Biomarker
연구설계와 분석

가천의대 예방의학교실
고광필

목차

- Biomarker의 정의와 개념
- Biomarker 개발 단계와 연구디자인
- Biomarker 의 정확성 측정 및 marker 간 비교
- 예측모델에서 Biomarker 의 Incremental value 평가

Biomarker(biological marker)

- characteristic
  - ✓ objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention

NIH Biomarkers Definitions Working Group. Clin Pharmacol Ther, 2001

- any substance, structure, or process
  - ✓ can be measured in the body or its products

    influence or predict the incidence of outcome or disease

WHO International Programme on Chemical Safety. 2001

- 예
  - ✓clinical sign/symptom, laboratory test, gene expression technology, proteomics 등, 또는 이들의 조합

Categorization of different biomarkers

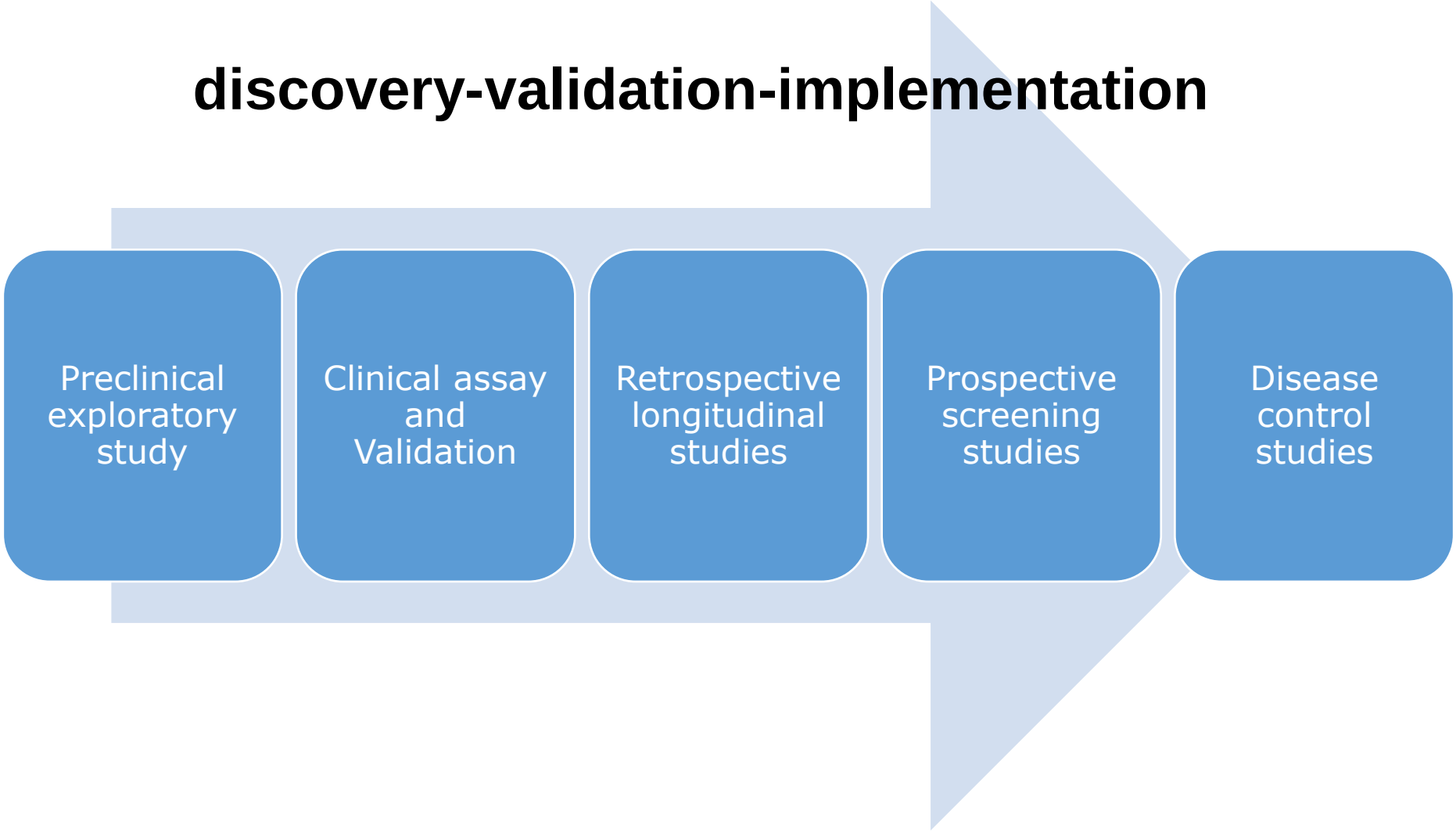
Table 1. Categorization of different biomarkers.

Term	Definitions	Properties	Examples
Preventative biomarkers	Biological traits that can help in patient-specific disease prevention	Often DNA biomarkers, stable over time and across tissue types	BRCA1 and BRCA2 in breast cancer prevention
Diagnostic biomarkers	Clinical indications that help doctors screen, diagnose, or measure the severity for a certain type of disease	Often exist after onset of the disease and disappear after the disease is cured	Citrullinated peptides/proteins (anti-CCP antibodies) in rheumatoid arthritis diagnosis
Prognostic biomarkers	A biomarker that monitors the disease progression or predicts the disease outcome	Often exist after the onset of the disease and may change over time; belong to either DNA cancer biomarkers or general biomarkers	MammaPrint to predict the metastasis in breast cancer
Predictive biomarkers	A biomarker that predicts the treatment response of a certain disease	Same as prognostic biomarkers.	Human epidermal growth factor receptor 2 (HER2) for breast cancer treatment of trastuzumab and lapatinib

He. Clin Invest 2015

Phases of Biomarker Development

discovery-validation-implementation



Pepe et al. J Natl Cancer Inst. 2001

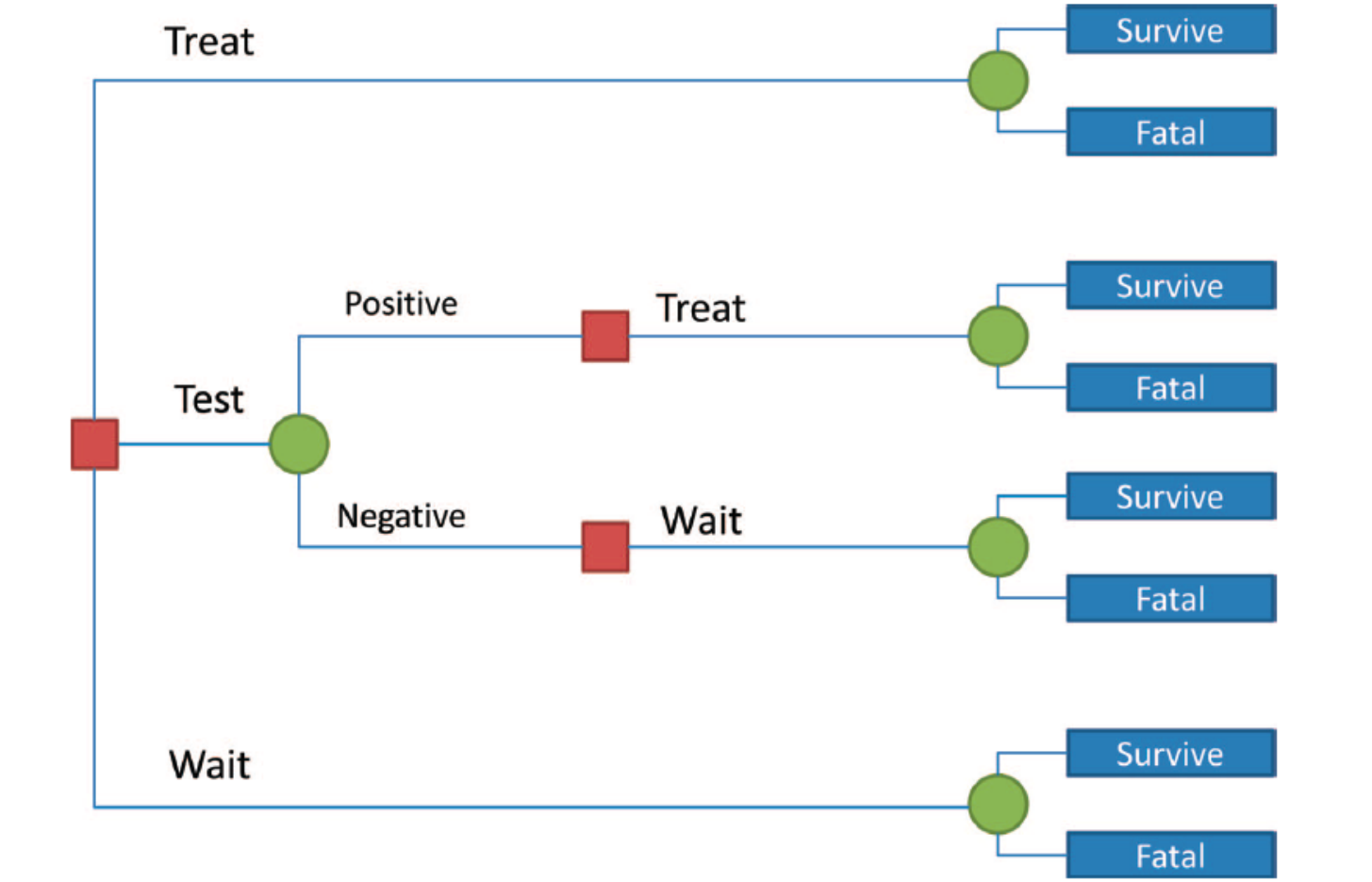
Biomarker 연구의 4단계 (Gludd 제안)

- Phase Ⅰ: Determining the **normal range** of values for a diagnostic test through **observational studies in healthy people**
- Phase Ⅱ: Determining the diagnostic **accuracy through case-control studies**, including healthy people and (a) people with known disease assumed by diagnostic standard and (b) people with suspected disease
- Phase Ⅲ: Determining the clinical consequences of introducing a diagnostic test through **randomized trials**
- Phase Ⅳ: Determining the **effects** of introducing a new diagnostic test into clinical practice by **surveillance in large cohort studies**

Biomarker evaluation: questions

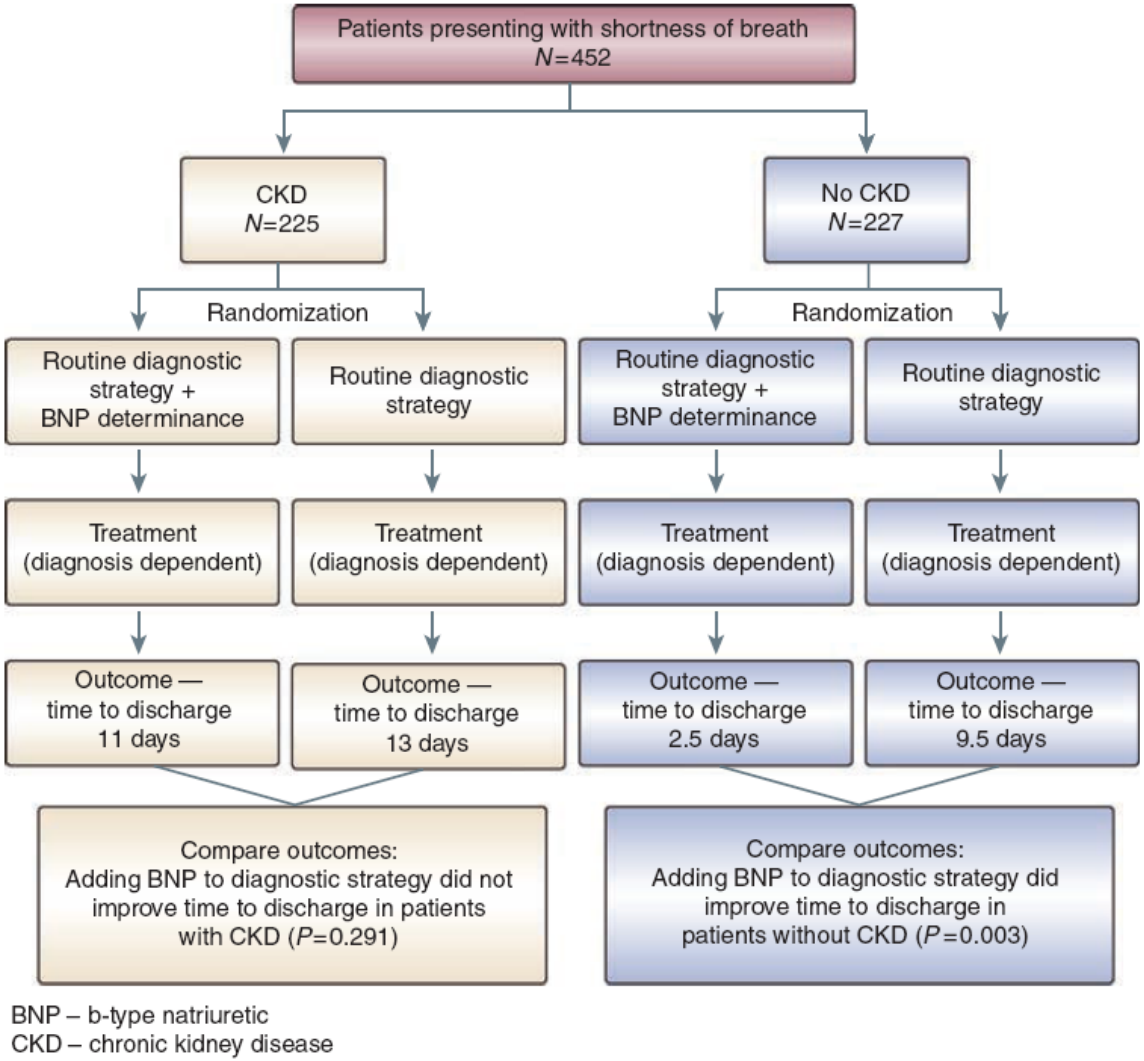
- analytical validity
  - ✓ "Can I trust the results of this marker?"
- clinical validity: accuracy of marker
  - ✓ "Are the results of this test meaningful?"
- clinical utility
  - ✓ "Is using the marker helpful in improving or maintaining the health of patients"

Clinical utility: 연구 설계



Bossuyt et al Clin Chem 2012

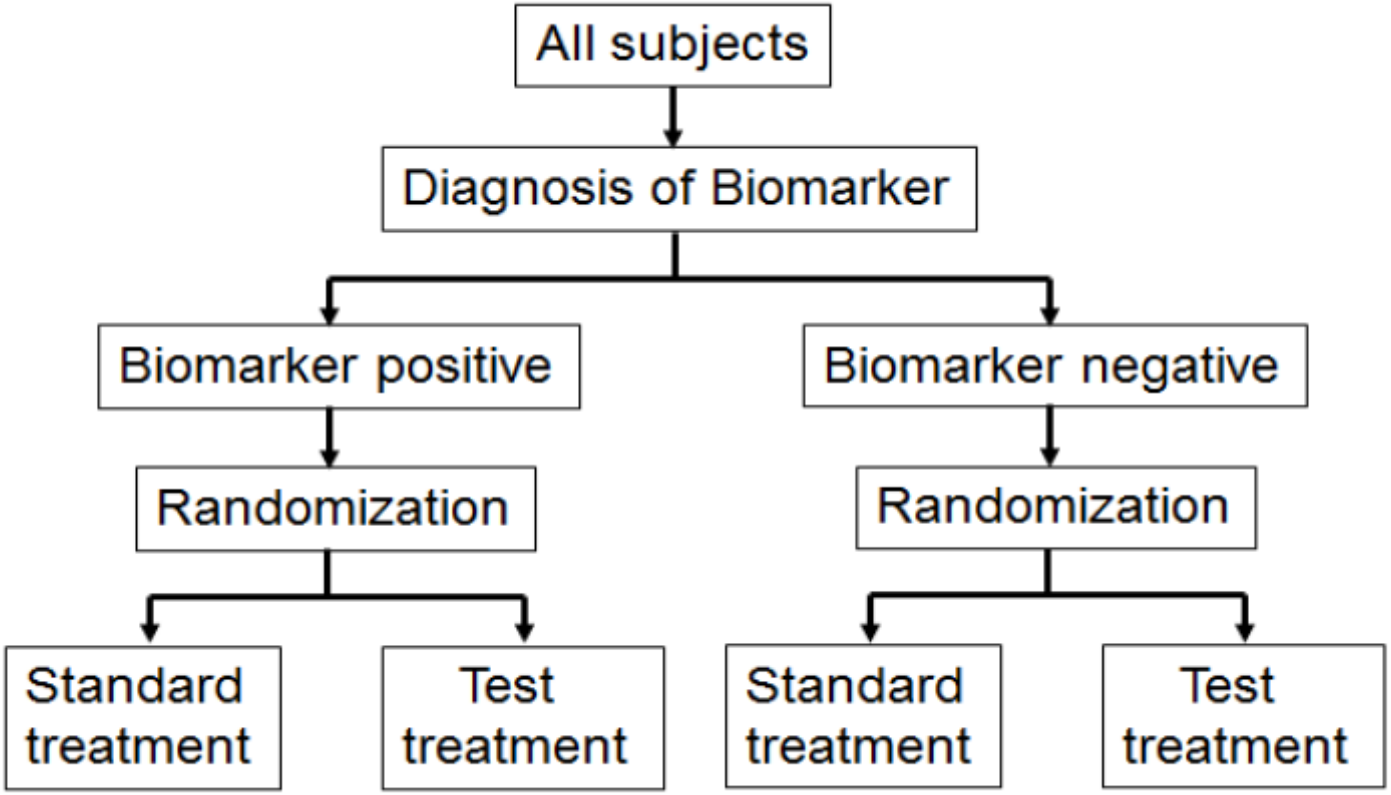
Clinical utility 예



BNP – b-type natriuretic
CKD – chronic kidney disease

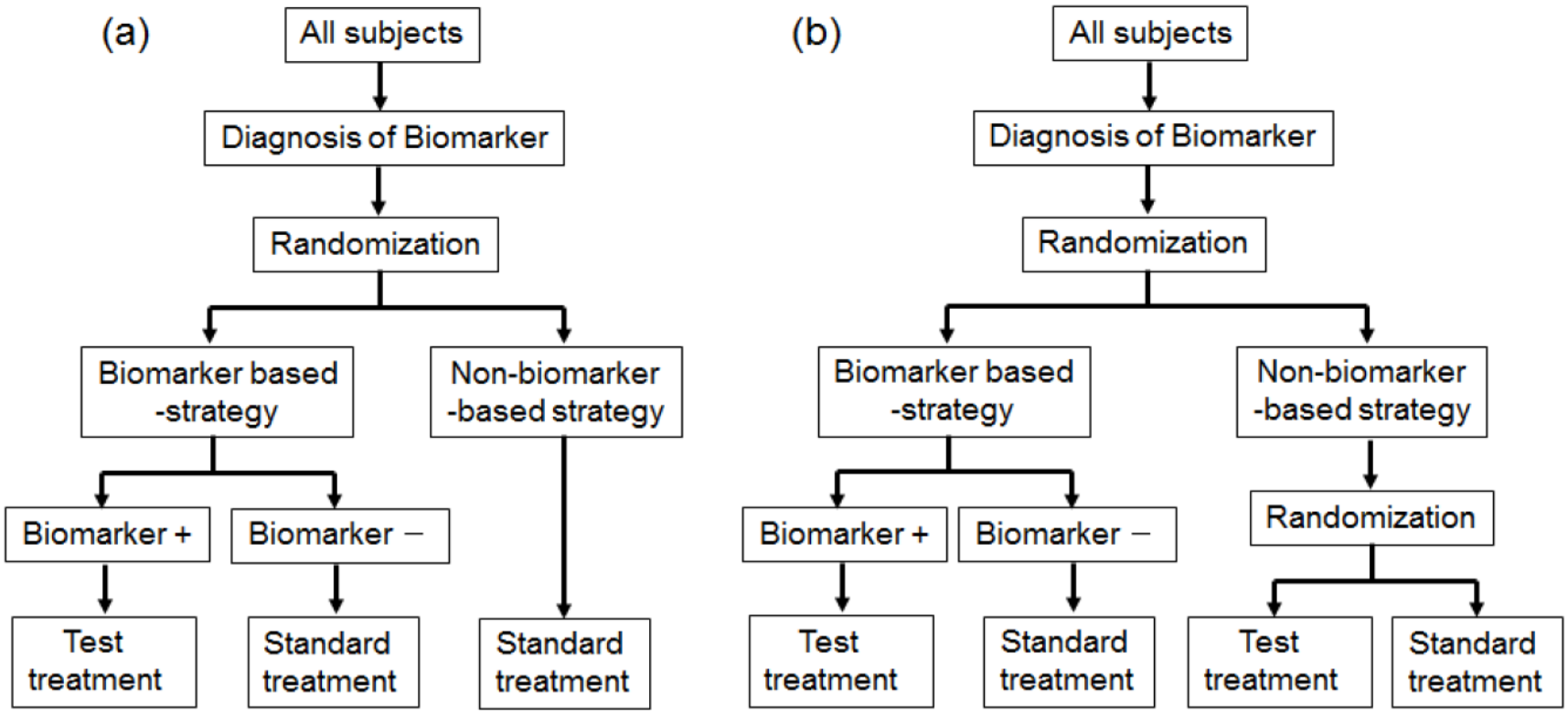
Stralen et al. Kidney Int 2009

Biomarker by treatment interaction design



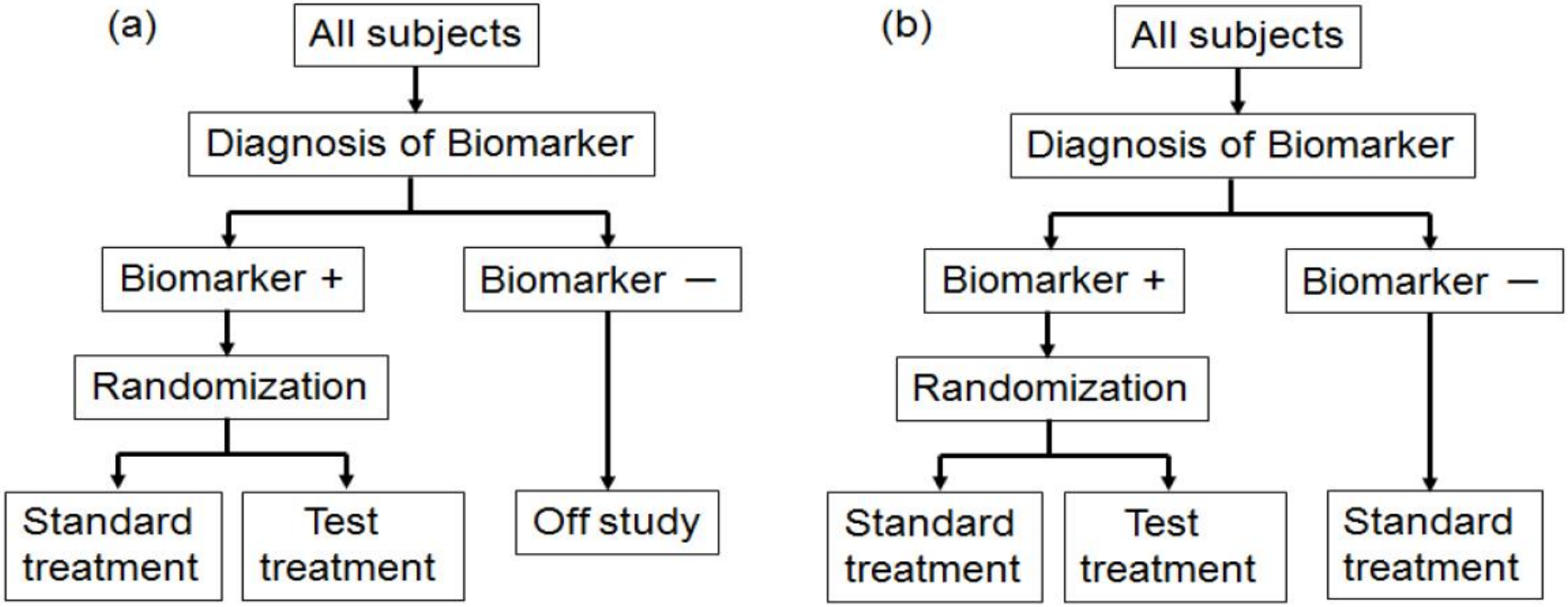
Gosho et al. Sensors 2012

Biomarker-strategy design



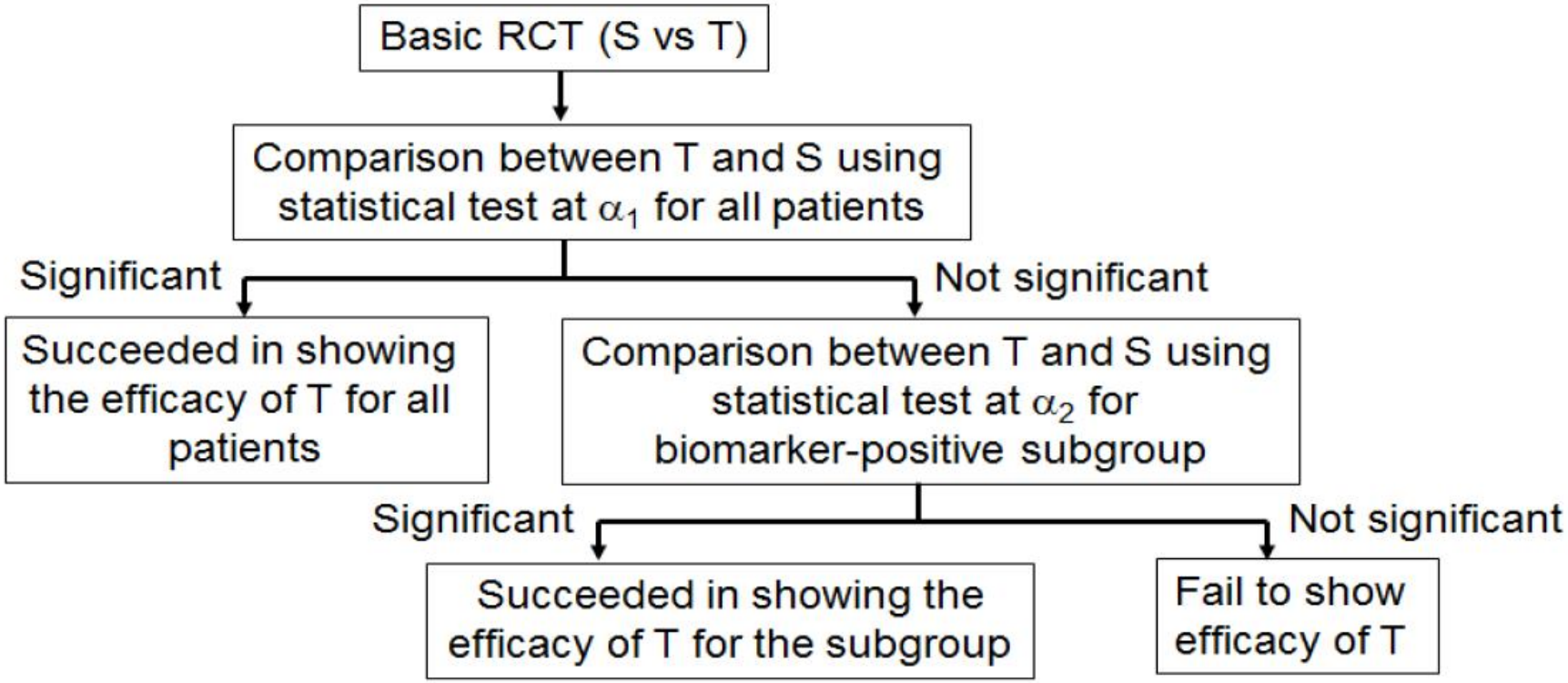
Biomarker strategy design	Patients are randomized to biomarker strategy group and control group	Complete information gathered; able to test companion diagnostic tools	Scale and cost might be large

Enrichment Design and Hybrid Design

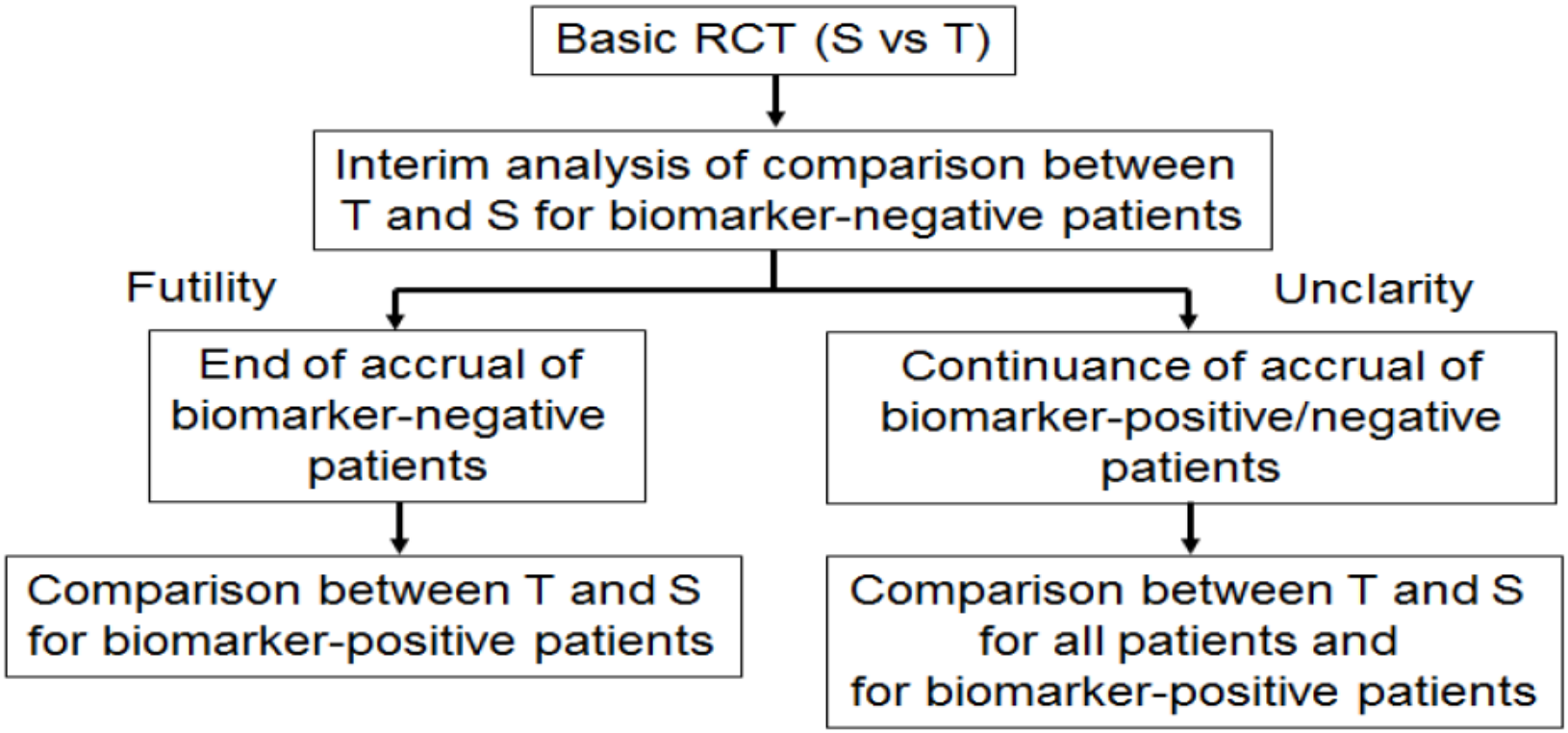


Designs	Properties	Advantages	Disadvantages
Enrichment design	Only enroll biomarker positive patients	Good for biomarker with clear evidence and/or low prevalence	Cannot gather treatment information for all population; cannot test for the companion diagnostic tool validity

adaptive signature and biomarker-adaptive threshold designs

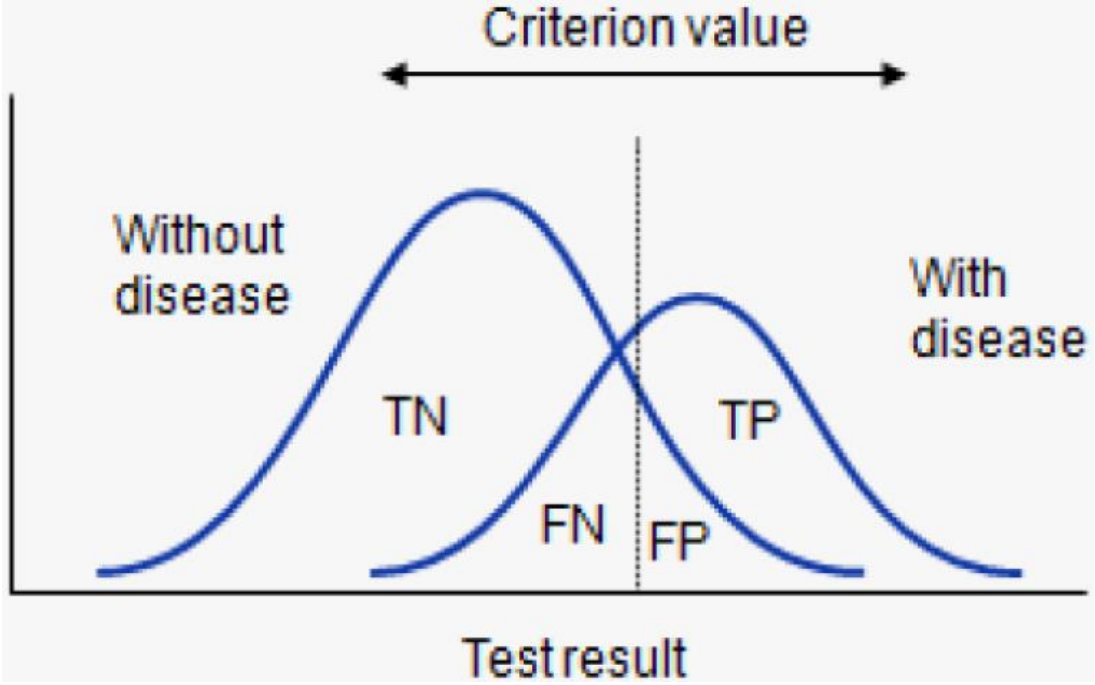


adaptive accrual design



Clinical validity 연구의 배경

- 환자와 정상인을 100% 정확히 가려내는 biomarker가 있다면? (Diagnostic accuracy)



Classification, Discrimination, Calibration, and Re-classification

- Classification
  - ✓ 생체지표가 대상자를 잘 정의된 인구집단 그룹으로 분류
- Discrimination
  - ✓ 생체지표가 관심 질병에 대한 고위험 대상자와 저위험 대상자를 구별하는 능력
  - ✓ 예: PPV, NPV, Mean Risk Difference (MRD), Net Benefit (NB), Overall concordance measures 등
- Calibration
  - ✓ 생체지표가 인구집단 내에서 관심 질병이 일어나는 것을 예측하는 능력
  - ✓ 예: Brier score
- Re-classification
  - ✓ 새로운 생체지표를 통해 기존의 분류에서 재분류하는 것

Binary Markers의 Accuracy 측정

biomarker	질병 있음	질병 없음
양성	**True positive (TP)**	**False positive (FP)**
음성	**False negative (FN)**	**True negative (TN)**

Accuracy 측정 지표

- Error-based measure
  - ✓Sensitivity, Specificity
  - ✓Overall fraction correct
  - ✓Youden's J index
    = true positive fraction – false positive fraction
  - ✓AUC

- Information-based measure
  - ✓Positive/Negative predictive value
  - ✓Positive/Negative likelihood ratio

- Association-based measure
  - ✓Diagnostic odds ratio = LR+ / LR-
    = (TP x TN) / (FP x FN)
  - ✓Kappa

예

- 결핵을 1시간 내에 진단할 수 있는 A 키트.
  A 키트를 validation 하기 위해 한국과 미국에서 각각 한국인과 미국인을 대상으로 임상에서 결핵이 의심되는 환자들에게 적용

- 한국에서와 미국에서의 A 키트의 민감도, 특이도, 양성예측도, 음성예측도를 비교해보면?

(단, 결핵 유병률은 한국이 더 높으며 A 키트의 진단능력에 인종, 성별 등이 영향을 주지는 않음)

Summary

		Condition (as determined by "Gold standard")			
	Total population	Condition positive	Condition negative	Prevalence = Σ Condition positive / Σ Total population	
Test outcome	Test outcome positive	**True positive**	**False positive** (Type I error)	Positive predictive value (PPV, Precision) = Σ True positive / Σ Test outcome positive	False discovery rate (FDR) = Σ False positive / Σ Test outcome positive
	Test outcome negative	**False negative** (Type II error)	**True negative**	False omission rate (FOR) = Σ False negative / Σ Test outcome negative	Negative predictive value (NPV) = Σ True negative / Σ Test outcome negative
	Positive likelihood ratio (**LR+**) = TPR/FPR	True positive rate (TPR, Sensitivity, Recall) = Σ True positive / Σ Condition positive	False positive rate (FPR, Fall-out) = Σ False positive / Σ Condition negative	Accuracy (ACC) = Σ True positive + Σ True negative / Σ Total population	
	Negative likelihood ratio (**LR−**) = FNR/TNR	False negative rate (FNR) = Σ False negative / Σ Condition positive	True negative rate (TNR, Specificity, SPC) = Σ True negative / Σ Condition negative		
	Diagnostic odds ratio (**DOR**) = LR+/LR−				

http://en.wikipedia.org/wiki/Sensitivity_and_specificity

Continuous Markers의 Accuracy 측정

RPHA 법을 통한 HBsAg 검출법의 정확도를 측정하기 위해 RIA 법을 황금기준으로 하여 다음과 같은 결과를 얻었다

RPHA 결과 (titer)	RIA 결과		합
	양성	음성	
2^0	2	15	17
2^1	12	94	106
2^2	6	8	14
2^3	7	4	11
2^4	52	5	57
합	79	126	205

RPHA 결과 (titer)	RIA 결과		합
	양성	음성	
2^0	2	15	17
2^1	12	94	106
2^2	6	8	14
2^3	7	4	11
2^4	52	5	57
합	79	126	205

민감도= 77/79
특이도 = 15/126

RPHA 결과 (titer)	RIA 결과		합
	양성	음성	
2^0	2	15	17
2^1	12	94	106
2^2	6	8	14
2^3	7	4	11
2^4	52	5	57
합	79	126	205

민감도= 65/79
특이도= 109/126

RPHA 결과 (titer)	RIA 결과		합
	양성	음성	
2^0	2	15	17
2^1	12	94	106
2^2	6	8	14
2^3	7	4	11
2^4	52	5	57
합	79	126	205

민감도= 59/79
특이도= 117/126

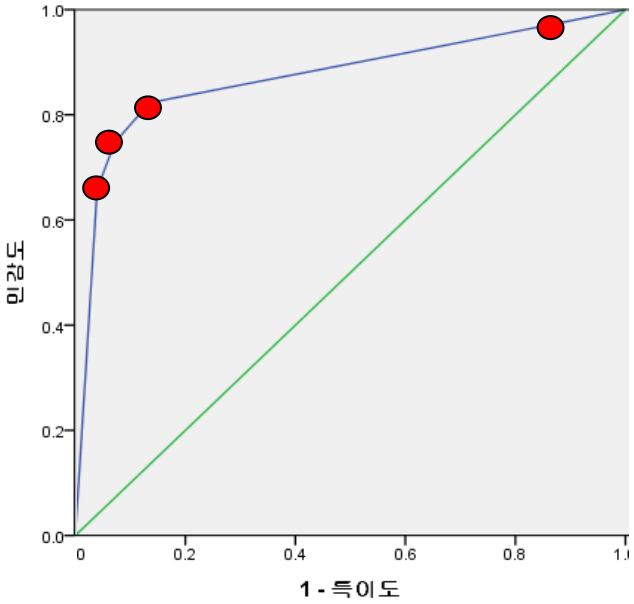
RPHA 결과 (titer)	RIA 결과		합
	양성	음성	
2^0	2	15	17
2^1	12	94	106
2^2	6	8	14
2^3	7	4	11
2^4	52	5	57
합	79	126	205

민감도= 52/79
특이도= 121/126

Continuous Markers의 Accuracy 측정: ROC

- 민감도와 특이도 간 trade-off

- AUC
  - ✓ROC 곡선 아래 면적
  - ✓진단도구의 정확도
  - ✓우연에 의한 값 : 0.5
  - ✓AUC 가 클수록 정확한 진단도구



- 왼쪽 위 지점과 가까운 점이 cut-off

AUC 산출: SAS 결과

The LOGISTIC Procedure

Odds Ratio Estimates

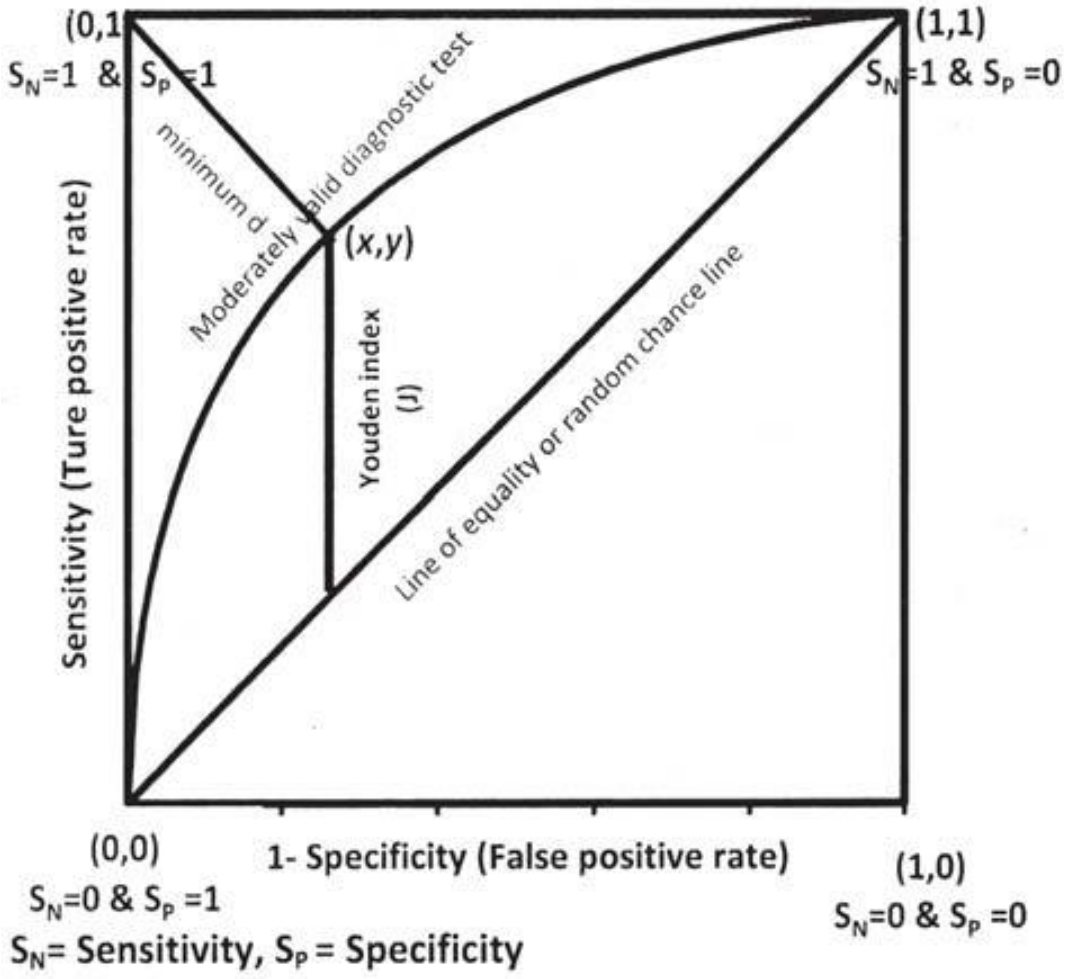
Effect	Point Estimate	95% Wald Confidence Limits	
hba1c	22.125	15.760	31.059

Association of Predicted Probabilities and Observed Responses

Percent Concordant	91.0	Somers' D	0.837
Percent Discordant	7.3	Gamma	0.852
Percent Tied	1.7	Tau-a	0.180
Pairs	716372	c	0.918

= AUC

Optimal cut-off point



Kumar et al. Indian Pediatr. 2011

Optimal cut-off point

- Euclidean method

$$\sqrt{(1-sensitivity)^2+(1-specificity)^2} \quad : \text{minimize}$$

- Youden Index

$$(sensitivity)+(specificity)-1 \quad : \text{maximization}$$

- Likelihood ratio

$$LR+=\frac{sensitivity}{(1-specificity)} \text{ :maximization, } LR-=\frac{(1-sensitivity)}{specificity} \text{ :minimization}$$

- Efficiency

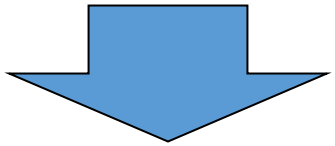
$$p\times sensitivity+(1-p)\times specificity, \quad p=prevalence\ rate$$

두 biomarker의 민감도 비교

독립된 두 집단의 경우 Chi-Square test / Fisher's exact test

검사결과	질병 있음	질병 없음
양성	80	10
음성	20	90

검사결과	질병 있음	질병 없음
양성	90	10
음성	10	90



검사결과	Test 1	Test 2
양성	80	90
음성	20	10

P value = 0.048

두 biomarker의 민감도 비교

동일 환자군에서 두 진단 검사 수행한 경우 McNemar's test

		Test 1	
	검사결과	양성	음성
Test 2	양성	**40**	**10**
	음성	**5**	**45**

P value = 0.302

AUC 비교

체질량지수(BMI)와 허리둘레 중 당뇨병 발생을 예측하는데 더 유용한 비만 관련 생체지표는?

체질량지수를 통해 당뇨병 발생을 예측할 때, 남성과 여성에서 AUC 차이는 ?

당뇨병 발생과 관련된 인자 중 나이, 성별, 입적 당시 혈당상태, 체질량지수로 구축된 모델에서 혈중 당화혈색소 농도를 추가하였을 때 AUC의 변화는 ?

두 ROC 비교

➢독립된 두 집단에서 각각 산출된 ROC 비교

$$z = \frac{AUC1 - AUC2}{\sqrt{SE_1{}^2 + SE_2{}^2}}$$

Z는 표준정규 분포 or
Z^2은 자유도 1인 카이제곱 분포를 따른다.
→ 3.841인 경우 P=0.05

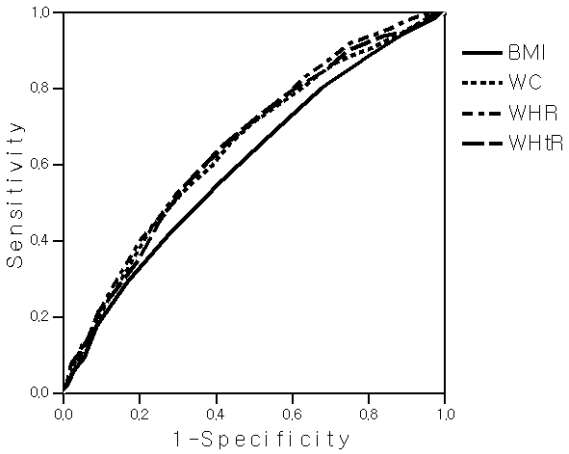
➢동일 집단에서 산출된 두 ROC 비교

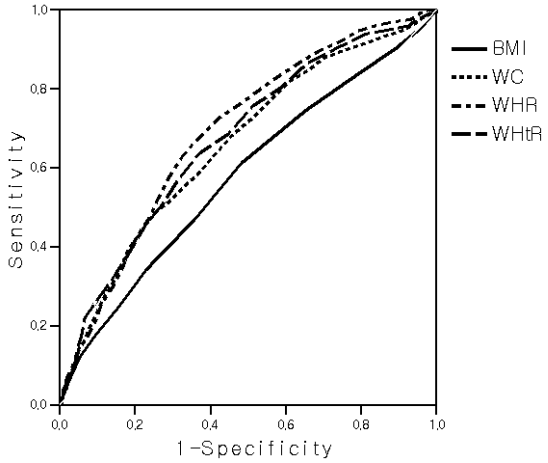
$$z = \frac{AUC1 - AUC2}{\sqrt{SE_1{}^2 + SE_2{}^2 - 2rSE_1SE_2}}$$

r: 두 진단결과값의 상관계수

두 ROC 비교 예

- ROC curves and AUC of obesity indices in relation to multiple metabolic risk factors





	Male			Female		
	AUC (95% CI)	P value	FDR	AUC (95% CI)	P value	FDR
BMI	0.605 (0.577-0.634)			0.581 (0.551-0.612)		
WC	0.646 (0.618-0.674)	0.046	0.091	0.657 (0.628-0.685)	<0.001	0.001
WHR	0.660 (0.633-0.687)	0.007	0.044	0.690 (0.663-0.717)	<0.001	<0.001
WHtR	0.651 (0.624-0.679)	0.025	0.075	0.673 (0.645-0.701)	<0.001	<0.001

Ko et al. J Epidemiol 2012

New Biomarker의 가치 평가

- 단순히 Outcome과의 association으론 부족
  - ✓ P-value, OR, HR 로는 불충분
- 모델의 Predictive accuracy 향상 정도
- Biomarker 추가 모델 구축에서의 이슈
  - ✓Internal validation
    - Cross-validation
    - Bootstrap
  - ✓External validation

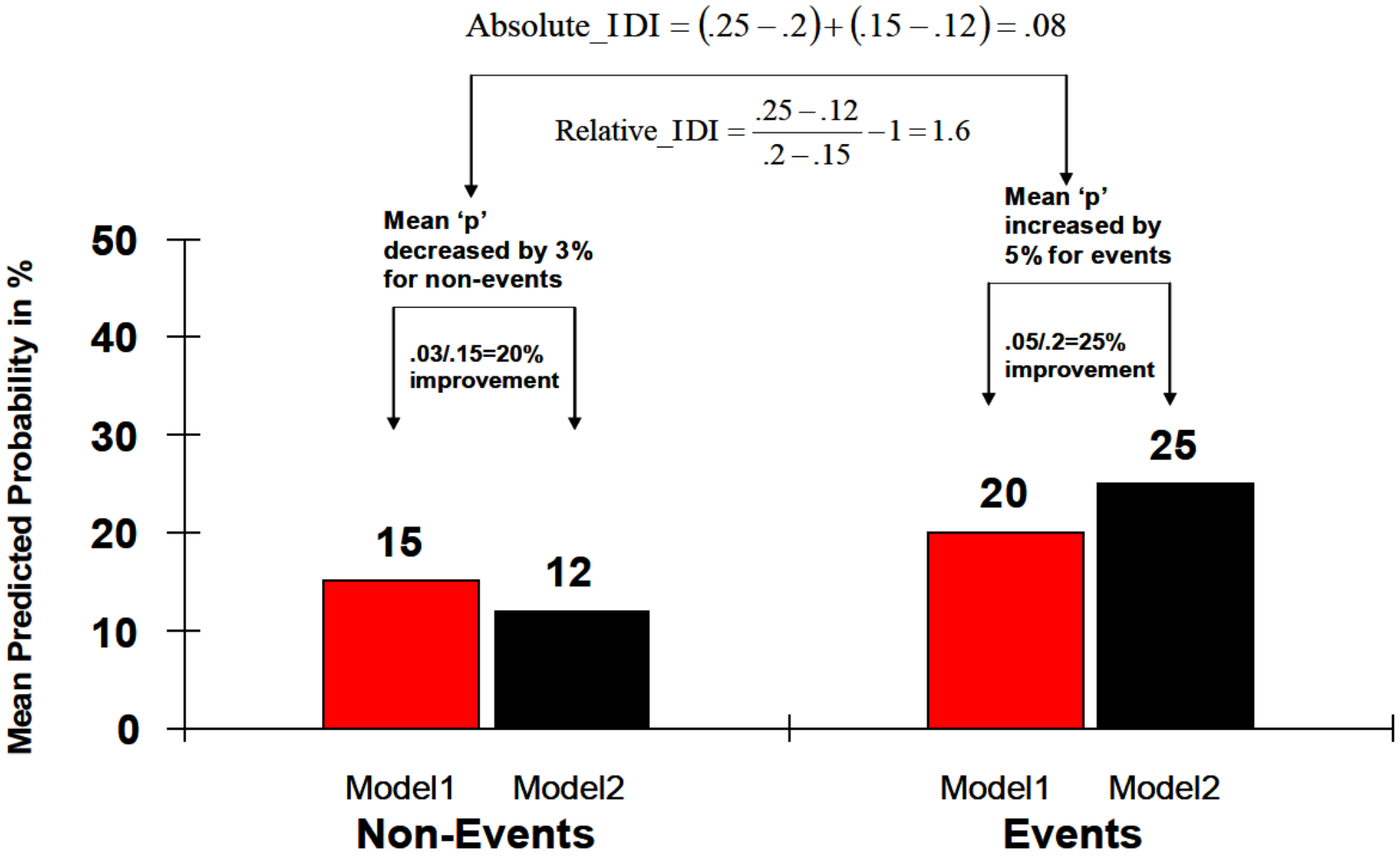
New Biomarker의 Incremental Value

$$Model1 : \text{logit}(y) = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n$$

$$Model2 : \text{logit}(y) = \alpha + \beta_1 x_1 + \ldots + \beta_n x_n + \beta_{n+1} x_{n+1}$$

- $Model_1$에 비해 X_{n+1}를 추가하였을 때 classification accuracy를 얼마나 향상시키는가
- β_{n+1} 값 추정
- 통계적 방법
  - ✓AUC 비교
  - ✓IDI (Integrated Discrimination Index)
  - ✓NRI (Net Reclassification improvement)
  - ✓Hosmer-Lemeshow Chi-Square test

IDI (Integrated Discrimination Index)



Kennedy et al

NRI(Net Reclassification improvement)

- predicted probabilities를 categorical scale 에서 평가

- 기존모델과 새모델에서의 Net Reclassification 된 개체 수 평가

$$NRI = \left[\left(\frac{\#\text{Events moving up}}{\#\text{of Events}}\right) - \left(\frac{\#\text{Events moving down}}{\#\text{of Events}}\right)\right] + \left[\left(\frac{\#\text{Non - events moving down}}{\#\text{of Non - events}}\right) - \left(\frac{\#\text{Non - events moving up}}{\#\text{of Non - events}}\right)\right]$$

NRI 예

- 3그룹으로 정의 시 low, moderate, high를 각각 probability가 10% 미만, 10~20% 미만, 20% 이상으로 정의

Table1: Crosstab for Events (n=100)

Model 1 \ Model 2	Low	Mod	High
Low	10	8	2
Mod	3	30	10
High	2	5	30

20 Events moving up

10 Events moving down

70 Events not moving

Net of 10/100 (10%) of events getting reclassified correctly

Table2: Crosstab for Non-events (n=200)

Model 1 \ Model 2	Low	Mod	High
Low	50	5	0
Mod	20	40	10
High	5	10	60

15 Non-events moving up

150 Non-events moving down

35 Non-events not moving

Net of 20/200 (10%) of non-events getting reclassified correctly

$$NRI = \left[\left(\frac{20}{100}\right) - \left(\frac{10}{100}\right)\right] + \left[\left(\frac{35}{200}\right) - \left(\frac{15}{200}\right)\right] = .2$$

- 그룹 정의에 따라 달라지므로 자료 분석 전에 미리 결정

Kennedy et al

Biomarker 연구에서의 bias

- Selection bias
  - ✓ 활용 가능한 specimen 에서 편의 추출한 경우
- Spectrum bias
  - ✓ 건강한 대상자 ------- 경증 환자 ------ 중증 환자
- Verification bias
  - ✓ 모든 대상자가 reference standard로 입증된 것은 아님
- Imperfect Reference Standard Bias
  - ✓ reference standard가 완벽하지 않은 경우

Biomarker 연구에서의 bias

- Co-morbidity bias
  - ✓ 다른 동반질환에 의한 영향
- Ordering bias
  - ✓ 비교 검사와 표준 검사를 통해 얻어지는 결과의 순서가 무작위되어 있지 않음 검사 받는 질환자와 비질환자의 순서가 무작위되어 있지 않음
  - ✓ 예측을 위한 검사에서 검사결과는 목표질환이 생긴 이후에 검사결과를 얻는 경우
- Test interpretation, integrity, and context bias
  - ✓ 측정자가 true disease status에 대해 masking 되지 않음
  - ✓ 임상 현장에서 다른 임상적 정보에 대한 접근성이 항상 일치하지는 않음

The END !!!