

Beginners' Training Sheet for Diagnostic study

ver.6.4 by last updated on August 19, 2012

南郷 栄秀 Eishu NANGO, MD, PhD

The SPELL <http://spell.umin.jp>

このシートは初めて診断法の論文を読むためのものです。診断法の定義と論文の構造にも触れながら、論文を読む上でのポイントを解説しました。

なお、このシートに関する質問、改善点などは、制作者まで直接お願いします。また、制作者は著作権を保持し、無断転載を禁止します。再配布に制限はしないつもりですが、再配布する際は制作者までご一報ください。

採用論文

Reviewer:

年 月 日

author :

title :

citation :

PubMed PMID :

quick check list

1. 論文の PICO を探る
2. C(Comparison)は gold standard に近いものになっているか？
3. 研究で行われた診断法は、いずれも全ての患者で行われているか？
4. 研究で行われた診断法と最終診断の判定は、互いに独立に行われているか？
5. 研究で行われた診断法と最終診断の判定は、いずれもその方法が明確に記載されているか？
6. 研究で行われた診断法と最終診断の判定は、いずれも再現性があるか？
7. 結果の評価

0 診断法の性能を検証するための研究デザインとは？

0-1) 診断の研究デザインとは？

ある診断法の正確度を調べるための研究デザインとしては、**横断研究**が最も優れている。横断研究とは、ある一時点に調査を行う研究である。これに対して、ランダム化比較試験（以下 RCT）、コホート研究、症例対照研究などは、時間軸に沿って複数の測定点を持つので縦断研究と呼ばれる。近年、診断法を介入と見なして RCT が行われているが、これは横断研究により精度が高いと判定された診断法を健診などで用いることが予後を延長させるか検証するものであり、さらに上位の研究方法と言える。

診断以外の、治療や予後・原因、頻度などのカテゴリーを扱う場合は、横断研究が最も適切な研究デザインとはいえない（これらは縦断研究を用いるべきである）。

他のカテゴリーの場合：

治療→RCT

予後・原因→コホート(Cohort)研究, RCT, 症例対照研究

害→症例対照研究（稀な場合）, RCT（比較的頻度が多い場合）, コホート(Cohort)研究

頻度→横断研究

横断研究論文の構造

要約 abstract, summary

緒言 introduction

方法 methods

←チェックすべき項目はほとんどここにある！

結果 results

考察 discussion

研究が扱っている題材は、「要約 abstract, summary」に記載されており、これは PICO でまとめることができる。ただし、「要約 abstract, summary」の部分だけでは情報が不十分なことが多く、論文の「方法 methods」の項で詳細を確認することが必要である。

1 論文の PICO を探る

methods から PICO を読みとる

PICO とは疑問を定式化したものであり、どんな患者が (P; patient) , どんな治療や検査を受けるのは (I; intervention) , 何と比べて (C; comparison) , どうなるか (O; outcome) を一文にまとめたものである。

1-1. 論文の patient

患者 P(Patient) : エントリーの基準は何か?

inclusion criteria :

exclusion criteria :

対象患者の特徴を読みとる。

inclusion, exclusion criteria はどのように設定されているか?

inclusion の基準は何か? 以下の点に注目してみよう。

- ①年齢
- ②地域
- ③疾患の定義 (病期・重症度などを含む) : 特に, 実際の臨床上, その検査が使われる際に対象となる典型的な患者集団で評価されているか?

exclusion の基準は何か? その基準に含まれる項目は, 研究から除外することが妥当か?

1-2. 論文の intervention, comparison

介入 I(Intervention) :

比較 C(Comparison) :

診断に関する研究では,

intervention : 研究対象となる検査法, 診断法

comparison : gold standard に近い reference standard (後述)

の比較となる。

1-3. 論文の outcome

結果 O(Outcome) :

最も重要な結果は客観的でバイアスが避けられているか?ここで入りうる最も可能性の高いバイアスは**情報バイアス**と**測定バイアス**である。

診断の論文では, 対象疾患を Outcome とする。

したがって, 診断の横断研究における論文の PICO は, 以下のような形になることが多い。

P: どのような患者が

I: どのような検査を受けると

C: gold standard と比較して

O: どれだけ正確にその疾患の有無を診断できるか

2 C(Comparison)は gold standard に近いものになっているか？

- gold standard に近いものである
- gold standard に近いものではない
- それは何か？

記載のある可能性の高い場所

「要約 Abstract」の「方法 methods」

本文の「方法 methods」にある「研究デザイン study design, 臨床診断 clinical diagnosis」

記載箇所を見つけるためのキーワード

diagnosis/clinical diagnosis/final diagnosis/reference standard/gold standard(まれ)

PICO の Outcome の診断法を探せばよい。

<例>

The diagnosis of TB pleurisy was regarded as established when examination of pleural fluid or bronchial wash revealed the presence of AFB by microscopy or M. tuberculosis by culture or when pleural biopsy specimens yielded a positive culture or granulomatous inflammation with caseous necrosis on histology. Malignancy was diagnosed on pleural biopsy or bronchoscopy specimens. When no diagnosis could be made with all available information the effusion was deemed idiopathic.

gold standard とは、必ずその疾患の有無を見分けることができる診断法である。研究対象となる診断法がその疾患を診断するのに適切かどうかを検討するためには、PICO の C (Comparison) に gold standard を用いて、患者が真にその疾患に罹患しているかどうか、正しく確認されなければならない。仮に2つの診断法の優劣を比較したい場合には、直接比較せず、2つの診断法の各々について gold standard に対する正確度（感度・特異度、尤度比）を測定し、その結果を比較すべきである。

しかし、真実は神のみぞ知るものであって、gold standard は現実的には存在しない。したがって、gold standard の代わりに、既存の診断法の中でもっとも正確だと考えられているものを採用して研究が行われる。これを reference standard と呼ぶ。Reference standard として用いられる診断法は単独のものとは限らず、正しい診断を下すために、複数の診断法の組み合わせや複数の人によるコンセンサス、さらに疾患の有無を判断するために専門家の判定が必要な場合も多い。

論文中には、残念ながら gold standard の文字はないことが多い。基準となる診断法についての記述は比較的簡単に見つけることができるので、その診断法が gold standard に近いかどうかを判断するとよい。

なお、しばしば“golden standard”という語が用いられるが、誤りである。“gold standard”が正しい。

3 研究で行われた診断法は、いずれも全ての患者で行われているか？

- 全ての患者で行われている
- 一部の診断法を施行していない患者がいる
- それぞれの検査法を施行していない患者はどれくらいいるか？
- その患者はどのように扱われているか？

記載のある可能性の高い場所

「要約 Abstract」の「方法 methods」

患者の組み入れ、除外の流れがフローチャートで示されていることもある。

本文中に書かれていない場合は、研究に組み入れられた人数と、各診断法を行った人数を比較すればよい

研究対象となった診断法が有用か否かを評価する場合、その診断法を含む全ての診断法（対照となる reference standard も）が全ての患者で施行されるべきである。しかし、研究によっては一方しか行われていないことがある。仮に患者の一部がいずれかの診断法でしか評価されていないような場合は、他の診断法による結果が得られないため、結果の四分表から除外されてしまう。すると両者の診断法の比較にゆがみが生じてしまい、研究対象である診断法の正確度が過大評価されたり、過小評価されたりする。

4 研究で行われた診断法と最終診断の判定は、互いに独立に行われているか？

- 独立に確認されている
- 一方の診断法の結果を知らされた上で他方の診断法が施行されている
 - どのような順序で施行されたか？
 - その順序は、正確度の評価に影響を与えるか？
 - 与えない
 - 与える
 - どのような影響か？
- 独立に確認されたか否か不明

記載のある可能性の高い場所

「要約 Abstract」の「方法 methods」

本文の「方法 methods」にある「研究デザイン study design」

記載箇所を見つけるためのキーワード

Unaware/blinded/masked/without reference

これらのキーワードが存在しない場合も多い。記載があるとすれば、診断法の実施方法の箇所である。

研究で行われた診断法を評価する場合、一方の診断法の結果を知った状況で他方の診断法が施行されると、あとから行った診断法の結果が歪められてしまう（情報バイアス）。それぞれの診断法は、他の診断法の結果を知らされずに独立に行うべきである。

このとき、最初の検査の結果を知った上で最終診断を下す場合と、最終診断を知った上で診断法の結果を出す場合の2つのバイアスが起りうるが、例えば最終診断が合わせ技である場合には、その構成要素に最初の検査が入っていると、前者のバイアスは避けられない。

5 研究で行われた診断法と最終診断の判定は、いずれもその方法が明確に記載されているか？

- 研究対象となる診断法：
 - 実施方法が明確に記載されている
 - 実施方法が明確に記載されていない
- 最終診断（reference standard）：
 - 実施方法が明確に記載されている
 - 実施方法が明確に記載されていない

記載のある可能性の高い場所

「要約 Abstract」の「方法 methods」

本文の「方法 methods」にある「研究デザイン study design」

研究結果を自分の診療現場で応用するためには、研究で実施された診断法と最終診断（reference standard）が現場でも実施可能なように明確に、具体的に記載されている必要がある。

6 研究で行われた診断法と最終診断の判定は、いずれも再現性があるか？

- 再現性は明らかにある
- 再現性の評価はされている
 - 検査者内一致度，検査者間一致度は評価されているか？
 - 検査者内一致度は評価されている
 - 検査者間一致度は評価されている
 - κ 値：
- 再現性の評価はされていない

記載のある可能性の高い場所

「要約 Abstract」の「方法 methods」

本文の「方法 methods」にある「研究デザイン study design」, 「統計的解析 statistical analysis」

記載箇所を見つけるためのキーワード

reproducibility / κ / kappa

再現性についての記載はない場合も多い。

身体所見や画像診断を用いる場合など、結果の判定に経験を要するものや、評価者の主観が入ってしまうような診断法では、同じデータを評価しても、評価者によってその結果が一致しない可能性がある。また、たとえ同じ評価者でも、再び同じ診断法を行って同じ診断に至るとは限らない。したがって、このような診断法の性能を検討する場合は、複数の検査者間あるいは同一の検査者で繰り返し行った診断結果の一致度を評価しておく必要がある。

一致度

検査結果の一致度には検査者内一致度と検査者間一致度がある。検査者内一致度と検査者間一致度は全く異なるものである。

検査者内一致度：1人の検査者が2回検査を行って、その結果がどのくらい一致するかを示すものである。検査者内一致度が高い検査は、検査結果のブレが少ないことを意味する。検査者内一致度が低い検査では、1回の検査では結果の信頼性が低く、繰り返し検査を行って判定する必要がある。

検査者間一致度：複数の検査者が同じ検査を行ったときに、その結果がどのくらい一致するかを指すものである。検査者間一致度が高い検査は、個人個人の技量などの影響を受けず、誰が行っても同じ結果が出ることが期待される検査である。一方、検査者間一致度が低い検査は、自分がその検査を行った場合にどのような結果を出しやすいか、把握しておく必要がある。

なお、この場合の“検査者”とは、検査そのものを実施する人、検査結果を判定する人、さらに最終的な診断を下す人などを全て含む。

カッパ κ 値 (κ 統計量)

診断の一致度を評価する指標として、現在最も一般的に用いられているのは **κ 統計量 kappa statistic** と呼ばれるものである。これは、偶然によらずに診断が一致する割合と定義される。検査値などは評価者の主観が入りにくい、κ 統計量は検査機器による測定の実現性を評価する場合にも用いることができる。

$$\kappa = \frac{Po - Pe}{1 - Pe}$$

例：呼吸器症状のある患者 100 人の胸部単純 X 線写真を 2 名の医師が読影した結果の一致性を評価する。診断結果が表のようになったとする。

1 人目の医師	2 人目の医師			計
	肺炎	肺結核	肺癌	
肺炎	61	3	1	65
肺結核	10	7	3	20
肺癌	4	5	6	15
計	75	15	10	100

この表で、2人の医師の診断が完全に一致する割合 P_o は、単純に

$$P_o = (61+7+6) / 100 = 0.74$$

である。しかし、この割合の中には2人の医師がX線写真など関係なくでたために患者を分類したときに偶然診断が一致する確率 P_e も含まれている。すなわち、1人目の医師が肺炎と診断する確率が0.65 (=65/100)、2人目が0.75 (=75/100) であるから、偶然に2人の医師がともに肺炎と診断する確率は、 0.65×0.75 になる。肺結核と肺癌にもこれが当てはまるため、

$$P_e = (0.65 \times 0.75) + (0.20 \times 0.15) + (0.15 \times 0.10) = 0.5325$$

となる。 κ 値は偶然に寄らず診断が一致する割合であるから、この確率を差し引いて、

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.74 - 0.5325}{1 - 0.5325} = 0.44$$

となる。この κ 値の式は、判定者間一致度だけでなく、判定者内一致度にも適用することができる。 κ 値は0~1の値を取りうるが、その評価には以下のような目安がある。

- 0 : 偶然の一致
- 0~0.4 : わずかな一致
- 0.4~0.6 : 中等度の一致
- 0.6~0.8 : 相当の一致
- 0.8~1.0 : ほぼ完璧な一致

κ 値の信頼区間を求めることもでき、標準誤差 SE を

$$SE = \sqrt{\frac{P_o(1 - P_o)}{N(1 - P_o)^2}}$$

の様に近似して、

$$\kappa - 1.96 \times SE \sim \kappa + 1.96 \times SE$$

が κ 値の95%信頼区間となる。この範囲が0を跨がなければ、 κ の値は有意であると言える。ただし、これは κ 値の値そのものが有意であるかどうかを検定しているのであって、2つの診断が有意に一致しているかどうかを示しているわけではない。2つの診断の一致度は κ 値の大きさと評価する。

ただし、 κ 値には著しく診断に偏りのある場合は、 κ 値はかえって低くなるという問題がある。

例えば、先ほどの呼吸器症状のある患者100人の胸部単純X線写真を2名の医師が読影した結果の一致性を評価する場合で、そのほとんどが肺炎で、診断結果が表のようになったとする。

1人目の医師	2人目の医師			計
	肺炎	肺結核	肺癌	
肺炎	98	0	1	99
肺結核	0	0	0	0
肺癌	0	1	0	1
計	98	1	1	100

この表で、2人の医師の診断が完全に一致する割合 P_o は、単純に

$$P_o = (98+0+0) / 100 = 0.98$$

である。この割合の中には2人の医師がX線写真など関係なくでたために患者を分類したときに偶然診断が一致する確率 P_e は、

$$P_e = (0.99 \times 0.98) + (0.00 \times 0.01) + (0.01 \times 0.01) = 0.9703$$

となる。 κ 値は偶然に寄らず診断が一致する割合であるから、この確率を差し引くと、

$$\kappa = \frac{P_o - P_e}{1 - P_e} = \frac{0.98 - 0.9703}{1 - 0.9703} = 0.33$$

となる。この値は一致率が高いにもかかわらず、先ほどの κ 値よりも低い。

7 結果の評価

□有病割合 prevalence = $\frac{a+c}{a+b+c+d}$:

□感度 Sn = $\frac{a}{a+c}$:

□特異度 Sp = $\frac{d}{b+d}$:

□陽性的中率 PPV = $\frac{a}{a+b}$:

□陰性的中率 NPV = $\frac{d}{c+d}$:

□陽性尤度比 LR+ = $\frac{b+d}{a+c} \times \frac{a}{b} = \frac{Sn}{1-Sp}$:

□陰性尤度比 LR- = $\frac{b+d}{a+c} \times \frac{c}{d} = \frac{1-Sn}{Sp}$:

□検査前オッズ pre-test Odds = $\frac{\text{prevalance}}{1-\text{prevalance}}$:

□検査後オッズ post-test Odds = pre-test Odds × LR :

□検査後割合 post-test probability = $\frac{\text{post-test Odds}}{\text{post-test Odds}+1}$:

		疾患の有無		
		(+)	(-)	
診断法の結果	(+)	a	b	a+b
	(-)	c	d	c+d
合計		a+c	b+d	a+b+c+d

従来、「有病率」，「事前（後）確率」，「検査前（後）確率」の語が用いられていたが，これは正確には誤りである。「率」は時間の概念を含む言葉であり，「罹患率」や「死亡率」等で用いるのが正しい。一方，時間の概念を含まず全体の中のある部分を示す場合には，「割合」の語を用いるのが正しく，「有病割合」，「事前（後）割合」，「検査前（後）割合」などというべきである。

診断法の陽性・陰性，および疾患の有無 (gold standard による陽性・陰性) により，2×2 表 (四分表) が作成できる。論文中に掲載されている場合もあるが，感度・特異度のみを記載している論文でも，自分で2×2表を作成すると理解しやすい。

		疾患の有無		合計
		(+)	(-)	
診断法の結果	(+)	a	b	a+b
	(-)	c	d	c+d
合計		a+c	b+d	a+b+c+d

1) 感度・特異度 Sensitivity and Specificity

感度 sensitivity; Sn : 疾患を持つ群での診断法の陽性の割合 (真陽性割合)

特異度 specificity; Sp : 疾患を持たない群での診断法の陰性の割合 (真陰性割合)

感度が高いということは，疾患を持っている患者ではその診断の結果はほとんど陽性になるということである。ということは，感度が高い診断法を用いれば，疾患を持っている患者を拾い上げすぎることにはあっても，見落としがなくなる。つまり，**感度(Sn)が高い診断法で陰性(Negative)だったとき，疾患を除外(rule out)できる**ということ，SnNOuT と覚える。感度が高い診断法は，陰性の時に威力を発揮し，除外診断に有用である。

一方，特異度が高いということは，疾患を持っていない患者ではその診断の結果はほとんど陰性になるということである。ということは，特異度が高い診断法を用いれば，疾患を持っていない患者を拾い上げすぎることにはあっても，偽陰性は減らせる。つまり，**特異度(Sp)が高い診断法で陽性(Positive)だったとき，疾患を確定(rule in)できる**ということ，SpPIIn と覚える。特異度が高い診断法は，陽性の時に威力を発揮し，確定診断に有用である。

感度・特異度は縦方向に集計しているため有病割合の影響を受けない。このため，集団が異なっても，(原則として) **感度・特異度は診断法固有の特性を純粹に示す**ことができる。

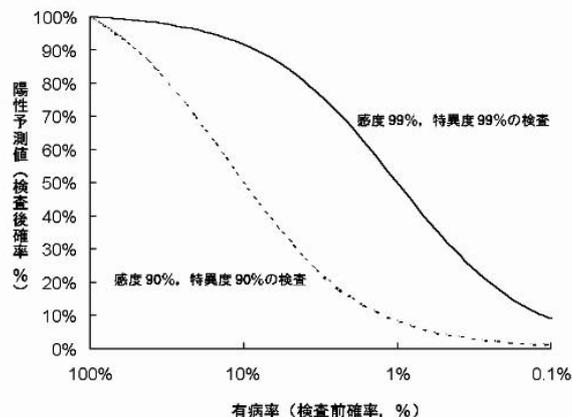
2) 的中率 Predictive value;PV

陽性的中率 positive predictive value;PPV : 診断結果が陽性の場合に疾患を持つ割合

陰性的中率 negative predictive value;NPV : 診断結果が陰性の場合に疾患を持たない割合

的中率(予測率, PV)は、横方向に集計するため、有病率の影響を受ける。従って、ある研究で得られた的中率は、その集団にしか適用できない。集団が変われば、同じ診断法を用いても中率は変わってくる。RCTにおけるARRとNNTの扱いに似ている。

臨床現場では検査結果から疾患の有無を推定するので、感度・特異度よりも陽性・陰性的中率の方が役に立つが、その臨床現場での有病率や患者自身の事前割合が分かっている必要がある。



3) 尤度比 Likelihood ratio;LR

陽性尤度比 positive likelihood ratio;LR+

陰性尤度比 negative likelihood ratio;LR-

尤度比は、その診断法を用いる前後で「疾患のある可能性」を変える便利な数字である。「疾患のある可能性」を確率ではなくオッズで示すと、以下のベイズの定理 Bayes' theorem と呼ばれる式が成り立つことが分かっている。

$$\text{検査前オッズ} \times \text{尤度比} = \text{検査後オッズ}$$

検査前オッズは検査前割合(有病割合)から、検査後割合は検査後オッズから計算できる。すなわち、検査前割合と尤度比が分かれば、検査後割合が求まる。

尤度比には陽性尤度比と陰性尤度比がある。診断法が陽性だったときに用いるのが陽性尤度比で、診断法が陰性だったときに用いるのが陰性尤度比である。

		Target disorder (iron deficiency anemia)		Totals
		(+)	(-)	
Diagnostic test result (serum ferritin)	(+) (<65mmo/L)	731	270	1001
	(-) (≥65mmo/L)	78	1500	1578
Totals		809	1770	2579

$$\text{感度 } Sn = \frac{a}{a+c} = \frac{731}{809} = 90\%$$

$$\text{陰性的中率 } NPV = \frac{d}{c+d} = \frac{1500}{1578} = 95\%$$

$$\text{特異度 } Sp = \frac{d}{b+d} = \frac{1500}{1770} = 85\%$$

$$\text{陽性尤度比 } LR+ = \frac{Sn}{1-Sp} = \frac{90\%}{15\%} = 6$$

$$\text{陽性的中率 } PPV = \frac{a}{a+b} = \frac{731}{1001} = 73\%$$

$$\text{陰性尤度比 } LR- = \frac{1-Sn}{Sp} = \frac{10\%}{85\%} = 0.12$$

特異度 100%の時には陽性尤度比が無限大、感度 100%の時には陰性尤度比が 0 となる。

ただし、現実的に感度や特異度が 100%になることはあり得ないので、偽陽性や偽陰性が 0 の場合には、症例数が少ないと考えるのが妥当である。この場合の 1 つの考えとして、以下のように四分表の全てのマスに 0.5 を加えて感度・特異度、尤度比を計算する方法がある。当然、症例数が多いほど、加えた 0.5 の影響は小さくなる。

		疾患の有無		合計
		(+)	(-)	
診断法の結果	(+)	21	105	126
	(-)	0	623	623
合計		21	728	749



		疾患の有無		合計
		(+)	(-)	
診断法の結果	(+)	21.5	105.5	127
	(-)	0.5	623.5	624
合計		22	729	751

$$\text{感度 } Sn = 21/21 = 100\%$$

$$\text{特異度 } Sp = 623/728 = 86\%$$

$$\text{陽性尤度比 } LR+ = 1/(1-0.86) = 7.1$$

$$\text{陰性尤度比 } LR- = (1-1)/0.86 = 0$$

$$\text{感度 } Sn = 21.5/22 = 98\%$$

$$\text{特異度 } Sp = 623.5/729 = 86\%$$

$$\text{陽性尤度比 } LR+ = 0.98/(1-0.86) = 7$$

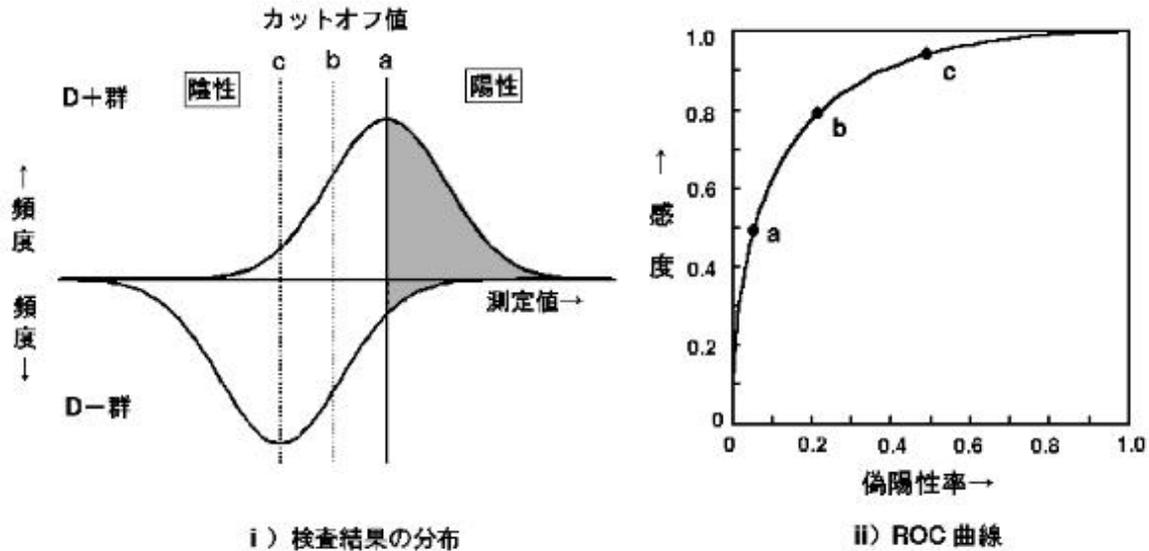
$$\text{陰性尤度比 } LR- = (1-0.98)/0.86 = 0.02$$

4) ROC 曲線 Receiver operating characteristic curve

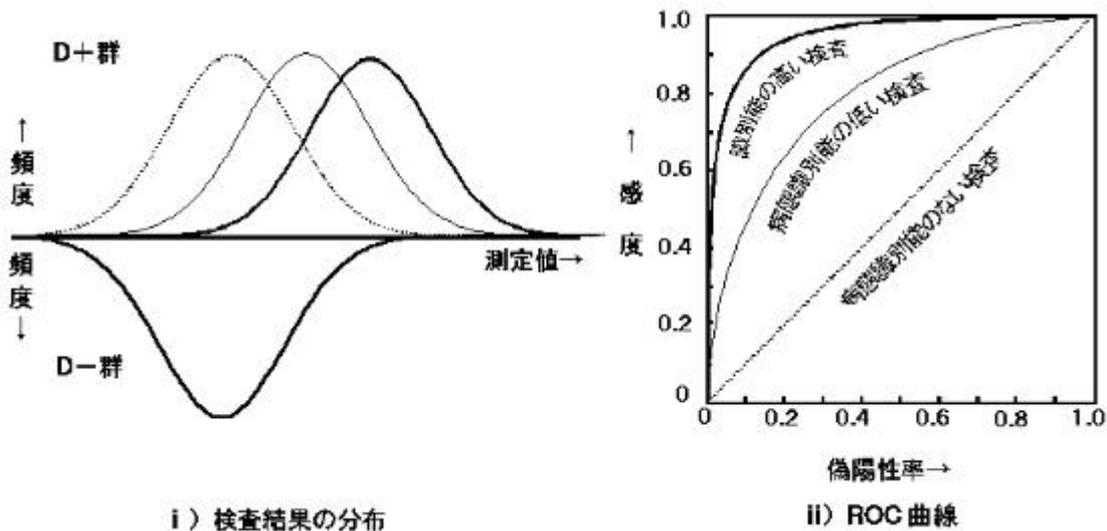
定量的な診断法の場合，疾患を持つ群（D+群）と持たない群（D-群）で測定値とそれに対応する頻度の分布を図示すると下図 i) の様に重なり合う．このとき，ある一定の値を cut off 値として検査陽性・陰性を 2 分する事ができ，定性的な診断法と同様にして感度・特異度が求められる．

定量的な診断法では，感度・特異度は固定した値ではなく，cut off 値を変更させれば変動する．通常，cut off 値を図の a→b→c 点と変更していくと，感度は高くなるが，特異度は低下する．

このように，感度と特異度には trade off 関係があつて，単一の cut off 値での感度・特異度だけでは，診断能を十分に評価できない．そこで，縦軸に真陽性率（感度），横軸に偽陽性率（=1-特異度）を取って，cut off 値を変更した場合の両者の変化を順次プロットしていくと下図 ii) に示すような曲線となる．これを ROC 曲線と呼ぶ．



ROC 曲線には，選択しうる全ての cut off 値での感度・特異度が描出される．そのため，任意の cut off 値を選択した場合の診断能を容易に知ることができ，また，最適な cut off 値を決定することにも用いられる．すなわち，感度が高く，偽陽性割合が低い点が最適な cut off 値となり，下図 ii) で最も左上隅に近い点を cut off 値として選ぶべきである．また，複数の診断法の優劣も ROC 曲線で評価することができ，優れた診断法は ROC 曲線が左上に寄っていく．このとき，曲線の下側の面積を **area under the curve; AUC** といい，さまざまな cut off 値をとるこの検査が総合的にみてどれだけ正しく診断できるか，すなわち**正診率**を表す．AUC は 0~1 の値を取りうるが，0.5 では五分五分であり，コイントスをして決めるのと同じことなので，診断を付けることに全く役立たない．通常は 0.5~1.0 の間で表現され，たとえば AUC=0.7 は正診率 70%を意味する．AUC が大きいほど優れた診断法となる．



5) 様々な事前割合と尤度比による事後割合の変化

下の表は、様々な事前確率について、異なる尤度比によって事後確率がどのように変化するかを、ベイズの定理を用いて計算した結果である。

事前割合 (%)	事前オッズ	尤度比										
		20	10	5	2	1.5	1	0.67	0.5	0.2	0.1	0.05
1%	1/99 = 0.01	16.8%	9.2%	4.8%	2.0%	1.5%	1.0%	0.7%	0.5%	0.2%	0.1%	0.1%
5%	5/95 = 0.05	51.3%	34.5%	20.8%	9.5%	7.3%	5.0%	3.4%	2.6%	1.0%	0.5%	0.3%
10%	10/90 = 0.11	69.0%	52.6%	35.7%	18.2%	14.3%	10.0%	6.9%	5.3%	2.2%	1.1%	0.6%
20%	20/80 = 0.25	83.3%	71.4%	55.6%	33.3%	27.3%	20.0%	14.3%	11.1%	4.8%	2.4%	1.2%
30%	30/70 = 0.43	89.6%	81.1%	68.2%	46.2%	39.1%	30.0%	22.3%	17.6%	7.9%	4.1%	2.1%
40%	40/60 = 0.67	93.0%	87.0%	76.9%	57.1%	50.0%	40.0%	30.9%	25.0%	11.8%	6.3%	3.2%
50%	50/50 = 1	95.2%	90.9%	83.3%	66.7%	60.0%	50.0%	40.1%	33.3%	16.7%	9.1%	4.8%
60%	60/40 = 1.5	96.8%	93.8%	88.2%	75.0%	69.2%	60.0%	50.1%	42.9%	23.1%	13.0%	7.0%
70%	70/30 = 2.33	97.9%	95.9%	92.1%	82.4%	77.8%	70.0%	61.0%	53.8%	31.8%	18.9%	10.4%
80%	80/20 = 4	98.8%	97.6%	95.2%	88.9%	85.7%	80.0%	72.8%	66.7%	44.4%	28.6%	16.7%
90%	90/10 = 9	99.4%	98.9%	97.8%	94.7%	93.1%	90.0%	85.8%	81.8%	64.3%	47.4%	31.0%
95%	95/5 = 19	99.7%	99.5%	99.0%	97.4%	96.6%	95.0%	92.7%	90.5%	79.2%	65.5%	48.7%
99%	99/1 = 99	99.9%	99.9%	99.8%	99.5%	99.3%	99.0%	98.5%	98.0%	95.2%	90.8%	83.2%

注) 事後割合 10%以下：青色，10～50%：水色，50～90%：黄色，90%以上：橙色で示している

ここで、例えば、陽性尤度比が2、陰性尤度比が0.2の検査があったとする。事前割合5%の患者がこの検査で陽性となった場合、上の表から、事後割合は9.5%となる。逆に、同じ患者がこの検査で陰性となった場合は、事後割合は2.6%と低くなる。さらに、この同じ検査について、事前割合が50%、90%の場合の事後割合を求めると、以下のようになる。事前割合が5%や90%とかなり低かったり高かったりする場合には、事後割合の変化は比較的小さいことが分かる。

事前割合	検査結果	事後割合
5%	陽性	9.5%
	陰性	2.6%
50%	陽性	66.7%
	陰性	16.7%
90%	陽性	94.7%
	陰性	64.3%

上の表を事前割合と事後割合の差に置き換えて、10～20%、20～30%、30%以上で色分けして示したのが下の表である。これを見ると、先程のように事前割合がかなり高い場合や、かなり低い場合には、検査を行っても事後割合に大きな変化はみられない（下表の上の方や下の方）。つまり、事前に疾患の有無が比較的高い可能性で予想される場合に敢えて検査を行う意味は乏しく、疾患があるかないか判断に迷っている場合にこそ、検査を行う価値が高いといえるのである。臨床上、検査の価値が大きいのは、下の図で色が付いている部分である。

事前割合 (%)	事前オッズ	尤度比										
		20	10	5	2	1.5	1	0.67	0.5	0.2	0.1	0.05
1%	1/99 = 0.01	16%	8%	4%	1%	0%	0%	0%	0%	-1%	-1%	-1%
5%	5/95 = 0.05	46%	29%	16%	5%	2%	0%	-2%	-2%	-4%	-4%	-5%
10%	10/90 = 0.11	59%	43%	26%	8%	4%	0%	-3%	-5%	-8%	-9%	-9%
20%	20/80 = 0.25	63%	51%	36%	13%	7%	0%	-6%	-9%	-15%	-18%	-19%
30%	30/70 = 0.43	60%	51%	38%	16%	9%	0%	-8%	-12%	-22%	-26%	-28%
40%	40/60 = 0.67	53%	47%	37%	17%	10%	0%	-9%	-15%	-28%	-34%	-37%
50%	50/50 = 1	45%	41%	33%	17%	10%	0%	-10%	-17%	-33%	-41%	-45%
60%	60/40 = 1.5	37%	34%	28%	15%	9%	0%	-10%	-17%	-37%	-47%	-53%
70%	70/30 = 2.33	28%	26%	22%	12%	8%	0%	-9%	-16%	-38%	-51%	-60%
80%	80/20 = 4	19%	18%	15%	9%	6%	0%	-7%	-13%	-36%	-51%	-63%
90%	90/10 = 9	9%	9%	8%	5%	3%	0%	-4%	-8%	-26%	-43%	-59%
95%	95/5 = 19	5%	4%	4%	2%	2%	0%	-2%	-5%	-16%	-29%	-46%
99%	99/1 = 99	1%	1%	1%	0%	0%	0%	0%	-1%	-4%	-8%	-16%

注) 事前割合と事後割合の差が10～20%：黄色または水色，20～30%：橙色または薄青色，30%以上：赤色または青色で示している

尤度比別の事前割合から事後割合への変化度について、オッズに変換してベイズの定理により事後割合を算出するのは煩雑であり、また一方で上記の表を覚えるのも現実的ではない。そこで McGee らは尤度比別の事前割合から事後割合への変化度について、近似式を用いて変化度を推定することを提唱している¹⁰⁾。

McGee らは、この方法によって事前割合が 10~90% の場合には、変化度は $0.19 \times \log LR$ で近似できるとしている。これを当てはめると、右表のように、尤度比 2 であれば +15%、尤度比 5 であれば +30%、尤度比 10 であれば +45% と近似できるとしている。一方、尤度比 0.5 (1/2) の場合は -15%、尤度比 0.2 (1/5) の場合は -30%、尤度比 0.1 (1/10) の場合は -45% と近似できることになる。

しかし、この方法には限界がある。前頁の表の通り、例えば尤度比 2 の場合に +15% と近似できるのはせいぜい事前割合が 20~70%、尤度比 5 の場合に +30% と近似できるのは 10~60%、尤度比 10 の場合に +45% と近似できるのは 10~50% の場合に限られる。考えても見れば分かることであるが、事前割合 90% の場合に尤度比 2 で +15% だと事後割合は 100% を超えてしまうのでナンセンスである。

もっとも、診断の場合には、とても可能性が高い場合ととても可能性が低い場合、つまり事前割合が 0% や 100% に近い場合は元から結果が分かっているわけで、病歴、身体診察、検査所見などといった情報が入ってもそれには影響されにくいのであまり問題にならない。

ところで、陽性尤度比の場合は、尤度比が大きくなる毎に近似できる事前割合は下がる。一般的に鑑別診断は複数あるわけだが、事前割合は 50% よりも小さいことの方が多いため、陽性尤度比を考える場合にはあまり問題にはならない。

一方、陰性尤度比の場合には、尤度比 0.5 の場合に -15% と近似できるのはせいぜい事前割合が 30~80%、尤度比 0.2 の場合に -30% と近似できるのは 40~90%、尤度比 0.1 の場合に -45% と近似できるのは 50~90% の場合に限る。先に述べたように一般的に事前割合は小さいことが多いので、この近似を適用するのは困難である。

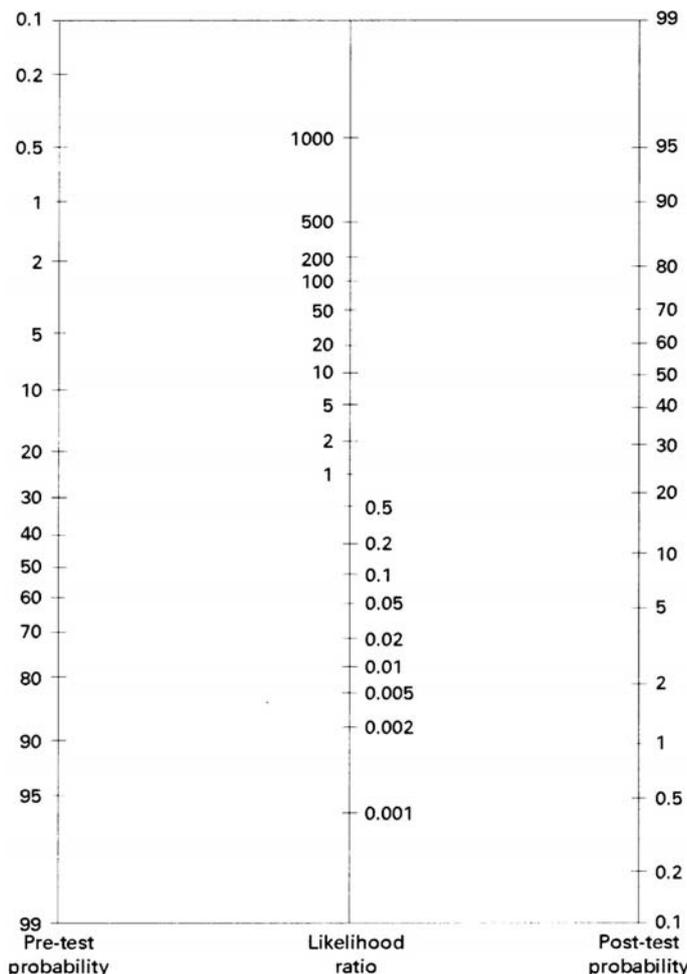
尤度比	事前割合から事後割合へ予想される変化度
0.1	-45%
0.2	-30%
0.3	-25%
0.4	-20%
0.5	-15%
1	±0%
2	+15%
3	+20%
4	+25%
5	+30%
6	+35%
7	
8	+40%
9	
10	+45%

6) Fagan nomogram

Fagan nomogram はある診断法を適用した結果のあとで、患者が疾患を持つ可能性がどれだけ変わったかを推定するための図式的なツールである (NEJM 1975;293:257)。

目の前の患者において、診断法を適用する前の、疾患可能性を見積もり、検査法の尤度比が分かれば、その点を結んで延ばしたところに診断法を適用したあとの疾患可能性が求まる。

このノモグラムの優れたところは、いちいちオッズを計算しなくても、尤度比さえ分かっているならば、事前割合から事後割合を直接知ることができるというところである。



参考文献

- 1) Jaeschke R, Guyatt G, Sackett DL, for the Evidence-Based Medicine Working Group: Users' guides to the medical literature. III: how to use an article about a diagnostic test, A: are the results of the study valid? JAMA 1994;271:389-391.
- 2) 開原成允, 浅井泰博, 診断用検査に関する文献の使い方, JAMA 医学文献の読み方, 中山書店 2001 年, 37-64.
- 3) Sackett DL et al. Evidence-Based Medicine, How to Practice and Teach EBM. Churchill Livingstone 2000, 67-93.
- 4) Bedenoch D 他著, 齊尾武郎監訳, EBM の道具箱. 中山書店 2002 年, 25-32.
- 5) 名郷直樹著, 続 EBM 実践ワークブック. 南江堂 2002 年, 80-140.
- 6) 日本大学医学部公衆衛生学教室, EBM のための臨床疫学入門講座, 診断の適否. http://www.med.nihon-u.ac.jp/department/public_health/ebm/ce201.html
- 7) 日本臨床検査医会. http://www.jaclap.org/seminar_qm1.html#QM0011_hint2
- 8) Fagan nomogram. Fagan TJ. Letter: Nomogram for Bayes theorem. N Engl J Med 1975;293(5):257.
- 9) Claassen JAHR. The gold standard: not a golden standard. BMJ. 2005;330(7500):1121-1121.
- 10) McGee S. Simplifying likelihood ratios. J Gen Intern Med. 2002 Aug;17(8):646-649.

改訂履歴

- | | |
|--|--|
| <p>1.0→1.1 (第 2 期第 4 回 pES club で使用後の改訂)</p> <ul style="list-style-type: none"> ・採用論文を記載する欄を設置 ・尤度比の項の表を修正 ・Fagan nomogram を拡大 <p>1.1→1.2 (2003.7.2)</p> <ul style="list-style-type: none"> ・gold standard の採用についての記述を追加 <p>1.2→2.0 (2004.2.28)</p> <ul style="list-style-type: none"> ・書き込み用 CAT sheet の改変 (1 ページ化) ・κ 統計量の記載を微修正 ・結果の評価の項を微修正, 2×2 表の追加 ・ROC 曲線の項を微修正 ・ノモグラムの項に特性についてを追加 <p>2.0→2.1 (2004.2.29, 第 2 期第 2 回 pES club で使用後の改訂)</p> <ul style="list-style-type: none"> ・誤字修正 ・論文の PECO を探るの項を改変 ・ノモグラムの拡大 <p>2.1→3.0 (2007.2.28)</p> <ul style="list-style-type: none"> ・レイアウトの変更 ・記述の探し方のガイドを新設 <p>3.0→3.1 (2007.4.4)</p> <ul style="list-style-type: none"> ・「3. 研究対象となる診断法と gold standard は全ての患者で独立に行われているか?」の項を, 「3. 研究で行われた診断法は, いずれも全ての患者で行われているか?」と「4. 研究で行われた診断法と最終診断の判定は, いずれも独立に行われているか?」に分割 ・「6. 研究で行われた診断法と最終診断の判定は, いずれも再現性があるか?」の項に, 一致度の項を新設し, 検査者内一致度と検査者間一致度について言及 | <p>3.1→4.0 (2007.7.9)</p> <ul style="list-style-type: none"> ・「7. 結果の評価」に「5) 様々な事前確率と尤度比による事後確率の変化」の項を新設 <p>4.0→4.1 (2008.2.29)</p> <ul style="list-style-type: none"> ・記述の探し方のガイドを新設 ・診断法と最終診断の判定は互いに独立に行われているかの項に, 不明の選択肢を追加 ・偽陽性や偽陰性がなかった場合のデータの処理の仕方についての記載を追加 <p>4.1→5.0 (2010.3.25)</p> <ul style="list-style-type: none"> ・Gold standard についてのチェック項目名, 記述を変更. Gold standard は存在しないことを追加. ・的中率, ROC 曲線についての説明を修正. <p>5.0→6.0 (2011.2.25)</p> <ul style="list-style-type: none"> ・PECO を PICO に変更 ・誤字修正. <p>6.0→6.1 (2011.5.8)</p> <ul style="list-style-type: none"> ・「率」と「割合」の違いについての記載を追加. <p>6.1→6.2 (2012.2.19)</p> <ul style="list-style-type: none"> ・「研究で行われた診断法と最終診断の判定は, 互いに独立に行われているか?」のキーワードに “without reference” を追加. ・誤字修正 <p>6.2→6.3 (2012.3.19)</p> <ul style="list-style-type: none"> ・“golden standard” という言葉は存在しない旨を追記. <p>6.3→6.4 (2012.8.19)</p> <ul style="list-style-type: none"> ・尤度比毎の事前割合から事後割合への変化度の表を修正. ・McGee らによる事前割合から事後割合への変化度の推定についての記述を追加. ・Fagan nomogram の表を変更. |
|--|--|

Critically appraised topic for Diagnostic study

Reviewer: _____ 年 月 日

authors : _____
 title : _____
 citation : _____
 PubMed PMID : _____

1. 論文の PICO を探る

P : _____

 I : _____

 C : _____

 O : _____

2. C(Comparison)は gold standard に近いものになっているか？

gold standard に近いものである gold standard に近いものではない
 →それは何か？ (_____)

3. 研究で行われた診断法は、いずれも全ての患者で行われているか？

全ての患者で行われている
 一部の診断法を施行していない患者がいる
 →それぞれどのくらい？ (_____)
 →どのように扱われている？ (_____)

研究対象となる診断法や gold standard を施行していない患者がいる

4. 研究で行われた診断法と最終診断の判定は、互いに独立に行われているか？

独立に確認されている
 一方の診断法の結果を知らされた上で他方の診断法が施行されている
 →どのような順序？
 →その順序は、正確度の評価に影響を与えるか？ 与えない 与える
 独立に確認されたか否か不明

5. 研究で行われた診断法と最終診断の判定は、いずれもその方法が明確に記載されているか？

研究対象となる診断法： 記載されている 記載されていない
 最終診断 (reference standard)： 記載されている 記載されていない

6. 研究で行われた診断法と最終診断の判定は、いずれも再現性があるか？

再現性は明らか
 評価はされている
 →検査者内一致度、検査者間一致度は評価されているか？ 検査者内一致度 検査者間一致度
 κ 値：
 再現性の評価はない

7. 結果の評価

有病割合 = $(a+c) / (a+b+c+d) = (\quad \%)$

感度 = $a / (a+c) = (\quad \%) = Sn$

特異度 = $d / (b+d) = (\quad \%) = Sp$

陽性的中率 = $a / (a+b) = (\quad \%) = PPV$

陰性的中率 = $d / (c+d) = (\quad \%) = NPV$

陽性尤度比 = $LR+ = Sn / (1-Sp) = (\quad)$

陰性尤度比 = $LR- = (1-Sn) / Sp = (\quad)$

その他の評価方法：

		疾患の有無		
		(+)	(-)	
診断法の 結果	(+)	a	b	a+b
	(-)	c	d	c+d
		a+c	b+d	a+b+c+d