

Çok Testli Çok Değerlendiricili ROC Çalışmalarında Kullanılan Yöntemlerin Karşılaştırılması

Comparison of the Methods Used in Multi-Case Multi-Reader ROC Studies

Gülhan OREKİCİ TEMEL¹, Emine Arzu KANIK¹

¹Biyoistatistik AD, Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin

Özet

Amaç: Bu çalışmanın amacı hem çoklu değerlendiricinin hem de çoklu tanı testlerinin var olduğu diagnostik doğruluk çalışmalarındaki yöntemleri karşılaştırmak ve yöntemler hakkında öneride bulunmaktır. **Gereç ve Yöntem:** Tanı testleri tek bir değerlendirici ile yapıldığında değerlendiricinin kişisel kararlarına bağlı olarak subjektiflik içerebilir. Değerlendiricilerin performansı değerlendiricilerin deneyimleri, eğitimleri ve vakaların mevcut durumlarıyla ilgili ilgilidir. Bu nedenle diagnostik testlerin doğruluğu tartışılırken çoklu değerlendiricilerin raporları da modelde ele alınmalıdır. Çok testli çok değerlendiricili diagnostik doğruluk çalışmalarında kullanılabilecek beş farklı yöntem vardır. **Bulgular:** Hem çoklu değerlendirici hem de çoklu tanı testi sonuçlarının ele alındığı doğruluğu ortaya koyan modeller mevcuttur. Bu modeller karşılaştırılmış ve tablo halinde sunulmuştur.

Sonuç: Bu modellerden en yaygın kullanılanları DBM (Dorfman Berbaum Metz) ve OR (Obuchowski ve Rockette) metodudur.

Anahtar Kelimeler: DBM; OR; çok testli çok değerlendiricili eğri altında kalan alan

Abstract

Objective: The goals of this study are to compare the methods used in diagnostic accuracy studies in which both multi-case and multi-reader methods are employed, and to make useful suggestions for the proper use of these methods.

Material and Methods: When diagnosis tests are carried out by a single reader, they might have subjectivity based upon the reader's personal decisions. The readers' performance is related to their experiences, education and present states of the cases. Therefore, while the accuracy of the diagnostic tests is discussed, the reports of multi-readers should be taken into consideration within the scope of the model. There are five different methods used in multi-case and multi-reader tests of diagnostic accuracy studies.

Results: There are models dealing with the results of both multi-reader and multi diagnosis tests demonstrating the accuracy. These models were compared and presented in the tables.

Conclusion: The most common models are DBM (Dorfman Berbaum Metz) and OR (Obuchowski ve Rockette) methods.

Key Words: DBM; OR; multi reader multi case receiver operating curve

Mersin Univ Sağlık Bilim Derg, 2010;3(3):9-14

Geliş Tarihi : 29.03.2011

Kabul Tarihi : 15.09.2011

Yazışma Adresi : Gülhan OREKİCİ TEMEL, Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Mersin.

Tel: 0324 341 28 18/1089

Fax: 0324 341 24 00

E-posta: gulhan_orekici@hotmail.com

Giriş

ROC (Receiver Operating Curve) analizinde geleneksel metotlar ya tek bir değerlendiricinin bir grup vaka üzerindeki sonuçlarını ya da birden fazla değerlendiricinin yine bir grup vaka üzerindeki sonuçları üzerinden yapılır (1). Geleneksel ROC analizinde amaç değerlendiricilerin performansını değerlendirmek değildir. Fakat geleneksel ROC analizinden farklı olarak hem tanı testlerinin vaka gruplarını ayırmadaki performansını hem de değerlendiricilerin performansını Çok Testli Çok Değerlendiricili ROC analizleriyle çözümlenmek mümkündür. Kısaca ÇTÇD ROC analizi olarak ifade edilebilir. Bu yöntem literatürde MRMC (Multi Reader Multi Case) ROC analizi olarak ifade edilmektedir. Bu analiz verilerinin toplanması ile başlar ve bir çeşit faktöriyel deneme tasarımı oluşturur. Bu tür çalışmaların temel amacı sadece eğri altında kalan alan hesaplaması yapıp bu alanları karşılaştırmak değildir. Amaç, değerlendiricilerin ya da değerlendiricilerin

ortalamalarından yararlanılarak tanı testlerinin ve değerlendiricilerin ortak bir modelle diagnostik performansını ortaya çıkarmaktır (2).

Çok Testli Çok Değerlendiricili ROC Analizi Deneme Tasarımı

ÇTÇD ROC çalışmalarında en genel deneme tasarımı R değerlendiricinin C vakayı K tanı testini birlikte değerlendirdiği faktöriyel deneme tasarımıdır (2,3).

Kabul edelim ki, R değerlendirici, C vaka ve K tanı testi var olsun. Bu durumda her değerlendiricinin C*K tane sonucu ve bütün çalışmanın ise R*K*C tane sonucu vardır. ÇTÇD analizlerinin deneme düzenleri bir tür faktöriyel deneme düzenidir. Her değerlendirici bir ya da birden fazla test sonucunu değerlendirir. Tablo 1' de ÇTÇD ROC analizi için deneme tasarımı görülmektedir.

Tablo 1: ÇTÇD Doğruluk Analizleri için Faktöriyel Deneme Tasarımı

	TANI TESTLERİ											
	1			2			...	K				
	Değerlendirici			Değerlendirici			...	Değerlendirici				
Vaka	R ₁	R ₂	R _r	R ₁	R ₂	R _r				R ₁	R ₂	R _r
(Hasta/Sağlam)	Y ₁₁₁	Y ₁₂₁	Y _{1r1}	Y ₁₁₂	Y ₁₂₂	Y _{1r2}				Y _{11k}	Y _{12k}	Y _{1rk}
2	Y ₂₁₁	Y ₂₂₁	Y _{2r1}	Y ₂₁₂	Y ₂₂₂	Y _{2r2}				Y _{21k}	Y _{22k}	Y _{2rk}
.
.
C	Y _{c11}	Y _{c21}	Y _{cr1}	Y _{c12}	Y _{c22}	Y _{cr2}				Y _{c1k}	Y _{c2k}	Y _{crk}

Çok değerlendiricili ROC çalışmalarının analizinde birkaç farklı istatistik analiz yöntemi vardır. Metotlardan en sık kullanılanı Dorfman ve ark. (4) tarafından önerilen DBM (Dorfman Berbaum Metz) yöntemi ve Obuchowski ve Rockette (5) tarafından önerilen düzeltilmiş F metodudur. DBM sözde değerler için üç yönlü ANOVA yaparken, OR doğruluk tahminleri için ilişkilendirilmiş hataların iki yönlü ANOVA analizini yapar (4,5). ANOVA analizlerinde çalışma sonuçları çalışma örneğine benzer hastalara ya da çalışma örneğine benzer değerlendiricilere genellenebilir ise hem hastalar hem de değerlendiriciler için rasgele etkili model kullanılır. ÇTÇD çalışmaları değerlendiriciler örnekleme ve hasta örnekleme olarak iki örnekleme içerir. Eğer çalışma sonuçları çalışma örneğine benzer hastalara ya da çalışma örneğine benzer değerlendiricilere genellenebilir ise istatistik analizlerde hem hasta hem de değerlendiriciler için rasgele etkili model kullanılır.

Çok testli çok değerlendiricili çalışmalarında hastalar ve değerlendiriciler için rasgele etkili model kullanılır. Tanı testleri için ise etkilerin faktörler arasında değişmediğini varsayan sabit etkili model kullanılır.

Değerlendiricilerin doğruluğu ve güvenilirliği, hastaların özelliklerinden (yaş, cinsiyet, hastalık aşaması, ölüm oranı) etkilenme durumlarıyla, değerlendiricilerin deneyim ve eğitimleriyle ilgilidir (2,6). Bu değişkenler kovaryat olarak adlandırılır. Bazı istatistiksel metotlar kovaryatları içerebilir. Bu modeller hastalar arasındaki farklılığı ve değerlendiriciler arasındaki farklılığı ortaya çıkarır. Bu metotlar aşağıdaki bölümlerde açıklanmıştır.

DBM Metodu Teorisi

Çok değerlendiricili ROC çalışmalarında en sık kullanılan metot Dorfman-Berbaum-Metz (DBM) metodudur. Modelin testinde amaç "değerlendiricilerin doğrulukları arasında fark yoktur", "diagnostik testlerin doğruluğu arasında fark yoktur" ve "değerlendiricilerin ortalama doğruluğu bütün diagnostik testler için aynıdır" ifadeli yokluk hipotezlerini test etmektir. Yokluk hipotezinin testinde Quenouille-Tukey Jackknife sözde değerleri (pseudo values) ile ANOVA hesaplaması yapılır. Karışık etkili ANOVA, sözde değerler ile yapılır. Sözde değerler orijinal verinin bir çeşit transformasyonu gibi düşünülür.

t tane tanı şekli ($i=1,2,\dots,t$), r tane değerlendirici ($j=1,2,\dots,r$) ve c tane de vakanın bulunduğu bir veri seti ($k=1,2,\dots,c$) olsun. Eğri altında kalan alan hesaplaması yapılırken başlangıçta her değerlendirici için Maksimum Likelihood (ML) tahmin edicisi kullanılıp, her değerlendirici-tanı kombinasyonu için binormal ROC eğrisi elde edilir. Her değerlendirici-tanı kombinasyonu için ROC eğrisinden bir A_z indeksi elde edilir. ML tahmini j değerlendirici tarafından i tanı değişkeni için bütün vakalar üzerinden yapılır ve A_{ij} ile gösterilir.

Sonra 1. vaka silinir ve ML ile A_{ij} nin yeni bir tahmini hesaplanır. Sonra 1. vaka örnekleme yeniden dahil edilir ve 2. vaka silinir ve A_{ij} nin yeni bir tahmini hesaplanır ve böylece bu işlemler bir döngü şeklinde devam eder. Bu durum örneklemedeki her vaka bir kez silininceye kadar devam eder. Bu işlemle, eğri altında kalan alan c kez hesaplanır. Çünkü örnekleme üzerinde c tane vaka vardır.

Bu durumda $A_{ij(k)}$ k. vaka silindiğinde A_{ij} nin ML tahminini gösterir. J değerlendirici ve i tanı için k tane sözde değer A_{ij*k} ile gösterilir ve Denklem 1 ile hesaplanır.

$$A_{ij*k} = cA_{ij} - (c-1)A_{ij(k)} \quad (1)$$

Burada c ve c-1 faktörleri k. vaka silinerek oluşturulan alt kümede k. vakanın ağırlığını gösterir. Bu ilişki, bütün vakalar üzerinde A_{ij*k} k. vakanın ortalama etkisi olarak düşünülebilir. Silinen vakanın hasta ya da sağlam olup olmamasına bağlı olarak, negatif ya da pozitif vaka sayısından bir vaka düşülür.

DBM metodu (4) karma etkili doğrusal ANOVA modeli olarak kabul edilir. Bu metodun modeli Denklem 2' de verilmiştir.

$$Y_{ijk} = \mu + \tau_i + R_j + C_k + (\tau R)_{ij} + (\tau C)_{ik} + (RC)_{jk} + (\tau RC)_{ijk} + \varepsilon_{ijk} \quad (2)$$

Burada

μ : Populasyon ortalamasını,

τ_i : i tanı testinin sabit etkisini

R_j : j değerlendiricinin tesadüfi etkisini

C_k : k vakanın tesadüfi etkisini

ε_{ijk} : k vaka, j değerlendirici ve i. tanının hata miktarını gösterir.

Parantez ile verilen ifadeler interaksiyon etkisini gösterir.

İnteraksiyon terimlerinin (tanı testi×değerlendirici, tanı testi×vaka, değerlendirici×vaka ve tanı testi×vaka×değerlendirici) hepsi rasgele etkilidir.

Bir tanı testinin performansı için F istatistiği eşitlik 3' de verilmiştir.

$$F_{DBM} = \frac{MS(T)_{pseudo}}{MS(T \times R)_{pseudo} + MS(T \times C)_{pseudo} - MS(T \times R \times C)_{pseudo}} \quad (3)$$

MS(T): Tanı etkisinin kareler toplamını

MS(T×R): Tanı-değerlendirici interaksiyon etkisinin kareler toplamı

MS(T×C): Tanı-vaka interaksiyon etkisinin kareler toplamı

MS(T×R×C): Tanı-değerlendirici-vaka interaksiyon etkisinin kareler toplamını gösterebilir.

Karar eğer F değeri ; $(F_{df_1, df_{sun}})$; $(1-\alpha)F$ tablo değerinden büyükse tanı testlerinin diagnostik doğrulukları arasındaki farklılık anlamlıdır. F istatistiğine karşılık gelen kritik tablo değerlerinin hesaplaması eşitlik 4 ve eşitlik 5' de verilmiştir.

$$df_1 = (t-1) \quad (4)$$

$$df_{satt} = \frac{[MS(T \times R)_{pseudo} + MS(T \times C)_{pseudo} - MS(T \times R \times C)_{pseudo}]^2}{\frac{MS(T \times R)_{pseudo}^2}{(t-1)(r-1)} + \frac{MS(T \times C)_{pseudo}^2}{(t-1)(c-1)} + \frac{MS(T \times R \times C)_{pseudo}^2}{(t-1)(r-1)(c-1)}} \quad (5)$$

ÇTÇD ROC analizlerinde performans değeri eğri altında kalan alandır (4,7-11).

Obuchowski ve Rockette Metodunun (Düzeltilmiş F Metodu) Teorisi

Obuchowski ve Rockette tarafından OR metodu olarak kısaltılan ve düzeltilmiş F testi olarak da ifade edilen matematiksel bir model önerilmiştir. Modelde değerlendiriciler sabit etkili, tanı testleri ve vakalar rasgele etkili kullanılarak tahmin yapılır. Buradaki doğruluk değeri, DBM metodundaki gibi sözde değerler kullanılarak değil, gerçek değerler kullanılarak yapılır. DBM metodu aynı vakalardan çoklu değerlendiricilerin her birinin performansının çoklu tekrarlarına izin verir. Fakat düzeltilmiş F testi faktoriyel deneme tasarımı için yalnız bir replikasyonda hesaplanır ve Denklem 6' daki gibi yazılabilir.

$$\hat{\theta}_{ij:gercek} = \mu + \tau_i + R_j + (\tau R)_{ij} + \varepsilon_{ij} \quad (6)$$

Burada

i tanı testi ($i=1,2,\dots,t$) j değerlendirici ($j=1,2,\dots,r$) olmak üzere

$\hat{\theta}_{ij:gercek}$ j.değerlendirici, i. tanı test için eğri altında kalan alanın tahmin değerini,

μ : Populasyon ortalamasını,

τ_i : i tanının sabit etkisini

R_j : j değerlendiricinin rasgele etkisini

$(\tau R)_{ij}$: Değerlendirici-tanı testi interaksiyon teriminin rasgele etkisini

ε_{ij} : j değerlendirici ve i tanının hata terimini gösterir.

R_j ve $(\pi R)_{ij}$ ortalaması sıfır varyansları σ_R^2 ve $\sigma_{\pi R}^2$ olan ortak bağımsız normal dağılım gösterdikleri varsayılır. Eşitlik 2.6 daki ϵ_{ij} 'de ortalaması sıfır ve varyansı sabit (σ_ϵ^2) normal dağıldığı varsayılır.

Ayrıca ϵ_{ij} ; R_j ve $(\pi R)_{ij}$ da bağımsız olduğu varsayılır. Fakat her tanı için her değerlendiricinin aynı vakayı okuması sonucunda, ϵ_{ij} 'nin ortak bağımsız olduğu varsayımını kabul etmez. Bunu yerine testler ve değerlendiriciler arasındaki hataların kovaryansları oluşturulur ve Denklem 7'deki gibi bu kovaryanslar ifade edilir.

$$\text{Cov}(\epsilon_{ij}, \epsilon_{i'j'}) = \begin{bmatrix} \text{Cov1} & i \neq i' & j = j' \\ \text{Cov2} & i = i' & j = j' \\ \text{Cov3} & i \neq i' & j \neq j' \end{bmatrix} \quad (7)$$

Obuchowski ve Rockette (15,16) bu kovaryansları Denklem 8'deki gibi sıralamışlardır.

$$\text{Cov1} \geq \text{Cov2} \geq \text{Cov3} \geq 0 \quad (8)$$

Bu kovaryans değerlerinden; Denklem 9 farklı tanı testleri-aynı değerlendirici için oluşturulan Cov_1 'nin, Denklem 10 aynı tanı testi-aynı değerlendirici için oluşturulan Cov_2 'nin, Denklem 11'de farklı tanı testi-farklı değerlendirici için oluşturulan Cov_3 değerinin hesaplanmasını göstermektedir.

$$\text{Cov1} = \text{Cov}(\epsilon_{ij}, \epsilon_{ij}) = \text{Cov}(\hat{\theta}_{ij}, \hat{\theta}_{ij} / R, \pi R) \quad (9)$$

$$\text{Cov2} = \text{Cov}(\epsilon_{ij}, \epsilon_{i'j'}) = \text{Cov}(\hat{\theta}_{ij}, \hat{\theta}_{i'j'} / R, \pi R) \quad (10)$$

$$\text{Cov3} = \text{Cov}(\epsilon_{ij}, \epsilon_{i'j'}) = \text{Cov}(\hat{\theta}_{ij}, \hat{\theta}_{i'j'} / R, \pi R) \quad (11)$$

Buradaki kovaryans $\text{Cov}(\cdot / R, \pi R)$; şartlı değerlendirici ve tanı-değerlendirici etkisinin kovaryansıdır. Eğer hatalar bağımsız ise bir tanının etkisini test etmek için ($H_0 : \tau_1 = \tau_2 = \dots = \tau_t$) Denklem 12'de kullanılan F istatistiği kullanılır.

$$F = \frac{r \sum_{i=1}^t (\hat{\theta}_i - \hat{\theta}_{..})^2 / (t-1)}{\sum_{i=1}^t \sum_{j=1}^r (\hat{\theta}_{ij} - \hat{\theta}_i - \hat{\theta}_j + \hat{\theta}_{..})^2 / [(t-1)(r-1)]} = \frac{MS(T)_{\hat{\theta}_{ij}, \text{orjinal}}}{MS(T * R)_{\hat{\theta}_{ij}, \text{orjinal}}} \quad (12)$$

Fakat Denklem 12'deki F istatistiği Denklem 6'daki modelde hatalar korelasyonlu olduğu için geçersizdir. Bu nedenle Obuchowski ve Rockette (9) Denklem 6'daki modele bir düzeltme yaparak Denklem 13'deki F istatistiğini oluşturmuşlardır.

$$F = \frac{MS(T)_{\hat{\theta}_{ij}, \text{orjinal}}}{MS(T * R)_{\hat{\theta}_{ij}, \text{orjinal}} + r(\text{cov2} - \text{cov3})} \quad (13)$$

Uygulamada Cov2 ya da Cov3 bilinmez bu nedenle ya pilot bir çalışma verilerinden ya da önceki çalışmalardan tahminleri kullanılır. Bu nedenle F istatistiği Denklem 14 'deki gibi yazılır.

$$F = \frac{MS(T)_{\hat{\theta}_{ij}, \text{orjinal}}}{MS(T * R)_{\hat{\theta}_{ij}, \text{orjinal}} + r(\text{cöv2} - \text{cöv3})} \quad (14)$$

Karar eğer F değeri ($F_{df_1, df_2; (1-\alpha)}$) tablo değerinden büyükse tanı testlerinin diagnostik doğrulukları arasında farklılık anlamlıdır. F istatistiğine karşılık gelen kritik tablo değerlerinin hesaplaması Denklem 15 ve Denklem 16'da verilmiştir (7,9,12-14).

$$df_1 = (t-1) \quad (15)$$

$$df_2 = (t-1) * (r-1) \quad (16)$$

Çok Değişkenli Wilcoxon-Mann-Whitney Yöntemi

Song (13) çok değerlendiricili çalışmalarda nonparametrik bir yaklaşım önermiştir. Önerilen modelin herhangi bir matematiksel modeli yoktur ve model tamamen non-parametriktr. Model her vaka için bir güven skoru oluşturularak kurulu. Bu modelde de değerlendiriciler modele, sabit etkili olarak katılırlar. Bu metotta diagnostik doğruluk olarak eğri altında kalan alan kullanılır. Bu yöntem için herhangi bir bilgisayar yazılımı yoktur.

ROC Deneme Tasarımları İçin Hiyerarşik Sıralı Regresyon Metodu

Ishwaran ve Gatsonis (14) çoklu değerlendiricili ROC deneme tasarımlarında Bayesci hiyerarşik sıralı regresyon modelini (HROC) geliştirmişlerdir. Radyoloji verilerinde çok düzeyli kümelmiş veri yapısında kullanılmışlardır. HROC modelleri her vaka için bir gizli değişken oluşturur. Sıralı bağımlı değişkenin değerleri çok değişkenli gizli değişkenlerin aralıklarının bölünmesi ile belirlenir. Modelde değerlendiriciler modele rasgele ya da sabit etkili olarak girebilirler. Ayrıca modele kovaryat koymak mümkündür. HROC modellerinde diagnostik doğruluk ölçümü için eğri altında kalan alan kullanılamaz, sadece test sonuçları kullanılarak analiz yapılır (15-16).

Varyans Unsurlarının Bootstrap Tahmini

Beiden, Wagner ve Campbell (17) çok değerlendiricili ROC çalışmaları için DBM metodundaki jackknife yeniden örnekleme yöntemine alternatif olarak bir bootstrap yeniden örnekleme metodu önermişlerdir. Bu metoda göre iki farklı şekilde veriler yeniden örnekleme yapılabilmektedir. Birincisi sıralı veriyi vakalar ve değerlendiriciler üzerinde; ikincisi sıralı veriyi sadece vakalar üzerinde bootstrap yeniden örnekleme yöntemini kullanarak uygulama şeklindedir (17-19).

Modelde veriler bootstrap yeniden örnekleme metodu kullanılarak elde edilir ve varyans analizindeki varyasyon kaynakları bu yeni verilerle hesaplanır.

Oluşturulan yeni veriler yine eğri altında kalan alan verileridir. Bu veriler ile varyasyon kaynakları hesaplanırken, varyansı; vaka ve değerlendiricilerden kaynaklı, her değerlendirici*taşı testi kombinasyonundan kaynaklı ve değerlendirici*vaka kombinasyonlarından kaynaklı varyasyon kaynaklarının toplamı olarak varsayar. Ayrıca varyans tahminlerinde dağılım serbestliği vardır. Değerlendiriciler modele rasgele ya da

sabit etkili olarak girebilirler. Model kovaryat içermez. Model için herhangi bir bilgisayar yazılımı yoktur.

Çok testli çok değerlendiricili ROC analizlerinde kullanılan beş yöntem model, değerlendiricilerin modele etkisi, kullanılan modelin herhangi bir kovaryatta izin verip vermemesi, modelde kullanılan doğruluk ölçümü ve bir bilgisayar yazılımlarının olup olmaması yönünden Tablo 2’de karşılaştırılmıştır (20-21).

Tablo 2. ÇTÇD ROC Yöntemlerinin Karşılaştırılması (22-24)

Özellikler	DBM	OR	Çok değişkenli WMW	BWC	HROC
Model Kısmı	Sözde Değerler	Doğruluğun özet ölçümü	Her hasta için güven skoru	Doğruluğun özet ölçümü	Her hasta için gizli değişken
Etki	Değerlendirici-ler modelde rasgele ya da sabit etkili olabilir.	Değerlendiriciler modelde rasgele ya da sabit etkili olabilir.	Değerlendiriciler modelde sabit etkilidir.	Değerlendiriciler modelde rasgele ya da sabit etkili olabilir.	Değerlendiriciler modelde rasgele ya da sabit etkili olabilir.
Kovaryat	Modelde kovaryatlar yer alabilir.	Modelde kovaryatlar yer almaz.	Modelde kovaryatlar yer almaz.	Modelde kovaryatlar yer almaz.	Modelde kovaryatlar yer alabilir.
Doğruluk ölçümü	Diagnostik istatistiklerden herhangi biri kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.	Eğri altında kalan alan kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.	Diagnostik istatistiklerden herhangi biri kullanılır.
Yazılım	LABMRCM (22) DBM MRMC 2.0 (23)	DBM MRMC 2.0 (23) OBUMRM (24)	Bilgisayar yazılımı yoktur.	Bilgisayar yazılımı yoktur	Bilgisayar yazılımı yoktur

Sonuç

Çok testli çok değerlendiricili diagnostik doğruluk çalışmalarında kullanılabilecek beş farklı yöntemle rastlanmıştır. Bu metotlardan DBM metodu veri setini sözde değerler kullanarak yeniden örnekleme yapması, modele kovaryatları dahil edebilmesi, diagnostik performans istatistiklerinden herhangi birini

kullanabilmesi açısından diğer metotlara üstünlük sağlamaktadır. DBM ve OR metotlarının uygulamada kullanılan birden fazla yazılımına internetten ücretsiz erişmek mümkündür. Çok değişkenli WMW, BWC, HROC metotlarının herhangi bir bilgisayar yazılımlarının olmaması bu yöntemlerin dezavantajıdır. Bu üç yöntemin bilgisayar yazılımlarının oluşturulması önerilmektedir.

Kaynaklar

- Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;143:29-36.
- Zhou X, Obuchowski N, McClish D. Statistical Methods in Diagnostic Medicine, 1st Ed., John Wiley&Sons.Inc. 2002:274-307.
- Obuchowski NA. Multireader Receiver Operating Characteristic Studies: A Comparison of study Designs. *Acad Radiol* 1995;2:709-16.
- Dorfman D, Berbaum K, Metz C. Receiver operating characteristic rating analysis: generalization to the population of readers and patients with jackknife method. *Investigate Radiology* 1992;27:723-31.
- Obuchowski N. Multi-Reader Multi Modality ROC Studies: Hypothesis Testing and Sample Size Estimation Using an ANOVA Approach with Dependent Observations., With rejoinder. *Academic Radiology* 1995;2:522-9.
- Obuchowski NA. Fundamentals of Clinical Research for Radiologists. *American Roentgen Ray Society* 2005;184:364-72.
- Hillis SL, Obuchowski NA, Schartz KM, Berbaum KS. A Comparison of the Dorfman-Berbaum-Metz and Obuchowski-Rockette methods for Receiver Operating Characteristic (ROC) data. *Statistics in Medicine* 2005;24:1579-607.

8. Dorfman DD, Berbaum KS, Lenth RV, Chen YF, Donaghy BA. Monte Carlo validation of a multireader method for receiver operating characteristic discrete rating data: factorial experimental design. *Academic Radiology* 1998;5:591-602.
9. Obuchowski NA, Rockette HE. Hypothesis testing of the diagnostic accuracy for multiple diagnostic tests: an approach with dependent observations. *Commun Stat Simul Coomputation* 1995;24:285-308.
10. Hillis SL, Berbaum. KS. Power estimation for the Dorfman-Berbaum-Metz method. *Academic Radiology* 2004;11:1260-73.
11. Hillis SL. Monte Carlo validation of the Dorfman-Berbaum-Metz Method using normalized pseudovalues and less data-based model simplification. *Academic Radiology* 2005;12:1534-41.
12. Hillis LS, Berbaum KS., Metz CE. Recent developments in the dorfman-berbaum-metz procedure for multireader roc study analysis. *Academic Radiology* 2008;15:647-61.
13. Song HH. Analysis of correlated ROC areas in diagnostic testing. *Biometrics* 1997;53:370-82.
14. Ishwaran H, Gatsonis CA. A general class of hierarchical ordinal regression models with applications to correlated ROC analysis. *Can J Stat* 2000;28:731-50.
15. Wang F, Gatsonis CA. Hierarchical models for roc curve summary mesures: design and analysis of multi-reader, multi-modality studies of medical tests. *Statistics in Medicine* 2008;27:243-56.
16. Toledano AY. Three methods for analysing correlated roc curves: a comprasion in real data sets from multi-reader, multi-case studies with a factorial design. *Statistics in Medicine* 2003;22:2919-33.
17. Beiden S, Wagner, Campbell G. Components-of-variance models and multiple-bootstrap experiments: an alternative method for random-effects, receiver operating characteristic analysis. *Academic Radiology* 2000;7:341-9.
18. Beiden S, Wagner R, Campbell G, Metz CE, Jiang Y. Components-of-variance models for random effects roc analysis: the case of unequal variance structures across modalities. *Academic Radiology* 2001;8:605-15.
19. Beiden S, Wagner R, Campbell G, Chan HP. Analysis of uncertainties in estimates of components of variance in multivariate ROC analysis. *Academic Radiology* 2001;8:616-22.
20. Obuchowski NA, Beiden SV, Berbaum SK, Hills SL, Ishwaran H, Song HH, Wagner RF. Multireader multicase receiver operating characteristic analysis: an empirical comparison of five methods. *Academic Radiology* 2004;11:980-95.
21. Obuchowski NA. New methodological tools for multiple-reader ROC studies. *Radiology* 2007;243:10-2.
22. LABMRMC. Erişim adresi: <http://metz-roc.uchicago.edu/MetzROC/software> Erişim Tarihi: 1 Temmuz 2011.
23. DBM MRMC.Erişim Adresi: <http://perception.radiology.uiowa.edu/Software/ReceiverOperatingCharacteristicROC/DBMMRMC/tabid/116/Default.aspx> Erişim Tarihi: 1 Temmuz 2011.
24. OBUMRM. Erişim adresi: <http://www.lerner.ccf.org/qhs/software/obumrm.php> Erişim Tarihi: 1 Temmuz 2011.