齢を通じてほとんど喀痰からは発見されていない。

### 2. 都道府県単位の解析

都道府県別の過去5年間の受診者数の推移を見ると、東日本では千葉・埼玉など東京のベッドタウンでの受診者数の急増が著しいが、西日本では漸減している府県も見られた。肺癌発見率を都道府県別に比較すると、最低で10万対20を切る府県と、最高で100を越える府県があり、その格差は6.2倍であった (Figure 3)。

個別方式と集団方式との間で、発見率を比較すると、集団方式の発見率は、各府県間であまり差がなかったが、個別方式では格差が極めて大きかった。

また要精検率とがん発見率を比較すると、個別方式では相関があり、集団方式では相関は全くなかった。精検受診率は個別方式で大きくバラツキがあり、集団方式ではほぼ全国一定であった。

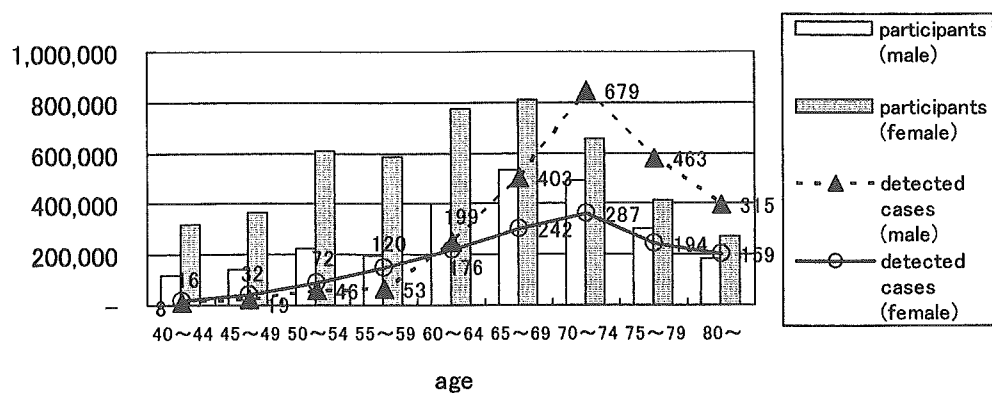


Figure 2. The number of screening participants and detected lung cancer by gender and age group.

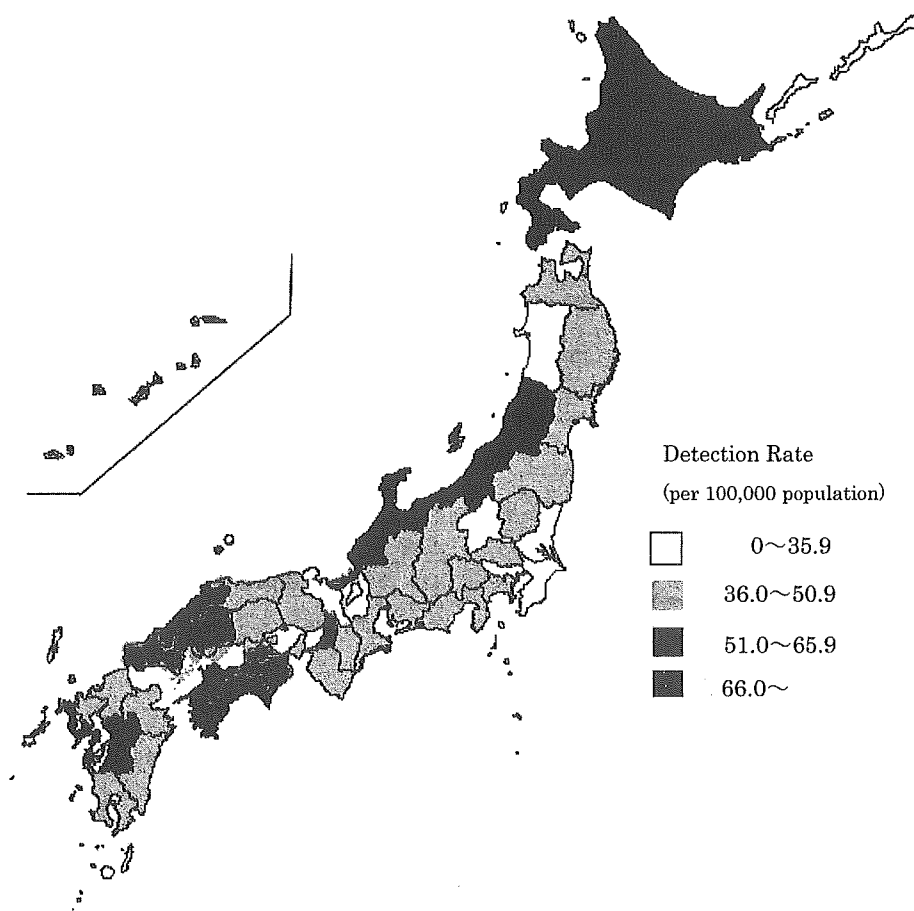


Figure 3. Detection rate of lung cancer by prefectures.

### 3. 市町村別集計

胸部 X 線の要精検率が 0.5% を下回るあるいは、20% を上回る市町村が併せて 6.4% 存在した (Table 2)。精検受診率はおおむね 80% を上回っていたが、30% を下回

る市町村も存在し、10% を下回る市町村もわずかながら存在した。精検完了率は 30% を下回る市町村が 60 市町村以上もあった。年間の平均受診者数を 1000~5000 人未満と 5000 人以上に分けて比較すると、5000 人未満では

**Table 2.** Quality Control Index of Lung Cancer Screening by Municipalities

Positive rate	(%)	The rate of compliance of diagnostic work-up	(%)	The rate of completion of diagnostic work-up	(%)
0- 0.4	173	0- 9.9	7	0- 9.9	27
0.5- 0.9	219	10.0-29.9	7	10.0-29.9	34
1.0- 4.9	2201	30.0-49.9	38	30.0-49.9	115
5.0- 9.9	369	50.0-79.9	534	50.0-79.9	1026
10.0-19.9	60	80.0-99.9	1767	80.0-99.9	1541
20.0-	20	100.0	648	100.0	258

**Table 3.** Detection Rate of Lung Cancer Cases by the Number of Participants of Lung Cancer Screening in Each Municipalities

Detection rate *	Participants (1000-4999)	(%)	Participants (5000-)	(%)
0	392	25.3	18	6.3
1.0-29.9	214	13.8	78	27.2
30.0-49.9	353	22.8	77	26.8
50.0-69.9	167	10.8	57	19.9
70.0-99.9	216	13.9	41	14.3
100-	207	13.3	16	5.6

\*, per 100000 population.

**Table 4.** The Rate of Sputum Cytology Screening Participants in Chest X-ray Screening Participants in Each Municipalities

	Male	(%)	Female	(%)
0- 4.9	680	22.4	2669	88.2
5.0- 9.9	684	22.6	140	4.6
10.0-14.9	553	18.3	65	2.1
15.0-29.9	798	26.4	45	1.5
30.0-49.9	191	6.3	34	1.1
50.0-99.9	106	3.5	61	2.0
100-	13	0.4	11	0.4

4分の1の市町村、5000人以上でも6.3%の市町村において過去2年間の肺癌発見率は0であった(Table 3)。また発見率100を越える市町村は、受診者数5000未満の市町村の13.3%を占めたが、5000以上の市町村では5.6%にすぎなかった。受診規模の小さい市町村で、発見率の高い精度の高い検診が行われていることが示された。

また喀痰細胞診を、女性受診者の50%以上に行っている市町村が72(2.4%)あった(Table 4)。女性の喫煙率はせいぜい15%にすぎず、この実施率は喀痰細胞診の対象を受診者も市町村も理解していないものと考えられる。

## 考 察

現状の肺癌検診の精度の問題点はあまりにも多い。

他のがん検診に比べて、機材が普及しており、処理能力が高く、安価であることなどから、本検診は急速に普及した。しかし直近5年の成績では受診者は増加を続けながらも肝心の肺癌発見数が伴っていない。すなわち、内容(精度)が吟味されないまま、ただ漫然と事業が拡大されているにすぎない。平成10年以降、がん検診は一般財源化され、市町村の「事業」として位置づけられている。市町村では、緊縮財政と市町村合併の動きの中で、全事業の見直しが行われている。がん検診事業については、予算や実施数という面ではなく、精度という面での見直しが行われるべきであり、精度が確保できない検診に公費をつぎ込んで漫然と継続していくことは大いに問題である。日本肺癌学会編集の「肺癌取扱い規約」の「肺癌集団検診の手引き」には、肺癌検診の対象者が定義されている。しかし喀痰細胞診に関しては、受診者のほぼ全数に行われている市町村がいまだに存在し、学会編集の手引きが守られていない。胸部X線での見落としを喀痰細胞診で拾い上げることを目的としていると善意に解釈することも可能であるが、残念ながらほとんどがんは発見されていない。また対象者として「肺癌取扱い規約」に定義されていても、若年女性の受診者が多いことは、効率性の観点から問題である。40歳代の女性受診者における肺癌発見率は10万人あたり10人を下回っている。「肺癌集団検診の手引き」は、老健法導入時から20年間判定基準をのぞけばほとんど改訂されておらず、現状に即したものとは言い難い。

検診の方式については、受診者の利便性を鑑み、各種がん検診において集団方式から個別方式への転換が図られている。しかし今回の検討で明らかなように、肺癌検診においては、個別検診は精度のバラツキが大きすぎる。専門医療機関と提携し、精度管理委員会を設けて、熱心に個別方式を行っている一部の地区医師会のように、すぐれた精度を報告している地区も見られるが、ほとんど肺癌が発見されていない地区も多く見られる。読影医側の問題ももちろんあるが、医師会や市区町村に集計や精度管理を行う機能がないため、実際は肺癌が発見されていても、集計もれが生じるといった問題も大きい。個別検診に関しては、担当医師会での事務局機能の整備を行っていただきたい。また集団方式に関しては、精度は全国ほぼ一定であったが、要精検率と発見率は相関していなかった。標準化発見比ではないため、正確な評価が難しいが、間接 X 線での限界ではないかと考えられる。

さて、これら肺癌検診に関わる様々な問題を生み出す原因は、肺癌検診を管理する指針や組織がないことに起因する。欧米では乳癌のマンモグラフィ検診や、子宮頸がん検診など有効性が確立された検診に関しては、法律を設け、中央で精度を管理した organized screening が行われ、高い受診率と精度が維持されている。しかし、日本では多種類の検診が市町村の自由裁量で行われ、希望する受診者に野放図に提供されており、結果として低い受診率と低い精度のまま放置されているのが現状である。各府県におかれた成人病管理指導協議会の部会が、その責務を担っているはずであるが、現在では、年 1 回（もしくは数年に 1 回）形式上開かれるだけのものになっている。その報告は、知事に対して行われているのであって、一般の住民に対してはほとんど公開されていない。

現状の肺癌検診の精度に問題があることが、社会問題化されていないことから、協議会を主催する行政側もドラスティックな変化を起こしたくないのであろうが、透明性を確保するという現在の流れには全く即していない。

今後、肺癌検診の精度向上の対策としては、検診を管理するための運用指針を、国として新たに検討し、作成することが現実的な選択であろう。20 年前に定められた現行の指針（肺癌集団検診の手引き）には、問題があることは、今回の検討で明らかであり、また市町村の現場では決して遵守されていないことも事実である。適切な対象者の選択や、精度管理指標の基準値、などを盛り込んだものを、作成することが必要であろう。また情報の徹底した公開を第三者機関で行うべきである。現状では老人保健事業報告という形で、実施成績は報告されており、厚生労働省統計情報部のデータベースからは市町村単位での成績もダウンロードすることは可能ではある。しかしその存在を知らない人が容易に探しだせる状況にはなく、また要精検率や発見率さえ計算されていない単純集計表のままである。このようなものではなく、専門家の解釈・解説付きの集計結果を、一般の住民が容易にアクセスし、理解できるものが必要である。

一般医療や行政サービスのすべての面で、見直しや情報公開が急速に進んでいる。肺癌検診は開始後約 20 年が経過したものの、いまだ検診黎明期と同じ様相で行われている。今後速やかな運営の見直しと情報公開が進むことを期待する。

本研究は、厚生労働科学研究費第 3 次対がん総合戦略研究事業「革新的な診断技術を用いたこれからの肺癌検診手法の確立に関する研究」の支援を受けた。

---

# Evaluation of Automated Lung Nodule Detection on Low-dose Computed Tomography Scans From a Lung Cancer Screening Program<sup>1</sup>

Samuel G. Armato III, PhD, Arunabha S. Roy, PhD, Heber MacMahon, MD, Feng Li, MD, PhD, Kunio Doi, PhD, Shusuke Sone, MD, Michael B. Altman, BA

---

**Rationale and Objectives.** The purpose of this study was to evaluate the performance of a fully automated lung nodule detection method in a large database of low-dose computed tomography (CT) scans from a lung cancer screening program. Because nodules demonstrate a spectrum of radiologic appearances, the performance of the automated method was evaluated on the basis of nodule malignancy status, size, subtlety, and radiographic opacity.

**Materials and Methods.** A database of 393 thick-section (10 mm) low-dose CT scans was collected. Automated lung nodule detection proceeds in two phases: gray-level thresholding for the initial identification of nodule candidates, followed by the application of a rule-based classifier and linear discriminant analysis to distinguish between candidates that correspond to actual lung nodules and candidates that correspond to non-nodules. Free-response receiver operating characteristic analysis was used to evaluate the performance of the method based on a jackknife training/testing approach.

**Results.** An overall nodule detection sensitivity of 70% (330 of 470) was attained with an average of 1.6 false-positive detections per section. At the same false-positive rate, 83% (57 of 69) of the malignant lung nodules in the database were detected. When the method was trained specifically for malignant nodules, a sensitivity of 80% (55 of 69) was attained with 0.85 false-positives per section.

**Conclusion.** We have evaluated an automated lung nodule detection method with a large number of low-dose CT scans from a lung cancer screening program. An overall sensitivity of 80% for malignant nodules was achieved with 0.85 false-positive detections per section. Such a computerized lung nodule detection method is expected to become an important part of CT-based lung cancer screening programs.

**Key Words.** computer-aided diagnosis (CAD); computed tomography (CT); image processing; lung neoplasms; lung nodule; cancer screening; lung CT.

© AUR, 2005

---

*Acad Radiol* 2005; 12:337-346

<sup>1</sup> From the Department of Radiology, MC 2026, University of Chicago, 5841 South Maryland Avenue, Chicago, IL 60637 (S.G.A., A.S.R., H.M., F.L., K.D., M.B.A.); and J. A. Azumi General Hospital, Nagano, Japan (S.S.). Received July 22, 2004; revision requested October 5; revision received October 15; revision accepted October 16. Presented in part at the 45th Annual Meeting of the AAPM, San Diego, California, 2003, and the 89th Scientific Assembly and Annual Meeting of the RSNA, Chicago, IL, 2003. Supported in part by U.S. Public Health Service grants CA83908, CA62625, and CA64370 and funding from The University of Chicago Cancer Research Center. Address correspondence to S.G.A. e-mail: s-armato@uchicago.edu.

© AUR, 2005

doi:10.1016/j.acra.2004.10.061

Lung cancer screening has been the subject of considerable controversy in the United States (1,2). Since the results of screening trials based on chest radiographs and sputum cytology conducted in the 1970s (3-5) were interpreted as failing to demonstrate a mortality benefit (6,7), the radiographic evaluation of asymptomatic individuals for lung cancer has received little attention. While some investigators challenged the results of these earlier screening trials and the conclusions drawn from them (8-10), another imaging modality was gaining recognition as a potentially viable screening tool. Helical computed tomography (CT) with a low-dose imaging protocol was

demonstrating an improved lung nodule detection sensitivity relative to radiography (11–13). Although the impact of earlier lung cancer detection on patient mortality remains to be proved, the efficacy of low-dose helical CT protocols (14–16) has revived interest in, and demand for, lung cancer screening.

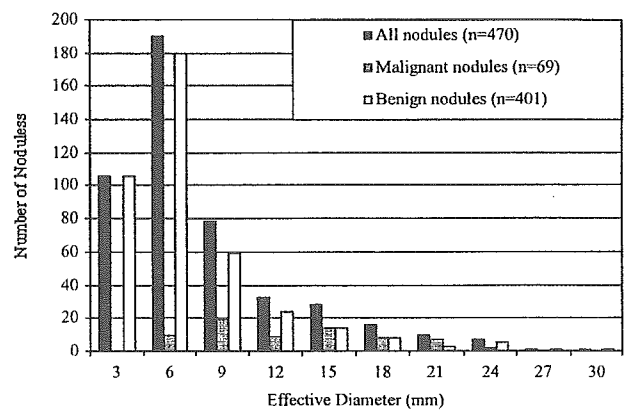
The expanding role of CT has increased the workload of radiologists, and it is clear that some form of automation will be required to assist radiologists in the task of lung nodule detection. As the use of low-dose CT for screening proliferates, the burden imposed on radiologists continues to increase. Thus, an automated nodule detection scheme such as the one reported here could become an integral component of future lung cancer screening protocols.

Investigators have developed a number of computerized nodule detection techniques (17–32). One of the first reported methods was that of Giger et al. (17), who developed an automated nodule detection scheme based on geometric feature analysis in conjunction with multiple gray-level thresholding. Other investigators have used gray-level thresholding–based methods for the automated detection of lung nodules (24,26–29). Fuzzy clustering algorithms (18), spatial filtering (20), template matching (30), object-based deformation procedures (21), morphologic analysis (23), and model-based techniques (19,25) have also been reported as viable means of detecting lung nodules. We previously reported our automated lung nodule detection method (24,33,34). The present study applies this method to a large database of thick-section (10 mm), low-dose CT scans acquired from a lung cancer screening program. Because nodules demonstrate a spectrum of radiologic appearances, the performance of the automated method was evaluated on the basis of nodule malignancy status, size, subtlety, and radiographic opacity.

## MATERIALS AND METHODS

### Database

The database used in this study consisted of 393 non-contrast, low-dose thoracic helical CT scans acquired from 393 different patients who participated voluntarily in a lung cancer screening program between 1996 and 1998 in Nagano, Japan (11,35). These cases were obtained from a larger collection of CT scans from the Nagano lung cancer screening program and were selected because they contained only lung nodules with a “confirmed” diagnosis. Informed consent was obtained from all patients.



**Figure 1.** Histograms representing the distribution of nodule effective diameters for all 470 nodules (69 malignant and 401 benign).

Of the 393 patients, 172 were female and 221 were male. The age range of the patients at the time of each scan was 30–88 years, with a mean age of 63 years. The CT examinations were performed on a mobile CT scanner (CT-W950SR; Hitachi Medical, Tokyo, Japan) and were acquired with a low-dose protocol of 25 mA (232 scans) or 50 mA (161 scans) (as the scanning protocol evolved over time), 2:1 helical pitch, 120 kVp, and 10-mm collimation with a 10-mm reconstruction interval (35). The pixel size was 0.586 mm for 278 scans and 0.684 mm for 115 scans. Each reconstructed CT section had an image matrix size of  $512 \times 512$  pixels. The 393 scans in the database consisted of 11,029 section images after sections representing anatomy inferior to the lung bases were manually excluded from each scan; the number of sections per scan ranged from 21 to 33 (mean, 28 sections per scan). The digital CT image data were transferred from optical disk storage media to a research computer (SGI Onyx; Silicon Graphics, Inc., Mountain View, CA).

The 393-case database contained a total of 470 lung nodules. Each case demonstrated at least one lung nodule, with 54 cases containing two nodules, seven cases containing three nodules, and three cases containing four nodules. Of the 470 lung nodules, 69 were malignant as determined by biopsy and 401 were benign as determined by biopsy or by follow-up over a period of 2 years. The purpose of the present study was to automatically detect lung nodules in thoracic CT scans rather than to classify such nodules as benign or malignant; however, interesting differences were observed between the ability of the automated method to detect lung nodules representing benign and malignant lesions. A serial approach to the automated

**Table 1**  
**Malignancy Status, Subtlety, and Radiographic Opacity Distributions for the 470 Nodules in the Database**

		Malignant (No.) (n = 69)	Benign (No.) (n = 401)	Total No.
Subtlety	Not subtle	28 (41%)	78 (19%)	106 (23%)
	Subtle	23 (33%)	184 (46%)	207 (44%)
	Very subtle	18 (26%)	139 (35%)	157 (33%)
Radiographic opacity	Solid	21 (30%)	152 (38%)	173 (37%)
	Part solid	31 (45%)	142 (35%)	173 (37%)
	Nonsolid	17 (25%)	107 (27%)	124 (26%)

detection and classification of lung nodules in this database has been reported previously (36).

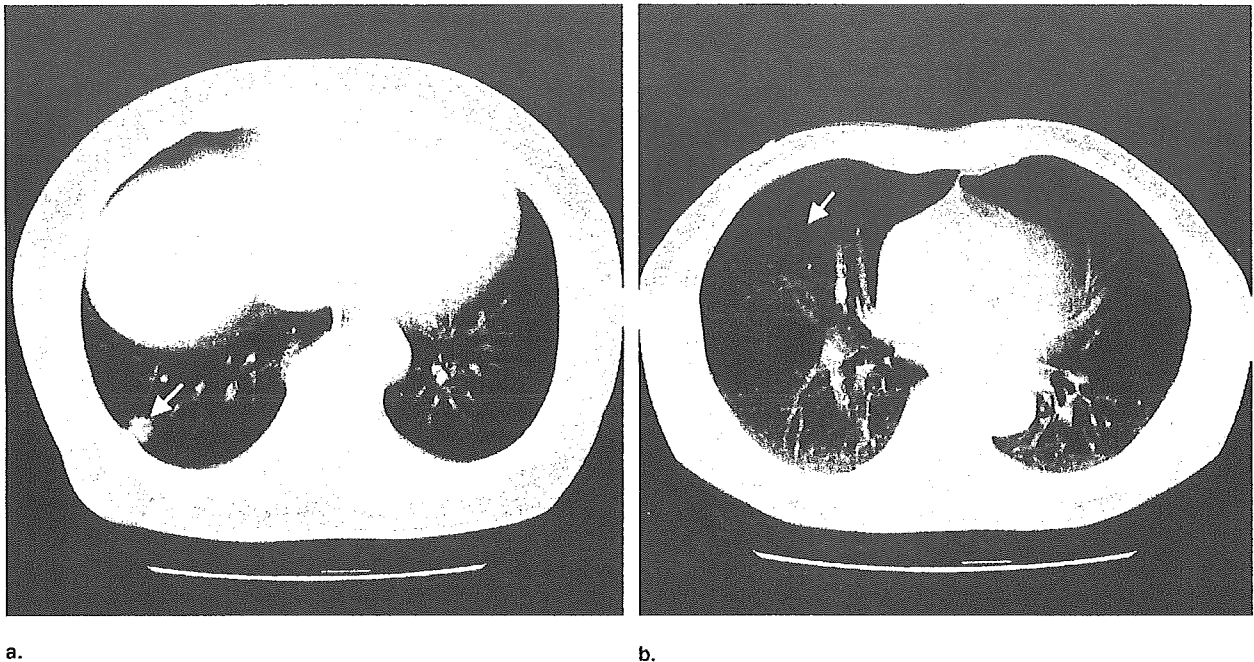
An experienced thoracic radiologist (F.L.) indicated the location of each nodule based on the radiology reports. The radiologist used a computer interface to construct rectangular bounding regions that encompassed a nodule on each CT section in which it appeared. Geometric parameters of the bounding regions for each lesion were then recorded. Figure 1 presents the distributions of effective diameters of the nodules in the database. The effective diameter of each nodule was computed as the mean bounding region dimension (one-half of the sum of the short- and long-axis lengths). When a nodule was present in more than one section, the bounding region with the greatest area was used for the effective diameter calculation. The geometric center-of-mass of each radiologist-defined bounding region and the location of the maximum-gray-level pixel within each bounding region were used to determine whether a structure detected by the automated nodule detection method correctly identified a lung nodule.

The subtlety and radiographic opacity of the nodules in this database had been evaluated by three experienced thoracic radiologists (including H.M. and F.L.) for an earlier study. Each radiologist independently rated the subtlety of each nodule on a 1–3 scale (1 = “very subtle,” 2 = “subtle,” 3 = “not subtle”) and the radiographic opacity of each nodule on a separate 1–3 scale (1 = “pure GGO (nonsolid),” 2 = “mixed GGO (part solid),” 3 = “solid”), where GGO is “ground-glass opacity.” A nodule was assigned the median rating of the three radiologists for each category. These data are summarized in Table 1. Figure 2 shows an example of a nodule categorized as “solid” and “not subtle” and an example of a nodule categorized as “part solid” and “very subtle.”

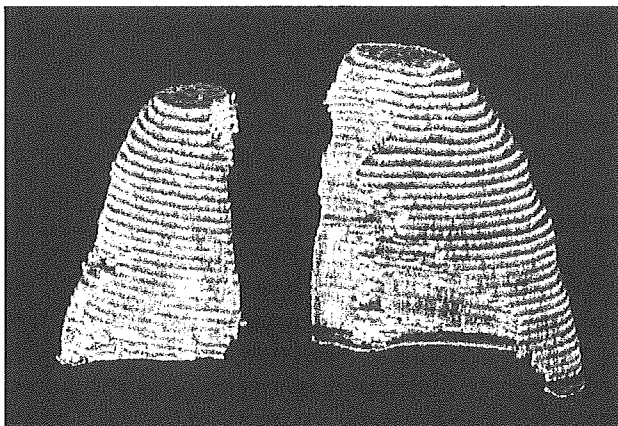
### Automated Nodule Detection

Technical details of the automated lung nodule detection method evaluated in this study have been reported previously (24,33,34). First, segmentation of the lung regions in each CT section is performed based on gray-level thresholds selected separately for each CT section from gray-level histogram analyses. After morphologic modifications are made to the segmented lung regions, a segmented lung volume is constructed from the individual segmented lung regions (Fig. 3). Next, a series of gray-level thresholds is applied to the segmented lung volume. This three-dimensional, multiple-gray-level–thresholding technique effectively represents a reverse region-growing technique that may be conceptualized as “region melting,” in which smaller volumetric structures dissociate from larger structures as the gray-level threshold gradually increases. Because a nodule is defined radiologically as any well-demarcated, soft-tissue focal opacity with a diameter of less than 3 cm (37), the stopping criterion for region melting is based on structure volume: a structure is designated a “nodule candidate” if its volume decreases below the volume of a model 3-cm-diameter sphere. The rationale for this approach is that structures too large to become nodule candidates at initial gray-level thresholds will, at higher thresholds, decompose into multiple smaller structures, many of which will qualify as nodule candidates (Fig. 4).

The categorization of nodule candidates as “nodule” or “non-nodule” is based on a combination of rule-based and linear discriminant (38) classifiers applied to a set of nine two- and three-dimensional features (volume, sphericity, mean gray level, gray-level standard deviation, gray-level threshold at which the structure became a nodule candidate, radius of the equivalent sphere, maximum eccentricity, maximum circularity, and maximum compactness)



**Figure 2.** Original computed tomography sections from the database demonstrating (a) a nodule categorized as "solid" and "not subtle" (arrow) and (b) a nodule categorized as "part solid" and "very subtle" (arrow).



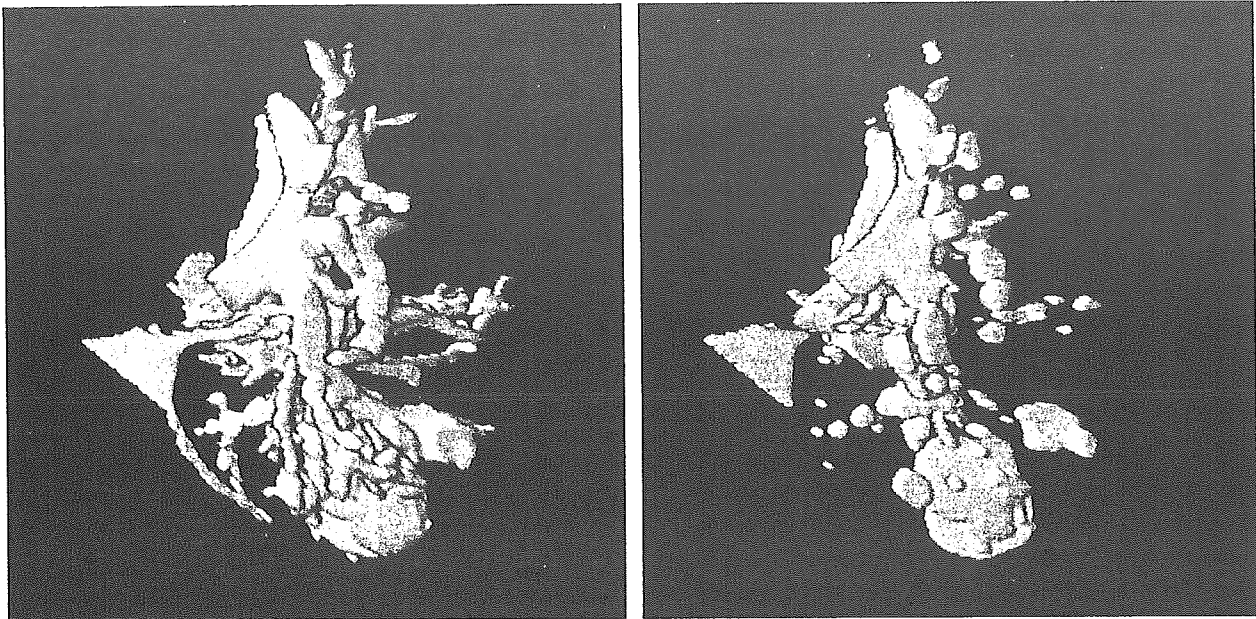
**Figure 3.** Segmented lung volume obtained from the application of gray-level thresholding techniques to the individual section images that comprise a computed tomography scan.

extracted from each nodule candidate (24,39). First, a rule-based approach is established based on three features: the maximum eccentricity and gray-level standard deviation features and an additional feature, average area, which is computed from the average two-dimensional area of a candidate in each section image in which it appears. These rules are applied to the initial set of lung nodule candidates that results from the multiple-gray-level-thresholding technique. Next, the nodule candidates

that pass the rule-based classifier (the "post-rules nodule candidates") are subjected to linear discriminant analysis (LDA) based on the nine nodule candidate features. The nodule candidates that emerge from the LDA with values below a specified LDA output value are considered "detected nodules."

None of the scans used in this study were used during the development of the method described in our earlier references, although the classifier was trained specifically on scans from the present database. Unlike our previous studies and the work of many other investigators, the combination of rule-based and linear discriminant classifiers was trained and tested through a jackknife approach in which the rules were established and the classifier was trained with nodule candidates from half of the scans, randomly chosen, and the combined classifier was tested on the candidates from the remaining scans in the database. Final performance was assessed based on the average result of multiple such jackknife runs. The jackknife approach provides a more robust evaluation of method performance. The ability of the combined classifier to differentiate nodule candidates that correspond to actual nodules from those that correspond to normal anatomy was evaluated by using receiver operating characteristic (ROC) analysis (40,41), and the overall performance of





A.

B.

**Figure 4.** (a) A single three-dimensional structure identified at a low gray-level threshold. (b) The three-dimensional structures that disassociate from the single structure in a as identified at a higher gray-level threshold. Many of these smaller structures will satisfy the volume criterion to become nodule candidates. (Reprinted from Armato et al.<sup>44</sup> with permission.)

the automated nodule detection method was evaluated through free-response receiver operating characteristic (FROC) curves (42), which are generated by plotting overall nodule detection sensitivity as a function of the number of false-positive structures identified per section. ROC and FROC curves were obtained by incrementally altering the operating point represented by the selected LDA output threshold. Both types of analysis are required because not all actual lung nodules in the database become members of the set of nodule candidates, a fact not captured by ROC analysis but reflected in FROC analysis.

#### Analysis of Complete Database

The automated lung nodule detection method was applied to the complete 393-case database of low-dose, lung cancer screening CT scans. For this analysis, all 470 lung nodules were the detection targets, regardless of nodule malignancy status, subtlety, or radiographic opacity. The initial set of nodule candidates generated by the method through the multiple gray-level thresholding stage included 415 of the 470 nodules (88.3%) contained within the database along with 211,813 non-nodule structures. Rule-based and linear discriminant classifiers were then applied to this initial set of nodule candidates to distin-

guish candidates that corresponded to actual nodules from those that corresponded to non-nodules (i.e., false positives). A jackknife training/testing approach was used in which the nodule candidates from 197 randomly chosen cases were used (1) to develop the thresholds for the rule-based classifier and (2) to establish the linear discriminant function. The rule-based and linear discriminant classifiers trained in this manner were applied to the candidates of the remaining 196 cases, and overall nodule detection results were obtained for these cases. The "training" and "testing" of the automated nodule detection method were thus performed separately on subsets of nodule candidates from independent patients.

Lung nodules demonstrate a spectrum of radiologic appearances (43). Accordingly, the performance of the automated detection method, as trained and tested on the complete 393-case database, was decomposed into performance on the basis of nodule malignancy status, size, subtlety, and radiographic opacity.

#### Analysis of Malignant Nodules

The methods reported in the previous section are based on the task of optimizing detection performance for all 470 lung nodules, regardless of nodule malignancy status,

size, subtlety, or radiographic opacity. A method that is trained and tested specifically on malignant nodules would be expected to yield improved performance on the more important clinical task of automated detection of *malignant* nodules. Among the 393 cases in the complete database, 66 cases contained at least one confirmed malignant nodule. These 66 cases contained all 69 malignant nodules in the database, along with 7 of the 401 benign nodules. When applied to this subset of cases, the initial set of nodule candidates generated by the method through the multiple gray-level thresholding stage included 66 of the 69 malignant nodules (95.7%) along with 35,447 non-nodule structures. Rule-based and linear discriminant classifiers were then applied to this initial set of nodule candidates to distinguish candidates that corresponded to actual malignant nodules from those that corresponded to non-nodules (i.e., false positives). A jackknife training/testing approach was used in which the nodule candidates from 33 randomly chosen malignant cases were used (1) to develop the thresholds for the rule-based classifier and (2) to establish the linear discriminant function. The rule-based and linear discriminant classifiers trained in this manner were applied to the candidates of the remaining 33 malignant cases, and detection results for malignant nodules were obtained for these cases. The "training" and "testing" of the automated nodule detection method were thus performed independently on independent subsets of nodule candidates from malignant cases. Detection of all malignant nodules was evaluated, and then this performance was decomposed into performance by radiographic opacity.

## RESULTS

### Complete Database

The overall performance of the method was obtained from the average performance of 10 random partitions of the nodule candidate set into training and testing subsets and is presented in the first row of Table 2. At an overall nodule detection sensitivity of 70% regardless of nodule malignancy status, size, subtlety, or radiographic opacity, a mean false-positive rate of  $1.6 \pm 0.2$  per section was obtained across all 10 partitions. This sensitivity level was selected for consistency with our other published automated nodule detection studies, which did not report performance on the basis of malignancy status, size, subtlety, or radiographic opacity.

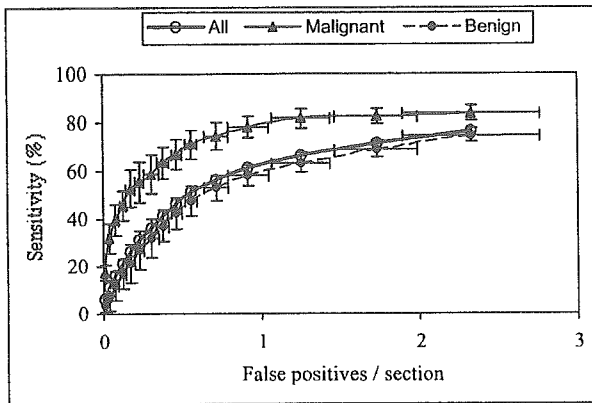
**Table 2**  
Overall Performance of the Automated Nodule Detection Method Obtained From the Average of 10 Jackknife Analyses of the Entire 393-Case Database With a Mean False-Positive Rate of 1.6 per Section

Nodule Subset	Nodule Detection Sensitivity (%)
All nodules	70
Malignant	$83 \pm 5.0$
Benign	$68 \pm 0.9$
Not subtle	$91 \pm 2.5$
Subtle	$73 \pm 2.1$
Very subtle	$52 \pm 2.9$
Solid	$76 \pm 2.0$
Part solid	$78 \pm 2.5$
Nonsolid	$51 \pm 3.7$
Effective diameter >5 mm	$74 \pm 2.7$

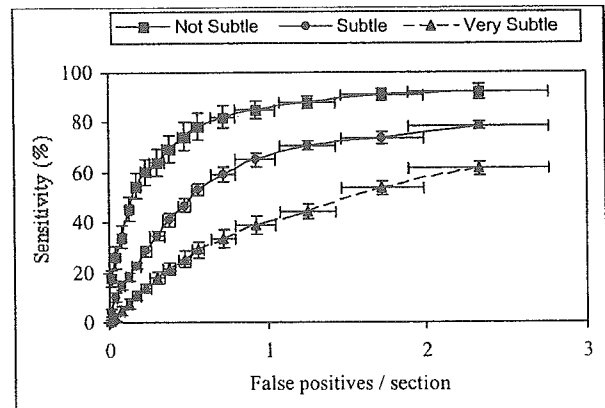
Reported sensitivities are at a mean false-positive rate of 1.6 per section. For this analysis, the method was trained to detect all nodules regardless of malignancy status, size, subtlety, or radiographic opacity.

Table 2 decomposes this *overall* performance into performance by nodule malignancy status, size, subtlety, and radiographic opacity. Note that the automated detection method does not *classify* nodules on the basis of malignancy status, size, subtlety, or radiographic opacity, but rather, in this analysis, the ability of the method to *detect* nodules within these various categories was evaluated. For example, given that 70% of all nodules in the test sets were detected,  $83\% \pm 5.0\%$  of the malignant nodules among the nodules of the test sets were detected (at the same false-positive rate of 1.6 per section), whereas only  $68\% \pm 0.9\%$  of the benign nodules among the nodules of the test sets were detected. As expected, nodule detection sensitivity increased as nodule subtlety decreased, with a detection sensitivity of  $91\% \pm 2.5\%$  for nodules rated as "not subtle" and a detection sensitivity of only  $52\% \pm 2.9\%$  for "very subtle" nodules. The detection sensitivity for solid and part-solid nodules were comparable ( $76\% \pm 2.0\%$  and  $78\% \pm 2.5\%$ , respectively), whereas the detection sensitivity for nonsolid nodules was substantially lower at  $51\% \pm 3.7\%$ . The detection sensitivity for nodules with an effective diameter greater than 5 mm was  $74\% \pm 2.7\%$  at the same false-positive rate.

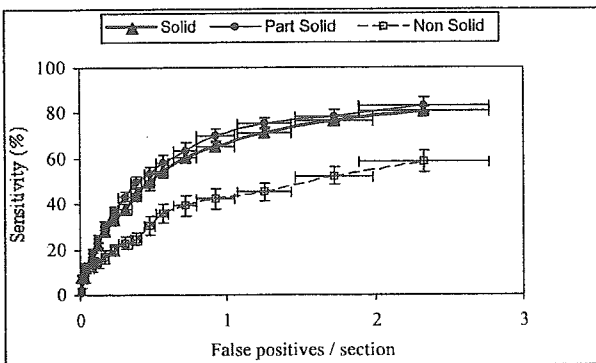
Figure 5 presents FROC curves that represent automated nodule detection sensitivity as a function of false-positive detections per section obtained from the average of 10 jackknife analyses of the entire 393-case database as described above. FROC curves are shown for (a) all



a.



b.



c.

**Figure 5.** FROC curves representing nodule detection sensitivity as a function of false-positive detections per section obtained from the average of 10 jackknife analyses of the entire 393-case database. For this analysis, the method was trained to detect all nodules regardless of malignancy status, size, subtlety, or radiographic opacity. FROC curves are shown for (a) all nodules and nodules divided into malignancy status, (b) nodules divided into subtlety, and (c) nodules divided into radiographic opacity. The data in Table 2 correspond to a single operating point from each of these nine curves.

nodules (which this analysis was intended to detect) and nodules divided into malignancy status, (b) nodules divided into subtlety, and (c) nodules divided into radiographic opacity. The data in Table 2 correspond to a single operating point from each of these nine curves. These curves illustrate the disparity in the ability of the method (when trained to detect all nodules regardless of malignancy status, size, subtlety, or radiographic opacity) to detect malignant and benign nodules; not subtle, subtle, and very subtle nodules; and solid or part-solid nodules and nonsolid nodules. The ability of the method to detect nodules that have been diagnosed as malignant is better than its ability to detect the benign nodules in the database (Fig. 5a), the ability of the method to detect nodules in the database decreases as nodule subtlety increases (Fig. 5b), and whereas the ability of the method to detect

solid and part-solid nodules is comparable, its ability to detect nonsolid nodules is lower (Fig. 5c).

### Scans with Malignant Nodules

The results reported in the previous section were based on the task of optimizing detection performance for all 470 lung nodules, regardless of nodule malignancy status, size, subtlety, or radiographic opacity. These results exhibited an interesting differentiation: 83% of the malignant nodules were detected, whereas only 68% of the benign nodules were detected. The preponderance of benign nodules in the database (401 of 470, or 85%, of the nodules were benign), combined with this malignancy status-based differential in detection performance, then yielded an overall nodule detection performance of 70% with an average of 1.6 false-positives per section. With

**Table 3**  
**Performance of the Automated Nodule Detection Method Specifically for Malignant Nodules Obtained From the Average of 20 Jackknife Analyses of the 66 Cases That Contained at Least One Malignant Nodule**

Malignant Nodule Subset	Nodule Detection Sensitivity (%)
All malignant nodules	80
Solid	90 ± 7
Part solid	80 ± 3
Nonsolid	56 ± 9
Not subtle	94 ± 5
Subtle	78 ± 6
Very subtle	46 ± 10

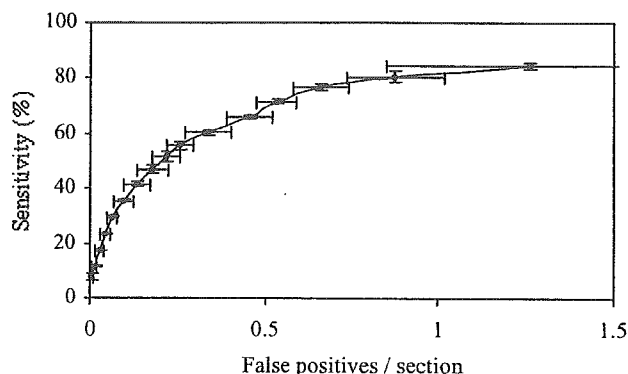
Reported sensitivities are at a mean false-positive rate of 0.85 per section. For this analysis, the method was trained to detect malignant nodules regardless of size, subtlety, or radiographic opacity.

the method trained and tested specifically on malignant nodules, improved performance was observed for the automated detection of malignant nodules.

The overall performance of the method for the detection of malignant nodules was obtained from the average performance of 20 random partitions of the nodule candidate set (obtained from the 66 cases with at least one malignant nodule) into training and testing subsets and is presented in the first row of Table 3. At a malignant nodule detection sensitivity of 80% ± 2.0%, a mean false-positive rate of 0.85 ± 0.14 per section was obtained across all 20 partitions. Figure 6 shows the corresponding FROC curve. This performance may be decomposed into performance by nodule radiographic opacity so that, on average, 90% ± 7% of the solid malignant nodules were detected, 80% ± 3% of the part-solid malignant nodules were detected, and 56% ± 9% of the nonsolid malignant nodules were detected at the same mean false-positive rate of 0.85 per section (Table 3).

## DISCUSSION

The automated method presented in this study was developed to identify any focal abnormality that generally may be regarded as a lung nodule. While lung cancers represent the most clinically important subset of such focal abnormalities, the automated identification of any focal abnormality for radiologists' consideration may be useful. In the clinical practice of the future, automated nodule *detection* would likely be integrated with auto-



**Figure 6.** FROC curve representing *malignant* nodule detection sensitivity as a function of false-positive detections per section obtained from the average of 20 jackknife analyses of the 66 cases with at least one malignant nodule. For this analysis, the method was trained to detect all malignant nodules regardless of size, subtlety, or radiographic opacity.

mated *classification* to distinguish between malignant and benign detected nodules as part of a comprehensive CAD system (36).

A limitation of the results obtained from this study is the 10-mm collimation of the CT scans that comprised the database. Such thick sections hinder the ability of the method to more accurately capture the three-dimensional contiguity of anatomic and pathologic structures. Moreover, the structures themselves are more sharply represented with state-of-the-art scanners that generate images with submillimeter axial resolution. We believe that thinner-section CT scans inherently will provide for improved performance of our method. A database of lung nodule cases from such scanners that is as well characterized as the database used in this study is certainly required to establish the clinical utility of these methods.

Despite the suboptimal nature of thick-section CT scans, a key observation from this study is that, as expected, the ability of a single automated method to detect the complete spectrum of lung nodules may be limited. Consequently, the importance of this study is the *relative* performance of the method in the detection of lung nodules in the various malignancy status, size, subtlety, and radiographic opacity categories. Based on this analysis, the automated method requires modification to improve the detection sensitivity of those lesions most likely to be missed by radiologists, namely, very subtle nodules and nonsolid nodules. The category-based performance analysis we present serves to underscore the importance of a full characterization of the lung nodules used by investigators when reporting results of CAD methods.