

診断研究における検査性能の比較

高橋 佳苗 (大阪大学医学部附属病院)

山本 紘司 (大阪大学大学院医学系研究科)

1. はじめに

医療において、スクリーニング検査は疾患の早期発見・早期治療のために重要な役割を果たしている。スクリーニング検査の性能をあらわす指標としては、感度 (sensitivity : SE)、特異度 (specificity : SP)、陽性適中率 (positive predictive value : PPV) および陰性的中率 (negative predictive value : NPV) がしばしば用いられている。SE は罹患している対象者をスクリーニング検査によって正しく罹患していると判断する確率であり、SP は罹患していない対象者を正しく罹患していないと判断する確率である。一方、PPV は対象者のスクリーニング検査の結果が陽性であるとき、当該対象者が実際に疾患に罹患している確率であり、NPV はスクリーニング検査の結果が陰性であった対象者が実際に疾患に罹患していない確率である。SE、SP、PPV、NPV は臨床的に有用であり、患者の治療方針の決定に影響を及ぼす。

2つのスクリーニング検査法の性能を比べる目的で上記の指標を比較する際、SE および SP については McNemar 検定が一般的に用いられている。しかし、PPV および NPV については主流となる解析手法は確立しておらず、いくつかの手法が提案されているのみである (Leisenring et.al, 2010, Moskowitz and Pepe, 2006, Kosinski, 2013)。さらに、提案されているいずれの手法も、中心極限定理が用いられており、対象者の少ない小規模臨床試験への適用に問題がないかは未だ検討されていない。

さらに、スクリーニング検査の性能については、本来1つの指標のみで判断するよりも、複数の指標から総合的に判断されるべきである。2つのスクリーニング検査間で複数の指標を比較する際は、すべての指標の優越性を確認するよりも、1つの指標の優越性と他の全ての指標での非劣性を組み合わせたデザインの方が実臨床の場に即していると考えられる。

そこで本講演では、①PPV および NPV の比較に対する解析手法を小規模臨床試験へ適用した場合の性能評価、②いずれか1つの指標の優越性が認められれば他の指標は非劣性を確認するのみでよいとする、優越性と非劣性を組み合わせた多重エンドポイントの解析手法 (Bloch et al, 2007, Muscha and Turan, 2012) の性能評価について報告する。

2. 方法および結果

①PPV および NPV の解析手法を小規模臨床試験へ適用した場合の性能評価

多項分布をもとに中心極限定理とデルタ法を利用したもの (恒等変換、対数変換、ロジット変換)、回帰分析の枠組みを利用したもの (一般化スコア統計量、重み付き一般化スコア

統計量)の解析手法がすでに提案されている。Kosinski (2013)で報告されている方法を用いて、表1の設定でシミュレーション用データを生成し、症例数を10例から50例とした場合の各手法の Type 1 error rate 及び検出力を求めた。結果を図1, 2に示す。

2つの診断検査間のPPV、NPVに差がないとしたとき(図1)、名目有意水準に最も近い検出力となったのは重み付き一般化スコア統計量を用いた方法であった。重み付き一般化スコア統計量を用いた方法は、2つの診断検査間のPPV、NPVに差があるとしたとき(図2)の検出力も高く、小規模臨床試験においては最も性能の良い方法と考えられる。

表 1

		真の確率					
		疾患に罹患している			疾患に罹患していない		
		test2			test2		
		+	-		+	-	
test1	+	π_a	π_b	π_1	π_e	π_f	π_5
	-	π_c	π_d	π_2	π_g	π_h	π_6
		π_3	π_4	π_{D+}	π_7	π_8	π_{D-}
➤	test1のPPV (PPV1) = 0.75				➤	有病割合 $\theta = \pi_{D+} : 0.4$	
➤	test2のPPV (PPV2) = 0.75, 0.85				➤	$OR_{D+} = \pi_a \pi_d / \pi_b \pi_c : 3.0$	
➤	test1のNPV (NPV1) = 0.80				➤	$OR_{D-} = \pi_e \pi_h / \pi_f \pi_g : 1.5$	
➤	test2のNPV (NPV2) = 0.80, 0.90				➤	名目有意水準: 0.05 (両側)	

図 1

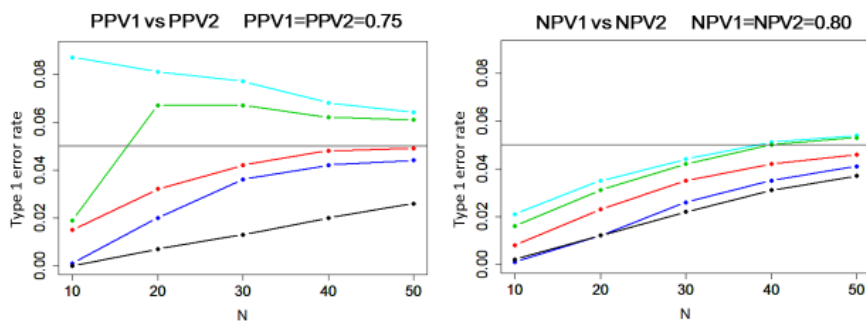
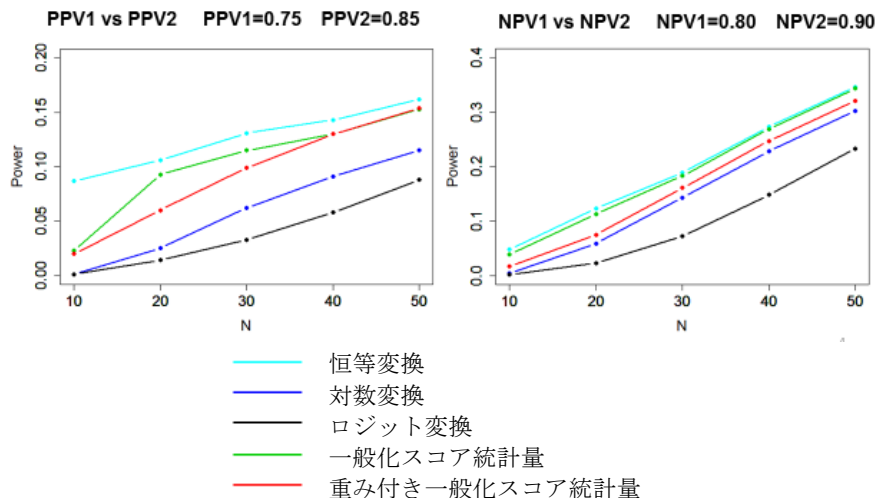


図 2



- 恒等変換
- 対数変換
- ロジット変換
- 一般化スコア統計量
- 重み付き一般化スコア統計量

②優越性と非劣性を組み合わせた多重エンドポイントの解析手法の性能評価

表 2 に示すシナリオについて、①と同様に Kosinski (2013)の方法を用いてシミュレーション用データを作成した。有病割合は 0.4、非劣性マージンは 0.1、名目有意水準は 0.05、症例数は 50 例から 500 例とした。SE および SP の解析には McNemar 検定を、PPV および NPV の解析には重み付き一般化スコア統計量を用いた方法を適用した。結果を図 3,4,5 に示す。

Bloch et al. (2007)の方法と Muscha and Turan (2012)の方法間で結果に大きな違いは見られなかった。また、優越性と非劣性の両方を評価するデザインの検出力が優越性の評価のみと比較して大きく変わらない場合もあることから、いずれか 1 つの指標の優越性が認められれば他の指標は非劣性を確認するのみでよいとする、優越性と非劣性を組み合わせた多重エンドポイントの枠組みでデザイン・解析することの有用性が示唆された。

表 2

Figure	SE1	SE2	SP1	SP2	PPV1	PPV2	NPV1	NPV2
3	0.75 0.85	0.75	0.80 0.90	0.80	-	-	-	-
4	-	-	-	-	0.75 0.85	0.75	0.80 0.90	0.80
5	0.75 0.85	0.75	0.80 0.90	0.80	SE, SPから導出			

図 3 SE & SP

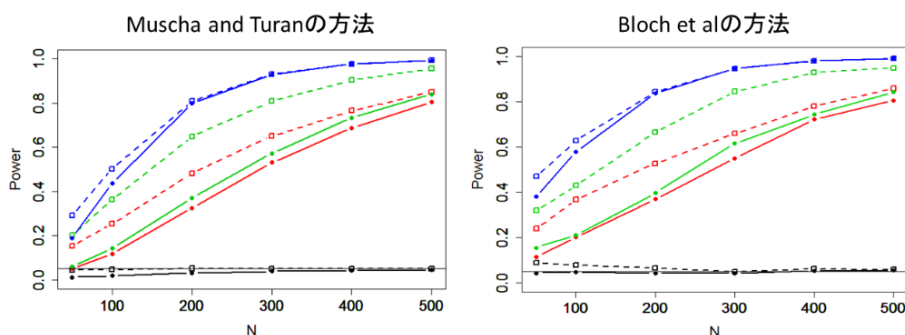


図 4 PPV & NPV

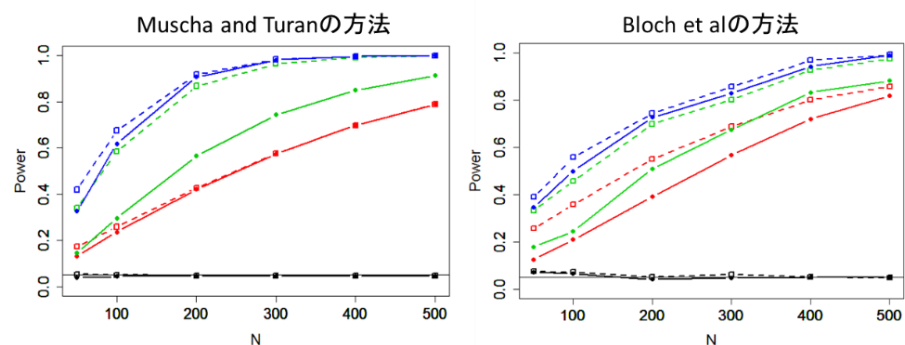
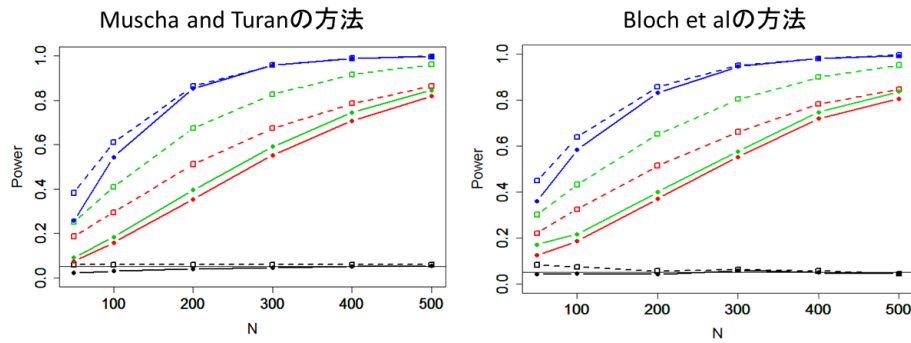


図 5 SE & SP & PPV & NPV



黒線 : SE1=SE2, SP1=SP2, PPV1=PPV2, NPV1=NPV2
 赤線 : SE1>SE2, PPV1>PPV2
 緑線 : SP1>SP2, NPV1>NPV2
 青線 : SE1>SE2, SP1>SP2, PPV1>PPV2, NPV1>NPV2
 実線 : 優越性と非劣性を組み合わせた場合
 点線 : 優越性のみ

参考文献

- Bloch, D. A. et al. A combined superiority and non-inferiority approach to multiple endpoints in clinical trials. *Statistics in medicine* 2013, **26**:1193-1207.
- Kosinski, A. S. A weighted generalized score statistic for comparison of predictive values of diagnostic tests. *Statistics in medicine* 2013, **32**:964-977.
- Leisenring W, Alonzo T, Pepe MS. Comparisons of predictive values of binary medical diagnostic tests for paired designs. *Biometrics* 2000; **56**:345-351.
- Mascha, E. J. and Turan, A. Joint hypothesis testing and gatekeeping procedures. *Anesthesia and Analgesia* 2012, **114**:1304-1317.
- Moskowitz, C. S. and Pepe M. S. Comparing the predictive values of diagnostic tests: sample size and analysis for paired study designs. *Clinical Trials* 2006; **3**:272-279.