GEÇERLİLİK & GÜVENİLİRLİK ARAŞTIRMALARI

Dr. Meltem Şengelen
HÜTF Halk Sağlığı AD
19 Şubat 2015

Epidemiyolojik Araştırmalar

I. GÖZLEMSEL
- 1. TANIMLAYICI
  - 1.1. Toplumsal
    - "Populations" (ekolojik çalışmalar, vb.)
    - Korelasyon çalışmaları
  - 1.2. Kişisel
    - Vaka takdimi
    - Vaka serileri
  - 1.3. Kesitsel
- 2. ANALİTİK
  - 2.1. Kesitsel
  - 2.2. Vaka-Kontrol
  - 2.3. Kohort

II. DENEYSEL/MÜDAHALE
- 1. KLİNİK (RANDOMİZE) DENEYLER
- 2. SAHA DENEYLERİ

III. METODOLOJİK
- **1. GEÇERLİLİK**
- **2. GÜVENİLİRLİK**

Geçerlilik & Güvenilirlik Araştırmaları

- Hastalıkların teşhisinde, tarama programlarında **hastalarla sağlamların ayırt edilmesinde** çeşitli fizik muayene yöntemleri (gözlem, palpasyon, perküsyon vb), çeşitli laboratuvar yöntemleri (hematolojik, mikrobiyolojik, radyolojik ve nükleer teknikler vb) kullanılır.

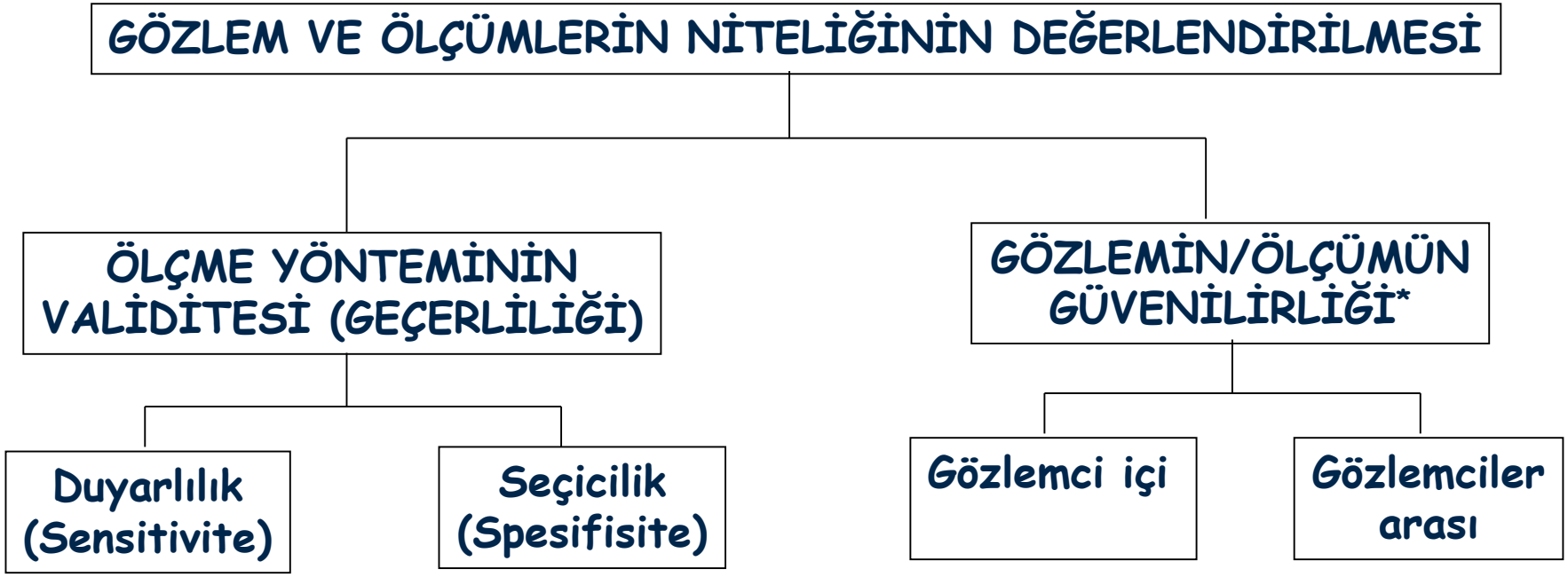
- Bu yöntemlerin kişiye **doğru tanı koymada ne kadar geçerli** oldukları ve bu yöntemleri kullanan kişilerden kaynaklanan **gözlem-ölçüm hatalarının** boyutunun ne olduğunu, yani gözlem ve ölçümlerin **kalitesi**ni belirlemek için yapılan araştırmalardır.

Tanı Testleri

- Klinik tanı
- Vaka bulma
- Erken Tanı-Tarama
- Epidemiyolojik çalışma
  - insidans
  - prevalans

Geçerlilik & Güvenilirlik Araştırmaları

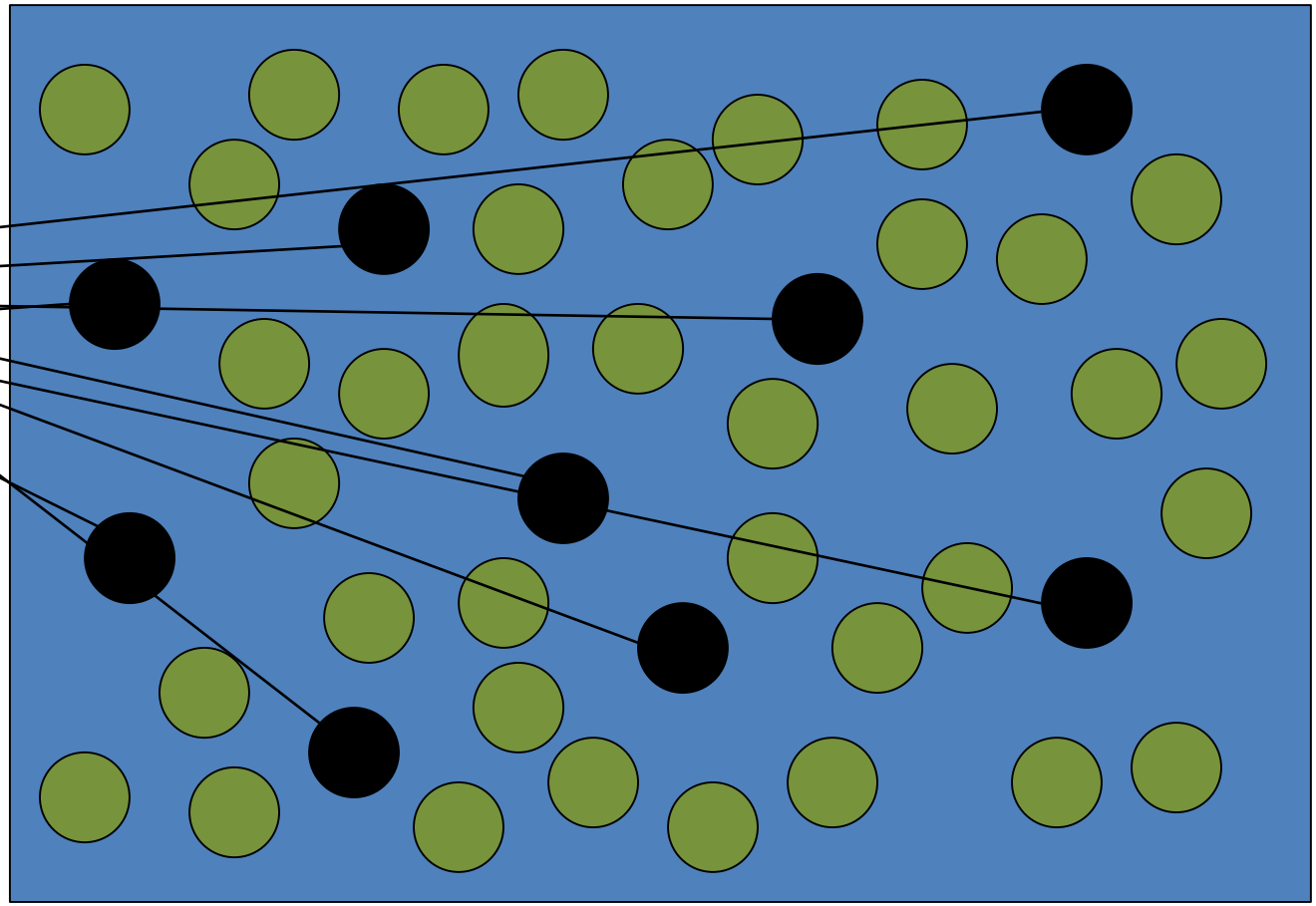
- **Geçerlilik-validite araştırmaları**: tanı yönteminin ne kadar doğru tanı koyduğunu belirler
- **Güvenilirlik-tutarlılık araştırmaları:** gözlem ve ölçümü yapan kişilerin, gözlemcilerin kendi kendileri veya birbirleriyle ne kadar tutarlı gözlem ve ölçüm yaptıklarını belirler



*Tutarlılık (consistency)
Uyum (agreement)
Aynı sonucun elde edilebilirliği (reproducibility)
Tekrar edilebilirlik (repeatability)

Sensitivite (Duyarlılık)

Gerçekten hasta olanların ne kadarını hasta olarak saptayabiliyor?



9/9=%100

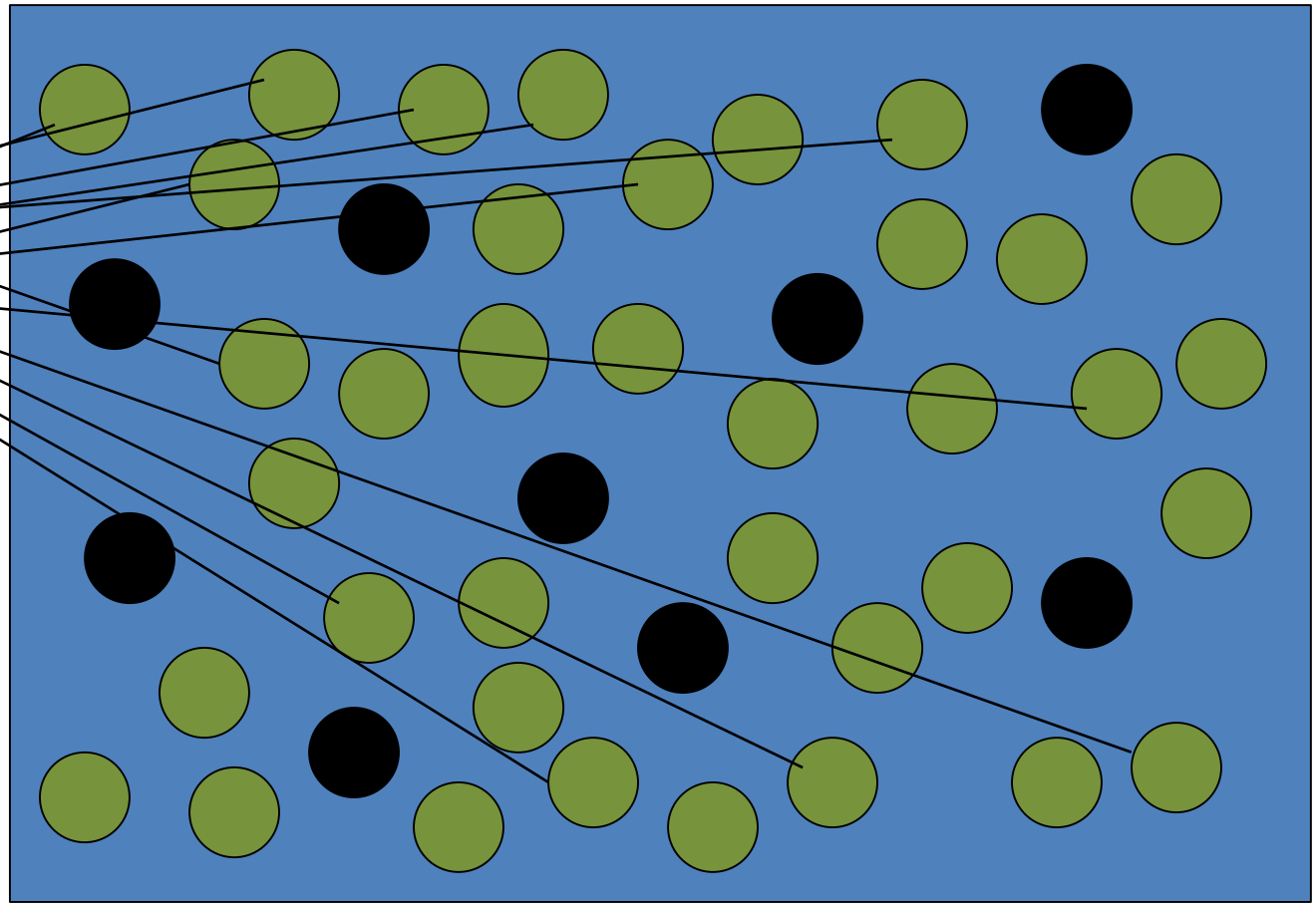
Spesifisite (Seçicilik)

- Validitesi saptanacak olan yeni ölçüm yönteminin, ***sağlam olanlardan ne kadarını doğru olarak (sağlam) saptayabildiğ***ini gösterir.

- Yeni test ile de sağlam bulunanların (***doğru negatifler***) referans teste göre sağlam olduğu bilinen kişilere oranı yeni testin seçiciliğini verir.

Spesifisite (Seçicilik)

Gerçekten sağlam olanların ne kadarını sağlam olarak saptayabiliyor?



13/37=%35

Referans Test

- Bir hastalık için kesin hasta ve sağlam tanısı koydurabilecek, geçerliliği daha önceden saptanmış ve yüksek olan yöntemlere **referans yöntem** adı verilir.

- Bir testin/yöntemin geçerliliği gerçek durumu yansıtan bu yönteme göre belirlenir.

		GEÇERLİ REFERANS TEST	
		Hasta	Sağlam
YENİ TEST	Hasta	a **Doğru Pozitifler**	b Yanlış Pozitifler
	Sağlam	c Yanlış Negatifler	d **Doğru Negatifler**
Toplam		a+c (Toplam hasta)	b+d (Toplam sağlam)

$$\text{Sensitivite} = \frac{a}{a+c} \times 100 = \frac{\text{Yeni testin saptadığı hastalar}}{\text{Toplam hasta sayısı (Referans teste göre)}}$$

$$\text{Spesifisite} = \frac{d}{b+d} \times 100 = \frac{\text{Yeni testin saptadığı sağlamlar}}{\text{Toplam sağlam sayısı (Referans teste göre)}}$$

Serviks Kanseri Taramalarında Pap Smear Tekniğinin Validitesi

		GEÇERLİ REFERANS TEST (Biyopsi)		
		Kanser var	Kanser yok	**Toplam**
YENİ TEST (SMEAR)	Kanser var	**96**	250	346
	Kanser yok	4	**250**	254
Toplam		100	500	600

Sensitivite = $\frac{96}{100}$ x 100 = %96 = $\frac{\text{Doğru Pozitifler}}{\text{Toplam hastalar (Referans teste göre)}}$

Spesifisite = $\frac{250}{500}$ x 100 = %50 = $\frac{\text{Doğru Negatifler}}{\text{Toplam sağlamlar (Referans teste göre)}}$

Prediktif Değerler

- Bir hastada tarama testi sonucu pozitif ise bu kişinin hasta olma olasılığı nedir?
- Tarama testinin pozitif olduğu kişilerin ne kadarı gerçekten hasta?

Testin Pozitif Prediktif Değeri

Uygulanan yeni tanı testine pozitif sonuç veren kişilerin ne kadarı gerçekten hasta (pozitif)?

Testin Negatif Prediktif Değeri

Uygulanan yeni tanı testine negatif sonuç veren kişilerin ne kadarı gerçekten sağlam (negatif)?

Yeni Tanı Testinin Pozitif Prediktif Değeri = $\frac{\text{Doğru pozitifler}}{\text{Toplam pozitifler (Yeni teste göre)}}$

Yeni Tanı Testinin Negatif Prediktif Değeri = $\frac{\text{Doğru negatifler}}{\text{Toplam negatifler (Yeni teste göre)}}$

Prediktif değerler, genellikle prevalanstan etkilenirler.

Prevalans ↑, pozitif prediktif değer ↑

Spesifisite ↑, pozitif prediktif değer ↑

	REFERANS TEST		
YENİ TEST	Hasta	Hasta değil	Toplam
Pozitif	6	8	14
Negatif	34	152	186
Toplam	40	160	200

Prevalans = 40/200 = **%20**

PPD $= \frac{6}{14} =$ **%42.9**

NPD $= \frac{152}{186} =$ **%81.7**

	REFERANS TEST		
YENİ TEST	Hasta	Hasta değil	Toplam
Pozitif	18	4	22
Negatif	102	76	178
Toplam	120	80	200

Prevalans = 120/200 = **%60**

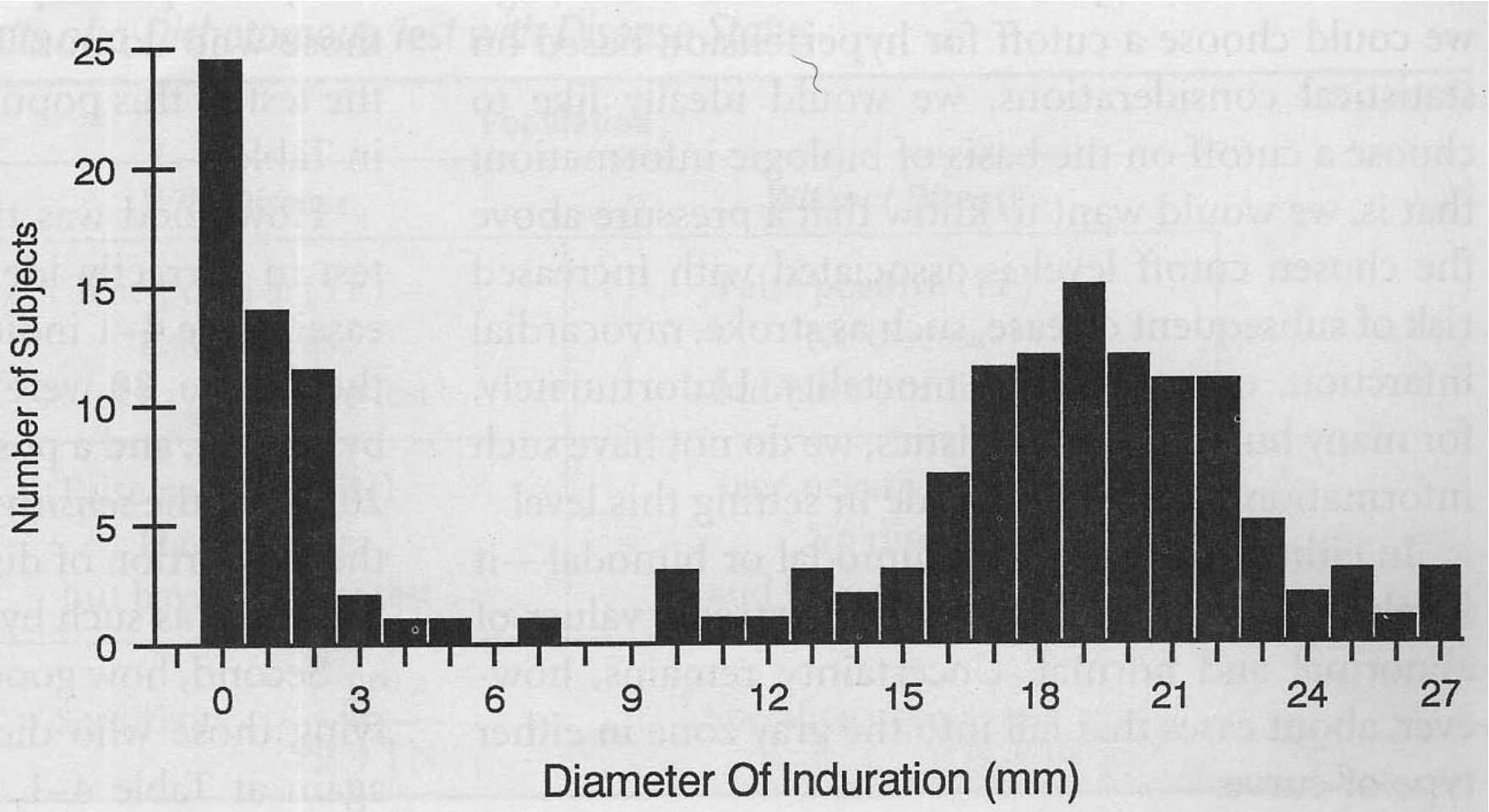
PPD $= \frac{18}{22} =$ **%81.8**

NPD $= \frac{76}{178} =$ **%42.7**

Solunum Sistemi Yakınmalarının Spirometrik Değerlendirme Bulgularına Göre Validitesi

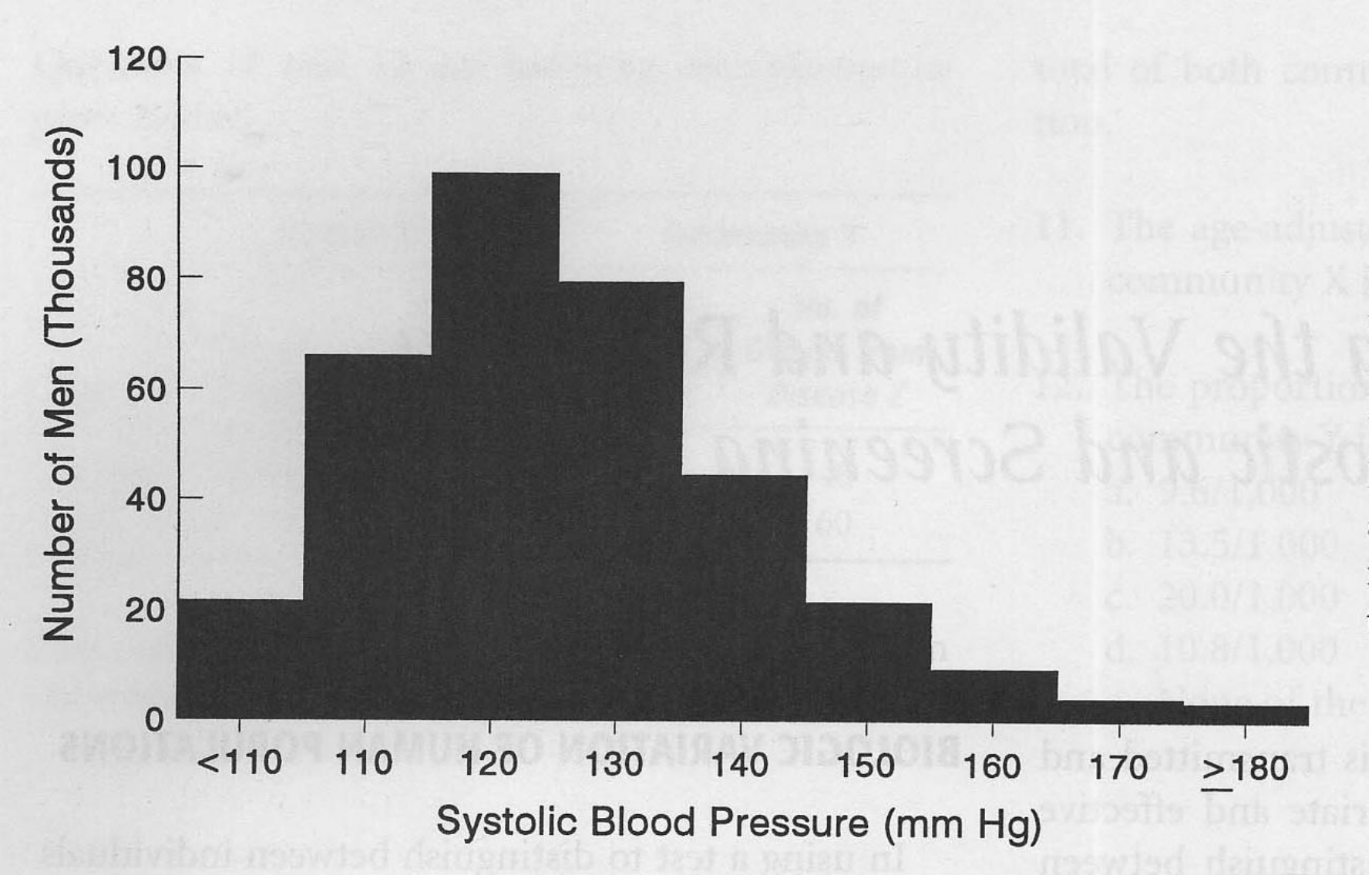
Solunum sistemi yakınmaları	SPİROMETRİK MUAYENE			
Öksürük	**Obstruksiyon (+)**	**Obstruksiyon (-)**	**Toplam**	
+	212	304	516	Sensitivite: %91 Spesifisite: %79 PPD: %41
-	21	1164	1185	
Balgam				
+	195	265	460	Sensitivite: %84 Spesifisite: %82 PPD: %42
-	38	1203	1241	
Nefes darlığı				
+	173	220	393	Sensitivite: %74 Spesifisite: %86 PPD: %44
-	60	268	1328	
Toplam	233	1468	1701	

Tüberkülin Reaksiyonu Dağılımı

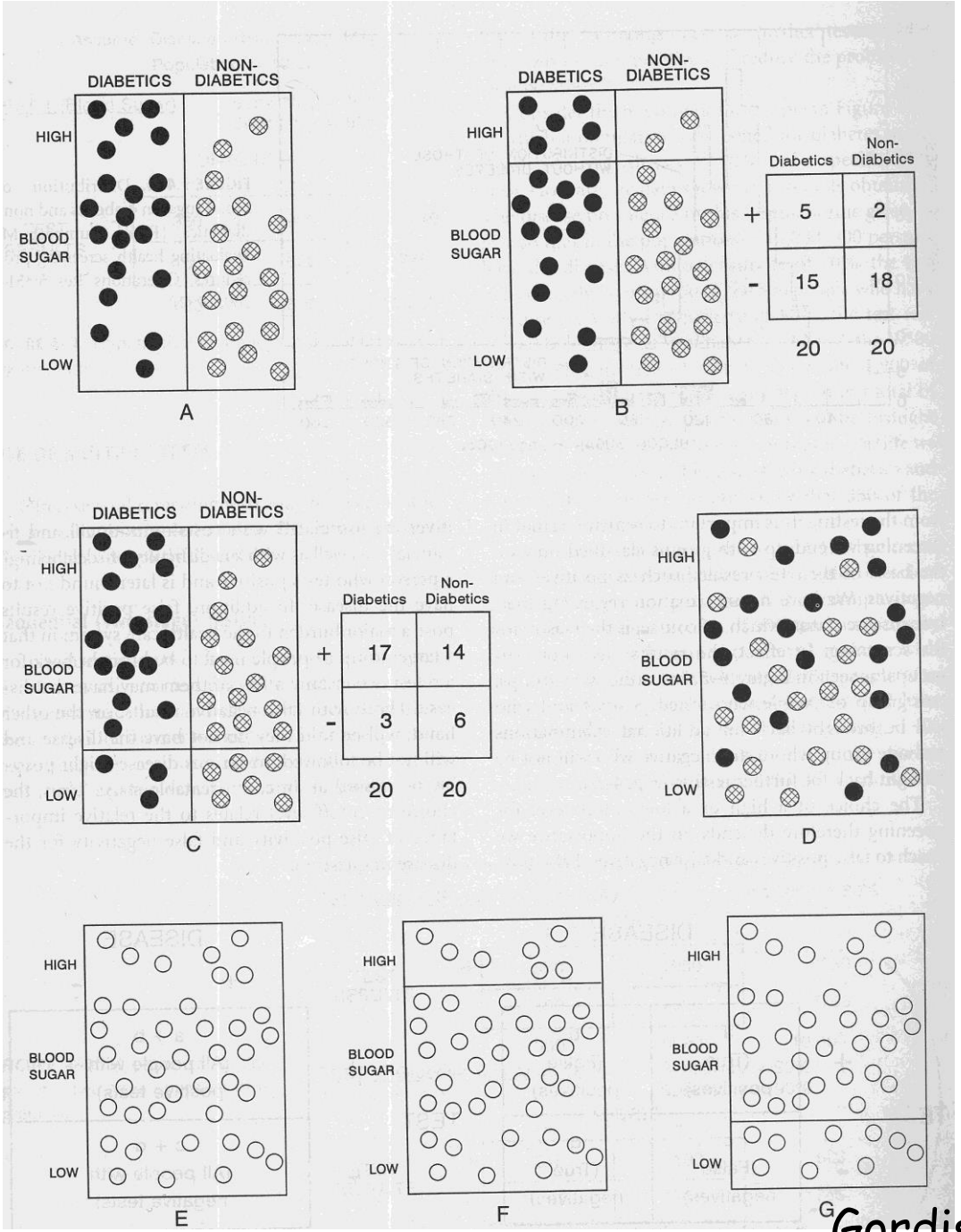


Gordis L. Epidemiology. Saunders, 1996.

Sistolik Kan Basıncı Dağılımı



Gordis L. Epidemiology. Saunders, 1996.



DM

Prevalans=%50

Gordis L. Epidemiology. Saunders, 1996

İdeal Tarama Testi

- %100 duyarlı
- %100 seçici olmalıdır.
- Ancak pratikte zor

- Genellikle bir testin seçiciliği yükseldikçe duyarlılığı, duyarlılığı yükseldikçe seçiciliği düşer.

- Duyarlılığı ↑ bir test
  - hastaların çoğunu belirler
  - ↑ negatif prediktif değer
  - **Duyarlı bir test negatifse hastalığı ekarte ettirir**
- Seçiciliği ↑ bir test
  - ↓ yalancı pozitif
  - ↑ pozitif prediktif değer
  - **Seçici bir test pozitifse hastalığı gösterir**

Sensitivitenin Önemli Olduğu, Yüksek Olması İstenen Durumlar

- Toplumda taranan hastalık erken teşhis edildiğinde yeterli tedavisi varsa veya ilerlemesi durdurulabiliyor, komplikasyonları önlenebiliyorsa,

- Yanlış olarak hasta tanısı konulanların (yalancı pozitifler) kesin tanıları için gerekli ileri tetkikleri sağlık kuruluşları için bir yük değilse, onların işlerim aksatmıyorsa ve tetkikler çok pahalı değilse,

- Toplumda sık sık tarama yapma olanakları yok veya çok sınırlı ise taramada kullanılacak tanı testinin sensitivitesinin yüksek olması istenir.

Spesifisitenin Önemli Olduğu, Yüksek Olması İstenen Durumlar

- Taranan hastalığın tedavi olanakları sınırlı veya tedavisi yoksa,
- Şüpheli vakaların ileri tetkikleri için uygun yöntemler yok ve ayrıntılı tetkikleri sağlık kuruluşları için büyük yük olacak, onların işlerini aksatacaksa veya bu testler çok pahalı ise,
- Yanlış olarak hasta tanısı konulması (yalancı pozitif) kişiyi ve çevresini sosyal yönden rahatsız, huzursuz edecekse (Örneğin sifilis, tüberküloz, lepra, AIDS gibi),
- Hastalığa erken tanı konulmasının, tedavi ve prognoz üzerinde olumlu bir etkisi yoksa,
- Tam ve tedavi izlemelerini yapacak yeterli sayıda/nitelikte personel ve kuruluşlar

ROC Eğrisi

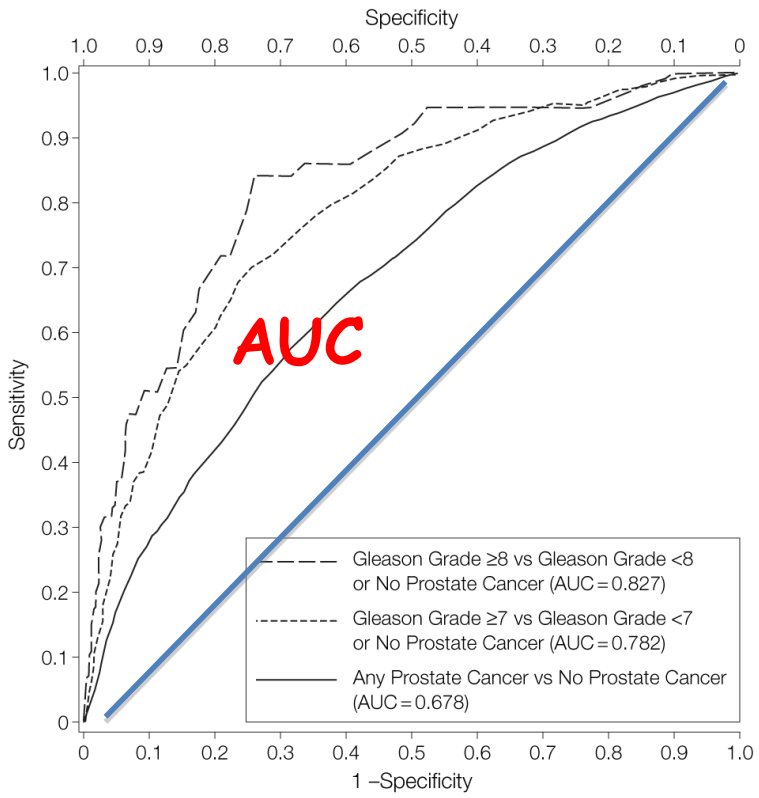
- *Receiver Operating Characteristics*
- Sensitivite vs 1-Spesifisite
- Eğri Altındaki Alan (*Area Under Curve -AUC*)

ROC Eğrisi

Prostat Ca PSA Düzeyleri

	PSA	
PSA	Sensitivite	1-Spesifisite
1,1	0,83	0,57
1,6	0,67	0,38
2,1	0,55	0,27
2,6	0,45	0,19
3,1	0,37	0,15
4,1	0,28	0,08
6,1	0,06	0,03
8,1	0,03	0,01
10,1	0,01	0,01

PSA ROC Eğrisi



Diagonal: sensitivite=1-spesifisite

Thompson et al. JAMA 2005;294:66-70.

Güvenilirlik (Tutarlılık)

Herhangi bir araştırmada veya kurumda (laboratuvarlar, saha çalışmaları, vb.) yapılan ölçümler, gözlemler, muayeneler aynı kişiler üzerinde aynı koşullarda, aynı gözlemciler tarafından **tekrar edildiklerinde aynı sonuçları elde etme boyutu** saptanır.

Böylece aynı zaman kesitinde yapılan gözlemlerin bulguları arasında fark olup olmadığı, varsa bu farkın boyutu, gözlem ve ölçümlerin güvenilirliği belirlenmiş olur

Güvenilirlik

- Tekrar edilebilirlik (repeatability),
- Uyum (agreement),
- Tutarlılık (consistency),
- Aynı sonucun elde edilebilirliği (reproducibility)

gibi çeşitli terimler de kullanılabilir.

Aynı teknikle yapılan çalışmalarda tutarlılığın %100 olması beklenir.

Çeşitli Değişkenlere İlişkin Gözlem ve Ölçümler Aynı Koşullarda Tekrar Edildiğinde Aynı Sonuçların Alınmasını Etkileyen Bazı Faktörler

- Ölçme için kullanılan araç-gereçlerin doğruluğu (tıbbi alet ve solüsyonlar vb. bozuk olmaması, anket, vb. veri toplama araçlarının iyi hazırlanması).

- Gözlenen/ölçülen değişkenin biyolojik olarak veya fizik koşullara göre varyasyon gösterip göstermemesi, aynı hastalığın farklı şekillerde klinik seyir göstermesi ve bilgi veren kişilerden kaynaklanan hatalar

- Gözlem ve ölçümleri yapan kişilerden (gözlemci) kaynaklanan varyasyon

Biyolojik Varyasyon

- Gözlem ve ölçümlerdeki tutarsızlığın bir kaynağı da bazı değişkenlerin biyolojik olarak varyasyon göstermesidir.
- Yemeklerle, dinlenme durumu ile, kişinin pozisyonu ile değişmesi (kan basıncı gibi)
- Gece-gündüz (diurnal varyasyon) arasında farklılık (Kortizol düzeyi gibi)
- Hastalık farklı klinik şekillerde seyredebilir (ağır ve tipik vakaların tanınması kolay, hafif ve atipik vakaların tanısı güç)
- Anketle bilgi toplanması gereken durumlarda ise anket formunun anlaşılamaması, özellikle yanlış cevaplanması gibi

Gözlemcilerden Kaynaklanan Hatalar

Gözlemciler-arası tutarlılık

- Aynı kişiler üzerinde bazı değişkenlere ilişkin gözlem ve ölçümler aynı koşullarda, ancak farklı gözlemciler tarafından yapıldığı zaman sonuçlar arasındaki benzerliğin boyutudur.
- Aynı hastayı ait bir EKG iki ayrı hekim (gözlemci)
- Gözlemciler (hekimler) aynı sonuçlara varmışsa bu değerlendirmede gözlemci varyasyonu yoktur, gözlemciler arası tutarlılık vardır.

Gözlemci-içi tutarlılık

- Aynı kişiler üzerinde, aynı koşullarda, aynı gözlemci tarafından tekrar edilen gözlem ve ölçümlerin sonuçları arasındaki uyumdur.
- Aynı patoloji preparatını farklı zamanlarda değerlendiren bir patolog (gözlemci) aynı sonuçlara varıyorsa kendi kendisiyle tutarlıdır, yani gözlemci-içi tutarlılık tamdır.

Kardiyolojiye Başvuran 150 Hastanın EKG Sonuçları

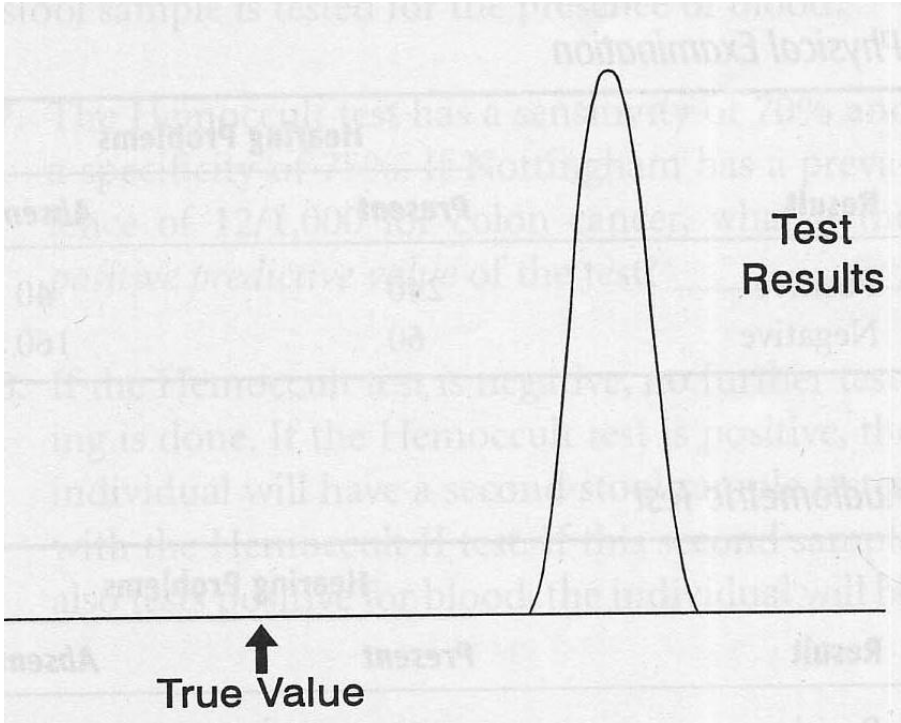
		2. Gözlemci					
Patoloji grupları		A1	A2	A3	A4	A5	Toplam
1. Gözlemci	A1	**20**	3	3	4	-	30
	A2	5	**30**	1	-	-	36
	A3	2	1	**25**	1	1	30
	A4	1	-	2	**35**	2	40
	A5	-	-	2	2	**10**	14
	Toplam	28	34	33	42	13	**150**

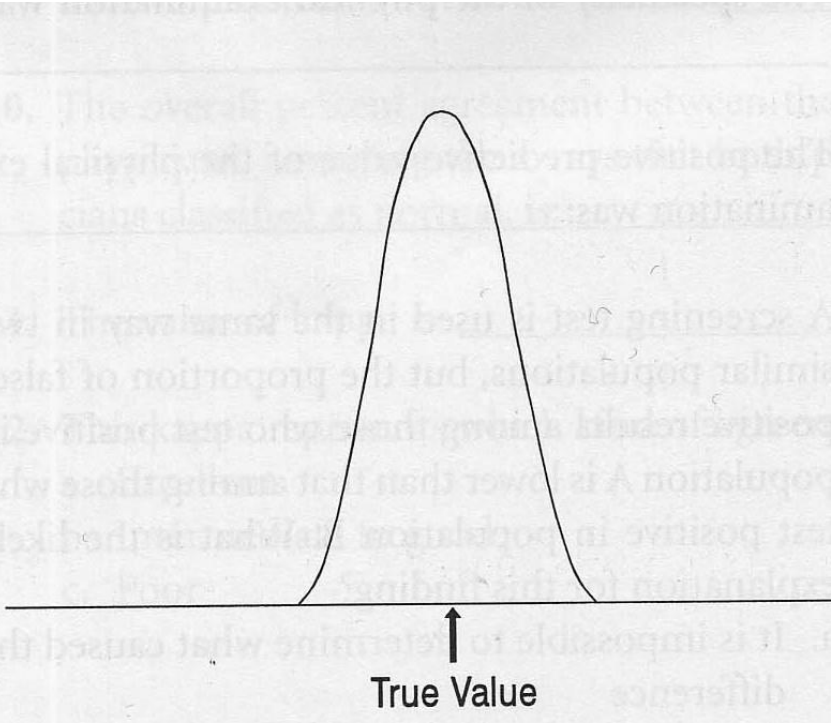
1 ve 2. gözlemci arasındaki tutarlılık $= \frac{A1+A2+A3+A4+A5}{\text{Toplam EKG}} \times 100 = \frac{20+30+25+35+10}{150} \times 100$

=% 80

Kappa İstatistiği

- Cohen, 1960
- Gözlenen uyum, şansa bağlı olarak beklenenden istatistiksel olarak iyi mi???
  - >0,80: Çok iyi
  - 0,60-0,80: İyi
  - 0,40-0,60: Orta
  - <0,40: Zayıf
- Prevalansdan etkilenir
  - prevalans ↓, κ ↓
- Güven aralıkları





Güvenilir ancak geçerli değil

Geçerli ve güvenilir

Gordis L. Epidemiology. Saunders, 1996.

Search for reporting guidelines

Use your browser's Back button to return to your search results

Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy

Reporting guideline provided for? (i.e. exactly what the authors state in the paper)	Studies of diagnostic accuracy
Full bibliographic reference	Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for Reporting of Diagnostic Accuracy. Clin Chem. 2003; 49(1):1-6. PMID: 12507953 BMJ. 2003; 326(7379):41-44. PMID: 12511463 Radiology. 2003; 226(1):24-28. PMID: 12511664 Ann Intern Med. 2003; 138(1):40-44. PMID: 12513043 Am J Clin Pathol. 2003; 119(1):18-22. PMID: 12520693

Key reporting guidelines

CONSORT	Full Record | Checklist | Flow Diagram
STROBE	Full Record | Checklist
PRISMA	Full Record | Checklist | Flow Diagram
STARD	Full Record | Checklist | Flow Diagram
COREQ	Full Record
ENTREQ	Full Record
SQUIRE	Full Record | Checklist
CARE	Full Record | Checklist
SAMPL	Full Record
SPIRIT	Full Record | Checklist
PRISMA-P	Full Record

Translations

Some reporting guidelines are also available in languages other than English. Find out more in our Translations section.

About the Library

STARD checklist for reporting of studies of diagnostic accuracy
(version January 2003)

Section and Topic	Item #		On page #
TITLE/ABSTRACT/ KEYWORDS	1	Identify the article as a study of diagnostic accuracy (recommend MeSH heading 'sensitivity and specificity').	
INTRODUCTION	2	State the research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups.	
METHODS			
Participants	3	The study population: The inclusion and exclusion criteria, setting and locations where data were collected.	
	4	Participant recruitment: Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?	
	5	Participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.	
	6	Data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?	
Test methods	7	The reference standard and its rationale.	
	8	Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.	
	9	Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.	
	10	The number, training and expertise of the persons executing and reading the index tests and the reference standard.	
	11	Whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.	
Statistical methods	12	Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals).	
	13	Methods for calculating test reproducibility, if done.	
RESULTS			
Participants	14	When study was performed, including beginning and end dates of recruitment.	
	15	Clinical and demographic characteristics of the study population (at least information on age, gender, spectrum of presenting symptoms).	
	16	The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or the reference standard; describe why participants failed to undergo either test (a flow diagram is strongly recommended).	
Test results	17	Time-interval between the index tests and the reference standard, and any treatment administered in between.	
	18	Distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition.	
	19	A cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard.	
	20	Any adverse events from performing the index tests or the reference standard.	
Estimates	21	Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).	
	22	How indeterminate results, missing data and outliers of the index tests were handled.	
	23	Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers, if done.	
	24	Estimates of test reproducibility, if done.	
DISCUSSION	25	Discuss the clinical applicability of the study findings.	

Okuma için Kaynaklar

1. Rothman KJ. Epidemiology. An Introduction. Oxford University Press , 2012.
2. Kramer MS. Clinical Epidemiology and Biostatistics. Springer-Verlag, 1988.
3. Gordis L. Epidemiology. Saunders, 1996.
4. Saracci R. Epidemiology. A Very Short Introduction. Oxford University Press, 2010.
5. Tezcan S. Epidemiyoloji: Tıbbi Araştırmaların Yöntem Bilimi. Hacettepe Halk Sağlığı Vakfı Yayını, 1992.