

生物統計基礎セミナー 第10回

2014年6月11日

診断法評価の統計的側面

国立がん研究センター 生物統計部門

野村 尚吾

OUTLINE

- 結果の予測に関心がある場合
 - 二値マーカーの場合
(感度/特異度/PPV/NPV)
 - 連続マーカーの場合(ROCとc統計量)
 - 生存時間の場合
 - 複数マーカーで予測する場合
- 診断法間の一致度に関心がある場合
 - McNemar検定、 κ 係数
- 事例紹介

全体を通して

- 本セミナーでは2×2分割表を頻用します

		真の分類		計
		陽性	陰性	
検査 (Marker)	陽性	a	b	a+b
	陰性	c	d	c+d
計		N_1	N_2	N

- 「行：検査(診断)の結果、列：真の分類」
として統一しています



肺がんの遺伝子検査の例

- ある遺伝子変異(変異割合5%)を1000例で検査

		真の分類		計
		陽性	陰性	
検査	陽性	45	50	95
	陰性	5	900	905
計		50	950	1000

- この検査の診断能を評価する方法は？



最もシンプルな評価指標

※「1 - 誤分類率」は「正診率(正しく診断できた割合)」と呼ばれることがあります

■ 誤分類率(Misclassification rate)

－ 定義：真の分類と一致しなかった割合

		真の分類		計
		陽性	陰性	
検査	陽性	45	50	95
	陰性	5	900	905
計		50	950	1000

$$\text{誤分類率} = \frac{5 + 50}{1000} = 5.5\%$$



誤分類率の問題点

- 450例(全例陽性)を追加検査した場合
 - 検査結果が再現されるとした場合の2×2表

ここが単純に10倍になるだけ

		真の分類		計
		陽性	陰性	
検査	陽性	450	50	500
	陰性	50	900	950
計		500	950	1450

$$\text{誤分類率} = \frac{50 + 50}{1450} = 6.9\%$$

- 検査対象によって結果が変わってしまう
- 陽性例・陰性例の検査を個別に評価できない

感度と特異度

感度をsensitivity、特異度をspecificityという

■ 真の分類を正しく予測できた割合

表を縦にみる ↓↓↓↓↓↓↓↓

		真の分類		
		陽性	陰性	計
検査	陽性	45	50	95
	陰性	5	900	905
計		50	950	1000

$$\text{感度} = \frac{45}{45 + 5} = 90.0\%$$

$$\text{特異度} = \frac{900}{50 + 900} = 94.7\%$$



追加検査後のデータ

※ 追加症例も同じ精度(感度・特異度)で検査できたと仮定しています

		真の分類		
		陽性	陰性	計
検査	陽性	450	50	500
	陰性	50	900	950
計		500	950	1450

- 感度 = 90.0% (=450/500) ・ 特異度は不変
- データによらず(真の陽性割合によらず)、検査の診断能を評価できる



PPVとNPV

- 検査結果が的中したか否かを測る指標
 - PPV(Positive predictive value; 陽性的中度)
 - 検査陽性例のうち、真に陽性だった症例の割合
 - NPV(Negative predictive value; 陰性的中度)
 - 検査陰性例のうち、真に陰性だった症例の割合





表を横にみる



		真の分類		計
		陽性	陰性	
検査	陽性	45	50	95
	陰性	5	900	905
計		50	950	1000

$$\text{PPV} = \frac{45}{45 + 50} = 47.4\%$$
$$\text{NPV} = \frac{900}{5 + 900} = 99.4\%$$



真の陽性割合とPPV・NPVの関係

■ 感度90.0%・特異度94.7%の2×2表データ

(1) 真の陽性割合5%

		真の分類		計
		陽性	陰性	
検査	陽性	45	50	95
	陰性	5	900	905
計		50(5%)	950	1000

$$\text{PPV} = \frac{45}{45 + 50} = 47.4\%$$
$$\text{NPV} = \frac{900}{5 + 900} = 99.4\%$$

(2) 真の陽性割合20%

		真の分類		計
		陽性	陰性	
検査	陽性	180	43	223
	陰性	20	757	777
計		200(20%)	800	1000

$$\text{PPV} = \frac{180}{180 + 43} = 80.7\%$$
$$\text{NPV} = \frac{757}{20 + 757} = 97.4\%$$



- 同じ精度(感度・特異度)の検査であっても、**PPV/NPVは真の陽性割合に影響を受ける**
 - 検査の「精度」は感度/特異度で評価すべき
 - 真の陽性割合に影響を受ける指標と受けない指標を区別して用いる必要がある



OUTLINE

■ 結果の予測に関心がある場合

- 二値マーカーの場合
(感度/特異度/PPV/NPV)
- 連続マーカーの場合(ROCとc統計量)
- 生存時間の場合
- 複数マーカーで予測する場合

■ 診断法間の一致度に関心がある場合

- McNemar検定、 κ 係数

■ 事例紹介

連続量マーカーによる検査

■ 年齢が高いと合併症の発生頻度が高い？

※ 200例のデータ(年齢を昇順に並び替えたもの)

ID	年齢	合併症
1	22	なし
2	24	なし
3	32	なし
...		
100	46	あり
101	48	なし
102	49	あり
103	50	あり
...		
199	68	あり
200	74	あり

合併症低リスク群
(検査陰性)

閾値 = 47歳

合併症高リスク群
(検査陽性)

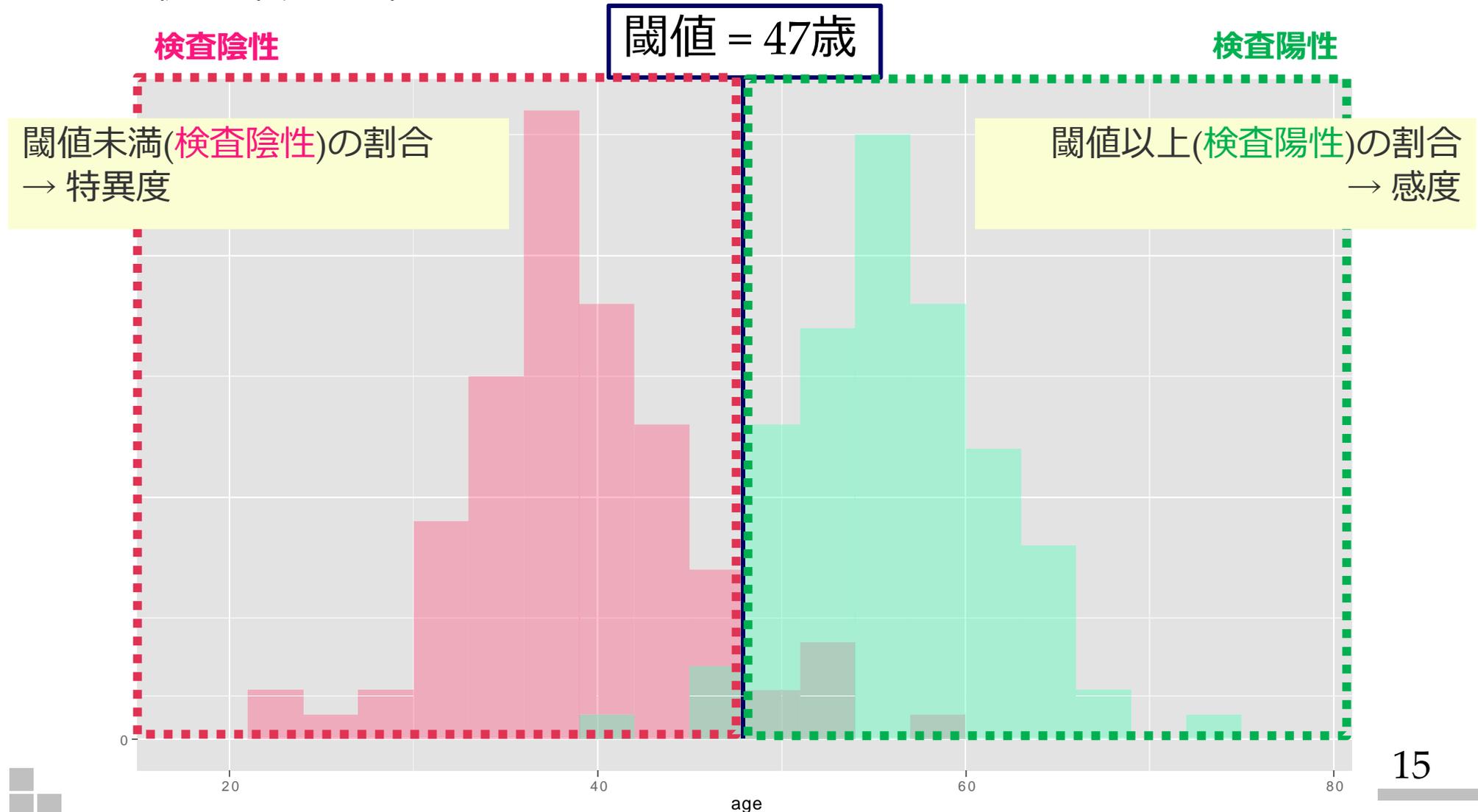


閾値と感度・特異度の関係

注：仮想例

合併症なし
合併症あり

※ 合併症の有無別の年齢ヒストグラム

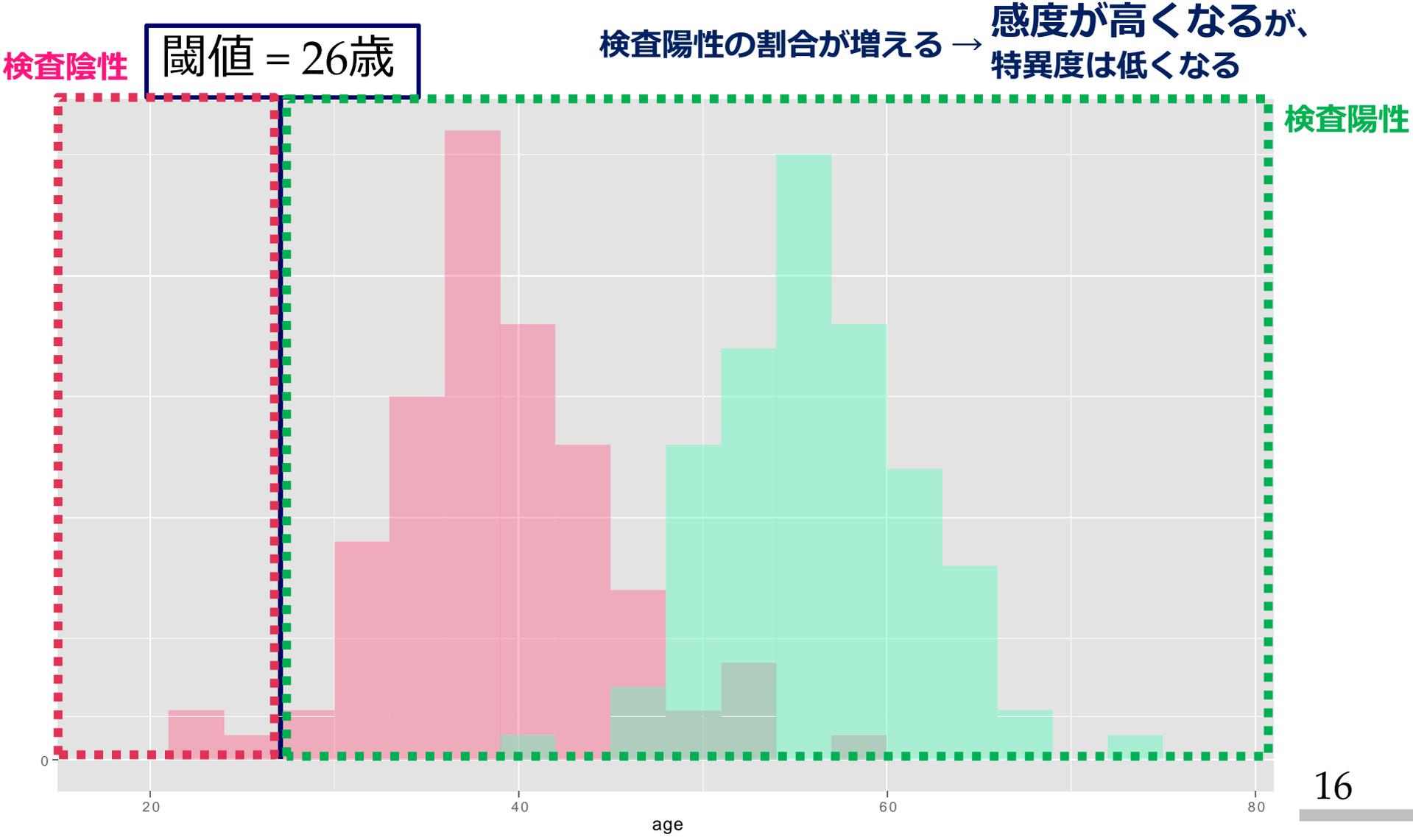


注：仮想例



閾値を低くすると・・・

合併症なし
合併症あり

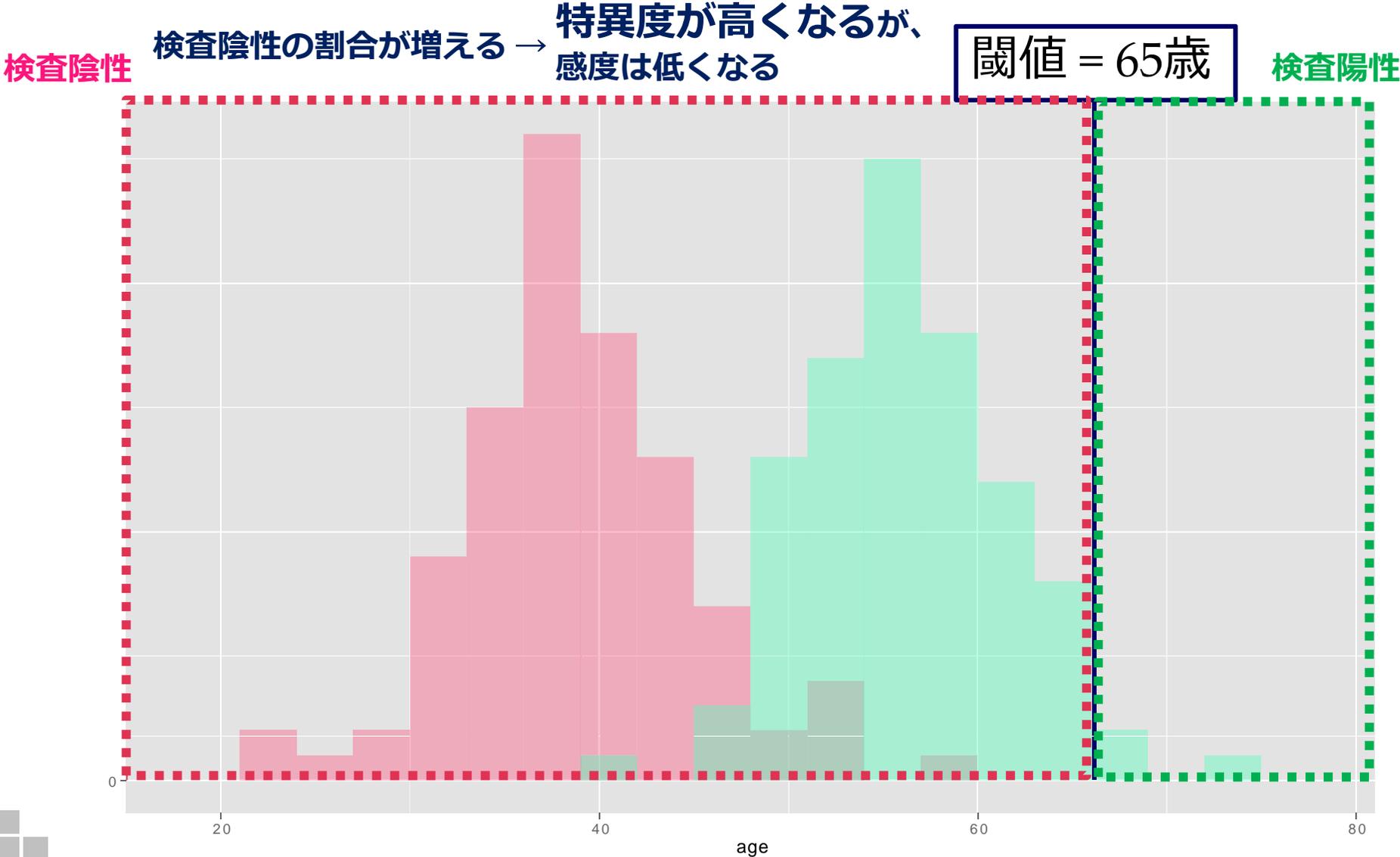


注：仮想例



閾値を高くすると・・・

合併症なし
合併症あり



注意点



■ 感度・特異度はトレードオフの関係

- 連続マーカーの場合は閾値の選択によって、診断能の評価が変わってしまう
- 閾値の違いが感度・特異度に与える影響を評価するためにROC曲線が頻用されている



感度/特異度の高い閾値の探索法

■ ROC (Receiver Operating Characteristics) 曲線

– ある閾値(c)を決めたとき 2×2 分割表が作れる

		合併症あり	合併症なし	
検査	年齢 $> c$ (陽性)	30	10	40
	年齢 $< c$ (陰性)	70	90	160
計		100	100	200

- この表から感度・特異度が計算できる
- 塗りつぶした部分が c の値によって変化する

– 閾値(c)の取り得る値すべてで感度・特異度を計算

– 横軸を1-特異度、縦軸を感度とした曲線を作成

例：先ほどの例から10例だけ抽出

年齢	合併症	
24	なし	閾値①
28	なし	閾値②
32	あり	閾値③
35	なし	閾値④
40	あり	閾値⑤
45	なし	閾値⑥
50	あり	閾値⑦
52	なし	閾値⑧
55	あり	閾値⑨
57	あり	閾値⑩
		閾値⑪

・ 閾値 c の値によって網掛け内の人数構成が変化する

		合併症あり	合併症なし	
検査	年齢 $> c$ (陽性)			
	年齢 $< c$ (陰性)			
計		5	5	10

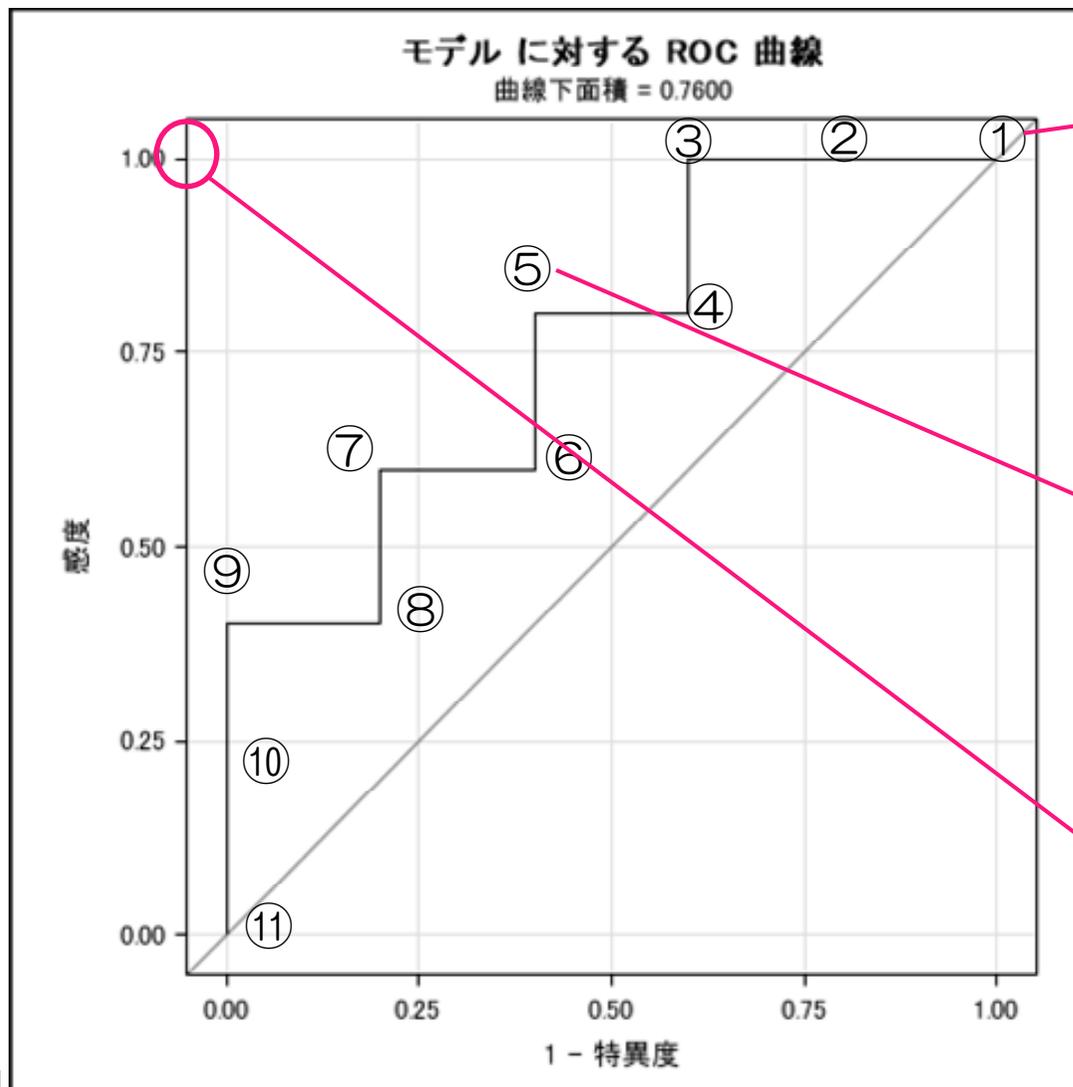
・ 閾値①の場合(感度：100%、特異度：0%)

		合併症あり	合併症なし	
検査	24歳以上	5	5	10
	24歳未満	0	0	0
計		5	5	10

・ 閾値⑤の場合(感度：80%、特異度：60%)

		合併症あり	合併症なし	
検査	40歳以上	4	2	6
	40歳未満	1	3	4
計		5	5	10

作成したROC曲線



閾値①	合併症あり	合併症なし	
24歳以上	5	5	10
24歳未満	0	0	0
計	5	5	10

(1-特異度：100%、感度：100%)

閾値⑤	合併症あり	合併症なし	
40歳以上	4	2	6
40歳未満	1	3	4
計	5	5	10

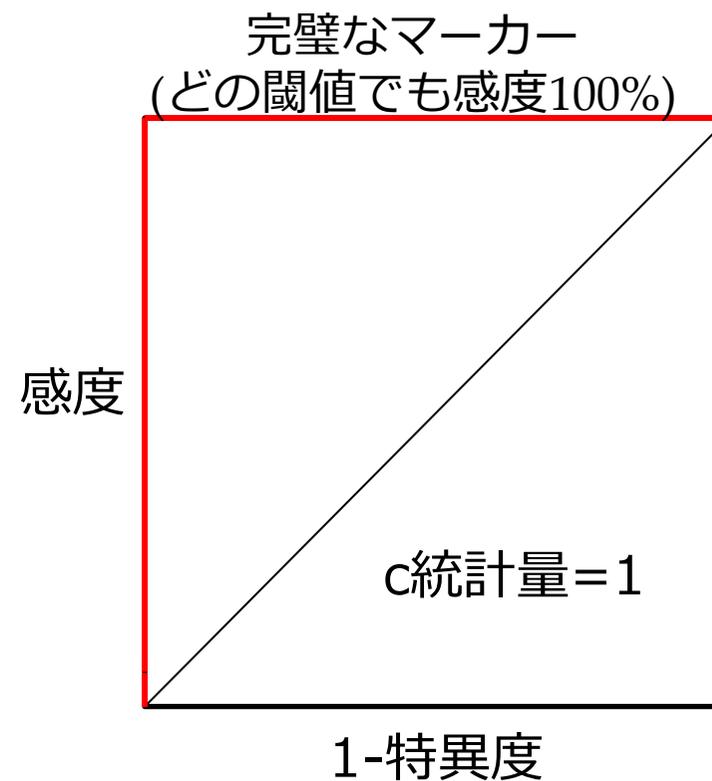
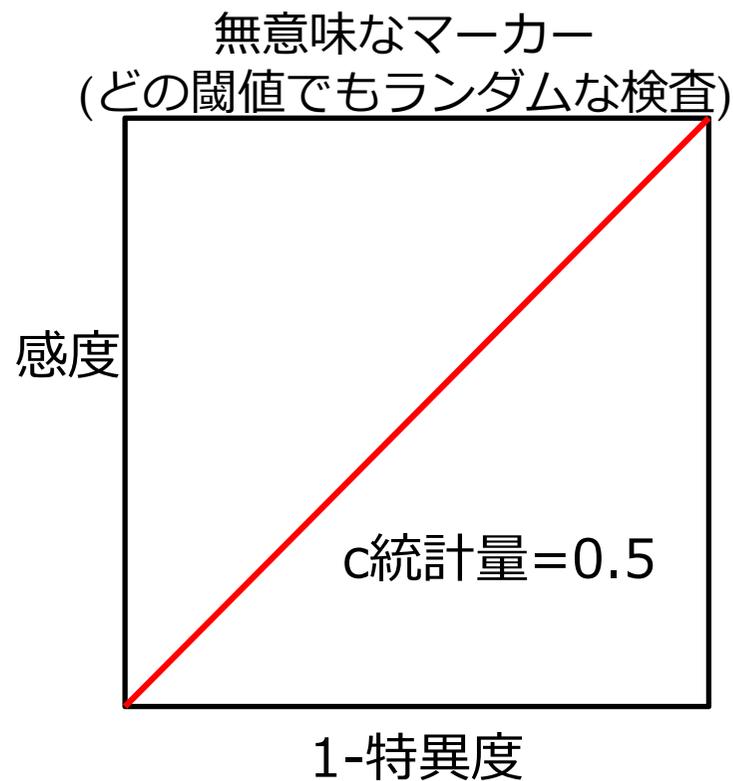
(1-特異度：40%、感度：80%)

対角線より下にいくことはない

左上の点に近い曲線であれば良い閾値
感度100%、特異度100%の点

連続マーカーの診断能

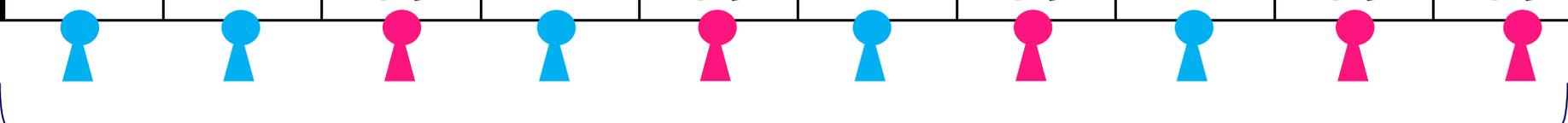
- ROC曲線の曲線下面積(Area Under the Curve)
 - c統計量(Concordance probability)ともいう



c統計量とは？

- 連続マーカーで閾値以上を陽性とする場合
 - 「陽性例のマーカー値が陰性例のマーカー値を上回る確率」と定義される
 - つまり、**検査結果が真実と一致した割合**

年齢	24	28	32	35	40	45	50	52	55	57
合併症	なし	なし	あり	なし	あり	なし	あり	なし	あり	あり



2例のペアについて全パターン検討する

🔑：合併症なし、🔑：合併症あり

括弧内は年齢を表しています

- 🔑 (35) < 🔑 (50)
 - 年齢が高いと合併症あり → 「一致」
 - 🔑 (32) < 🔑 (35)
 - 年齢が高いと合併症なし → 「不一致」
 - 🔑 (30) = 🔑 (30) ※ 前ページの例では該当症例なし
 - 年齢が同じ → 「タイデータ」
 - 🔑 (32) < 🔑 (50)
 - 両者ともに合併症あり(または合併症なし)
 - 判断がつかないため「情報なし」
- 「情報あり」

$$c\text{統計量} = \frac{\text{一致ペア数} + 0.5 \times \text{タイデータ数}}{\text{一致ペア数} + \text{不一致ペア数} + \text{タイデータ数}}$$

■ 全ての情報のあるペア=25通り

– 24歳 – 32,40,50,55,57

赤：一致19, 青：不一致6

– 28歳 – 32,40,50,55,57

– 32歳 – 35,45,52

– 35歳 – 40,50,55,57

– 40歳 – 45,52

– 45歳 – 50,55,57

– 50歳 – 52

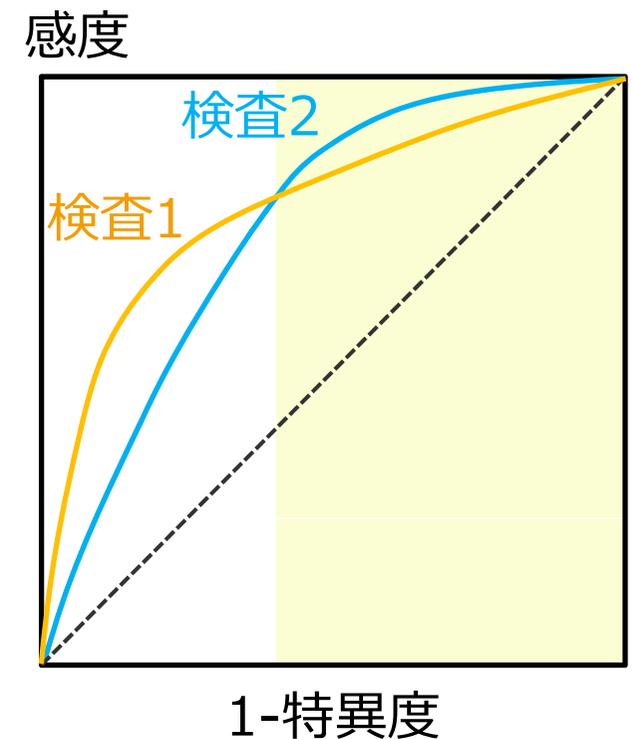
– 52歳 – 55,57

– 55歳 – なし

c統計量 = 19/25 = 0.76
P20のROC曲線のAUCと一致

検査法を比較する上での注意点

- 検査同士の優劣はc統計量の大小では不明
 - ROC曲線が完全に上回っているとは限らない
 - 網掛け部分より左では**検査1**の方が**検査2**よりも感度が高い
 - **実際は閾値を固定した検査を行う**
 - **感度/特異度のバランスが良い検査が選択されるべき**
 - **c統計量の大小では判断できない**

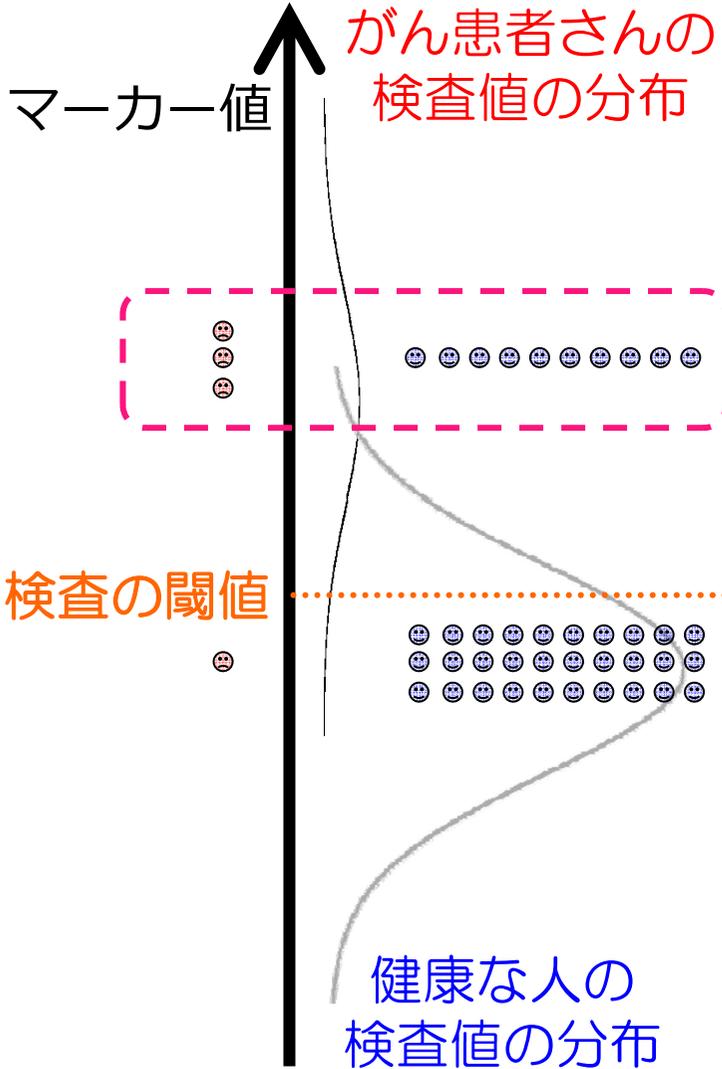


補足：検診の診断能評価

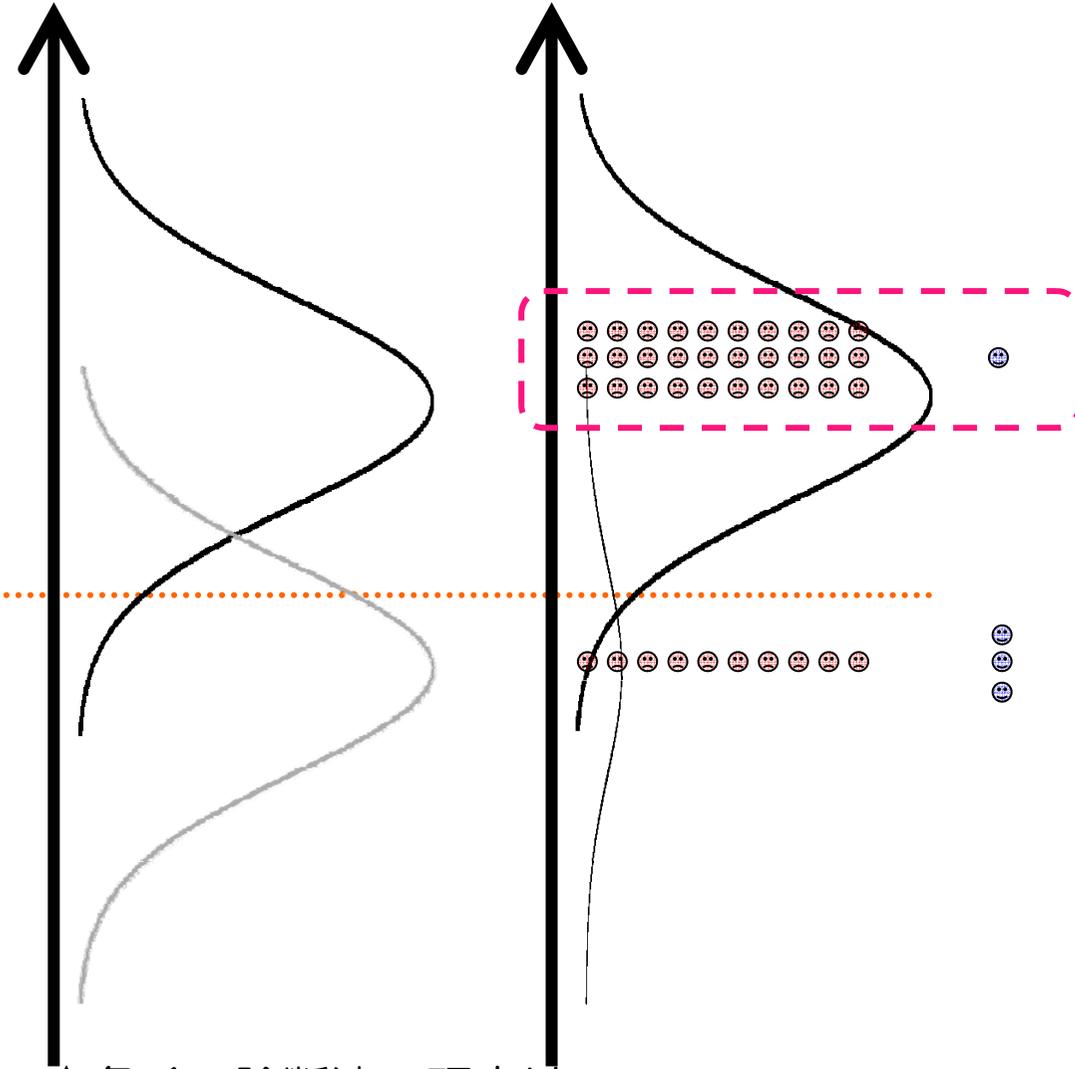
※ 陽性例をがん患者、陰性例を健常人としています

- 診断時期によって考え方が異なる
- スクリーニング時：**罹患割合が低い**
 - 真の陰性例が追加検査に進むリスクを減らしたい
 - 追加検査は侵襲性が高いことが多いため
 - 偽陽性を減らす、つまり**NPVが高いことが理想**
- 追加検査・確定診断時：**罹患割合が高い**
 - 真の陽性例を誤って非癌とするリスクを減らしたい
 - 偽陰性を減らす、つまり**PPVが高いことが理想**

スクリーニング（がん検診）の場合：
検査陽性の多くは健康な人



確定診断に使う場合：
検査陽性の多くはがんの人



↑ 多くの診断法の研究は
患者さんと健康な人の人数に近い

話の整理

■ c統計量が大きい検査が良いとは限らない

– 感度/特異度のバランスが重要

- 検査によって過小/過剰治療となってしまう場合、感度/特異度に優先順位がつけられることがある

■ 閾値の外挿性【検診の場合】

– 検査時期に応じた診断能の評価が不可欠

- (感度が高いが)特異度が低い検査をスクリーニング時に行っても無駄な追加検査を増やすだけ
- 閾値設定時の対象と検査対象の罹患割合が異なると (PPV/NPVの値が変わり)診断能を見誤る恐れがある

OUTLINE

■ 結果の予測に関心がある場合

- 二値マーカーの場合
(感度/特異度/PPV/NPV)
- 連続マーカーの場合(ROCとc統計量)
- 生存時間の場合
- 複数マーカーで予測する場合

■ 診断法間の一致度に関心がある場合

- McNemar検定、 κ 係数

■ 事例紹介

アウトカムが生存時間の場合

- 年齢が予後を予測するかどうか？
(死亡=0の場合は打切り)

年齢	24	28	32	35	40	45	50	52	55	57
死亡	<u>0</u>	<u>0</u>	1	<u>0</u>	1	<u>0</u>	1	<u>0</u>	1	1
生存期間(日)	22	25	20	19	21	29	28	12	10	8

- マーカーの診断能をどうやって評価する？
 - アウトカムが死亡/生存ではなく生存期間であることに注意しないとイケない

打切り例を適切に取り扱う必要あり

生存時間に対する診断能評価

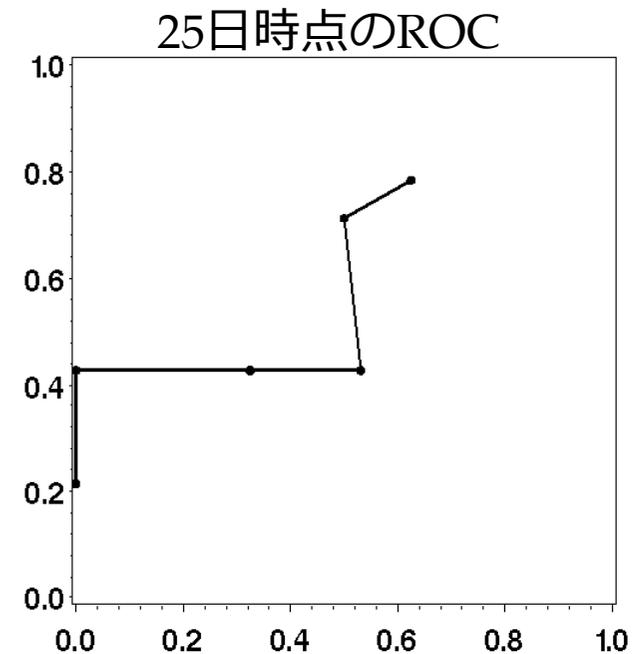
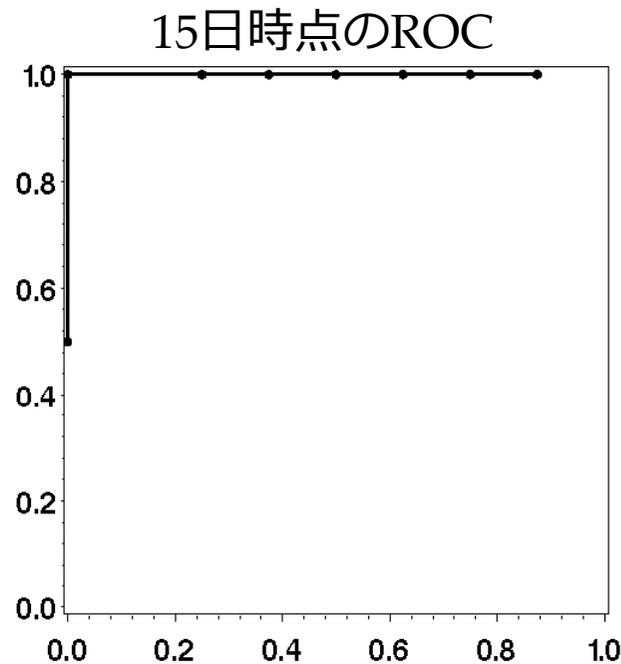
- 特定の時点でイベント発生を予測する場合
 - 2×2表が作れるので前述の方法論が利用可能
 - 特定時点まで追跡できなかった打ち切り例を「生存例」とするバイアスが入ってしまう
 - 選択した時点によって結果が変わってしまう
 - 複数時点で結果を定性的に評価することも可能
 - 時間依存性ROC曲線による評価
- Harrellのc-index
 - c統計量の生存時間への拡張

時間依存性ROC曲線

注：仮想例

■ 特定時点におけるROC曲線

- 感度/特異度をKaplan-Meier法で推定
- 打ち切りの影響で単調増加にならない場合がある



Harrellのc-index

■ 40歳以下を陽性とする場合

- 「40歳以下の生存期間 > 40歳超の生存期間」
となる確率をc統計量の場合と同様に計算する

40歳以下 (Young)	40歳超 (Old)	生存時間の関係 y:40歳以下,o:40歳超	一致	備考
死亡	死亡	$T_y > T_o$	1	
死亡	生存	$T_y > T_o$?	T_o が打切りのため
生存	死亡	$T_y > T_o$	1	
生存	生存	$T_y > T_o$?	両者とも打切りのため
死亡	死亡	$T_y < T_o$	0	
死亡	生存	$T_y < T_o$	0	
生存	死亡	$T_y < T_o$?	T_y が打切りのため
生存	生存	$T_y < T_o$?	両者とも打切りのため

?を除いて一致ペア数の割合を算出

■ 全ての情報のあるペア = 27通り

– 24歳 – 32,40,55,57

赤：一致22，青：不一致5

– 28歳 – 32,40,55,57

– 32歳 – 40,45,50,55,57

– 35歳 – 55,57

– 40歳 – 45,50,55,57

– 45歳 – 50,55,57

– 50歳 – 55,57

– 52歳 – 55,57

– 55歳 – 57

$$\text{c-index} = 22/27 = 0.8148$$

当然、時間依存性ROC曲線のAUC \neq c-index
(ROC曲線のAUCは“ある時点”に対するものだが、
c-indexは時点を特定しないので)

OUTLINE

■ 結果の予測に関心がある場合

- 二値マーカーの場合
(感度/特異度/PPV/NPV)
- 連続マーカーの場合(ROCとc統計量)
- 生存時間の場合
- 複数マーカーで予測する場合

■ 診断法間の一致度に関心がある場合

- McNemar検定、 κ 係数

■ 事例紹介

複数マーカーによる予測

■ 複数マーカーによる予測に関心がある場合

– 術後合併症/奏効/予後の予測

- 背景因子や術時因子、複数遺伝子の発現有無を組み合わせて予測能の高いモデルを構築する
(詳しくは第6回セミナー 多変量解析2を参照下さい)

– 2×2表が作れないため、 今までの評価指標が(容易に)計算できない

– 多変量解析の手法を用いて評価することが一般的

例：肝切除後の合併症予測

※ p は合併症発生割合を表す

■ ロジスティック回帰モデルの利用

– 以下の背景・術時因子が候補変数

- 年齢、併存症、術式、手術部位、切除断片数
- stepwise法で年齢と切除断片数が残ったとする

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \times \text{年齢} + \beta_2 \times \text{切除断片数} = \eta$$

– 構築モデルから合併症の予測発生確率が計算可能

$$\text{予測確率 } (\hat{p}) = \frac{e^\eta}{1 + e^\eta}$$

※ \hat{p} は p の予測値という意味です



■ モデル構築後のデータ

ID	年齢	切除断片数	合併症	\hat{p}
1	21	2	なし	0.232
2	54	3	なし	0.142
3	35	2	あり	0.532
...				
N	46	4	あり	0.624

- 「予測確率 \hat{p} が合併症を予測するか」に着目
 - 複数変数を1つのスコアに要約し、その予測能を評価 (第9回セミナー 傾向スコアの考え方と同じ)
 - **複数マーカーが1つの予測スコアに集約されたのでROC曲線やc統計量を作図・計算することができる**



補足



■ アウトカムが生存時間の場合

- c-indexをCox回帰モデルに拡張した指標が提案されている(Gonen and Heller, Biometrika 2005.)
- 複数変数の組合わせでprognostic scoreを作成し、そのscoreを用いてc-indexを計算することも可能
 - もちろん、scoreが二値で、特定時点のイベント予測に関心があるなら2×2表に基づく評価が可能



OUTLINE

- 結果の予測に関心がある場合
 - 二値マーカーの場合
(感度/特異度/PPV/NPV)
 - 連続マーカーの場合(ROCとc統計量)
 - 生存時間の場合
 - 複数マーカーで予測する場合
- 診断法間の一致度に関心がある場合
 - McNemar検定、 κ 係数
- 事例紹介



ここまで取り上げたデータ

- マーカーが単体/複数の場合、アウトカムが二値/生存時間の場合に対応する診断能の評価指標を解説した

		真の分類		
		陽性	陰性	計
検査 (Marker)	陽性	a	b	a+b
	陰性	c	d	c+d
計		N_1	N_2	N



Matched-pairなデータ

- 新規診断法を評価する状況において、
真の分類との一致よりも、
既存の診断法との一致に興味がある場合

		検査2(Marker2)		計
		陽性	陰性	
検査1 (Marker1)	陽性	n_{11}	n_{12}	n_{1+}
	陰性	n_{21}	n_{22}	n_{2+}
計		n_{+1}	n_{+2}	n_{++}

- 同一症例に複数の検査をしたデータ

Matched-pairデータの解析法

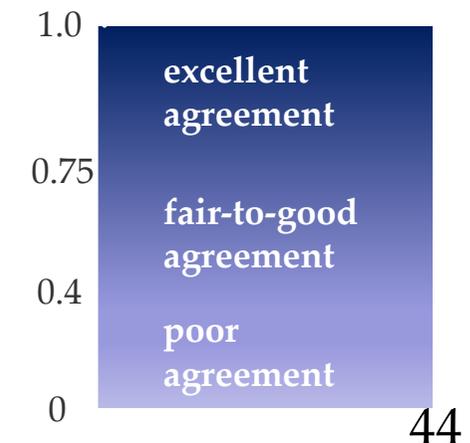
■ McNemar検定

- 「 H_0 : 検査1の陽性割合 = 検査2の陽性割合」に対する検定
- 有意 = 検査間で陽性割合に差がある
 - 検定結果は症例数に依存するため評価が難しい

■ κ 係数

- 検査間の一致度を定量的に示す指標
 - 「偶然の一致」を除いた上で一致度を評価
 - 1-誤分類率とは値が異なるので注意が必要

Landis & Kochの分類



一致度を評価する上での注意点

- 診断能を評価する指標ではない
 - 真の分類との一致度を評価する指標ではない
 - 「 κ 係数が0.99 = 検査精度が高い」とは限らない
 - 感度/特異度の評価が必要



OUTLINE

- 結果の予測に関心がある場合
 - 二値マーカーの場合
(感度/特異度/PPV/NPV)
 - 連続マーカーの場合(ROCとc統計量)
 - 生存時間の場合
 - 複数マーカーで予測する場合
- 診断法間の一致度に関心がある場合
 - McNemar検定、 κ 係数
- 事例紹介

事例1：JCOG0201

■ 背景と目的

- 病理学的診断に基づき早期肺がんの手術法が決定
 - 病理学的非浸潤がん → 縮小切除適応
 - 病理学的浸潤がん → 肺葉切除
- 目的：術前の胸部薄切CT画像で(縮小手術の適応である)病理学的非浸潤癌を術前に同定すること

■ ここでは病理学的診断が真の分類

感度と特異度のどちらを優先？

		病理学的診断	
		縮小手術適応(+) (非浸潤癌)	縮小手術非適応(-) (浸潤癌)
画像診断	縮小切除適応(+) (非浸潤癌)	感度	
	縮小切除非適応(-) (浸潤癌)		特異度

- 感度が低い場合：過剰治療
 - 縮小手術を受けるべき患者が肺葉切除される
- 特異度が低い場合：過小治療
 - 肺葉切除を受けるべき患者が縮小手術される
- 過小治療の方がリスクが高い

■ primary endpointは特異度

- 合併症のリスクや肺機能のデータから、「特異度は97%以上が必要」と判断された
 - 「95%CIの下限が97%以上」になれば妥当と判断
- なぜNPVにしなかったのか？
 - (前述の通り)浸潤癌の割合に依存してしまうため

■ 肺葉切除 vs 縮小手術のRCTが進行中(JCOG0802)

事例1の補足

- 臨床的観点から感度と特異度に優先順位がつけられることがある
 - 最適な閾値をROC曲線で探索する場合
 - たとえば「特異度(or 感度)が〇〇%以上の閾値のうち、感度(or 特異度)が最大」という基準は有用
- どうしても決められない場合
 - 統計的(幾何学的)に決めることは可能
 - 「左上点からの距離が最も近い閾値」
 - 「対角線からの距離が最も長い閾値」

事例2：TRANSBIG & MINDACT trial

■ MammaPrint®

– 早期乳がん患者の術後5年以内の転移再発を予測

- 多変量解析の手法に基づき抽出した231遺伝子のうち、Top 70遺伝子を用いてgene signatureを構築・検証

- 2007年2月、FDAにより承認された(510k) (Van'tVeer, et al. Nature 2002.)

■ Adjuvant! (Adjuvant Online) (Ravdin PM. J Clin Oncol 2001)

– 患者背景因子に基づき作成した予測モデル

- age, tumor size, nodal status, histological tumor grade, ER, comorbidityから患者が層別化される

■ TRANSBIG trial (Buyse M, et al. JNCI 2006. / FDA Decision Summary k062694. 2007)

– MammaPrint®やAdjuvant!などの再発予測モデルを
欧州の独立したコホート(n=302)で比較した試験

留意点



- MammaPrint[®]もAdjuvant!も患者を high or low riskに層別する
- low risk患者にはホルモン療法単独を、 high risk患者には化学療法を推奨している
- 感度/特異度の定義
 - 感度：転移再発例におけるhigh risk患者の割合
 - 特異度：転移再発なし例におけるlow risk患者の割合



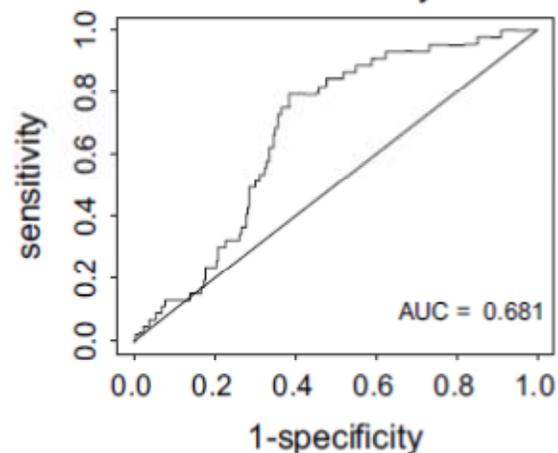
感度・特異度

Risk classification	Number of patients in risk groups		Metastases within 5 years	
	High risk	Low risk	Sensitivity	Specificity
Gene signature	194	113	0.90 (0.78 to 0.95)	0.42 (0.36 to 0.48)
Adjuvant! software	222	80	0.87 (0.75 to 0.94)	0.29 (0.24 to 0.35)

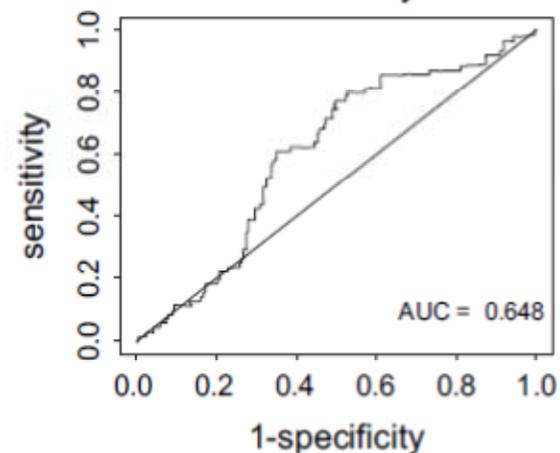
- 5年時点における転移・再発の予測能
 - 感度は9割程度、特異度は3割～4割しかない
 - Adjuvant!の特異度は低いこと以外、大差はない

時間依存性ROC曲線

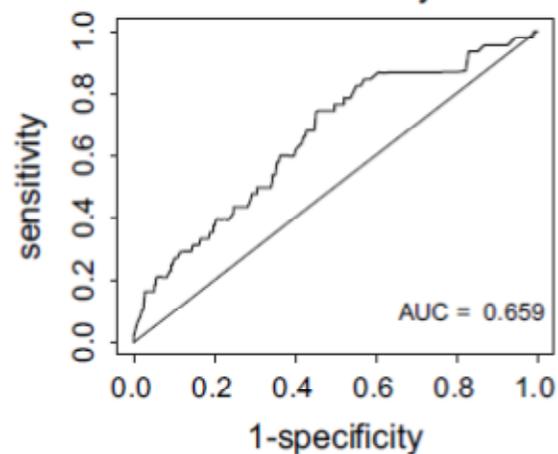
Gene signature score, for time to distant metastases at 5 years



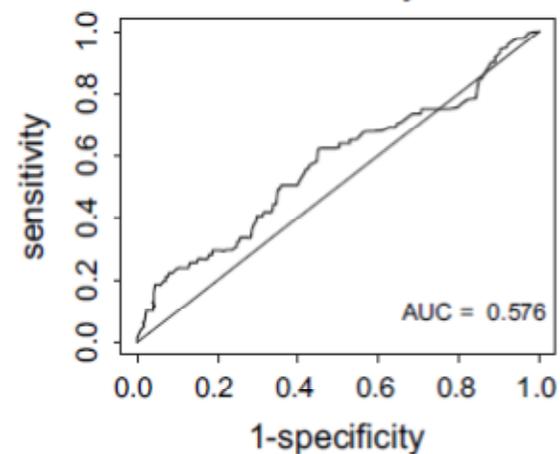
Gene signature score, for overall survival at 10 years



Adjuvant! score, for time to distant metastases at 5 years



Adjuvant! score, for overall survival at 10 years



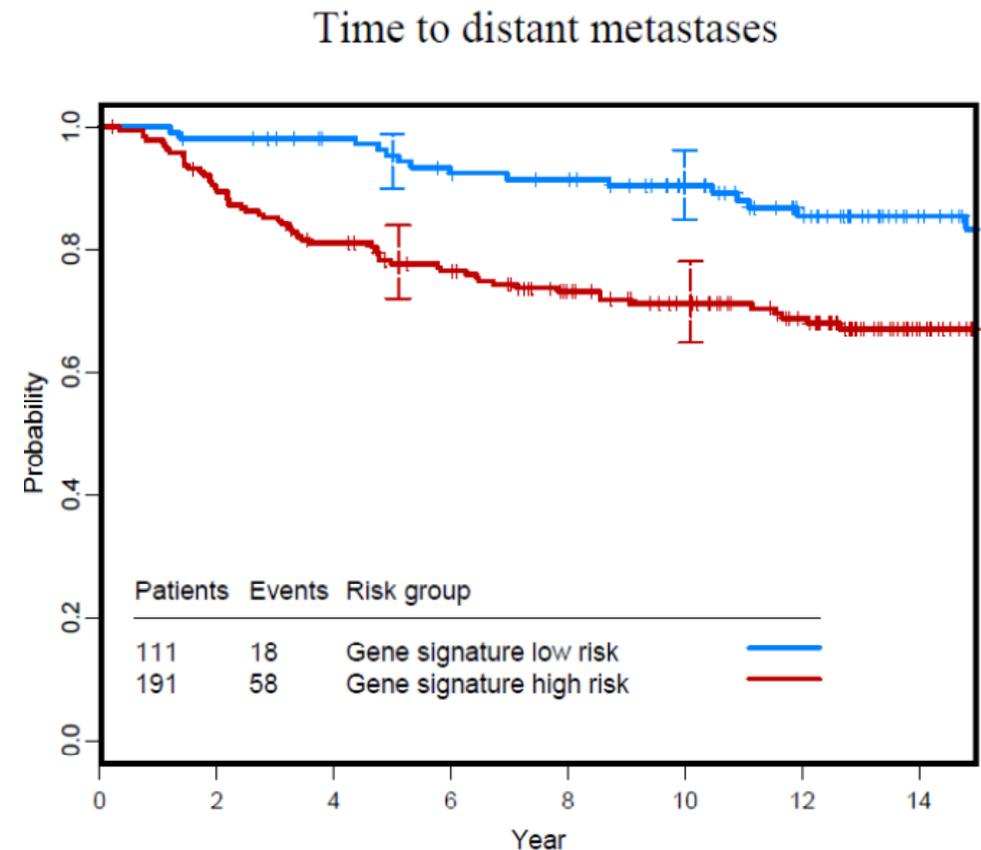
PPV・NPV

■ 10年時点の転移再発を真の分類とした場合

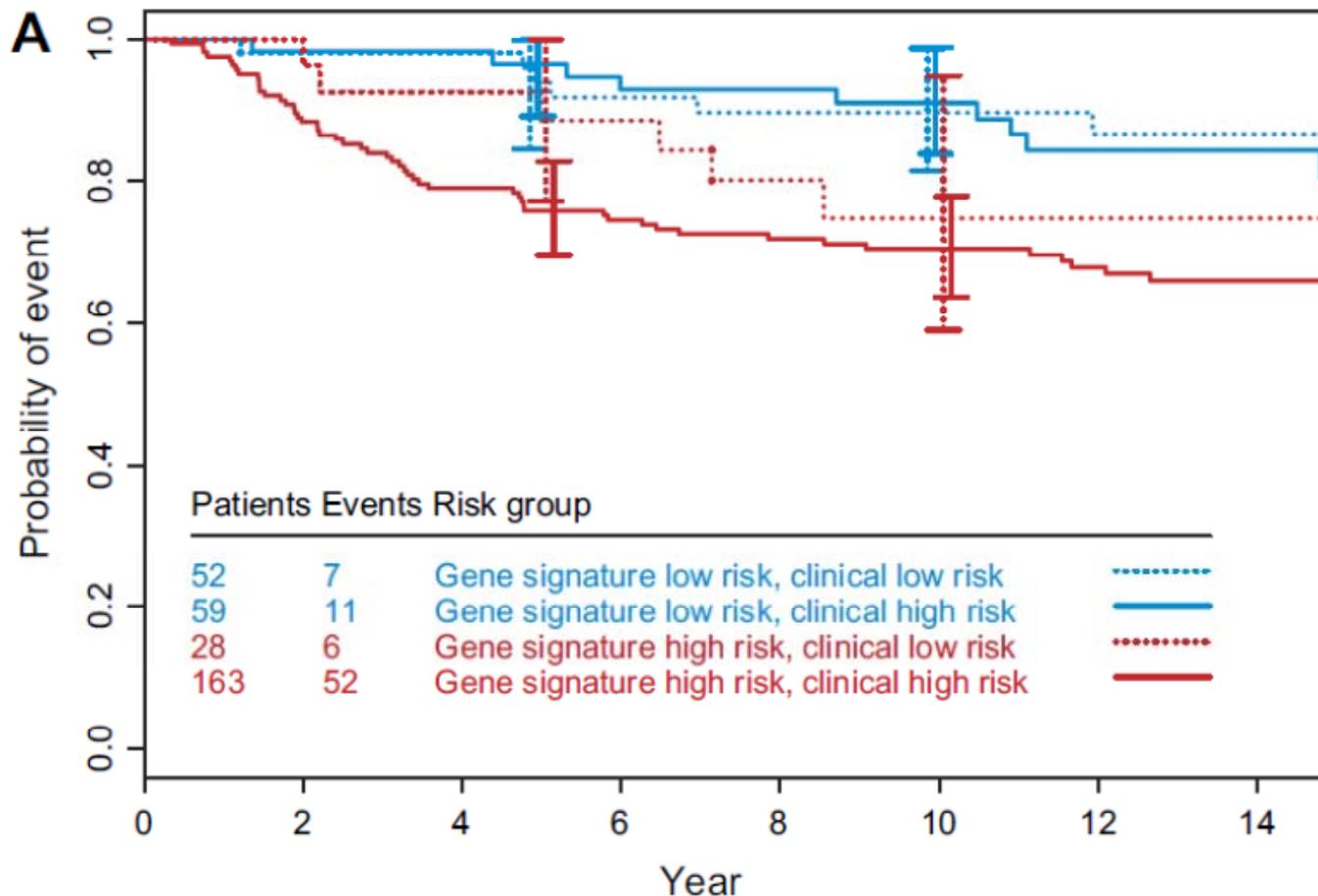
– MammaPrint[®]のPPVは極めて低い

- PPV = 0.29 (0.22-0.35)
- NPV = 0.90 (0.85-0.96)

– Adjuvant!は報告なし



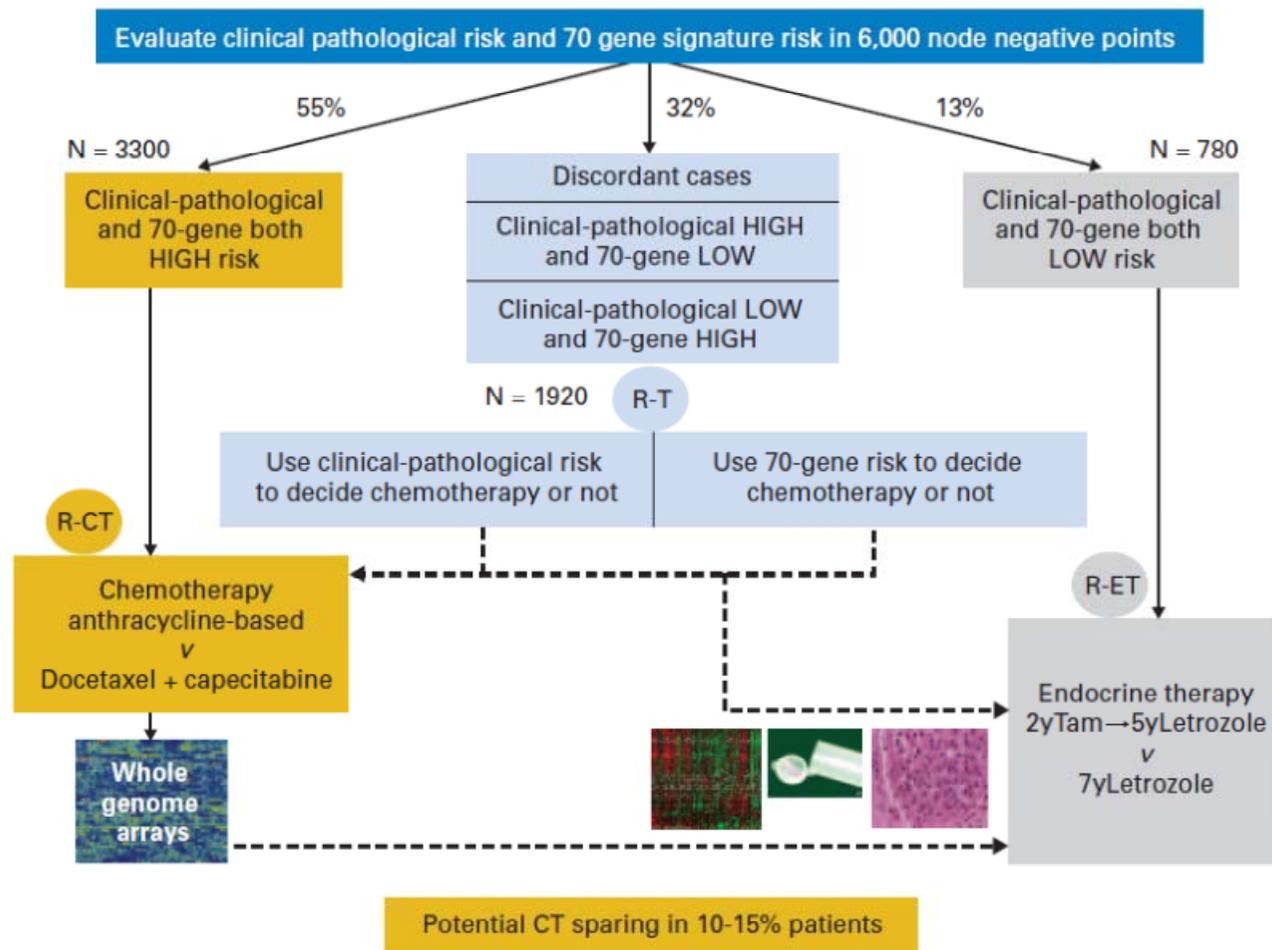
Adjuvant!との不一致(n=87; 29%)



- 青実線は予後良好だが赤点線は予後不良
 - 不一致の場合はMammaPrintの方が良い？

不一致例に対してRCTが進行中

■ MINDACT trial



事例3：ALK遺伝子転座に対するCrizotinib

■ CrizotinibはEML4-ALK遺伝子転座を有する非小細胞肺癌患者に高い有効性を示す

(Kwak EL, et al. N Engl J Med 2010/Shaw AT, et al. N Engl J Med 2013)

– 第I/II相試験

- FISH法(FDA承認を受けた検査法)でALK転座陽性と判定されたn=82の結果
- 奏効割合：57%[46%-48%]

■ 検査法による奏効割合の違い (Chihara D and Suzuki R. N Engl J Med 2011)

- IHCやRT-PCRでも転座を確認した症例：81%(22/27)
- FISH単独でしか転座を確認しなかった症例：48%(25/52)
- FISH検査には偽陽性例が含まれる？

FISHとIHC法の不一致例

- 承認から薬価収載までの期間(2012/4/4-5/28)でFISHとIHCが両方検査された2337検体を調査

		高感度IHC法		(日本肺癌学会 2012)
		陽性	陰性	
FISH	陽性	213	36	
	陰性	12	2076	

- McNemar検定 : $p < 0.0005$
 - κ 係数 : 0.89 [0.86-0.92]
 - 不一致例(n=48)の奏効割合が低かったことなどから、「IHCによるスクリーニング後、FISH法で確認」とする手順が推奨されている
- 基準に従えば一致度は高いが、陽性割合に有意差はある (Nが大きいため?)

(日本肺癌学会 「ALK診断アルゴリズムについての注記」 2012)

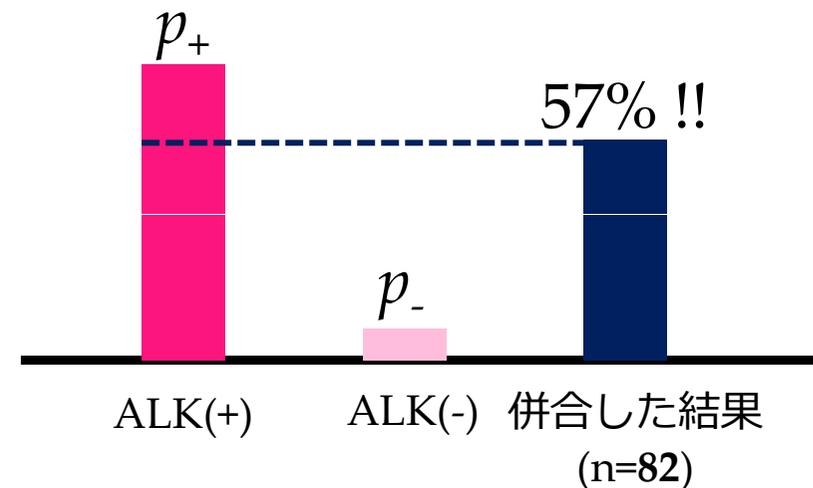
事例3の補足

■ Enrichmentデザインの問題点

- 診断法の精度により効果が希釈されうる
- Crizotinibの単群第II相試験の例

表を横にみる

		真の分類		計
		陽性	陰性	
FISH	陽性	??	??	82
	陰性	??	??	1418
	計	??	??	1500



$$57\% = PPV \times p_+ + (1 - PPV) \times p_-$$

※ RAS野生型大腸がんに対するCetuximabの効果などはNPVの重み付き和



- 検査により対象を限定した場合の治療効果は感度/特異度ではなくPPV/NPVに影響される
 - 感度・特異度が高い検査でenrichしても、治療効果が正しく評価できるとは限らない
- 検査と治療効果の関係を検討する上ではPPV/NPVの評価は重要



まとめ



■ 診断能を評価する指標の紹介

- 感度/特異度/PPV/NPV/ROC曲線/c統計量
 - 真の陽性割合に影響される指標には注意が必要
- アウトカムの種類・予測因子の個数に応じて最適な評価指標を選択すべき

■ 検査法の一致を評価する場合

- McNemar検定や κ 係数を利用する

BACKUP

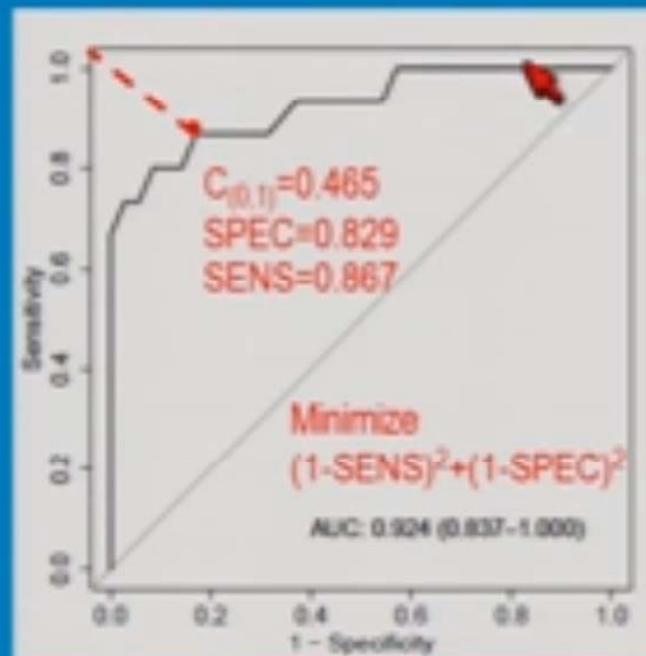
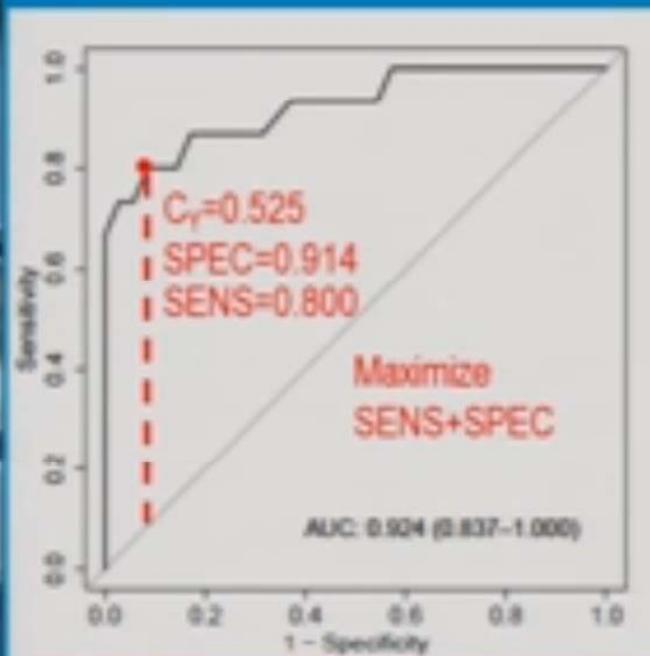


統計的な基準で閾値を探索する方法

Optimality criteria for cut-off selection

Youden index: Select cut-off c_Y that maximizes vertical distance from diagonal line to ROC curve

Closest to (0,1): Select cut-off $c_{(0,1)}$ that minimizes distance from point (0,1) to ROC curve



32

National Cancer Institute

23:37 / 28:02

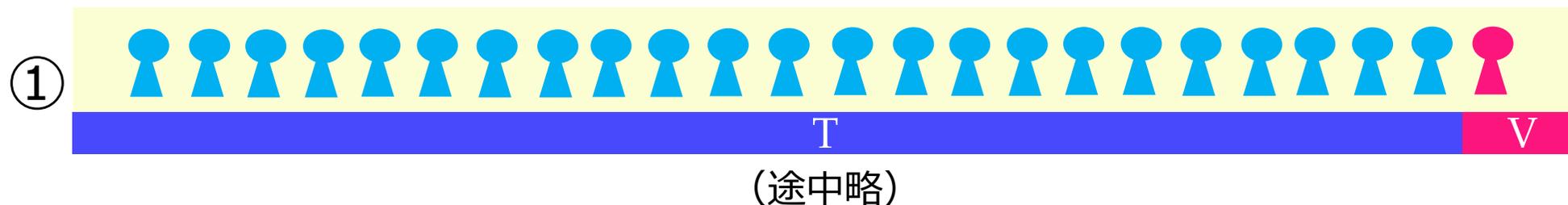
64

補足-1

※ N : 症例数、 T : Training data、 V : Validation data

■ 構築モデルのバリデーション

– Leave-one-out cross validationを実施した場合



- 全validation症例(👤)について
予測確率(\hat{p})と合併症有無が分かる
- ROC曲線・c統計量を用いて予測能が評価できる

McNemar検定

検査1	検査2		合計
	B ₁	B ₂	
A ₁	n_{11}	n_{12}	n_{1+}
A ₂	n_{21}	n_{22}	n_{2+}
合計	n_{+1}	n_{+2}	n_{++}

$$\pi_{11} = n_{11}/n_{++}, \pi_{12} = n_{12}/n_{++}$$
$$\pi_{21} = n_{21}/n_{++}, \pi_{22} = n_{22}/n_{++}$$

- 「 $H_0: \pi_{1+} = \pi_{+1}$ 」に対する検定
 - 非対角成分における割合の差の検定

$$\chi^2 = \frac{(n_{12} + n_{21})^2}{n_{12} + n_{21}}$$

- カイ二乗検定については第2回セミナーを参照下さい

κ係数

		検査2		計
		B ₁	B ₂	
検査1	A ₁	p_{11}	p_{12}	p_{1+}
	A ₂	p_{21}	p_{22}	p_{2+}
計		p_{+1}	p_{+2}	1

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e}$$

$$p_o = \sum_{i=1}^2 p_{ii}, \quad p_e = \sum_{i=1}^2 p_{i+} \cdot p_{+i}$$

Landis & Kochの分類



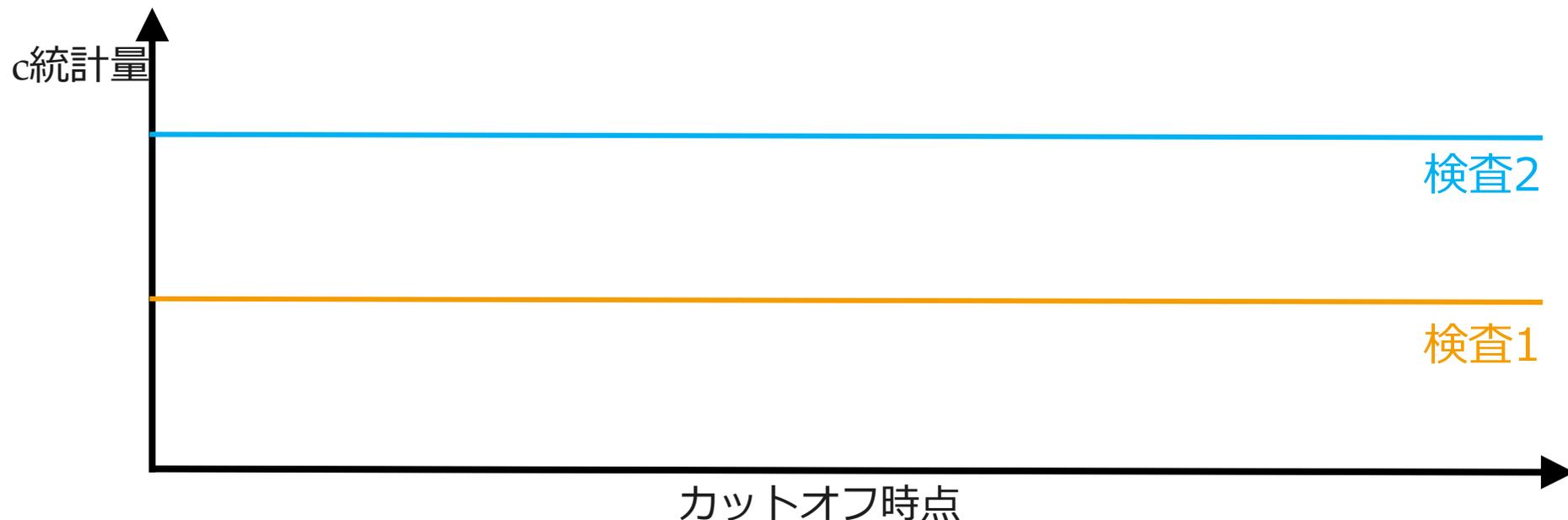
○ $\hat{\kappa} = 1$: 完全に一致

○ $\hat{\kappa} = 0$: 診断が独立



時間依存性ROC曲線の群間比較

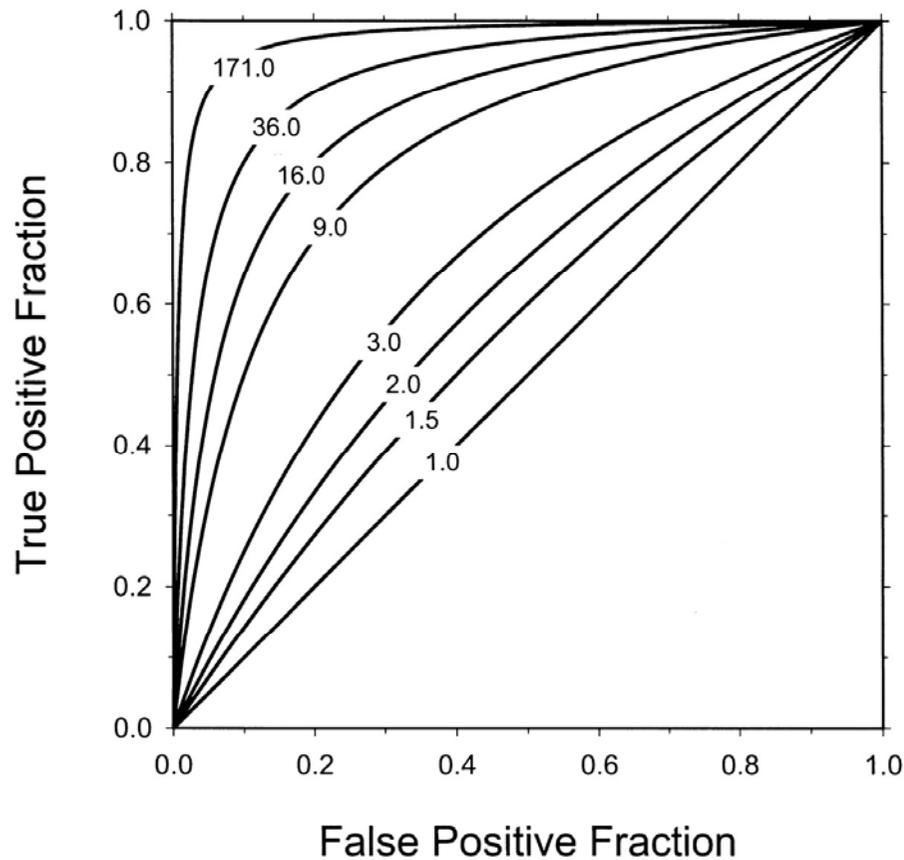
- すべての時点で時間依存性ROC曲線を作成
 - 全時点のc統計量を計算
 - 検査法間で比較



※ 全時点でc統計量が大きい検査2の方が予測能が高いと言える

c統計量を大きくすることは難しい(1)

■ 各ROC曲線を得るために必要なオッズ比



Pepe M S et al. Am. J. Epidemiol. 2004;159:882-890

c統計量を大きくすることは難しい(2)

- オッズ比と陽性群(D)/陰性群(\bar{D})の分布の模式図
 - Nとは独立(Nを増やしても分布の形状は変わらない)

