

検診の効果とバイアス

祖父江友孝¹

要旨 目的・方法・検診の効果とバイアスについて、これまでの知見を整理して解説する。結果・結論・検診の効果については、対象とするがんの死亡率を評価指標とするのが一般的である。検診発見例における病期分布や生存率を症状発見例と比較することは、一連の検診評価プロセスの中で重要な評価指標ではあるが、種々のバイアス (self-selection bias, lead time bias, length bias, overdiagnosis bias) の影響を受ける可能性が高く、死亡減少効果に代わりうるものではない。Overdiagnosis は、バイアスとしての意義に加えて、検診による不利益としての意義が大きくなりつつある。死亡率減少効果の評価方法としての国際水準はランダム化比較試験 (Randomized Controlled Trial) であり、諸外国では RCT 以外の研究デザインは証拠として軽視される傾向にある。しかし、RCT の中でも計画・実行・解析が適切にされたかの吟味が重要であり、効果に関する証拠のまとめを作成するには、研究デザインだけでなく、研究の質のチェックが手順の中に組み込まれつつある。(肺癌. 2003;43:1013-1017)

索引用語 がん検診, 肺がん, 効果, バイアス, 過剰診断

Tomotaka Sobue¹

ABSTRACT *Objective and Methods.* This aims to overview the effect of screening and related biases by reviewing previous findings. *Results and Conclusion.* The effect of screening is usually measured by site-specific mortality rate as an endpoint. Although comparisons of stage distribution or survival between screen-detected and symptom-detected cases are important measures in the entire process of screening evaluation, they will not be surrogates for mortality due to high possibility of suffering from various biases, such as self-selection bias, lead time bias, length bias. In particular, overdiagnosis has become of great concern not only as bias but also as harm due to screening. For evaluating mortality reduction, randomized controlled trial is a gold standard method. In the US and European countries, the results from any study design other than RCT, tend to be considered less valuable. However, it is important to evaluate how well the study designed, conducted and analyzed even for RCT. In the process of summarizing the evidence, checking internal validity is now concerned essential part of ranking the quality of study results. (JLCC. 2003; 43:1013-1017)

KEY WORDS Cancer screening, Efficacy, Bias, Overdiagnosis

はじめに

がん検診には、利益 (= 効果) もあるが不利益もある。がん検診を推奨する際には、対象となる集団について利益と不利益のバランスを考慮し、利益が不利益を上回ると判断される場合のみ、がん検診を推奨すべきである。この場合の利益とは、主として死亡率減少効果を示す。これらのことは、1983 年に開催された UICC のがん検診に関するワークショップの会議記録の中にも、すでに明確に記載されている！

検診の効果

検診の効果については、対象とするがんの死亡率を評価指標とするのが一般的である。対象とするがん死亡率ではなく全死因死亡率を評価指標にすべきとの議論もあるが、がん検診の全死因死亡率減少に対する効果の大きさが小さいために数百万人レベルの対象者数が必要となり、現実的には実行可能性に欠ける。ただし、全死因死亡率を副次的評価指標 (むしろ有害事象としての) とすべきであることには異論はない。

¹ 国立がんセンターがん予防・検診研究センター情報研究部。

¹ Statistics and Cancer Control Division, Research Center for Cancer Prevention and Screening

National Cancer Center, Japan.

© 2003 The Japan Lung Cancer Society

1. Nonmalignant or borderline lesions erroneously classified as malignancy
 - Tumors with extremely long sojourn time
2. Cancers that would never have become manifest clinically before the patients die of other causes
 - All Tumors when the patients die of other causes during lead-time period
 - Competing risk

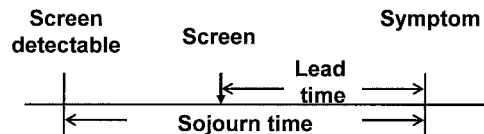


Figure 1. Definition of overdiagnosis.

わが国では、死亡率減少効果を中心とする有効性を評価した厚生労働省研究班（主任研究者 久道茂）による報告書が1998年、1999年、2001年と過去3回報告されている³

生存率を指標とする際のバイアス

上記報告書でも解説されているように、一般に用いられてきた発見動機別（検診発見と症状発見）の生存率比較による評価の際には、以下の4つのバイアスの影響のために、実際の死亡率減少効果を過大に推定することになる^{4,5}

1. 先行時間による偏り(Lead time bias)：検診により早期に発見された期間(lead time)だけ生存率の計算の始点が前にずれるために生ずる偏り。

2. 滞在時間の長さによる偏り(Length bias)：検診では進行の早いがんよりもゆっくりしたがんを発見しやすく、進行のゆっくりしたがんの方が生存時間が長いために生ずる偏り。

3. 過剰診断による偏り(Overdiagnosis bias)：検診では、受診者が余命を全うするまでには臨床的に問題にならない(症状などを呈しない)がんも発見してしまうために生ずる偏り。

4. 自己選択による偏り(Self-selection bias)：検診受診者の方が健康意識も高く、健康管理に十分な注意を払うために、生存時間が長いと考えられるために生ずる偏り。

このうち、lead time bias, length bias, overdiagnosis biasは理論上必ず存在する偏りだが、self-selection biasは必ず存在するというものではない。発見動機別の生存

率比較は、有効性評価のプロセスの中で重要な情報ではあるが、死亡率減少効果に代わりうる指標とはならない。

死亡率を指標とする評価方法とバイアス

死亡率減少効果の評価方法としての国際水準はランダム化比較試験(Randomized Controlled Trial)である。RCTは、きちんと計画・実行・解析されれば、バイアスが入り込む危険性が最も少なく、その意味で最も質の高い証拠と考えられる。諸外国ではRCT以外の研究デザインは証拠として軽視される傾向にある。

一方、これまでRCT以外に死亡率を指標とした観察的研究が行われてきた。個人の受診歴に基づいた研究としては比較として症例対照研究が数多く行われ、集団の受診歴に基づいた研究として、地域相関・時系列研究(高受診率地区と低受診率地区との死亡率および罹患率の比較など)が行われてきた。症例対照研究は、死亡率を評価指標とするので、lead time bias, length bias, overdiagnosis biasの影響は受けませんが、self-selection biasの影響を受ける可能性がある。このself-selection biasは、生存率による評価の際のself-selection biasとは異なり、受診群と非受診群の間で罹患率自体が異なる可能性を意味している。この場合、受診者は非受診者に比べて健康意識が高くリスク要因を避ける傾向にあるため罹患率が低い可能性もあるが、逆に、何らかの体調不調があるために受診をした場合には罹患率が高くなる可能性もあり、どちらの方向に傾くかは予期しがたい。地域相関・時系列研究の場合、上で述べた特定のバイアスの影響を受けることはないが、集団単位での比較となるため、交絡要因の調整が難しく、因果関係を示すための証拠とし

Table 1. Measuring the magnitude of overdiagnosis

Compare incidence between screened and non-screened groups in randomized controlled trials - Needs long follow-up phase after screening phase - Data can be collected within fixed time - At least maximum magnitude can be estimated Compare survival for early stage non-treated cases between screen-detected and symptom-detected groups - Uncertain to collect sufficient number of cases within fixed time - Data quality may be poor - Difficult to determine the cause of deaths

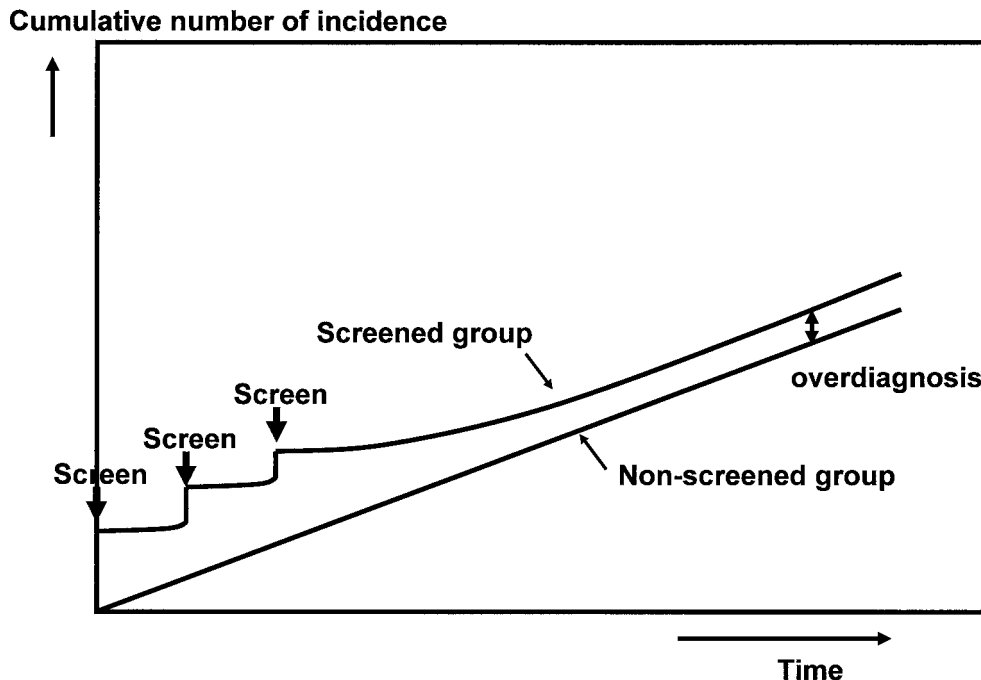


Figure 2. Compare incidence between screened and non-screened groups in randomized controlled trials.

てはレベルが低くなる。

ガイドライン作成過程における証拠の質のランク付け

これまでのガイドライン作成過程では、証拠の質のランク付けは、専ら研究デザインに基づいて行われてきた。しかし、きちんと計画・実行・解析されていないRCTよりも、きちんと計画・実行・解析された症例対照研究の方が質の高い証拠を提示しうる場合もある。そこで、研究がどの程度きちんとされたか（内部妥当性）のチェックした上でランク付けを行う動きが出てきた。US Preventive Services Task Force, NCI-PDQ, CDC-Guide to Community Preventive Services などにおいても内部妥当性のチェックリストを設けている。^{6,8}

過剰診断の定義と測定法

過剰診断は、生存率による評価の際のバイアスの1つとして数えられているが、最近ではそれ以上に受診者に対する不利益の1つとして重要視されてきている。過剰診断の定義としては、従来は「境界領域の病変を誤ってがんと診断した例」といった意味合いであったが、最近では「受診者が他の死因で死亡する以前に、臨床的に意味のある状態（症状を呈するといった）にはならない病変をがんと診断した例」といった意味合いになってきている（Figure 1）。後者の定義は、病変そのものの要因だけで決定されるものではなく、宿主側の要因も考慮する必要がある。例えば、通常の成長速度を有するがんでも、

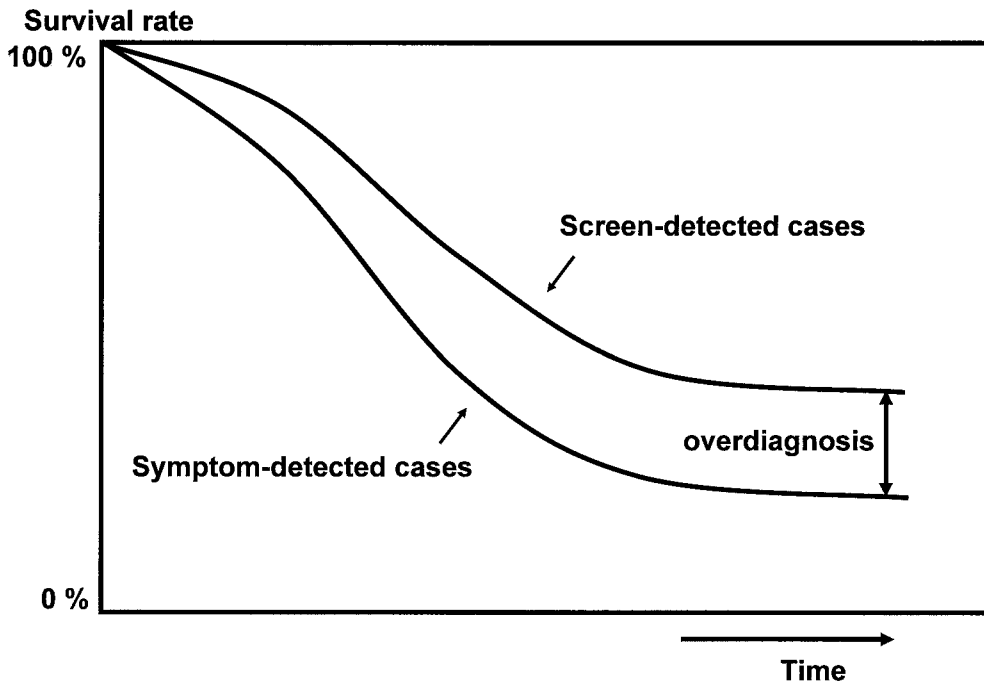


Figure 3. Concept of comparing survival for early stage non-treated cases between screen-detected and symptom-detected groups.

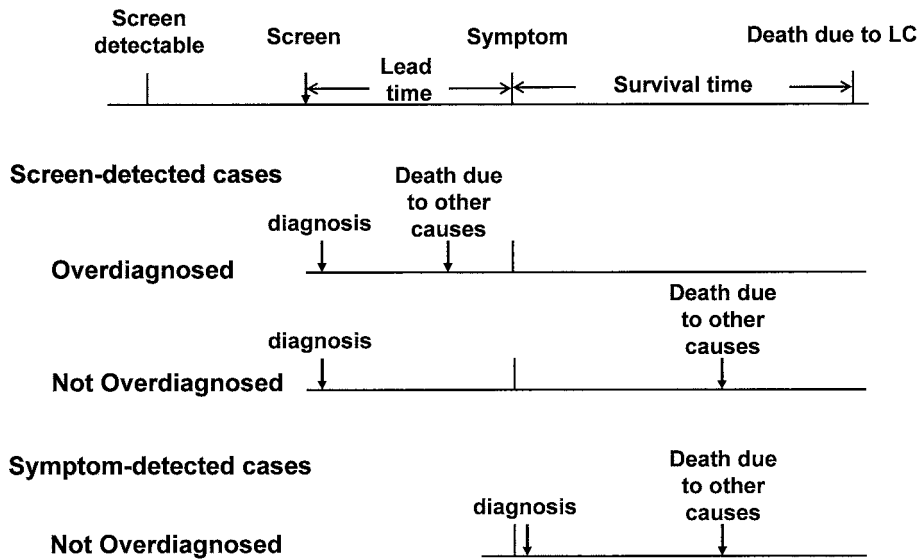


Figure 4. Overdiagnosis defined as “ cancers that would never have become manifest clinically before the patients die of other causes ”.

余命1ヶ月の人に無症状の状態で見つかった場合、過剰診断となりうる。ただし、個々の例について過剰診断が否かを判断するには、無治療でかつ詳細な臨床経過を観察する必要があり、一般には相当の困難を伴う。

集団において検診による過剰診断の大きさを測定する

方法としては (Table 1), 第一に, RCT において罹患率を検診群と対照群で比較することが考えられる (Figure 2). また, 無治療例の生存率を発見動機別に比較する (Figure 3) ことも考えられるが, この場合, 検診発見例の他死因死亡だけでなく, 症状発見例の他死因死亡を把

握して差を求めることにより，過剰診断の大きさを計測できる（Figure 4）。

過剰診断は，検診技術の精度が向上すればするほど，不利益として大きな問題となる可能性がある．検診の利益と不利益を定量的に評価するために，利益だけでなく不利益を定量的に評価することが重要となる．

REFERENCES

- 1 . Prorok PC, Miller AB, eds. Screening for cancer. *UICC Technical Report Series*. Vol 78. Genova: UICC; 1983.
- 2 . Black WC, Haggstorm DA, Welch HG. All-cause mortality in randomized trials of cancer screening. *J Natl Cancer Inst*. 2002;94:167-173.
- 3 . がん検診の適正化に関する調査研究事業「新たながん検診手法の有効性の評価」報告書(主任研究者 久道 茂). 東京：財団法人日本公衆衛生協会；2001.
- 4 . Kramer BS, Gohagan JK, Prorok PC, eds. *Cancer screening*. New York: Marcel Dekker; 1999.
- 5 . 大野良之，柳川 洋，編集．生活習慣病予防マニュアル．東京：南山堂；1999.
- 6 . U.S. Preventive Services Task Force (USPSTF) Methods and Background .(<http://www.ahrq.gov/clinic/uspstmeth.htm>] last accessed Oct 2003]
- 7 . PDQ Cancer Information Summaries: Screening /Detection (Testing for Cancer) Levels of Evidence for Cancer Screening and Prevention Studies.(<http://www.nci.nih.gov/cancerinfo/pdq/screening/levels-of-evidence>] last accessed Oct 2003]
- 8 . Guide to Community Preventive Services . Methods . (<http://www.thecommunityguide.org/>] last accessed Oct 2003]