

メタ・アナリシスの手法を用いた 肺がん検診の有効性評価

ナカヤマ トミオ クスノキ ヨウコ スズキ タカイチロウ
中山 富雄*¹ 楠 洋子*² 鈴木 隆一郎*³

目的 現行の肺がん検診の有効性に関しては、否定的な意見が多く、わが国以外では公的資源を投入しての肺がん検診は行われていない。本論文では、国内外の研究成績をレビューし、メタ・アナリシスの手法を用いて解析することで、肺がん検診に関する総合的な評価を行う。

方法 肺がん検診の死亡率について検討した14の研究のうち、1970年代に行われた4つのランダム化比較試験と70年代以降に行われた8つの症例対照研究の成績を用いた。解析には固定効果モデルを用い、評価測定指標としてランダム化比較試験は累積死亡率を、症例対照研究ではオッズ比をそれぞれ用いた。

結果 すべての研究を含めたsummarized relative risk (SRR) は0.780 (95%信頼区間0.71-0.857)であったが、同質性が棄却された($p=0.003$)。新潟・宮城の研究を除くと、 $SRR=0.859(0.776-0.952)$ となり、同質性は保たれた。検診と無検診を比較した9つの研究に限ると、 $SRR=0.701(0.626-0.784)$ となり、肺がん検診に約30%の死亡率減少効果があることが示唆された。

結論 研究として偏りが少ないとされるランダム化比較試験と、偏りが混入しやすいとされる症例対照研究の間で結果が異なることに関しては、議論が必要である。しかし、日本で現在行われている肺がん検診に30%程度の死亡率減少効果があることに関しては、かなり信頼性が高いものと考えられる。ただし、この効果の大きさは、他のがん検診に比べると満足すべきものではなく、費用効果分析等の検討も必要である。

キーワード 肺がん検診, メタ・アナリシス, 死亡率減少効果

I 緒 言

わが国では、昭和58年以降老人保健法に基づいて胸部X線撮影と喀痰細胞診を用いた肺がん検診が広く行われてきた。平成11年に、国の老人保健事業から市町村の事業として、その予算が一般財源化されたものの、年間200万人以上が受診している。しかし、他のがん検診よりも多くの受診者があるにもかかわらず、その有効性を示す証拠は世界的にも乏しく、わが国以外では公的資源を投入しての肺がん検診は行われて

いない。ここ数年一部のマスコミ報道で、肺がん検診は無効との批判をうけてきたなかで、厚生省がん研究助成金「肺がん検診の有効性評価に関する研究」班(藤村重文班長)において、4つの地域(宮城・新潟・群馬・岡山)で症例対照研究を行い、その結果が報告された。今回、従来報告されてきた諸外国での成績と併せて、メタ・アナリシスの手法を用い、肺がん検診の評価を試みた。

* 1 大阪府立成人病センター調査部疫学課主査 * 2 同疫学課長 * 3 同研究所特別研究員

表1 肺がん検診の有効性評価を試みたランダム化比較試験

	引用	開始年	検診方法		参加者数(人)		RR(95%CI)
			検診群	対照群	検診群	対照群	
JHLP	1	1973年	胸部X線年1回 +喀痰4か月毎	胸部X線年1回	5 250	5 171	0.91(0.72-1.16)
MSKLP	2	1974	胸部X線年1回 +喀痰4か月毎	胸部X線年1回	4 968	5 072	0.92(0.67-1.26)
MLP	3	1971	X線と喀痰 4か月毎	X線と喀痰 年1回を推奨	4 618	4 593	1.06(0.82-1.36)
CSLP	4	1976	X線と喀痰 6か月毎	無検診	3 171	3 174	1.36(0.94-1.97)

注 JHLP: Johns Hopkins Lung project, MSKLP: Memorial Sloan-Kettering Lung project
MLP: Mayo Lung project, CSLP: Czechoslovakian Lung project
RR: Relative risk, 95%CI: 95% confidence interval

表2 肺がん検診の有効性評価を試みた症例対照研究

	引用	開始年	検 診 方 法	参加者数(人)		OR(95%CI)
				症例群	対照群	
GDR-1	5	1974年	胸部X線2年に1回	130	260	1.07(0.67-1.78)
GDR-2	6	1974	胸部X線2年に1回	278	967	0.93(0.65-1.33)
成毛班	7	1977	胸部間接X線+喀痰細胞診年1回	273	1 269	0.72(0.50-1.03)
金子班	8	1985	胸部間接X線+喀痰細胞診年1回 (個別検診)	193	579	0.53(0.30-0.94)
群馬	9	1989	胸部間接X線(結核検診)	121	536	0.68(0.44-1.05)
宮城	10	1990	胸部間接X線+喀痰細胞診年1回	328	1 886	0.54(0.41-0.73)
新潟	11	1990	胸部間接X線+喀痰細胞診年1回	174	801	0.40(0.27-0.59)
岡山	12	1991	胸部間接X線+喀痰細胞診年1回	412	3 490	0.59(0.46-0.74)

注 GDR: German Democratic Republic
OR: Odds ratio, 95%CI: 95% confidence interval

II 方 法

(1) 肺がん検診の有効性評価研究の概要

がん検診の有効性を評価する手法、すなわち当該がんの死亡率減少効果をエンドポイントとする研究にはランダム化比較試験、コホート研究、症例対照研究などがある。これらの研究手法には、一長一短があるが、理論上ランダム化比較試験には最もバイアス(偏り)が混入しがたく、その研究成果には国際的評価が高いとされている。

諸外国では、新しい検診技法が開発されると、通常複数のランダム化比較試験を実施し、適切な事前評価を行った上で、公的に実施するか否かが決められている。

胸部単純X線写真と喀痰細胞診を用いた肺がん検診について行われた評価研究^{1)~12)}の概要を示す(表1, 2)。有名な米国で行われた3つのランダム化比較試験^{1)~3)}のうち2つは、喀痰細胞診の上乗せ効果を見たものであり、胸部X線+喀痰細胞診と無検診の比較をデザインしたのはMLPのみである³⁾。しかし実際のMLPでは本来無検診であるはずの対照群の半数以上が、少なくとも年1回胸部X線撮影を受診していたことが報告されており、純粋な意味での検診と無検診の比較がなされたわけではない。

わが国では結核検診がすでに普及していたこともあり、ランダム化比較試験が事前になされないまま、昭和58年に老人保健法に基づくがん検診に導入された。その後6つの症例対照研究がわが国で報告されており^{7)~12)}、そのうち4つは統計学的に有意に逐年検診が死亡率減少効果

のあることを示している。

(2) 研究の選択

同一のテーマに関して研究を行ったとしても、標本数が少ないことや、研究デザインの差から必ずしも、決定的な結論を示し得ないことがある。メタ・アナリシスはこのような場合に、複数の研究を収集吟味し、総合的に評価して、一つの結論を導き出すための研究手法である。肺がん検診においては、前述したようにわが国で行われた症例対照研究のみで有効性を示唆する結果が得られているが、欧米での研究では全く有効性を示唆する結果は得られていない。これらの研究に関する詳細な文献レビューは、厚生省老人保健推進費等補助金「がん検診の有効性評価に関する研究」班(久道 茂班長)においてなされたので、詳細は参照されたい。

今回、解析に用いる研究として、①1970年代に行われた4つのランダム化比較試験(Johns Hopkins Lung project, Memorial Sloan-

Kettering Lung project, Mayo Lung project, Czechoslovakian project), ②1970年代以降に行われた8つの症例対照研究 (GDR-1, GDR-2, 成毛班の研究, 金子班の研究, 藤村班の4つの研究) を選んだ。1960年に開始されたNorth west London study (集団単位ランダム化比較試験)¹³⁾は研究開始時期が古く, レントゲン機器や診療技術の差が著しいこと, 追跡期間がなく死亡の把握が不十分なことを理由として, 解析から除外した。1972年に開始されたErfurt study (集団単位非ランダム化比較試験)¹⁴⁾は, 検診精度が低く(感度47%), 切除率も著しく低く(28%), かつ追跡期間の記載がないため, 除外した。

GDR-1に関しては⁵⁾, 住民ベースの対照Aと外来患者ベースの対照Bの二つの成績が報告されているが, 対照Aでは交絡因子である喫煙歴が不明のため, 対照Bの成績を利用した。

(3) 解析方法

解析には, 主として固定効果モデルである general valiance-based methodを用い, ランダム効果モデルである DerSimonian - Laird methodで得られた値を参考として示した。ランダム化比較試験には, 観察人年が報告されていないものもあるため累積死亡率を, 症例対照研究ではオッズ比をそれぞれ評価測定指標とした。研究間のバラツキを評価するための同質性の検定に関しては, χ^2 検定で統計値Qが $p < 0.05$ の場合に同質性が棄却されるものとし, Q値が高いものから順に解析から外して検定を繰り返した。研究手法別・性別に層別解析を行った。金

子班の研究は性別の解析結果が示されていないため, 性別の解析から除外した。

III 結 果

表3にすべての研究を含めた解析の結果を示した。4つのランダム化比較試験と8つの症例対照研究をすべて含めたsummarized relative risk (SRR) は, 0.780 (95%信頼区間0.711-0.857) であったが, 同質性が棄却された ($p = 0.003$)。そこで, 統計値Qが高い順から新潟, 宮城の研究を除くと, SRRは0.859 (0.776-0.952) となり, 同質性は保たれた。ランダム効果モデルでもSRRは0.854 (0.719-1.014) とあまり変化はなかった。Johns-Hopkins studyとMemorial Sloan Ketterings studyは喀痰細胞診の上乗せ効果を見たもので, 検診と無検診を比較したものではない。よってこの二つの研究を除いた解析を行った(表4)。2つのランダム化比較試験と8つの症例対照研究では, SRRは0.741 (0.666-0.825) となったが, 同質性が棄却 ($p = 0.017$) されるため, Czechoslovak studyを除くと, SRRは0.701 (0.626-0.784) となった。ランダム効果モデルでもSRRは0.739 (0.600-0.911) と大きな変化はなく, 検診受診が無検診よりも30%弱の死亡率減少効果があることが示唆された。

ランダム化比較試験のような実験的研究と, 症例対照研究のような観察的研究を, 一つの解析に含めることには, 方法として望ましいものではない。そこで研究デザイン別の解析を表5に示した。ランダム化比較試験のSRRは

表3 すべての研究を含めた解析

	Summarized relative risk	95%信頼区間	
		下限	上限
4つのランダム化比較試験と8つの症例対照研究*			
固定効果モデル	0.780	0.711	0.857
ランダム効果モデル	0.813	0.681	0.971
4つのランダム化比較試験と6つの 症例対照研究(新潟、宮城を除く)			
固定効果モデル	0.859	0.776	0.952
ランダム効果モデル	0.854	0.719	1.014

注 * : 同質性の χ^2 検定: $p = 0.003$

表4 検診と無検診を比較した解析 (JHLPとMSKLPを除いた解析)

	Summarized relative risk	95%信頼区間	
		下限	上限
2つのランダム化比較試験と8つの症例対照研究*			
固定効果モデル	0.741	0.666	0.825
ランダム効果モデル	0.789	0.631	0.986
1つのランダム化比較試験と8つの 症例対照研究 (Czechoslovak studyを除く)			
固定効果モデル	0.701	0.626	0.784
ランダム効果モデル	0.739	0.600	0.911

注 * : 同質性の χ^2 検定: $p = 0.017$

1.013(0.881-1.165)、検診と無検診の比較のみ(JHLPとMSKLPを除く2つのランダム化比較試験のみ)に限ると1.144(0.930-1.409)で、検診の有効性は全く認められない。一方症例対照研究では、SRRは0.570-0.667の間にあり、いずれの場合においても統計学的に有意に検診の有効性を示す結果であった。

性別の解析では、興味深いことにランダム化比較試験を含めても、症例対照研究のみでも、男性のSRRは0.684-0.881という範囲にあり、統計学的に有意ではあるが、検診の効果が10～30%という小さな効果にすぎないことを示している(表6)。一方女性では日本の症例対照研究での評価に限られるが、SRRは0.474(0.344-0.652)と、検診の効果が50%程度の大きな死亡率減少効果があることが示された。

IV 考 察

米国では国をあげた喫煙対策が進み、1990年代には肺がんを中心とした喫煙関連がんの年齢調整死亡率が減少の方向に進んできた¹⁵⁾。米国が今直面している問題は、タバコをやめた人間に対して更に死亡リスクを軽減させる試みとして、検診を行う意義はあるだろうか?ということである。一方わが国のがん対策は対称的に、二次予防である検診が重視され、一次予防をなおざりにしてきたことは言うまでもない。このような方針が決して適切ではなかったことは、未だに年齢調整肺がん死亡率が増加し続けていることから明らかである。さて、それでは今までの二次予防(肺がん検診)は全く無駄であったのだろうか?この疑問に答えるべく、われわれは厚生省がん研究助成金藤村班において、4つの地区で症例対照研究を行い、逐年検診に30～60%の肺がん死亡リスク減少効果があることを示唆した。しかし、これらの結果は従来の欧米の研究成績と異なり、その相違をどう解釈

表5 研究デザイン別の解析

	研究対象者数(人)		Summarized relative risk	95%信頼区間	
	検診群 (症例群)	対照群		下限	上限
ランダム化比較試験	18 007	18 010	1.013	0.881	1.165
JHLP, MSKLP除く	肺がん死亡 388	382			
	7 789	7 767	1.144	0.930	1.409
肺がん死亡	186	162			
症例対照研究*	1 909	9 788	0.633*	0.559	0.717
新潟の研究を除く	1 735	8 987	0.667	0.584	0.761
日本の研究	1 501	8 561	0.570	0.496	0.656

注 * : 同質性の χ^2 検定: $p=0.017$

表6 男女別の解析

	Summarized relative risk	95%信頼区間	
		下限	上限
男			
全研究*	0.840	0.761	0.927
新潟の研究を除く	0.881	0.796	0.975
症例対照研究のみ	0.753	0.650	0.873
日本のみ	0.684	0.575	0.812
女	0.474	0.344	0.652

注 * : 同質性の χ^2 検定: $p=0.017$

するのかという疑問が残った。

今回の検討は、メタ・アナリシスの手法を用い、これらの成績の総合評価を試みたものである。70年代に行われた4つのランダム化比較試験は、デザインがそれぞれ異なるが、結果としていずれも肺がん死亡率減少を示さなかった。一方症例対照研究では、2年に1回の検診を評価したGDR-1と2をのぞき、すべて肺がん検診受診により肺がん死亡率減少効果があることを示す成績であり、今回の検討からは、毎年検診受診により約40%の死亡率減少効果があることが示唆された。この二つの相反する結果を説明するには、次のような仮説が考えられる。1) 症例対照研究に偏りが混入したため、検診の効果を過大に評価した。2) ランダム化比較試験に偏りが混入したため、検診の効果を過小に評価した。3) ランダム化比較試験が行われた70年代の医療レベル(X線撮影技術・読影技術・確定診断技術)では、検診の効果は認められなかったが、80年代以降改善された。4) 欧米の肺がんは検診の結果が現れにくく、日本では効果が現れやすい生物学的差異があった。

まず、個々の症例対照研究については、バイ

アスを制御する様々な試みがなされている。しかし観察的研究で最も問題となるself-selection bias (検診を受診する人は、健康意識が高いため、罹患率・死亡率ともに、非受診者よりも低くなるという偏り) については、必ずしも完全に制御された訳ではない。とはいえ、40%の死亡率減少効果をこれだけで説明できるものではなく、6つの研究がほぼ同じ結果を示したことから、信頼性は高いものと考えられる。

次にランダム化比較試験については、最近のMayo Lung projectの長期追跡研究により¹⁶⁾、その妥当性が一見証明されたかに見える。しかし、その結果を詳細に検討すると、検診群の過剰診断のみでは説明できない腺がんの罹患率の差等々疑問が残っており、結果として割付が均等には行えなかった可能性が示唆される¹⁷⁾。

また研究手法としての妥当性に加えて、医療レベルの進歩には相当なものがある。Mayo clinicにおいても、当時は末梢性肺がんの術前診断が確立されておらず、X線撮影で異常が認められた場合、X線での経過観察か開胸生検が用いられていた¹⁸⁾。現在のCT装置や気管支ファイバースコープの普及した医療状況と当時を比較することには、はなはだ問題がある。X線の撮影・読影技術にも大きな進歩があることは言うまでもないが、日本で肺がん検診の撮影・読影に従事しているX線技師・読影医の技術が一樣に進歩していることを決して保証するものではない。

最後に欧米と日本の肺がんとの差異についてであるが、欧米では喫煙との因果関係が強い扁平上皮がんが多く、日本では喫煙との因果関係が弱い腺がんが多いという組織学的な割合の差が、以前より指摘されている¹⁹⁾。腺がんは末梢発生でX線での診断には適しており、進行速度も緩やかなものも多く、定期的検診で発見されやすい。この腺がんの割合が日本で多いことが、検診の効果として現れたのかもしれない。

いずれにせよ、一つの仮説だけでこの二つの異なる結果を説明できるものではなく、4つの仮説がすべて成り立ったところに、真実があるのかもしれない。肺がん検診の効果の大きさと

して、症例対照研究のみの解析では、約40%の死亡率減少効果を示す成績であったが、症例対照研究がbiasを制御できずに真の効果を過大評価しているとすれば、ランダム化比較試験(MLP)を加えて検診と無検診を比較した場合の30%減少が、真の効果であるのかもしれない。

ただし、たとえ統計学的な有意性を持って効果が証明されたとしても、この効果の大きさが許容できるか否かに関しては、費用効果分析等の別の分析が必要である。がん検診が市町村の事業となった現在、検診の規模縮小や中止を図る市町村が見られるが、このような議論をする際は、やはり科学的な根拠に基づくべきであり、マスコミの報道や風潮に惑わされて短絡的な決断はすべきでないことを最後に申し述べる。

文 献

- 1) Tockman MS. What did we learn from the Johns Hopkins Lung project? Proceedings of International conference for prevention and early diagnosis of lung cancer. 1998; Varese-Italy: 29-33.
- 2) Melamed MR, Flehinger BJ, Zaman MB, et al. Memorial Sloan-Kettering Study. Screening for early lung cancer. Results of Memorial Sloan-Kettering in New York. Chest 1984; 86: 44-53.
- 3) Fontana RS, Sanderson DR, Woolner LB, et al. Screening for lung cancer. A critique of the Mayo Lung Project. Cancer 1991; 67: 1155-64.
- 4) Kubik A, Parkin DM, Khlat M, et al. Lack of benefit from semi-annual screening for cancer of the lung: Follow-up report of a randomized controlled trial of a population of high-risk males in Czechoslovakia. Int J Cancer 1990; 45: 26-33.
- 5) Ebeling K, Nischan P. Screening for lung cancer. Results from a case-control study. Int J Cancer, 1987; 40: 141-4.
- 6) Berchdt R, Nischan P, Ebeling K. Screening for Lung cancer in the middle-aged. Int J Cancer, 1990; 45: 229-30.
- 7) Sobue T, Suzuki T, Naruke T, et al. A case control study for evaluating lung-cancer screen-

- ing in Japan. Int J Cancer, 1992 ; 50 : 230-7.
- 8) 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, 1998 ; 25 : 77-85.
- 9) 中山富雄, 金子昌弘, 馬場 孝, 他. 肺がん検診の有効性の評価: 群馬県における症例対照研究. 「肺がん検診は有効か?」厚生省藤村班 平成9~10年度研究報告書 (速報). 1999 ; 10-3.
- 10) 佐川元保, 斉藤泰紀, 藤村重文, 他. 宮城県での肺癌検診の死亡減少効果に関する症例対照研究. 「肺がん検診は有効か?」厚生省藤村班 平成9~10年度研究報告書 (速報). 1999 ; 6-9.
- 11) 塚田裕子, 栗田雄三, 三沢博人, 他. 症例対照研究の手法を用いた, 新潟県の肺がん検診の有効性の評価. 「肺がん検診は有効か?」厚生省藤村班 平成9~10年度研究報告書 (速報). 1999 ; 14-7.
- 12) 西井研治, 守谷欣明, 正影三恵子, 他. 岡山県における症例対照研究の手法を用いた肺癌検診の有効性評価に関する研究. 「肺がん検診は有効か?」厚生省藤村班 平成9~10年度研究報告書 (速報). 1999 ; 18-21.
- 13) Brett GZ. The value of lung cancer detection by six-monthly chest radiographs. Thorax, 1968 ; 23 : 414-20.
- 14) Wilde J. A 10 year follow-up semi-annual screening for early detection of lung cancer in the Erfurt County, GDR. Eur Respir J, 1989 ; 2 : 656-62.
- 15) Jemal A, Chu KC, Tarone RE, Recent Trends in lung cancer mortality in the United States. J Natl Cancer Inst, 2001 ; 93 : 277-83.
- 16) Marcus OM, Bergstralh EJ, Fagerstrom RM, et al. Lung cancer mortality in the Mayo Lung Project : Impact of extended follow-up. J Natl Cancer Inst, 2000 ; 92 : 1308-15.
- 17) Sobue T, Nakayama T. Re. Lung cancer mortality in the Mayo Lung Project : Impact of extended follow-up. J Natl Cancer Inst, 2001 ; 93 : 320-1.
- 18) Fontana RS, Sanderson DR, Taylor WF, et al. Early lung cancer detection : Results of the initial (Prevalence) radiologic and cytologic screening in the Mayo Clinic study. Am Rev Respir Dis, 1984 ; 130 : 561-5.
- 19) Hanai A, Benn T, Fujimoto I, et al. Comparison of lung cancer incidence rates by histologic type in high and low incidence countries, with special reference to the limited role of smoking. Jpn J Cancer Res (Gann), 1988 ; 79 : 445-52.

2001年 3 動向誌発行のお知らせ

表示は本体価格です。
定価は別途消費税が
加算されます。

- * 国民衛生の動向2,095円
8月31日発行予定
- * 国民の福祉の動向1,800円
10月中旬発行予定
- * 保険と年金の動向1,800円
11月中旬発行予定

財団法人 厚生統計協会

〒106-0032 東京都港区六本木5-13-14
TEL 03-3586-3361