

TANI TESTİ PERFORMANSLARININ DEĞERLENDİRİLMESİ VE KIYASLANMASI*

Ahmet DİRİCAN

Background and Design.- Diagnostic tests are widely used in many areas of our modern technological society, but they are of particular importance in medicine, where early and accurate diagnosis can decrease the morbidity and mortality of disease.

As new diagnostic tests are introduced, their impact on patients health must be determined. It is often important to compare various diagnostic tests with each other for specific clinical conditions in order to determine which one is the best to use.

A perfect diagnostic test would be one with no false positive or false negative results and would be represented by a line that started at the origin and went up the y-axis to a sensitivity of 1, and then across to false positive rate of 0.

The methods of ROC curve, Kappa coefficient, Youden index, and Mc-Nemar test are used to present and/or compare the performances of one or more diagnostic tests.

ROC curves plot sensitivity against 1-specificity for the different cut-off points. Most popular information is the area under the curve, which has been called the best index of detectability.

Results and Conclusion.- In this study, we expressed mathematical basics and calculations of ROC curve analysis and the other methods within hypothetical examples.

* *Anahtar Kelimeler:* Tanı testleri, ROC Eğrileri, Kappa katsayısı, Mc-Nemar testi; *Key Words:* Diagnostic tests, ROC (Receiver Operating Characteristics) Curves, Kappa coefficient, Mc-Nemar test; *Alındığı Tarih:* 2000; *Doç. Dr. Ahmet Dirican:* İÜ Cerrahpaşa Tıp Fakültesi Biyoistatistik Bilim Dalı; *Yazışma Adresi (Address):* Dr. A. Dirican, İÜ Cerrahpaşa Tıp Fakültesi, Biyoistatistik Bilim Dalı, 34303, Cerrahpaşa, İstanbul.
<http://www.ctf.istanbul.edu.tr/dergi/online/2001v32/s1/011a4.htm>

Dirican A. Evaluation of the diagnostic test's performance and their comparisons. *Cerrahpaşa J Med* 2001; 32: 25-30.

B iyoistatistiksel değerlendirmelerde, bireylerin ölçümlenebilen sınırlı sayıda özellikleri esas alınarak değişkenler arası farklılıklar veya bağıntılar araştırılır. Oysa, biyolojik olaylar neredeyse sonsuz sayıdaki iç ve dış nedenlerin ortak bir sonucu olarak karmaşık bir yapı göstermektedir. Bilinmeyen etkenler, katı yargılar yerine "belli değerleri taşıyabilme olasılığı"nın belirlenmesine neden olur. Özellikle tıp alanında yeterli güven düzeyinde erken, risksiz ve ekonomik tanılar çok büyük öneme sahiptir.

Koruyucu tıp veya klinik çalışmalarda bireylerin sağlam olup olmadıklarını belirlemek amacı ile kullanılan laboratuvar tekniklerine, klinik gözlemlere veya özgün gereç ölçümlerine bağlı olarak karara erişilen değerlendirme kurgularına "Tanı Testleri" denir.

Araştırmanın amacı; çok daha yalın, risksiz ve ekonomik bir şekilde, olabildiğince doğru sonuç verebilecek tanı testlerine ulaşmaktır. Eğer tanı testi yeterince yetkin ve güçlü ise kullanılmalıdır.

YÖNTEM VE GEREÇLER

Tanı testi sonuçlarının irdelenebilmesi için, testin üzerinde uygulandığı olguların "gerçek tanı sonuçları" bilinmelidir. Gerçek tanıya çok daha kapsamlı bir dizi tanı yöntemi (veya altın standart) kullanılarak erişilir. Tanı testi değerlendirmesinin üzerinde yapılacağı örneklem, ya rastlantısal olarak ele alınmış hasta veya sağlam bireylerden oluşur, yada gerçek bir toplum taraması ile elde edilir. Gerçek taramadan elde edilen veriler aynı zamanda hastalığın sıklığını da (prevalans) yansıtır.

Tanı Testlerinde Temel Kavramlar

A - ÖZGÜN ORANLAR

Tanı testinin denetimi ya gerçek hasta ve sağlamlar üzerinde yada gerçeği yansıttığından kuşku bulunmayan bir ana tanı testi (Altın standart) sonucuna göre yapılır. Bir tanı testinin değerlendirilmesi, aşağıdaki tablo (2x2) tipi çerçevesinde; üstte gerçek sonuç, yanda ise tanı testi olmak üzere gözlemlenen adlar belirlidir ve bundan sonraki açıklamalarda bu adlar kullanılacaktır.

Gerçek Sonuç (Altın Standart)

		H+	H-	TOPLAM
Tanı	H+	A (GP)	B (YP)	A+B
Testi	H-	C (GN)	D (GN)	C+D
Sonucu	Toplam	A+C	B+D	A+B+C+D

Gerçek hastalara konan tanı açısından; "A" gözü gerçek tanıya uygun olarak tanı testinin de hasta dediği olgular yani gerçek pozitifler (GP) olarak adlandırılır. Buna karşılık "C" gözündeki olgular, gerçekte hasta olup, tanı testinin hatalı olarak sağlam dediği yani yanlış negatif (YN) olgulardır. Bu çerçevede testin geneli için;

1) DUYARLILIK (*Sensitivity*): Testin, gerçek hastalar içinden hastaları ayırma yeteneğidir. $DUYARLILIK = A / (A+C) = GP / (GP + YN)$

Gerçek sağlamlara konan tanıları açısından; "D" gözü gerçek tanı sonucuna uygun olarak testinde sağlam dediği gerçek negatif (GN) olgulardır. "B" gözü ise gerçekte sağlam oldukları halde testin hatalı olarak hasta dediği yanlış pozitif (YP) olgulardır.

2) ÖZGÜLLÜK (*Specificity*) testin, gerçek sağlamlar içinden sağlamları ayırma yeteneğidir. $ÖZGÜLLÜK = D / (D + B) = GN / (GN + YP)$

Duyarlılık ve özgüllük oranlarının tersi olan, testin hatalı yargıları da hesaplanabilir.

3) YANLIŞ NEGATİF ORANI: Gerçek hastalar içinden testin hatalı olarak sağlam dediği olgulardır. $YN = (1-DUYARLILIK) = C / (A + C) = YN / (YN + GP)$

Testin hatalı olarak hasta bulduğu olguların oranı da, "yanlış pozitif oranı" adını alır.

4) YANLIŞ POZİTİF ORANI: Gerçek sağlamlar içinden testin hatalı olarak hasta dediği olgulardır. $YP = (1-ÖZGÜLLÜK) = B / (B+D) = YP / (YP+GN)$

Bu oranların yanı sıra çok önem taşıyan bir başka oran ise "test sonucu olasılık oranı (LIKELIHOOD RATIO)"dır.

5) Pozitif test sonucu olasılık oranı (L+): Testin, hastalığa var dediği zaman doğruyu bildirmesinin yanılmasına oranıdır. (hastalık tanısı koymamanın doğruluk oranı)

$$L+ = DUYARLILIK / (1-ÖZGÜLLÜK) = A / (B+D) / B = GP / (YP+GN) / YP = GP / (GP+YN)$$

Bu oran ne kadar yüksek olursa, gerçek hastalar o derecede iyi ayrımlanmaktadır

6) Negatif test sonucu olasılık oranı (L-): Sağlam tanısının doğruluk oranıdır.

$$L- = (1-DUYARLILIK) / ÖZGÜLLÜK = C / (B+D) / D = YN / (YP+GN) / GN = YN / (GP+YN)$$

Bu oran ne kadar küçük olursa, gerçek sağlamlar o kadar iyi ayrımlanabilmektedir.

7) DOĞRULUK (*Accuracy*): Gerçekte testin hasta ve sağlam olarak toplam doğru tanı oranına "doğruluk" denir. $DOĞRULUK = (A+D) / (A+B+C+D) = (GP+GN) / (GP+YP+YN+GN)$

Diğer oranlardan farklı olarak doğruluk, aynı duyarlılık-özgüllük için bile hastalık sıklığına bağlı olarak değişebilir. Örnek;

		Gerçek			Gerçek		
		H+	H-		H+	H-	
I.	Tanı	H+	75	300	375		
	Testi	H-	5	420	425		
			80	720	800		
II.	Tanı	H+	840	42	882		
	Testi	H-	60	58	118		
			900	100	1000		

$$DUYARLILIK=75/80=0.93 \quad DUYARLILIK=840/900=0.93$$

$$ÖZGÜLLÜK=420/720=0.58 \quad ÖZGÜLLÜK=58/100=0.58$$

$$DOĞRULUK=(75+420)/800=0.6188$$

$$DOĞRULUK=(840+58)/1000=0.898$$

$$Prevalans (P(H+))=75/80=0.094$$

$$Prevalans (P(H+))=900/1000=0.90$$

B - KESTİRİM GÜCÜ

Gerçekte araştırmacıların yanıt aradığı en önemli soru; "Tanı testi sonucu pozitif olanın, gerçek bir hasta olma olasılığı nedir? (veya Tanı testi sonucu negatif olanın gerçekten sağlam olma olasılığı nedir?)" sorusudur. Bu kav-

ram "KESTİRİM DEĞERİ" adını alır ve Bayes kuramı çerçevesinde çözümlenir.

1) POZİTİF SONUCUN KESTİRİM DEĞERİ (PKD): Tanı testi hasta yargısı verdiğinde, gerçekten hasta olma olasılığıdır. $PKD = P(H+/T+) = A / (A+B) = GP / (GP+YP)$

2) NEGATİF SONUCUN KESTİRİM DEĞERİ (NKD): Tanı Testi sağlam dediğinde gerçekten sağlam olma olasılığıdır. $NKD = P(H-/T-) = D / (D+C) = GN / (GN+YN)$

Özellikle iç dağılımın farklı özellikler açısından karmaşık olduğu durumlarda çeşitli nedenler aynı sonucu vermiş olabilirler. Sonucun hangi olasılıkla, hangi nedenlerden kaynaklandığını bulmak için "BAYES TEOREMİ" kullanılmaktadır. Bu kural ters olasılık "veya" nedenler olasılığı adını da almaktadır. Konuyu somutlaştırmak için bir örnekle açıklamak gerekirse; Gerçekten A hastası olduğu bilinen ve gerçekten sağlam olduğu bilinen kişilerden oluşan bir gruba, A hastalığı tanısı koymada kullanılan bir test uygulanarak aşağıdaki dağılım tablosu elde edilmiştir.

		H+	H-	Toplam
Tanı Testi	H+	75	18	93
	H-	10	97	107
Sonucu		85	115	200

i) Tanı testinin hasta dediği bir kişinin gerçekten de hasta olma olasılığı nedir? (PKD)

$$P(H+/T+) = [P(H+) * P(T+/H+)] / [P(H+) * P(T+/H+) + P(H-) * P(T+/H-)]$$

$$P(H+/T+) = (0.42 * 0.88) / [(0.425 * 0.88) + (0.575 * 0.157)] = 0.374 / (0.374 + 0.09) = 0.806$$

ii) Tanı testinin sağlam dediği bir kişinin gerçekte hasta olma olasılığı nedir?

$$P(H-/T-) = [P(H-) * P(T-/H-)] / [P(H-) * P(T-/H-) + P(H+) * P(T-/H+)]$$

$$= [0.425 * 0.12] / [(0.425 * 0.12) + (0.575 * 0.84)] = 0.051 / [0.051 + 0.483] = 0.096$$

Bu oran ne kadar küçük olursa tanı testi, sağlamları o derecede iyi ayırmaktadır.

Yöntem

Uygulama olarak Cerrahpaşa Tıp Fakültesi KBB Anabilim Dalı'nda tanı tedavi ve izle-

meleri yapılan Akustik Nörinom şüpheli 58 olgunun BERA testi ve CT (Computerize Tomografi) sonuçları ve bazı özel hipotetik sayılar kullanılmıştır.

Tanı testlerinin yargılama yetenekleri yani performanslarının denetlenmesi gerektiğinde, sonuçların geçerliliğinden emin olmak amacı ile öncelikle tanı testinin gerçek sonuçlarla "FARKSIZ" olup olmadığının denetlenmesi gerekir. Bu denetim Mc Nemar testi ile gerçekleştirilir ve tanı testi sonuçlarının, gerçek tanı sonuçlarından farksız olduğu sonucuna erişilirse o zaman tanı testinin geçerliliği kabul edilir.

Mc Nemar testi gerçekte iki ayrı yöntemin aynı özelliği saptama gücünün birbiri içindeki dağılımını yargular. Test aslında değişimlenmiş bir χ^2 testidir ve aynı alt başlıkları taşıyan iki ana özellik (tanı testi) için 2x2 tablolarında ana özelliklerin birbirlerine ters konumlarında bulunan değerlerine göre yargıya varır. Uygulama örneğimizde;

		H	S	Toplam
BERA	H	10	7	17
	S	0	24	24
Toplam		10	31	41

Mc Nemar $\chi^2 = [(|B-C|-1)^2] / [B+C+1] = [(|7-0|-1)^2] / [7+0+1] = 4,5$ p < 0,05 sonuçlarına göre iki tanı testi arasında anlamlı fark bulunmuştur. Bu durumda; Mc Nemar testi BERA ve CT tanı sonuçlarının uyumsuz olduklarını saptamıştır.

Sıklıkla karşılaşılan diğer bir kıyaslama yöntemi de "YOUDEN İNDEKSİ (J)"dir.

$J = 1 - (YN + YP)$ dır. J değeri; a) -1 ile +1 arasında değişir. b) Sıfırdan küçük ise, testin tanısal gücü olmadığını göstergesi olacaktır. c) İki ayrı tanı testinin güçleri kıyaslanırken ise, J değeri daha büyük olan tanı testinin kabaca daha etkin olduğu söylenebilir. Örnek;

		Gerçek		
		H+	H-	
I. Tanı Testi	H+	62	23	85
	H-	3	47	50
		65	70	135
		Gerçek		
		H+	H-	
II. Tanı Testi	H+	122	10	132
	H-	38	95	133
		160	105	265

$$J = 1 - (0.046 + 0.329) = 0.625 \quad J = 1 - (0.2375 + 0.095) = 0.6675$$

Bu konumda ikinci test genel anlamda daha başarılıdır denebilir, ancak görüldüğü gibi ikinci testin duyarlılık ve özgüllüğü belli düzeyde iken (0.7625 ve 0.905), birinci testin özgüllüğü düşük (0.67) fakat duyarlılığı çok yüksek (0.954) bulunmuştur.

Başka bir deyişle, birinci test gerçek hastalar içinden hastaları daha başarılı olarak ayırmakta. Bu gibi durumlar klinik açıdan da önem taşıyor ise birinci test salt duyarlılığına bağlı konumlar çerçevesinde ikinci teste tercih edilebilir. BERA ve CT tanı testi sonuçlarını youden indeksi ile kıyaslarsak;

		Gerçek				Gerçek			
		H+	H-			H+	H-		
BERA	H+	10	7	17	CT	H+	11	0	11
	H-	0	24	24		H-	0	47	47
Tanı					Testi				
Testi		10	31	41	Testi		11	47	58

$$J=1-(0.00+0.226)=0.774$$

$$J=1-(0.00+0.00)=1.00$$

Bu durumda CT testi genel anlamda daha başarılıdır. ancak görüldüğü gibi BERA testinin de duyarlılığı "1"dir, ve özgüllüğü ise 0.774'dür. Kabaca CT testi iyi testtir denmesinin yanısıra BERA testi de bazı özelliklerinden dolayı (maliyet düşüklüğü, vs) çok rahatlıkla kullanılmaktadır. Zaten youden indeksi de direkt olarak testleri kıyaslamak değil testlere ait bir oransal değer sunmaktadır.

Tanı uyumlarını denetleyen diğer bir yöntem ise "KAPPA KATSAYISI"dir. Bu denetleme tipinde şansa dayalı olarak beklenen uyum ile gözlenen uyum değerleri arasında bir bağıntı irdelemesi yapılmaktadır.

		CT		
		H	S	Toplam
BERA	H	10	7	17
	S	0	24	24
	Toplam	10	31	41

$$U_G = \sum_{i=1}^k n_{ii}, P_G = U_G / N, P_B = \sum_{i=1}^k \frac{T_{1i} * T_{2i}}{N^2}$$

Formülleri ile, Uyumlu ögelerin toplamı,

$U_G = 24 + 10 = 34$ ve gözlenen uyum oranı $P_G = 34/41 = 0.8292$ olarak bulunur. Beklenen uyum oranı da

$$P_B = \sum_{i=1}^k \frac{T_{1i} * T_{2i}}{N^2} = \frac{(10*17)}{41^2} + \frac{(31*24)}{41^2} = 0.5437$$

Elde edilen bu iki orana dayalı olarak "kappa değerlendirme katsayısı" hesaplanabilir.

$$\kappa = (P_G - P_B) / (1 - P_B) = (0.8292 - 0.5437) / (1 - 0.5437) = 0.6256$$

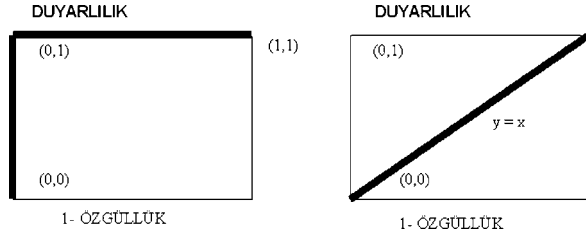
Bu katsayı iki testin orta güçte bir uyuma sahip oldukları yargısına varmaktadır.

Tanı Testi performanslarının değerlendirilmesi ve kıyaslanması için en yaygın kullanıma sahip olan yöntem ROC eğrisidir. Klinik çalışmalarda sürekli sayıların kullanıldığı ölçümlerde olguları ayırma (hasta/sağlam), çözümlenmeyi karmaşık hale getirir ve hata olasılığını yükseltir. Sonuçta, klinik şartlara bağlı olarak tanı testinin optimum etki noktası değişmektedir.

Bu çerçevede seçilen farklı eşik değerleri için bulunan farklı duyarlılık-özgüllük karakterlerine bağlı olarak ara seçenekler belirlenerek, ROC eğrileri (*Receiver operating characteristic curves*) oluşturulmuştur.

ROC eğrisi yöntemi; 1) Testin ayırt etme gücünün belirlenmesine, 2) Çeşitli testlerin etkinliklerinin kıyaslanmasına, 3) Uygun pozitiflik eşliğinin belirlenmesine, 4) Laboratuvar sonuçlarının kalitesinin izlenmesine, 5) Uygulayıcının gelişiminin izlenmesine ve 6) Farklı uygulayıcıların etkinliklerinin kıyaslanmasına olanak sağlar.

ROC eğrisinin oluşturulacağı koordinat sisteminin ordinatında tanı testinin gerçek pozitif değeri (duyarlılık), apsisinde ise yanlış pozitif değeri (1-özgüllük) yer alır. Tanı testi ne kadar iyi ise eğri o kadar yukarıya (yüksek duyarlılık bölgesi) ve sola (düşük yanlış pozitif oranı bölgesi) doğru kayar. Yanlış değerlere sahip olmayan ideal bir testte ROC eğrisi (0,0)-(0,1)-(1,1) noktalarını birleştirmektedir. Buna karşın ROC çizimi $y=x$ fonksiyonuna yaklaştıkça başarısız bir test ortaya çıkar. Çünkü bu testte yanlış değerlerin oranı en yükselmektedir. Bu fonksiyonun altındaki ROC eğrisine sahip test başarısızdır.



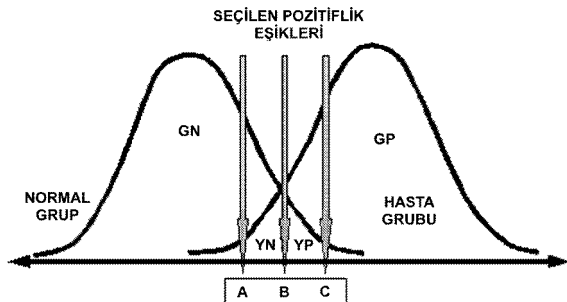
Şekil 1. İdeal ve kötü performans göstergesi olan ROC eğrileri

Böylece, bir tanı testi için ROC altında kalan alan etkinlik düzeyine bağlı olarak 0.50 ile 1.00 arasında değerler alabilecektir. Bu alan ne kadar büyükse, tanı testi o denli ayırım yeteneğine sahip olacaktır. Bu alan 0.975 ve daha üzerinde ise mükemmel sayılmaktadır.

ROC çözümlemesi kantitatif verilerin elde edildiği testlerin tanımlanabilmesi için matematiksel yaklaşımları belirleyerek farklı klinik durumlarda uygulanan tanı testlerini bütünüyle değerlendiren bir yöntemdir. Başarılı tanı testinin performansının tanımlayarak yeni bir tanı testinin gelişimini de açıklar.

Uygulamada karışık (hasta/sağlam) olgular ele alınarak, tanıları belirlenir ve iki ayrı grup meydana getirirler. Teorik olarak çalışma grubundaki sağlam ve hasta olguların normal dağılıma uygun olduğu kabul edilirse, ordinatta frekans, apsiste ise popülasyon standart sapmaları olduğu halde gruplara ait normal dağılım eğrileri çizilerek, eşik seçimi ve ROC eğrisinin fonksiyonunu daha basit bir şekilde açıklayabiliriz.

Şekil 2'deki grafik üzerinde görülen üç dikey çizgi seçilmesi olası üç farklı eşik düzeyini göstermektedir. Normal dağılım gösteren bu gruplarda, özellikle yanlış değerler olarak da adlandırdığımız, birbirlerine doğru bir girişim



Şekil 2. Normal dağılım eğrileri, girişim bölgesi ve pozitiflik eşik seçenekleri

bölgesi oluştururlar. İşte bu girişim bölgesinin en ideal yerinden bir ayırım sınırı seçerek maksimum doğrulukta bir sonuç elde edebiliriz.

Mümkün olabilecek tüm eşik düzeyleri için hesaplanan duyarlılık ve 1-özgüllük çakışım noktaları yani, $DUYARLILIK = f(1-ÖZGÜLLÜK)$ fonksiyonu "ROC EĞRİSİNİ" meydana getirir.

Testin performansı ve tanı sonucunun doğruluğu, testinin özgün oranları ve hastalığın prevalansına bağlıdır. Tarama konumlarında özellikle sağlamların belirlenmesi söz konusu olduğunda testin "NKD" önem kazanır. NKD'nin büyümesi için yanlış negatiflerin oranı azalmalı yani testin duyarlılığı büyümelidir. Buna karşılık tanıda, hastalığın varlığının doğrulanması gerekir, dolayısıyla PKD önem kazanır. Bu yanlış pozitiflerin oranı azaltılarak sağlanabilir.

Uygun bir tedavisi olan ve hasta olmayanlara (YP) boşuna uygulandığında ağır sonuçlar getirmeyen hastalık tanısına yönelik testlerde DUYARLILIK YÜKSEK tutulmalıdır. Buna karşın, daha az ağır sonuçlara sahip hastalıklarda hele yanlış pozitiflerin boşuna tedavisi ağır yan etkilere sahip ise ÖZGÜLLÜK YÜKSEK tutulmalıdır.

Pozitiflik eşiklerinin düşük olduğu durumlarda, testin duyarlılığı çok yüksek olacak, ancak çok sayıda yanlış pozitif elde edilecektir. Eşik yükseldikçe özgüllük yükselecek yanlış pozitifler azalacak, buna karşılık yanlış negatifler artacak ve böylece tedaviye alınmamış hastalar ortaya çıkacaktır.

Tüm teorik yaklaşımlar, doğru tanının niceliksel ifadesini çok popüler bir bulgu olan eğrisi altındaki alana indirgemeyi amaçlamaktadır. Bu alan bağımsız örneklemeler için non-parametrik *Wilcoxon Mann-Whitney* uyarlamasıyla hesaplanabilir.

ROC eğrisi altındaki alan, tanı testinin uygulandığı (hasta ve sağlam) iki veri dizisi için matematiksel olarak teorik bir z değeri ve bu z'nin standart normal değerleri tablosuna göre dönüşümü yapılarak elde edilmektedir.

Eğri altı alanının varyansı da aşağıdaki gibi hesaplanabilir.

$$SD_A^2 = \frac{A*(A-1) + (n_H - 1)*(Q_1 - A^2) + (n_S - 1)*(Q_2 - A^2)}{n_H * n_S}$$

Burada; A : Hesaplanan ROC ALANI n_H : Hasta denek sayısı, n_S : Sağlam denek sayısı $Q_1 = A / (2 - A)$ $Q_2 = 2A^2 / (1 + A)$ Bağımsız iki ROC alanı arasındaki farklılığın önemi aşağıdaki Z eşitliğine göre test edilebilir.

$$Z = \frac{A_1 - A_2}{\sqrt{SD_1^2 + SD_2^2}}$$

İncelenen olgular, iki ayrı grup olarak iki ayrı teknikte elde edilen veri dizileri olmalıdır ve çalışılan her bir vaka üzerinde karşılaştırılan her iki tekniğinde uygulanması durumunda eşli diziler elde edilmiş olur. Eşli dizilerde ROC eğrisi altındaki alanların

kıyaslanması aşağıdaki Z eşitliği kullanılarak yapılabilir.

$$Z = \frac{A_1 - A_2}{\sqrt{SD_1^2 + SD_2^2 - 2 * r_s * SD_1 * SD_2}}$$

Burada; r_s : üzerinde iki tanı testinin de uygulandığı denekler için hesaplanan Spearman korelasyon katsayısıdır. Alanlar arası korelasyon katsayısının yüksek olduğu örneklerde denek sayısı daha düşük alınabilmektedir.

BULGULAR VE TARTIŞMA

Uygulama örneğimizde BERA yönteminin duyarlılığı "1" özgüllüğü "0.77" ve doğruluğu "0.83", CT yönteminin ise duyarlılığı, özgüllüğü ve doğruluğu "1" olarak bulunmuştur. Yine eğri altı alanlar ve standart hata ölçütleri de sırasıyla 0.7733 ± 0.0624 ve 1.00 ± 0.00 'dır. Bu sonuçlar CT'nin altın standart ile aynı düzeyde ve BERA'nın ise anlamlı olarak farklı olduğunu ($p < 0.001$) göstermektedir. BERA testi bir ön test olarak kullanılmasına karşın sonuç yargılama için başarısız bulunmuştur.

Test uyumlarını araştırmak için "Mc Nemar testi" ve "Kappa Katsayısı" yöntemleri kullanılabilir. Mc Nemar yöntemlerin uyumsuz oldukları ($p < 0.05$) sonucuna varırken, Kappa denetlemesinde elde edilen katsayısı ($\kappa = 0.6259$) orta güçte bir uyum olduğunu vurgulamaktadır.

ÖZET

Tanı testleri modern teknolojinin kullanıldığı bir çok alanda yaygın olarak kullanılmaktadır. Erken ve doğru tanı ile önemli hastalıkların teşhis ve müdahalesine olanak sağladığı için tıpta çok önemli bir yeri vardır. Önerilen bir tanı testinin verdiği sonuçları kapsamlı ve güvenilir bir biçimde irdeleyebilmek için öncelikle tanı testinin gerçek etkinlik düzeyinin denetlenmesi gerekmektedir.

Temel tanı testi kurgusunda, testin üzerinde uygulandığı olguların hasta ve sağlam olarak "gerçek tanı sonuçları" da elde edilerek tanı testinin verdiği sonuçlar bu gerçek tanı sonuçlarına göre değerlendirilir.

Yargılamalardaki amaç, en az riskle ve ekonomik koşullara uygun, yüksek doğrulukta sonuç verebilecek yüksek performanslı tanı testlerine ulaşmaktır. Eğer tanı testi yeterince yetkin ve güçlü ise kullanılmalıdır.

Uygulama olarak hipotetik veriler üzerinde örnek olabilecek çözümlenmeleri ve çeşitli yöntemlerinin kıyaslamalardaki avantaj ve dezavantajları tartışılmaktadır.

KAYNAKLAR

1. Metz CE. Basic principles of ROC analysis. Stat Nucl Med 1978; 8: 283-298.
2. Hanley JA, McNeil BJ. The meaning and use of the area under a ROC curve. Radiology 1982; 143: 29-36.
3. McClish DK. Comparing the areas under more than two independent ROC curves. Med Decis Making 1987; 7: 149-155.
4. McNeil BJ. Guidelines for evaluating nev tests Nuclear medicine: Clin App 1988; 665-679.
5. McClish DK. Analysing a proportion of the ROC curve. Med Decis Making 1989; 9:190-195.
6. Somoza E et al. Evulation and optimization of diagnostic tests using ROC analysis and information theory. Int J Biomed Compute 1989; 24: 153-189.
7. Bergus GR. When is a test positive? The use of decision analysis to optimize test interpretation. Fam Med 1993; 25: 656-660.
8. Catalona WJ, Hudson MA et al. Selection of optimal prostate specific antigen cutoffs for early detection of prostate cancer: receiver operating characteristic curves. J Urol 1994; 152: 2037-2042.
9. Şenocak M. Biyoistatistik. İstanbul, İÜ Yayınları No. 4053-214, 1998; 237-245.