ARAŞTIRMA / RESEARCH REPORT
İnönü Üniversitesi Tıp Fakültesi Dergisi 2012;19(3):153-8.

İki Sonuçlu Tanı Testlerinde İki Hekim Arasındaki Uyum İstatistiklerinin Prevalanstan Etkilenme Durumları+

E. Arzu Kanik[1], Semra Erdoğan[1], Gülhan Orekici Temel[1]

[1] Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin

Başvuru Tarihi: 09.02.2012
Kabul Tarihi: 19.03.2012

DOI: 10.7247/jiumf.19.3.5

Amaç: İki kategorisi olan bir tanı testinin iki hekimin (değerlendiricinin) uyumunu (tutarlılığını) test edebilmek için literatürde pek çok uyum istatistikleri önerilmektedir. Literatürde çok yaygın olarak kullanılan Scott'ın π istatistiği ve Cohen'in kappa istatistiğidir. Bu istatistiklere alternatif olarak Gwet tarafından AC1 istatistiği geliştirilmiştir. Bu çalışmanın amacı, iki değerlendirici arasındaki uyumu test etmekte kullanılan uyum istatistiklerinin, prevalanstan etkilenip etkilenmediklerini ortaya koymaktır.
Gereç ve Yöntemler: Tüm uyum istatistiklerinin marjinal homojenliğini test edebilmek adına prevalansa göre kısmi türevleri alınarak formüle edilip hesaplanmıştır.
Bulgular: Prevalans değerinin 0.50'e eşit olduğu durumda π, kappa, G-index ve AC1 uyum istatistiklerinin benzer sonuçlar verdiği hatta G-index'in tüm prevalans değerleri için sabit bir değere eşit olduğu belirlenmiştir. Bunun yanı sıra prevalans değeri yüksek (1'e eşit) veya düşük (0'a eşit) olduğunda π ve kappa istatistiklerinin 0 değerini aldığı ve bu değerlerin değerlendiriciler arasındaki uyumu doğru bir şekilde yansıtmadığı gözlenmiştir.
Sonuç: Bu çalışmada, 2X2 deneme düzenlerinde, değerlendiriciler arasındaki uyum istatistikleri araştırılmış ve diğer uyum istatistiklerine nazaran G-index ve AC1 istatistiğinin duyarlılık, seçicilik ve prevelans değerinden etkilenmediği ve daha iyi bir performans gösterdiği sonucuna varılmıştır.
Anahtar Kelimeler: Cohen Kappa İstatistiği; AC1 İstatistiği; Prevalans; Değerlendiriciler Arası Uyum.

Agreement Statistics Impacts of Prevalence Between the Two Clinicians in Binary Diagnostic Tests

Objective: It has been recommended a lot of statistics in the literature to test the agreement (consistent) of both clinicians (the raters) for a diagnostic test with two categories. Scott's π statistics and Cohen's kappa statistics used most frequently in the literature. As an alternative to this statistics, the AC1 statistics have been developed by Gwet. The purpose of this study, to determine agreement statistics, used to test the agreement between two raters, whether affected by the prevalence. Material and Methods: For testing marginal homogeneity of all agreement statistics, it's formulated and calculated by taking partial derivatives on behalf of prevalence. Results: It's determined that π, kappa, G-index and AC1 gave similar results with prevalence case is equal to 0.50. Moreover G-index is determined equal to a fixed value for all prevalence values. In addition, π and kappa statistics are equal to 0 when prevalence values case is high (equal to 1) or low (equal to 0) and its observed these values are not accurately reflect the agreement between raters. Conclusion: In this study, it's researched that the agreement statistics between raters in the 2x2 trial designs. And, it's concluded that G-index and AC1 statistics not affected sensitivity, specify and prevalence value than other agreement statistics and showed a better performance.
Key Words: Cohen's Kappa Statistics; AC1 Statistics; Prevalance; Agreement Between Raters.

İletişim Adresi:
Dr. Semra ERDOĞAN
Mersin Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, MERSİN
Tel: 0 324 3412818- 1089
e-mail: semraerdogann@gmail.com

+Ankara-Kızılcahamam 12-14 Eylül 2011 XIII. Ulusal Biyoistatistik Kongresi'nde sunulmuştur.

Giriş

A ve B gibi iki farklı değerlendiricinin, pozitif ve negatif olmak üzere sadece iki kategorisi olan bir ölçek üzerinde güvenilirlik çalışmaları, eğitim bilimlerinde olduğu kadar sağlık alanında da çok yaygın olarak kullanılmaktadır. Literatürde değerlendiriciler arasındaki uyumu değerlendirmek için kullanılan iki popüler uyum katsayısı vardır. Bunlardan biri 1955 yılında Scott tarafından geliştirilmiş olan π istatistiği ve 1960 yılında Cohen tarafından geliştirilen kappa istatistiğidir. Çok yaygın olarak kullanılmasa da Holley ve Guildford tarafından 1964 yılında uyum katsayısı olarak G-index uyum katsayısı geliştirilmiştir. Genellikle araştırmacılar π istatistiğini kullanmakta ve bu istatistiği de hesaplayabilmek için kappa istatistiğine başvurmaktadırlar. Bu uyum istatistiklerinin sonuçları birbirine benzer olmakta birlikte genellikle karıştırılmaktadır.[1] Ayrıca bu sözü edilen istatistiklere alternatif olarak, Gwet tarafından 2001 yılında yeni bir uyum katsayısı olan AC1 istatistiği geliştirilmiştir.

Bu çalışmanın amacı, kategori sayısı 2 olduğunda değerlendiriciler arasındaki uyum istatistiklerinin duyarlılık, seçicilik ve prevelans değerinden etkilenme durumlarını ortaya koymaktır.

Yöntem

A ve B gibi iki farklı değerlendiricinin, cevap değişkeni olarak pozitif (+) ve negatiflerden (-) oluşan iki kategorili bir duruma ait veriler, Tablo 1'deki gibi bir çapraz tablo ile gösterilmektedir. Tablo 1

Tablo 1. İki değerlendirici ve iki kategorili duruma ait çapraz tablo.

		A		
		+	-	Toplam
B	+	A	B	B1=A+B
	-	C	D	B2=C+D
Toplam		A1=A+C	A2=B+D	N

Tablo 1'de, köşegen üzerindeki değerler her iki değerlendiriciye ait ölçümlerin uyumlu olduğunu gösterirken, köşegen haricindeki ölçüm sonuçları uyumsuzluğu göstermektedir. Diğer bir deyişle, A hücresi her iki değerlendiricinin de "+" kategorisindeki sınıflandırmalarının sayısını, D hücresi ise her iki değerlendiricinin de "-" kategorisindeki sınıflandırmalarının sayısını göstermekte iken, B ve C hücreleri her iki değerlendiricinin de uyumsuz olduğu durumlardaki sınıflandırma sayısını göstermektedir. A1 ve A2, A değerlendiricisine, B1 ve B2, B değerlendiricisine ait + ve - kategori içerisindeki sınıflandırma sayısını vermektedir.[1]

Değerlendirici sayısı iki, kategori sayısı iki olduğunda değerlendiriciler arası uyum için literatürde bazı katsayılar önerilmektedir.

Scott'ın π istatistiği

1955 yılında Scott tarafından geliştirilmiştir. Aşağıdaki şekilde hesaplanmaktadır.

$$PI = \pi = \frac{P_a - P_e(\pi)}{1 - P_e(\pi)} \quad (1)$$

Burada, P_a toplam uyum olasılığını göstermekte olup,

$$P_a = \frac{(A + D)}{N} \quad (2)$$

şeklinde hesaplanmaktadır. $P_e(\pi)$ ise şansa bağlı uyum olasılığını ifade etmekte olup,

$$P_e(\pi) = \left(\frac{(A1 + B1)/2}{N}\right)^2 + \left(\frac{(A2 + B2)/2}{N}\right)^2 \quad (3)$$

şeklinde hesaplanmaktadır. [1-3]

Scott'ın π istatistiğinin varyansı

$$Var(\pi) = \frac{P_a(1 - P_a)}{(N - 1)(1 - P_e(\pi))^2} \quad (4)$$

şeklinde, %95 güven aralığı ise, $\pi \pm 2\sqrt{Var(\pi)}$ şeklinde hesaplanmaktadır.[4]

Cohen'in Kappa istatistiği

İki değerlendirici arasındaki uyumun değerlendirilmesi için 1960 yılında Cohen tarafından kappa istatistiği geliştirilmiş ve aşağıdaki gibi formüle edilmiştir.[2-5]

$$\kappa = \frac{P_a - P_e(\kappa)}{1 - P_e(\kappa)} \quad (5)$$

Toplam uyum olasılığı Denklem 2'deki gibi, şansa bağlı uyum olasılığı ise,

$$P_e(\kappa) = \left(\frac{A1}{N}\right)\left(\frac{B1}{N}\right) + \left(\frac{A2}{N}\right)\left(\frac{B2}{N}\right) \quad (6)$$

şeklinde hesaplanmaktadır.[2-5]

π ve kappa istatistiklerindeki şansa bağlı uyum olasılıkları birbirinden farklılık göstermektedir. π istatistiğindeki şansa bağlı uyum istatistik hesaplamasının aksine kappa istatistiğindeki şansa bağlı uyum istatistiği değerlendiricilerin sınıflandırma oranlarının çarpımları ile elde edilir. Bu iki istatistiğin şansa bağlı uyum olasılıklarının tahminlerindeki farklılığa rağmen bu iki uyum istatistikleri pratikte genellikle benzer sonuçlar vermektedirler.[2]

Kappa istatistiğinin varyansı

$$Var(\kappa) = \frac{P_a(1 - P_a)}{(N - 1)(1 - P_e(\kappa))^2} \quad (7)$$

şeklinde, %95 güven aralığı ise, $\kappa \pm 2\sqrt{Var(\kappa)}$ şeklinde formüle edilmektedir.[4,5]

G-index

1954 yılında Bennet tarafından S puan olarak ortaya atılmış, 1964 yılında Holley ve Guildford tarafından G-index olarak adlandırılmıştır. Aşağıdaki şekilde formüle edilmektedir.

$$\hat{\gamma}_G = \frac{(P_a - P_{e/G})}{1 - P_{e/G}} \quad (8)$$

Şans uyum olasılığı, π ve kappa istatistiklerinden farklı olarak kategori sayısının tersi olarak hesaplanmaktadır ($P_{e/G} = \frac{1}{q}$, q; kategori sayısını vermektedir).[2]

İki kategori olduğu durumda G- index;

$$\hat{\gamma}_G = \frac{\left(P_a - \frac{1}{2}\right)}{\left(1 - \frac{1}{2}\right)} = 2P_a - 1 \quad (9)$$

şeklinde hesaplanmaktadır. G-index'in varyans değeri, $Var(\hat{\gamma}_G) = \frac{4}{(N-1)} P_a(1-P_a)$ (10) şeklinde, %95 güven aralığı ise, $\hat{\gamma}_G - 2\sqrt{Var(\hat{\gamma}_G)} \pm \frac{0.5}{N}$ şeklinde formüle edilmektedir.[2-4]

Gwet'in AC1 istatistiği

Gwet'in AC1 istatistiği olarak adlandırılmakta olup 2001 yılında Gwet tarafından ortaya atılmıştır. Literatürde birinci dereceden uyum katsayısı (the first order agreement coefficient) olarak da adlandırılmaktadır ve aşağıdaki gibi hesaplanır.[2,3,6]

$$AC1 = \gamma = \frac{P_a - P_e(\gamma)}{1 - P_e(\gamma)} \quad (11)$$

Şansa bağlı uyum olasılığı ise, $P_e(\gamma) = 2P_1(1-P_1)$ (12) şeklinde hesaplanır. Formülde P_1 olasılığı yaklaşık olarak bir şansı verir ki bu şans + kategori içerisindeki bir deneğin A veya B değerlendirici tarafından sınıflandırılmasındaki şansı ifade etmektedir ve aşağıdaki gibi formüle edilmektedir. [6,7]

$$P_1 = \frac{(A1 + B1)/2}{N} \quad (13)$$

Gwet'in AC1 istatistiğinin varyansı,

$$Var(\gamma) = \frac{P_a(1-P_a)}{(N-1)(1-P_e(\gamma))^2} \quad (14)$$

şeklinde, %95 güven aralığı ise, $\gamma \pm 2\sqrt{Var(\gamma)}$ şeklinde formüle edilmektedir.[4]

Uygulama

Pozitif ve negatif olmak üzere iki olası kategori içerisinde 100 bireyi değerlendirmek için A ve B gibi iki farklı değerlendiricinin bulunduğu ve her iki değerlendiricinin de 85 bireyi doğru sınıflandırdığı iki farklı deneme düzeni oluşturalım (Tablo 2 ve 3). Bu iki deneme düzenine ait uyum istatistiklerinin sonuçları Tablo 4'de verilmektedir. [1-3]

Tablo 2. 1. Deneme düzeni.

		Değerlendirici 1 (A)		
		+	-	Toplam
Değerlendirici 2 (B)	+	40	9	49
	-	6	45	51
	Toplam	46	54	100

Tablo 3. 2. Deneme düzeni.

		Değerlendirici 1 (A)		
		+	-	Toplam
Değerlendirici 2 (B)	+	80	10	90
	-	5	5	10
	Toplam	85	15	100

Tablo 4. Her iki deneme düzenine ait uyum istatistikleri sonuçları.

	P_a	π	Kappa	G-index	AC1
1. Deneme düzeni	0.85	0.70	0.70	0.70	0.70
2. Deneme düzeni	0.85	0.31	0.32	0.70	0.81

Tablo 2 ve 3'e ait deneme düzenlerinde, 100 bireyden toplam 85'inin pozitif ya da negatif olarak değerlendirilmesinde tutarlı olduğu görülmektedir. Buna göre her iki değerlendiricinin vermiş oldukları kararın uyumlu olması yani değerlendiriciler arası uyum katsayılarının doğal olarak yüksek olması beklenmektedir. Ancak değerlendiriciler arası uyum katsayılarından π ve kappa istatistik değerleri incelendiğinde, 1. deneme düzeninde yüksek iken 2. deneme düzeninde bu katsayıların oldukça düşük olduğu gözlenmektedir.

Bunun yanı sıra, deneme düzenleri iki kategorili cevap değişkenine sahip olduğundan G-index değeri her iki deneme düzeni için de eşit bulunmuştur. AC1 istatistiği ile hesaplamalar yapıldığında 1. deneme düzeninde diğer uyum istatistikleri ile benzer sonuçlar verirken 2. deneme düzeninde elde edilen AC1 istatistiği sonucu 2. deneme düzenine ait tablodaki sonuçlarla çok daha tutarlı bir sonuç vermiştir.

Bu paradoks, pek çok araştırmacı tarafından π ve kappa istatistiklerinin populasyondaki prevelanstan etkilenmesi şeklinde açıklanmaktadır. Bazı bilim adamları da kappa istatistiğinin yeterli bir şekilde tanımlayabilmek için marjinal homojenliğin test edilmesi gerektiğini önermektedirler.[1] Öyleyse uyum istatistiklerinin, değerlendiricilerin duyarlılık, seçicilik ve prevelanstan etkilenme durumları ortaya konulmaya çalışılmalıdır. Bu amaçla, tüm uyum istatistiklerinin prevelansa göre kısmi türevleri alınarak formülasyonlar yeniden düzenlenmelidir.

Değerlendiriciler arası uyum istatistiği olan $\gamma = \frac{P_a - P_e}{1 - P_e}$ eşitlik, P_r'ye (prevelans) göre kısmi türev alındığında, $\frac{\partial \gamma}{\partial P_r} = -\frac{1 - P_a}{(1 - P_{e\gamma})^2} \frac{\partial P_{e\gamma}}{\partial P_r}$ (15) şeklinde elde edilir. Denklem 15'de yer alan toplam uyum olasılığı (P_a), P_{++} ile P_{--} toplamından elde edilir. Burada P_{++}, bireyi her iki değerlendiricinin de + kategori içerisinde sınıflandırma olasılığını, P_{--} ise bireyi her iki değerlendiricinin de - kategori içerisinde sınıflandırma olasılığını göstermekte olup aşağıdaki gibi hesaplanmaktadır.[2,3]

$$P_{++} = \alpha_A \alpha_B P_r + (1 - P_r)(1 - \beta_A)(1 - \beta_B) = \alpha_A \alpha_B P_r + (1 - P_r)(1 - \alpha_A)(1 - \alpha_B)$$

(16) ve $P_{--} = 1 - (P_{A+} + P_{B+} - P_{++})$ (17) şeklinde elde edilir. Denklem 16'da yer alan α_A ve α_B, A ve B değerlendiricilerinin duyarlılık (sensitivity) değerlerini, β_A ve β_B, A ve B değerlendiricilerinin seçicilik (specifity) değerlerini göstermektedir. Denklem 17'de yer alan P_{A+}, A değerlendiricisi için pozitif kategori içerisindeki deneğin sınıflandırılma olasılığını P_{B+}, B değerlendirici için pozitif kategori içerisindeki deneğin sınıflandırılma olasılığı göstermekte ve aşağıdaki gibi formüle edilmektedir.[2,3,7]

$$P_{A+} = P_r \alpha_A + (1 - P_r)(1 - \beta_A) \quad (18)$$

$$P_{B+} = P_r \alpha_B + (1 - P_r)(1 - \beta_B) \quad (19)$$

Buradan, toplam uyum olasılığı

$$P_a = (1 - \alpha_A)(1 - \alpha_B) + \alpha_A \alpha_B \quad (20)$$

şeklinde elde edilir. Toplam uyum olasılığı eşitliğinden yola çıkarak $\frac{\partial P_a}{\partial P_r} = 0$ olur. Bu formülasyonlar elde edilirken iki yaklaşım göz önünde bulundurulmaktadır:[2,3]

1. Her iki değerlendirici için duyarlılığın ve seçiciliğin; $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ olduğu durum
2. Doğru sınıflandırmanın bağımsız olduğu durum. Yani, eğer α_A, + kategori içerisindeki bir bireyin A ve B değerlendiricileri tarafından doğru bir şekilde sınıflandırılma olasılığını gösteriyorsa $\alpha_{AB} = \alpha_A \alpha_B$ (bağımsızlık koşuludur) olur.

Denklem 20'ye göre, iki değerlendirici arasındaki toplam uyum olasılığının prevelansa değil değerlendiricilerin duyarlılık ve seçicilik değerlerine bağlı olduğunu göstermektedir.[2]

Denklem 1'de verilen Scott'ın π istatistiği için şansa bağlı uyum olasılığı aşağıda verilmektedir.

$$P_{e/\pi} = \pi_+^2 + (1 - \pi_+)^2 \quad (21)$$

$$\pi_+ = \frac{(P_{A+} + P_{B+})}{2} = \lambda P_r + (1 - \lambda)(1 - P_r)$$

ve $\lambda = \frac{(\alpha_A + \alpha_B)}{2}$ (22)

Bunlar formülasyonda yerine konulduğu zaman,

$$\hat{\gamma}_\pi = \frac{(2\lambda - 1)^2 P_r (1 - P_r) - \frac{(\alpha_A - \alpha_B)^2}{4}}{(2\lambda - 1)^2 P_r (1 - P_r) + \lambda (1 - \lambda)} \quad (23)$$

haline dönüşür. Denklem 23'de P_r=0 veya P_r=1 için,

$\hat{\gamma}_\pi = -\frac{(\alpha_A - \alpha_B)^2}{4\lambda(1 - \lambda)}$ (24) olur.

P_r=0.50 için, $\hat{\gamma}_\pi = 2P_a - 1 = 1 - 4\lambda + 4\alpha_A \alpha_B$ (25)

olur. π istatistiğine ait denklemlerde bir paradoks gözlenmektedir. Değerlendiricilerin duyarlılık değerleri ne olursa olsun, π istatistiği negatif bir güvenilirlik değeri verir. Dolayısıyla eğer prevelans değeri düşük veya yüksek ise, π istatistiği değeri değerlendiriciler arasındaki uyumu doğru bir şekilde yansıtmayacaktır. π istatistiğinin 0-0.50 arasındaki prevelans değerleri için P_r'nin artan bir fonksiyonunu göstermektedir ve P_r>0.50 için bir azalma olur, P_r=0.50 olduğu durumda da maksimum değere ulaşır. Çünkü $0 \leq P_r \leq 1$ alır. $\hat{\gamma}_\pi$, P_r=0 ve P_r=1'de en küçük değerini alır.[2]

Denklem 5'de verilen Cohen'in Kappa istatistiği için şansa bağlı uyum olasılığı aşağıda verilmektedir.

$$P_{e/\kappa} = 1 - (P_{A+} + P_{B+}) + 2P_{A+}P_{B+} \quad (26)$$

Kappa istatistiğinin genel denklemi ise,

$$\hat{\gamma}_\kappa = \frac{(2\alpha_A - 1)(2\alpha_B - 1)P_r(1 - P_r)}{(2\alpha_A - 1)(2\alpha_B - 1)P_r(1 - P_r) + \frac{(1 - P_a)}{2}} \quad (27)$$

şeklinde hesaplanır.[2]

Kappa istatistiği, prevelansın 0.50'den çok küçük olduğu durumda, A ve B değerlendiricilerinin duyarlılık değerlerine bağlı olan prevelansın azalan veya artan bir fonksiyonu olabileceğini göstermektedir. Yani, eğer bir değerlendiricinin duyarlılığı 0.50'den küçük, diğer değerlendiricinin duyarlılık değeri 0.50'den büyükse, bu durumda kappa istatistiği, P_r'nin azalan bir fonksiyonu aksi takdirde artan bir fonksiyonudur.

Bu durum prevelansın 0.50'den büyük olduğu durumla benzerdir. Kappa istatistiği prevelans 0.50'ye eşit olduğu durumda maksimum ya da minimum değer alır. Eğer değerlendiricilerden birinin duyarlılık değeri 0.50 ise bu durumda prevelans ne olursa olsun kappa istatistiği değeri de sıfır olur.[2]

Eğer P_r=0 veya P_r=1 ise $\hat{\gamma}_\kappa = 0$ olur. Eğer P_r=0.50 ise $\hat{\gamma}_\kappa = 2P_a - 1 = 1 - 4\lambda + 4\alpha_A\alpha_B$ olur. Scott'ın π istatistiğine benzer olarak, yalnızca prevelans değeri 0.50'ye yakın olduğunda, kappa istatistiği makul değerleri verir gibi görünmektedir. 0'a ya da 1'e yakın olan prevelans değeri, değerlendiriciler arası uyumun herhangi bir miktarını ifade eden kappa istatistiğinin değerini oldukça azaltacaktır.[2]

G-index ($\hat{\gamma}_G$), sadece P_a'nın bir fonksiyonu olduğundan $\frac{\partial \gamma_G}{\partial P_r} = 0$ olur. Bu yaklaşım koşulu altında, G-index, değerlendiricilerin duyarlılığına bağlı olarak, ($2P_a$-1) gibi sabit bir değeri verir.[2]

AC1 istatistiğine ait şansa bağlı uyum olasılığı aşağıda verilmektedir.

$$P_e(\gamma) = 2\pi_+(1-\pi_+) \quad (28)$$

Denklem 28'de

$$\pi_+ = \frac{(P_{A+} + P_{B+})}{2} = \lambda P_r + (1-\lambda)(1-P_r)$$

ve

$$\lambda = \frac{(\alpha_A + \alpha_B)}{2}$$

şeklinde hesaplanmaktadır. Aşağıda verilen AC1 istatistiğine ait genel denklemde tüm formülasyonlar yerine konularak hesaplanmaktadır.[2]

$$\hat{\gamma}_1 = \frac{P_a - P_e(\gamma)}{1 - P_e(\gamma)} \quad (31)$$

Bulgular

Uyum istatistiklerinin kısmi türevleri alındıktan sonra elde edilen formülasyonlardan yararlanılarak, duyarlılık ve seçicilik değerlerinin birbirine eşit olduğu 11 farklı (0; 0.10; 0,20; 0,30; 0,40; 0,50; 0,60; 0,70; 0,80; 0,90 ve 1) durum, 13 farklı (0; 0,05; 0.10; 0,20; 0,30; 0,40; 0,50; 0,60; 0,70; 0,80; 0,90; 0,95 ve 1) prevalans değerleri için hesaplanmış ve tablolar halinde sunulmuştur. Duyarlılık ve seçicilik değerlerinin 0 ve 1 olduğu durumda, 1 hariç tüm olası prevalans değerleri için tüm uyum istatistikleri 1 değerini alırken, prevalansın 1'e eşit olduğu durumda π ve kappa istatistikleri hesaplanamamıştır. Duyarlılık ve seçicilik değerlerinin 0.5 olduğu durumda, tüm olası prevalans değerleri için tüm uyum istatistiklerinin 0 değerini aldığı gözlenmiştir. Duyarlılık ve seçicilik değerlerinin 0.10 ve 0.90 olduğu durumlarda, uyum istatistik değerleri Tablo 5'de, 0.20 ve 0.80 olduğu durumlarda Tablo 6'da, duyarlılık ve seçicilik değerlerinin 0.30 ve 0.70 olduğu durumlarda Tablo 7'de, 0.40 ve 0.60 olduğu durumlarda ise uyum istatistikleri değerleri Tablo 8'de verilmiştir.

Tablo 5. $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.10; $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.90

P_r	P_a	π	Kappa	G-index	AC1
0	0.82	0	0	0.64	0.78
0.05	0.82	0.25	0.25	0.64	0.76
0.10	0.82	0.39	0.39	0.64	0.74
0.20	0.82	0.53	0.53	0.64	0.71
0.30	0.82	0.60	0.60	0.64	0.67
0.40	0.82	0.63	0.63	0.64	0.65
0.50	0.82	0.64	0.64	0.64	0.64
0.60	0.82	0.63	0.63	0.64	0.65
0.70	0.82	0.60	0.60	0.64	0.67
0.80	0.82	0.53	0.53	0.64	0.71
0.90	0.82	0.39	0.39	0.64	0.74
0.95	0.82	0.25	0.25	0.64	0.76
1.00	0.82	0	0	0.64	0.78

Tablo 6. $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.20; $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.80

P_r	P_a	π	Kappa	G-index	AC1
0.00	0.68	0	0	0.36	0.53
0.05	0.68	0.10	0.10	0.36	0.50
0.10	0.68	0.17	0.17	0.36	0.48
0.20	0.68	0.26	0.26	0.36	0.43
0.30	0.68	0.32	0.32	0.36	0.39
0.40	0.68	0.35	0.35	0.36	0.37
0.50	0.68	0.36	0.36	0.36	0.36
0.60	0.68	0.35	0.35	0.36	0.37
0.70	0.68	0.32	0.32	0.36	0.39
0.80	0.68	0.26	0.26	0.36	0.43
0.90	0.68	0.17	0.17	0.36	0.48
0.95	0.68	0.10	0.10	0.36	0.50
1.00	0.68	0	0	0.36	0.53

Tablo 7. $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.30; $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.70

P_r	P_a	π	Kappa	G-index	AC1
0.00	0.58	0	0	0.16	0.28
0.05	0.58	0.03	0.03	0.16	0.26
0.10	0.58	0.06	0.06	0.16	0.24
0.20	0.58	0.11	0.11	0.16	0.21
0.30	0.58	0.14	0.14	0.16	0.18
0.40	0.58	0.15	0.15	0.16	0.17
0.50	0.58	0.16	0.16	0.16	0.16
0.60	0.58	0.15	0.15	0.16	0.17
0.70	0.58	0.14	0.14	0.16	0.18
0.80	0.58	0.11	0.11	0.16	0.21
0.90	0.58	0.06	0.06	0.16	0.24
0.95	0.58	0.03	0.03	0.16	0.26
1.00	0.58	0	0	0.16	0.28

Tablo 8. $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.40; $\alpha_A = \beta_A$ ve $\alpha_B = \beta_B$ =0.60

P_r	P_a	π	Kappa	G-index	AC1
0.00	0.52	0	0	0.04	0.08
0.05	0.52	0.01	0.01	0.04	0.07
0.10	0.52	0.01	0.01	0.04	0.06
0.20	0.52	0.03	0.03	0.04	0.05
0.30	0.52	0.03	0.03	0.04	0.05
0.40	0.52	0.04	0.04	0.04	0.04
0.50	0.52	0.04	0.04	0.04	0.04
0.60	0.52	0.04	0.04	0.04	0.04
0.70	0.52	0.03	0.03	0.04	0.05
0.80	0.52	0.03	0.03	0.04	0.05
0.90	0.52	0.01	0.01	0.04	0.06
0.95	0.52	0.01	0.01	0.04	0.07
1.00	0.52	0	0	0.04	0.08

Sonuç

Sonuç olarak π ve kappa uyum istatistiklerinin, prevalanstan etkilendikleri gözlenmektedir. Prevalans değerinin 0.50'ye eşit olduğu durumda tüm uyum istatistiklerine ait değerlerin eşit olduğu gözlenmektedir. G-index, toplam uyum olasılığının bir fonksiyonu olduğu için prevelanstan değil, duyarlılık ve seçicilik değerlerine bağlı olarak değişiklik göstermektedir. Dolayısıyla, duyarlılık ve seçiciliğe bağlı olarak sabit bir değere eşit olduğu, duyarlılık ve seçicilik değeri arttıkça G-index değerinin de arttığı gözlenmektedir. Ayrıca, G-index hariç diğer uyum istatistiklerinin prevelansa göre simetrik bir dağılım gösterdiğini söyleyebiliriz. Bunun yanı sıra prevelans değeri yüksek (1) veya düşük (0) olduğunda π ve kappa istatistiği değerinin sıfır değerini almaktadır ki bu değerler değerlendiriciler arasındaki uyumu doğru bir şekilde yansıtmamaktadır. 0 ile 1 arasındaki tüm prevelans değerleri için bu iki uyum istatistiğinin benzer sonuçlar verdiği görülmektedir (Tablo 5-8). Dolayısıyla alternatif olarak Gwet tarafından ortaya atılan AC1 istatistiği, π ve kappa uyum istatistiklerine nazaran duyarlılık, seçicilik ve prevelans değerinden etkilenmediği ve daha iyi bir performans gösterdiği söylenebilmektedir.

Bu çalışma sonunda yaygın olarak kullanılan π ve kappa istatistiklerinin prevelanstan etkilendikleri ve güvenilirlik çalışmalarında uyum katsayısı olarak artık bu istatistiklerin kullanımlarında daha dikkatli olunması gerektiği ortaya konulmuştur. Bunun yanı sıra G-index ve AC1 istatistikleri sonucunda elde edilen gözlenen uyum değerlerinin, toplam uyum olasılığına göre daha tutarlı sonuçlar verdiği gözlenmiş ve araştırmacılara uyum çalışmalarında bu iki uyum istatistiğini tercih etmeleri önerilmektedir.

Kaynaklar

1. Gwet K. Kappa Statistics is not Satisfactory for Assessing the Extent of Agreement Between Raters. Series: Statistical Methods Inter-Rater Reliability Assesment 2002; 1: 1-5.
2. Gwet KL. Computing Inter-Rater Reliability and Its Variance in the Presence of High Agreement. Br J Mathem Stat Psychol 2008; 61: 29-48. DOI: 10.1348/000711006X126600
3. Gwet K. Inter-Rater Reliability: Dependency on Trait Prevelance and Marginal Homogeneity. Series: Statistical Methods for Inter-Rater Reliability Assessment 2002; 2:1-9.
4. Gwet K. Handbook of Inter-Rater Reliability; 1st rev ed. USA: STATAXIS Publishing Company; 2001.
5. Kundel HT, Polansky M. Measurement of Observer Agreement. Radiology: Statistical Concepts Series 2003; 228: 303-8. DOI: 10.1148/radiol.2282011860
6. Haley DT, Thomas P, Petre M, Roeck AD. Using a New Inter-Rater Reliability Statistics. Technl Rep 2008; 15: 14-23.
7. Blood ES, Kevin F. Disagreement on Agreement: Two Alternative Agreement Coefficients. SAS Global Forum 2007.
8. Feuerman M, Miller AR. Relationships Between Statistical Measures of Agreement: Sensitivity, Specificity and Kappa. J Eval Clin Pract 2008; 14: 930-3. DOI: 10.1111/j.1365-2753.2008.00984.x
9. Feuerman M, Miller AR. The Kappa Statistics as a Function of Sensitivity and Specifity. IJEMEST 2005; 36 (5): 517-27. DOI: 10.1080/002073905000063967

Bu makaleye atıf yapmak için: Kanik EA, Erdoğan S, Temel GO. İki Sonuçlu Tanı Testlerinde iki Hekim Arasındaki Uyum İstatistiklerinin Prevalanstan Etkilenme Durumları. JIUMF 2012; 19(3): 153-8. DOI: 10.7247/jiumf.19.3.5