

Farklı Örneklem Büyüklükleri Kullanılarak ROC Eğrisi Altında Kalan Alan İçin Bazı Parametrik Olmayan Yöntemlerin Güven Aralıklarının Karşılaştırılması: Bir Simülasyon Çalışması

Comparison of Confidence Intervals Some Nonparametric Methods for the Area Under the ROC Curve Using Different Sample Sizes: A Simulation Study

İsmet DOĞAN^a, Nurhan DOĞAN^a

^aAfyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, Afyonkarahisar, Türkiye

ÖZET Amaç: Bu çalışmanın amacı, alıcı çalışma karakteristiği [receiver operating characteristic (ROC)] eğrisi altında kalan alana [area under the curve (AUC)] ait güven aralıklarının tahmin edilmesinde kullanılan bazı parametrik olmayan yöntemleri güven aralıklarının genişliği bakımından karşılaştırmak, farklı örneklem büyüklüklerinin AUC güven aralıkları üzerindeki etkilerini araştırmaktır. **Gereç ve Yöntemler:** Farklı örneklem büyüklüklerinin varlığında AUC'nin güven aralığını oluşturmak için yaygın olarak kullanılan 8 farklı parametrik olmayan yöntem kullanılmıştır. Bu yöntemleri karşılaştırmak için bir simülasyon çalışması gerçekleştirilmiş, Python-random kütüphanesi kullanılarak $10 \leq n \leq 1.000$ aralığında yer alan 35 farklı n değeri için veri türetilmiştir. **Bulgular:** Güven aralıklarının genişliği dikkate alındığında, çalışmada dikkate alınan yöntemler iyiden kötüye doğru Logit Dönüşüm Tabanlı Mann-Whitney, Mann-Whitney, Agresti-Coull, Wilson, Frekansçı Yaklaşım, Süreklilik Düzeltmeli Wilson, Uyarlanmış Wald ve Süreklilik Düzeltmeli Uyarlanmış Wald yöntemi şeklinde sıralanabilir. Alt/üst sınırlara ait minimum-maksimum değerlerinin 0-1 aralığında kalması bakımından ise bu sıralama Logit Dönüşüm Tabanlı Mann-Whitney, Wilson, Frekansçı Yaklaşım, Süreklilik Düzeltmeli Wilson, Agresti-Coull, Uyarlanmış Wald, Süreklilik Düzeltmeli Uyarlanmış Wald ve Mann-Whitney şeklindedir. Güven aralıklarının AUC değerlerini kapsama eğilimi bakımından $n \geq 175$ için yöntemlerin çoğu benzer güven aralıkları sağlar. Bu nedenle tercih edilen bir yaklaşım yoktur. $n < 175$ durumunda ise AUC güven aralıklarının belirlenmesinde Logit Dönüşüm Tabanlı Mann-Whitney ve dönüşümsüz Mann-Whitney yöntemlerinin kullanılması gerektiği sonucuna ulaşılmıştır. Alt ve üst sınır değerlerinin < 0 ya da > 1 olmasına göre ise Logit Dönüşüm Tabanlı Mann-Whitney, Wilson ve Frekansçı Yaklaşımın diğer yöntemlere göre $0 \leq alt, üst sınır \leq 1$ aralığında değerler aldığı belirlenmiştir. **Sonuç:** Karşılaştırılan yöntemler arasında gerek kapsama eğilimi gerekse sınır değerlerin 0-1 aralığında değerler alması kriterleri dikkate alındığında Logit Dönüşüm Tabanlı Mann-Whitney yönteminin en iyi yöntem olduğu söylenebilir.

Anahtar kelimeler: Tanı testi; ROC analizi; eğri altında kalan alan; duyarlılık; özgüllük

ABSTRACT Objective: The aim of this study is to compare some non-parametric methods used to estimate the confidence intervals of the area under the receiver operating characteristic (ROC) curve (AUC) in terms of the width of the confidence intervals, and to investigate the effects of different sample sizes on AUC confidence intervals. **Material and Methods:** Eight different commonly used nonparametric methods were used to establish the confidence interval of the AUC in the presence of different sample sizes. A simulation study was carried out to compare these methods, and data was derived for 35 different n values in the $10 \leq n \leq 1.000$ range using the Python-random library. **Results:** Considering the width of confidence intervals, the methods considered in the study can be listed from best to worst as Logit Transformation Based Mann-Whitney, Mann-Whitney, Agresti-Coull, Wilson, Frequency Approach, Wilson with Continuity Correction, Adapted Wald and Adapted Wald method with Continuity Correction. In terms of the min-max values of the lower/upper limits remaining in the range of 0-1, this order is Logit Transformation Based Mann-Whitney, Wilson, Frequency Approach, Continuity Corrected Wilson, Agresti-Coull, Adapted Wald, Continuity Corrected Adapted Wald and Mann-Whitney. For $n \geq 175$, most methods provide similar confidence intervals in that confidence intervals tend to encompass AUC values. Therefore, there is no preferred approach. In case of $n < 175$, it was concluded that Logit Transformation Based Mann-Whitney and non-transformational Mann-Whitney methods should be used to determine AUC confidence intervals. Depending on whether the lower and upper limit values are < 0 or > 1 , it has been determined that the Logit Transformation Based Mann-Whitney, Wilson and Frequency Approach has values in the range of $0 \leq$ lower and upper limit ≤ 1 compared to other methods. **Conclusion:** Among the compared methods, it can be said that the Logit Transformation Based Mann-Whitney method is the best method, considering both the coverage tendency and the criteria for the border values to be in the range of 0-1.

Keywords: Diagnostic test; ROC analysis; area under curve; sensitivity; specificity

KAYNAK GÖSTERMEK İÇİN:

Doğan I, Doğan N. Farklı örneklem büyüklükleri kullanılarak ROC eğrisi altında kalan alan için bazı parametrik olmayan yöntemlerin güven aralıklarının karşılaştırılması: Bir simülasyon çalışması. Türkiye Klinikleri J Biostat. 2024;16(2):95-106.

Correspondence: Nurhan DOĞAN

Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD, Afyonkarahisar, Türkiye

E-mail: nurhandogan@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

Received: 16 Feb 2024 **Received in revised form:** 15 Mar 2024 **Accepted:** 25 Jun 2024 **Available online:** 03 Jul 2024

2146-8877 / Copyright © 2024 by Türkiye Klinikleri. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Alıcı çalışma karakteristiği [receiver operating characteristics (ROC)] eğrisi, II. Dünya Savaşı sırasında mühendisler tarafından savaş alanlarındaki düşman nesnelere tespit etmek için geliştirilmiş, diğer alanlara yayılması hızlı olmuştur. Tıpta, ROC analizinin ilk uygulamalarından biri 1960'larda yayınlanmış, ancak ROC eğrisi popülerliğini ancak 1970'lerde kazanmıştır. Günümüzde tıbbi teknolojiler, bir hastalığı teşhis etmek veya hastalığın ilerleyişini tahmin etmek için çok çeşitli yollar sunmakta ve yeni tanı testleri ve biyobelirteçler üzerinde sürekli olarak çalışılmaktadır. ROC analizi, bir tanı testini temsil eden sürekli bir değişkenin ayırt edici performansını değerlendirmek için yaygın olarak kullanılmaktadır. Farklı amaçlara göre ROC analizi:

- Sürekli bir belirtecin, iki grupta bir sınıflandırmaya doğru şekilde atama yapma konusundaki ayırt edici yeteneğini değerlendirmek,
- İki gruptaki deneklerin en az yanlış sınıflandırılması için en uygun kesme noktasını bulmak,
- İki (veya daha fazla) tanı testinin veya belirtecinin etkinliğini karşılaştırmak,
- İki veya daha fazla gözlemci aynı sürekli değişkeni ölçtüğünde gözlemciler arası değişkenliği incelemek

için faydalıdır.¹ ROC eğrisi, belirli bir eşik değerindeki tek bir duyarlılık ve özgüllük değeriyle karşılaştırıldığında:

- ROC eğrisi olası tüm kesme noktalarını gösterir ve daha sonra verilecek prosedüre göre hastalıklı veya hastalıklı olmayan deneklerin doğru şekilde tanımlanması için en uygun kesme noktası okunabilir,
- ROC eğrisi, hastalık prevalansından bağımsız olduğu bilinen duyarlılık ve özgüllüğe dayandığından hastalığın prevalansından bağımsızdır,
- İki veya daha fazla teşhis testi, tek bir şekilde aynı anda görsel olarak karşılaştırılabilir,
- Bazen duyarlılık, özgüllükten daha önemlidir veya tam tersi, ROC eğrisi, sabit özgüllük değerinde gerekli duyarlılık değerinin bulunmasına yardımcı olur,
- ROC eğrisinin altındaki ampirik alan, bir sabitin eklenmesine veya çıkarılmasına veya log veya karekök gibi dönüşümlere göre değişmez,
- Eğri altında kalan alan [area under the curve (AUC)] ve eğri altındaki kısmi alan gibi tanı testinin geçerliliğini belirlemek için yararlı ölçüm özeti elde edilebilir,

avantajlarına sahiptir.² Ancak anlamlı bir değerlendirme, testin bireysel hastalar için sonuçları nasıl etkilediğini, günlük uygulamaya uygulanabilir bir yaygınlıkta, yanlış negatif ve yanlış pozitif tanıların farklı yanlış sınıflandırma maliyetlerinin açık bir değerlendirmesini içermelidir. Ayrıca veriler, klinisyenlerin, hastaların ve sağlık hizmeti sağlayıcılarının seçimlerini kolaylaştırmak için anlaşılır ve sezgisel olmalıdır. AUC bu hedeflere kolayca ulaşamaz. ROC eğrisi klinik olarak ilgili potansiyel klinik yorum alanının ötesine uzanır. Hasta olan ve olmayan hastalar arasında mükemmel bir ayırım sağlayan bir tanı testinin AUC'sinin 0,5 olması mümkündür. İki tanı testinin doğruluğunu karşılaştırırken testlerin ROC eğrileri kesişebilir. Bu durumda testlerin yalnızca AUC değerlerine göre karşılaştırılması yanıltıcı olabilir. Son olarak, ROC eğrisini ve özellikle de AUC'yi hastalığın test öncesi olasılığını ve yanlış teşhislerin maliyetini hesaba katmadığı için eleştirilmektedir.^{3,4} Bir tanı testinin doğruluğunu ölçmenin kapsamlı bir yolu, özgüllük ve duyarlılık seviyelerini grafiksel olarak karşılaştıran ROC eğrisini kullanmaktır. ROC eğrisi spesifik olarak tüm olası eşik değerlerinde duyarlılığın, 1-özgüllüğe karşı grafiğidir. ROC eğrisini özetlemenin yaygın bir yolu ise ROC eğrisinin altındaki alanı (AUC) hesaplamaktır. AUC'nin değeri 0,5'ten (esasen anlamsız bir test) 1'e (mükemmel bir test) kadar değişir ve rastgele seçilen hasta bir denegün, rastgele seçilmiş sağlıklı bir denegünkünden daha yüksek bir test değerine sahip olma olasılığı olarak yorumlanabilir. AUC çoğu durumda örnek ROC eğrisinin altındaki alanın hesaplanmasıyla ampirik olarak hesaplanır ve yalnızca bir nokta tahmini yerine AUC'nin aralık tahmininin sunulması genellikle ilgi çekicidir.⁵ ROC eğrisi esasen hastalıklı ve sağlıklı deneklerin dağılımları arasındaki ayırımı dağılımdan bağımsız bir tanımını sağlar. Dolayısıyla özet ölçümlerin her biri bir

anlamda bu iki dağılım arasındaki mesafenin özetidir. Aslında, AUC'nin ampirik tahmincisi Mann-Whitney U istatistiğine eş değerdir ve rastgele seçilmiş hastalıklı bir deneğin, rastgele seçilmiş sağlıklı bir deneğinkinden daha yüksek bir test değerine sahip olma olasılığı olarak yorumlanabilir. AUC'nin değeri $0 \leq AUC \leq 1$ arasında değişir ve daha büyük AUC değerleri daha iyi doğruluğu ifade eder.⁶ Nokta tahminleri gerçek değere en iyi yaklaşımı sağlasa da ne kadar kesin oldukları konusunda bilgi vermezler. Bu güven aralıkları ile elde edilir. Gerçek değer bilinmediğinden, örneklem için tahmin edilen parametreler ile popülasyon için gerçek değer arasındaki farkın büyüklüğü hakkında kesin bir ifadeye bulunmak elbette imkânsızdır. Ancak nokta tahmininin gerçek değere yakın olduğuna dair bir miktar güven duymak istenir. Güven aralıkları, gerçek değerlerin belirli bir aralıkta olma olasılığını tanımlamak için kullanılabilir. Geniş bir güven aralığı kesin olmayan bir sonuca işaret eder ve istatistiksel anlamlılığa bakılmaksızın sonuçların dikkatle yorumlanması gerektiğini belirtir.⁷ AUC çoğu durumda ROC eğrisinin altındaki alanın hesaplanmasıyla ampirik olarak hesaplanır ve yalnızca bir nokta tahmini yerine AUC'nin aralık tahmininin sunulması genellikle ilgi çekicidir. AUC için bir aralık tahmin edicisinin hesaplanmasına yönelik önerilen bir takım yöntemler mevcuttur. Bamber, 1975'te AUC için bir varyans tahmincisi ve buna karşılık gelen güven aralığı önermiş, birçok araştırmacı için bu öneri başlangıç noktası olmuştur.⁸ Önerilen yöntemlerin çoğu, AUC ile Wilcoxon'un sıra sayıları toplamı test istatistiği arasındaki eş değerklik ilişkisine dayanarak türetilmiştir.⁵ Bu çalışmanın amacı, ROC eğrisi altında kalan alana (AUC) ait güven aralıklarının tahmin edilmesinde kullanılan bazı parametrik olmayan yöntemleri, güven aralıklarının genişliği bakımından karşılaştırmak, farklı örneklem büyüklüklerinin AUC güven aralıkları üzerindeki etkilerini araştırmaktır. Çalışmada Helsinki Deklarasyonu prensipleri dikkate alınmıştır.

GEREÇ VE YÖNTEMLER

Çalışmada Phyton-random kütüphanesi kullanılarak $10 \leq n \leq 1.000$ aralığında yer alan 35 farklı n değeri (10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 125, 150, 175, 200, 225, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1.000) için veri türetilmiştir. Veri türetimine ilişkin Python 3.9.10 (Python Software Foundation, ABD) programlama dili kullanılarak yazılan programda yer alan simülasyon kurgusunun detayları aşağıdaki gibidir:

Adım 1. Veri türetimi için gerekli değerler (toplam kaç adet veri türetilmek istendiği, kategorilere atanacak değerlerin toplamı, kategori sayısı) programa girilir.

Adım 2. Python random kütüphanesindeki choice fonksiyonu kullanılarak rastgele bir şekilde kategori (GP: Gerçek pozitif, YP: Yanlış pozitif, YN: Yanlış negatif, GN: Gerçek negatif) seçimi yapılır.

Adım 3. Choice fonksiyonu tekrar kullanılarak seçilen kategori için 0 ile istenen örneklem büyüklüğü arasında rastgele bir tam sayı seçimi yapılır. Örneğin istenen örneklem büyüklüğü 30 ise $[0, 30]$ aralığında rastgele bir tam sayı seçilir.

Adım 4. Tüm kategoriler için ikinci ve üçüncü adımlar tekrarlanır k 'nci kategori için bir tam sayı seçildikten sonra seçilen sayıların toplamına bakılır. Eğer bu toplam, istenilen toplamdan fazla ise bu geçersiz bir veri olacağından işleme baştan başlanır. Son kategoriye atanacak tam sayı ise istenilen toplamdan mevcut toplamın çıkarılmasıyla bulunur.

Adım 5. Türetilen verinin daha önce türetilip türetilmediğine bakılır. Aynı veri daha önce türetilmişse veri setine eklenmez.

Adım 6. İstenilen toplam veri sayısına ulaşana kadar birinci adımdan itibaren program bir döngü içinde tekrar çalıştırılır (Bu çalışmada standardizasyonu sağlamak amacıyla her bir örneklem genişliği için 1.000 adet veri türetilmesi planlanmış ancak, 4 kategori için \overline{AUC} 'nin hesaplanabildiği, tekrar etmeyen 1.000 tane veri bulmak imkânsız olduğundan farklı örneklem büyüklüklerinde veri türetilmiştir).

Adım 7. Son olarak türetilen verilerden basit bir Excel (Microsoft Corporation, USA) uygulaması kullanılarak duyarlılık, özgüllük, AUC ve güven aralıkları hesaplanır.

Bir tıbbi testin doğal istatistiksel geçerliliğinin iki popüler göstergesi, gerçek hastalıklı denekler ve gerçek hastalıklı olmayan denekler arasında testle doğru tanıyı tespit etme olasılıklarıdır. İkili yanıt için, test pozitif veya test negatif cinsinden sonuçlar [Tablo 1](#) ile verilen 2x2'lik bir durum tablosunda özetlenebilir.

[Tablo 1](#)'den yararlanarak duyarlılık ve özgüllük değerleri,

$$\text{Duyarlılık} = \frac{GP}{GP + YN} \quad (1)$$

$$\text{Özgüllük} = \frac{GN}{GN + YP} \quad (2)$$

eşitlikleri yardımı ile hesaplanır. İki sonuçlu bir sınıflandırıcı için gerek AUC'nin ampirik tahmincisinin Mann-Whitney U istatistiğine eş değer olması gerekse ROC eğrisinin duyarlılık ve özgüllük ile olan ilişkisinden dolayı AUC değerlerinin,

$$\widehat{AUC} = \frac{1}{n} (\bar{R}_1 + \bar{R}_0) + 0,5 \quad (3)$$

\bar{R}_1 : Birinci gruba ait gözlem değerlerinin sıra sayılarının ortalaması,

\bar{R}_0 : İkinci gruba ait gözlem değerlerinin sıra sayılarının ortalaması,

veya

$$\widehat{AUC} = \frac{1}{2} (\text{duyarlılık} + \text{özgüllük}) \quad (4)$$

eşitlikleri kullanılarak tahmin edilebileceği gösterilmiştir.⁹⁻¹¹ Çalışmada dikkate alınan yöntemlere ilişkin güven aralığı formülleri [Tablo 2](#)'de verilmiştir.

[Tablo 2](#)'de verilen Mann-Whitney ve Logit Dönüşüm Tabanlı Mann-Whitney yöntemleri için standart sapma (SS) değerleri;

$$SS = \sqrt{\frac{\sum_{i=1}^k (\widehat{AUC}_i - \widehat{AUC})^2}{k - 1}} \quad (5)$$

k = simülasyon sayısı,

\widehat{AUC}_i = i. simülasyondan elde edilen \widehat{AUC} değeri,

\widehat{AUC} = i. simülasyondan elde edilen \widehat{AUC} değerlerinin ortalaması,

eşitliği kullanılarak her bir örneklem büyüklüğü için ayrı ayrı hesaplanmıştır. Parametrik yöntemlerin kullanılması durumunda güven aralıklarının hesaplanmasında kullanılan standart sapma değerleri özel formüller kullanılarak hesaplandığı için parametrik yöntemlerle hesaplanmış bir AUC değeri için parametrik olmayan güven aralığı hesaplama yöntemleri kullanılmamalıdır.

TABLO 1: İki sonuçlu sınıflandırma test sonuçları için standart 2x2 durum tablosu.

		Gerçek durum	
		Hastalık pozitif	Hastalık negatif
Tanı testi sonucu	Test pozitif	GP	YP
	Test negatif	YN	GN

GP: Gerçek pozitif; YP: Yanlış pozitif; YN: Yanlış negatif; GN: Gerçek negatif.

TABLO 2: Çalışmada kullanılan yöntemlere ait güven aralıkları için formüller.

Yöntem	Sınır	Eşitlik
Mann-Whitney ⁹	Alt/Üst sınır	$\widehat{AUC} \mp z * \frac{SD}{\sqrt{n}}$
Logit Dönüşüm Tabanlı Mann-Whitney ⁹	Alt/Üst sınır	$Expit(logit(\widehat{AUC}) \mp z * \frac{SD}{\widehat{AUC}(1 - \widehat{AUC})\sqrt{n}})$ $logit(\widehat{AUC}) = \ln\left(\frac{\widehat{AUC}}{1 - \widehat{AUC}}\right)$ $Expit(.) = \frac{\exp(.)}{\exp(.) + 1}$
Wilson ^{9,12}	Alt/Üst sınır	$\frac{\widehat{AUC} + 0,5t}{(1 + t)} \mp \sqrt{\frac{\widehat{AUC}(1 - \widehat{AUC})t + 0,25t^2}{(1 + t)^2}}$ $t = \frac{z^2_{1-\frac{\alpha}{2}}}{n}$ $z_{1-\frac{\alpha}{2}} = 1,645 \text{ } \alpha = 0,10 \text{ için, } z_{1-\frac{\alpha}{2}} = 1,96 \text{ } \alpha = 0,05 \text{ için,}$
Süreklilik Düzeltmeli Wilson ^{9,12}	Alt sınır	$\frac{2n\widehat{AUC} + z^2 - 1 - z\sqrt{z^2 - 2 - \frac{1}{n} + 4\widehat{AUC}(n(1 - \widehat{AUC}) + 1)}}{2(n + z^2)}$
	Üst sınır	$\frac{2n\widehat{AUC} + z^2 + 1 + z\sqrt{z^2 + 2 - \frac{1}{n} + 4\widehat{AUC}(n(1 - \widehat{AUC}) + 1)}}{2(n + z^2)}$
Uyarlanmış Wald ⁹	Alt/Üst sınır	$\widehat{AUC} \mp z \sqrt{\frac{\widehat{AUC}(1 - \widehat{AUC})}{0,75n - 1}}$
Süreklilik Düzeltmeli Uyarlanmış Wald ⁹	Alt/Üst sınır	$\widehat{AUC} \mp z \sqrt{\frac{\widehat{AUC}(1 - \widehat{AUC})}{0,75 * n - 1}} + \frac{1}{2n}$
Agresti-Coull ^{9,13}	Alt/Üst sınır	$\widehat{AUC} \mp z \sqrt{\frac{\widehat{AUC}(1 - \widehat{AUC})}{n + 4}}$ $\widehat{AUC} = \frac{AUC * n + 2}{n + 4}$
Frekansçı Yaklaşım ¹⁴	Alt/Üst sınır	$\frac{\widehat{AUC} + \frac{z^2}{2n} \mp z \sqrt{\frac{\widehat{AUC}(1 - \widehat{AUC})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}}$

%95 güven aralığı için $z = 1,96$ değerine sahiptir. n örneklem büyüklüğünü göstermektedir.

BULGULAR

\widehat{AUC} değerleri Eşitlik 4 kullanılarak belirlenmiştir. Çalışmada farklı örneklem büyüklükleri için dikkate alınan simülasyon sayıları, örneklem büyüklüklerine göre \widehat{AUC} için minimum ve maksimum değerler ile elde edilen \widehat{AUC} değerleri [Tablo 3](#)'te verilmiştir.

TABLO 3: Farklı örneklem büyüklükleri için \widehat{AUC} değerlerine ait tanımlayıcı istatistikler.

N	Simülasyon sayısı	\widehat{AUC}	
		Minimum-Maksimum	$\bar{X} \pm SS$
10	264	0,00-1,00	0,5005±0,2568
15	784	0,00-1,00	0,5002±0,2415
20	979	0,00-1,00	0,5009±0,2304
25	978	0,00-1,00	0,5048±0,2265
30	989	0,00-1,00	0,4859±0,2199
35	993	0,00-1,00	0,4974±0,2172
40	993	0,00-1,00	0,5006±0,2129
45	995	0,01-1,00	0,5000±0,2178
50	997	0,00-1,00	0,5008±0,2201
55	995	0,00-1,00	0,5012±0,2203
60	993	0,00-1,00	0,5020±0,2129
65	998	0,00-1,00	0,4929±0,2104
70	998	0,01-1,00	0,4895±0,2155
75	998	0,00-1,00	0,5042±0,2172
80	999	0,00-1,00	0,4948±0,2107
85	1000	0,01-0,97	0,4987±0,2095
90	1000	0,00-0,99	0,5021±0,2135
95	998	0,00-1,00	0,5036±0,2141
100	998	0,00-1,00	0,4915±0,2126
125	999	0,01-1,00	0,5018±0,2100
150	1000	0,01-1,00	0,4990±0,2046
175	1000	0,00-0,98	0,4928±0,2102
200	1000	0,00-1,00	0,5121±0,2043
225	999	0,02-0,99	0,4920±0,2111
250	1000	0,02-0,99	0,4931±0,2074
300	1000	0,02-0,99	0,4918±0,2096
350	1000	0,00-0,98	0,4914±0,2099
400	1000	0,03-0,99	0,4975±0,2044
450	1000	0,00-0,99	0,5097±0,2053
500	1000	0,01-0,99	0,4948±0,1991
600	1000	0,01-0,98	0,5103±0,2094
700	1000	0,00-0,98	0,4996±0,2078
800	1000	0,03-0,97	0,5067±0,2083
900	1000	0,03-0,98	0,5016±0,1987
1.000	1000	0,01-0,95	0,4913±0,2040
Genel	33947	0,00-1,00	0,4987±0,2129

\widehat{AUC} : Area Under Curve; SS: Standart sapma.

[Tablo 3](#)'ten de görüldüğü üzere \widehat{AUC} değerleri örneklem büyüklüğünden etkilenmemektedir. %95 güven aralıkları örneklem büyüklüklerine göre incelenmiş ve elde edilen sonuçlar [Tablo 4](#)'te verilmiştir.

TABLO 4a: Yöntemlere göre \widehat{AUC} sınır değerleri için tanımlayıcı istatistikler.

N	Mann-Whitney Yöntemi			Logit dönüşüm tabanlı Mann-Whitney Yöntemi		
	Sınırlar	Alt sınır	Üst sınır	Sınırlar	Alt sınır	Üst sınır
10	-0,16-1,16	0,34±0,26	0,66±0,26	0,00-1,00	0,34±0,18	0,66±0,18
15	-0,12-1,12	0,38±0,24	0,62±0,24	0,00-1,00	0,37±0,19	0,63±0,19
20	-0,10-1,10	0,40±0,23	0,60±0,23	0,00-1,00	0,40±0,20	0,60±0,20
25	-0,09-1,09	0,42±0,23	0,59±0,23	0,00-1,00	0,41±0,20	0,59±0,20
30	-0,08-1,08	0,41±0,22	0,56±0,22	0,00-1,00	0,41±0,20	0,57±0,20
35	-0,07-1,07	0,43±0,22	0,57±0,22	0,00-1,00	0,42±0,20	0,57±0,20
40	-0,07-1,07	0,43±0,21	0,57±0,21	0,00-1,00	0,43±0,20	0,57±0,20
45	-0,05-1,06	0,44±0,22	0,56±0,22	0,00-1,00	0,43±0,21	0,56±0,21
50	-0,06-1,06	0,44±0,22	0,56±0,22	0,00-1,00	0,44±0,21	0,56±0,21
55	-0,06-1,06	0,44±0,22	0,56±0,22	0,00-1,00	0,44±0,21	0,56±0,21
60	-0,05-1,05	0,45±0,21	0,56±0,21	0,00-1,00	0,45±0,20	0,56±0,20
65	-0,05-1,05	0,44±0,21	0,54±0,21	0,00-1,00	0,44±0,20	0,55±0,20
70	-0,04-1,05	0,44±0,22	0,54±0,22	0,00-1,00	0,44±0,21	0,54±0,21
75	-0,05-1,05	0,46±0,22	0,55±0,22	0,00-1,00	0,45±0,21	0,56±0,21
80	-0,05-1,05	0,45±0,21	0,54±0,21	0,00-1,00	0,45±0,20	0,54±0,20
85	-0,04-1,02	0,45±0,21	0,54±0,21	0,00-0,99	0,45±0,21	0,54±0,20
90	-0,04-1,04	0,46±0,21	0,55±0,21	0,00-1,00	0,46±0,21	0,55±0,21
95	-0,04-1,04	0,46±0,21	0,55±0,21	0,00-1,00	0,46±0,21	0,55±0,21
100	-0,04-1,04	0,45±0,21	0,53±0,21	0,00-1,00	0,45±0,21	0,53±0,21
125	-0,02-1,04	0,47±0,21	0,54±0,21	0,00-1,00	0,46±0,21	0,54±0,21
150	-0,03-1,03	0,47±0,20	0,53±0,20	0,00-1,00	0,47±0,20	0,53±0,20
175	-0,03-1,01	0,46±0,21	0,52±0,21	0,00-0,99	0,46±0,21	0,52±0,21
200	-0,03-1,02	0,48±0,20	0,54±0,20	0,00-1,00	0,48±0,20	0,54±0,20
225	-0,01-1,02	0,46±0,21	0,52±0,21	0,00-1,00	0,46±0,21	0,52±0,21
250	-0,01-1,01	0,47±0,21	0,52±0,21	0,00-1,00	0,47±0,21	0,52±0,21
300	0,00-1,01	0,47±0,21	0,52±0,21	0,01-1,00	0,47±0,21	0,52±0,21
350	-0,02-1,00	0,47±0,21	0,51±0,21	0,01-1,00	0,47±0,21	0,51±0,21
400	0,01-1,01	0,48±0,20	0,52±0,20	0,02-1,00	0,48±0,20	0,52±0,20
450	-0,02-1,01	0,49±0,21	0,53±0,21	0,01-1,00	0,49±0,20	0,53±0,20
500	-0,01-1,01	0,48±0,20	0,51±0,20	0,00-1,00	0,48±0,20	0,51±0,20
600	-0,01-0,99	0,49±0,21	0,53±0,21	0,00-0,99	0,49±0,21	0,53±0,21
700	-0,01-0,99	0,48±0,21	0,51±0,21	0,00-0,99	0,48±0,21	0,52±0,21
800	0,02-0,99	0,49±0,21	0,52±0,21	0,02-0,98	0,49±0,21	0,52±0,21
900	0,01-0,99	0,49±0,20	0,51±0,20	0,02-0,99	0,49±0,20	0,51±0,20
1.000	-0,01-0,96	0,48±0,20	0,50±0,20	0,00-0,96	0,48±0,20	0,50±0,20
Genel	-0,16-1,16	0,45±0,21	0,54±0,21	0,00-1,00	0,45±0,21	0,54±0,21

Sınırlar: Alt sınır için Minimum, üst sınır için Maksimum değerlerini göstermektedir, Alt/Üst sınırları için değerler $\bar{X} \pm SS$ olarak verilmiştir,

TABLO 4b: Yöntemlere göre \widehat{AUC} sınır değerleri için tanımlayıcı istatistikler.

N	Wilson Yöntemi			Süreklilik Düzeltmeli Wilson Yöntemi		
	Sınırlar	Alt sınır	Üst sınır	Sınırlar	Alt sınır	Üst sınır
10	0,00-1,00	0,27±0,19	0,73±0,19	0,00-1,12	0,23±0,18	0,81±0,20
15	0,00-1,00	0,30±0,20	0,70±0,20	0,00-1,09	0,27±0,19	0,75±0,20
20	0,00-1,00	0,32±0,20	0,68±0,19	0,00-1,07	0,30±0,19	0,72±0,20
25	0,00-1,00	0,34±0,20	0,67±0,20	0,00-1,06	0,32±0,19	0,70±0,20
30	0,00-1,00	0,34±0,19	0,64±0,20	0,00-1,05	0,32±0,19	0,66±0,20
35	0,00-1,00	0,36±0,20	0,64±0,20	0,00-1,04	0,34±0,19	0,66±0,20
40	0,00-1,00	0,36±0,20	0,63±0,20	0,00-1,04	0,36±0,19	0,65±0,20
45	0,00-1,00	0,37±0,20	0,63±0,20	0,00-1,03	0,36±0,20	0,64±0,20
50	0,00-1,00	0,38±0,21	0,62±0,20	0,00-1,03	0,37±0,20	0,63±0,21
55	0,00-1,00	0,39±0,21	0,62±0,21	0,00-1,03	0,38±0,21	0,63±0,21
60	0,00-1,00	0,39±0,20	0,61±0,20	0,00-1,03	0,38±0,20	0,62±0,20
65	0,00-1,00	0,39±0,20	0,60±0,20	0,00-1,02	0,39±0,20	0,61±0,20
70	0,00-1,00	0,39±0,20	0,59±0,21	0,00-1,02	0,38±0,20	0,60±0,21
75	0,00-1,00	0,41±0,21	0,60±0,21	0,00-1,02	0,40±0,21	0,61±0,21
80	0,00-1,00	0,40±0,20	0,59±0,20	0,00-1,02	0,39±0,20	0,60±0,20
85	0,00-0,99	0,40±0,20	0,59±0,20	0,00-1,00	0,40±0,20	0,60±0,20
90	0,00-1,00	0,41±0,20	0,59±0,21	0,00-1,01	0,41±0,20	0,60±0,21
95	0,00-1,00	0,41±0,21	0,59±0,21	0,00-1,02	0,41±0,21	0,60±0,21
100	0,00-1,00	0,41±0,21	0,58±0,21	0,00-1,02	0,40±0,20	0,59±0,21
125	0,00-1,00	0,42±0,21	0,58±0,20	0,00-1,01	0,42±0,20	0,59±0,20
150	0,00-1,00	0,43±0,20	0,57±0,20	0,00-1,01	0,42±0,20	0,57±0,20
175	0,00-0,99	0,43±0,21	0,56±0,21	0,00-1,00	0,42±0,21	0,56±0,21
200	0,00-1,00	0,45±0,20	0,57±0,20	0,00-1,00	0,45±0,20	0,58±0,20
225	0,01-1,00	0,43±0,21	0,55±0,21	0,01-1,00	0,43±0,21	0,55±0,21
250	0,01-1,00	0,44±0,20	0,55±0,20	0,01-1,00	0,44±0,20	0,55±0,20
300	0,01-1,00	0,44±0,21	0,54±0,21	0,01-1,00	0,44±0,21	0,54±0,21
350	0,00-0,99	0,44±0,21	0,54±0,21	0,00-1,00	0,44±0,21	0,54±0,21
400	0,02-1,00	0,45±0,20	0,54±0,20	0,02-1,00	0,45±0,20	0,54±0,20
450	0,00-1,00	0,47±0,20	0,55±0,20	0,01-0,98	0,55±0,20	0,47±0,20
500	0,00-1,00	0,45±0,20	0,53±0,20	0,00-1,00	0,45±0,20	0,54±0,20
600	0,00-0,99	0,47±0,21	0,55±0,21	0,00-0,99	0,47±0,21	0,55±0,21
700	0,00-0,98	0,47±0,21	0,53±0,21	0,00-0,99	0,47±0,21	0,53±0,21
800	0,02-0,98	0,48±0,21	0,54±0,21	0,02-0,98	0,47±0,21	0,54±0,21
900	0,02-0,99	0,47±0,20	0,53±0,20	0,02-0,99	0,47±0,20	0,53±0,20
1.000	0,00-0,96	0,46±0,20	0,52±0,20	0,00-0,96	0,46±0,20	0,52±0,20
Genel	0,00-1,00	0,41±0,21	0,59±0,21	0,00-1,12	0,40±0,21	0,60±0,21

Sınırlar: Alt sınır için Min, üst sınır için Max değerlerini göstermektedir, Alt/Üst sınırları için değerler $\bar{x} \pm s$ olarak verilmiştir,

TABLO 4c: Yöntemlere göre \widehat{AUC} sınır değerleri için tanımlayıcı istatistikler.

N	Uyarlanmış Wald Yöntemi			Süreklilik Düzeltmeli Uyarlanmış Wald Yöntemi		
	Sınırlar	Alt sınır	Üst sınır	Sınırlar	Alt sınır	Üst sınır
10	-0,13-1,13	0,19±0,28	0,81±0,28	-0,08-1,18	0,24±0,28	0,86±0,28
15	-0,09-1,09	0,24±0,25	0,76±0,25	-0,05-1,12	0,27±0,25	0,79±0,25
20	-0,06-1,06	0,27±0,24	0,73±0,23	-0,04-1,09	0,30±0,24	0,75±0,23
25	-0,05-1,05	0,30±0,23	0,71±0,23	-0,03-1,07	0,32±0,23	0,73±0,23
30	-0,04-1,04	0,30±0,22	0,67±0,23	-0,03-1,06	0,32±0,22	0,69±0,23
35	-0,04-1,04	0,32±0,22	0,67±0,22	-0,02-1,05	0,34±0,22	0,68±0,22
40	-0,03-1,03	0,34±0,21	0,66±0,22	-0,02-1,04	0,35±0,21	0,68±0,21
45	-0,03-1,03	0,35±0,22	0,65±0,22	-0,02-1,04	0,36±0,22	0,66±0,22
50	-0,03-1,03	0,36±0,22	0,64±0,22	-0,02-1,04	0,37±0,22	0,65±0,22
55	-0,02-1,02	0,36±0,22	0,64±0,22	-0,01-1,03	0,37±0,22	0,65±0,22
60	-0,02-1,02	0,37±0,21	0,63±0,21	-0,01-1,03	0,38±0,21	0,64±0,21
65	-0,02-1,02	0,37±0,21	0,62±0,21	-0,01-1,03	0,37±0,21	0,63±0,21
70	-0,02-1,02	0,37±0,22	0,61±0,22	-0,01-1,02	0,37±0,22	0,62±0,22
75	-0,02-1,02	0,39±0,22	0,62±0,22	-0,01-1,02	0,39±0,22	0,63±0,22
80	-0,01-1,02	0,38±0,21	0,61±0,21	-0,01-1,02	0,39±0,21	0,62±0,21
85	-0,02-1,01	0,39±0,21	0,61±0,21	-0,01-1,02	0,39±0,21	0,62±0,21
90	-0,01-1,01	0,39±0,21	0,61±0,21	-0,01-1,02	0,40±0,21	0,61±0,21
95	-0,01-1,01	0,40±0,21	0,61±0,21	-0,01-1,02	0,40±0,21	0,61±0,21
100	-0,01-1,01	0,39±0,21	0,59±0,21	-0,01-1,02	0,39±0,21	0,60±0,21
125	-0,01-1,01	0,41±0,21	0,59±0,21	-0,01-1,01	0,41±0,21	0,60±0,21
150	-0,01-1,01	0,42±0,20	0,58±0,21	-0,01-1,01	0,42±0,20	0,59±0,21
175	-0,01-1,00	0,42±0,21	0,57±0,21	0,00-1,01	0,42±0,21	0,57±0,21
200	-0,01-1,01	0,44±0,21	0,58±0,20	0,00-1,01	0,44±0,21	0,59±0,20
225	0,00-1,00	0,42±0,21	0,56±0,21	0,00-1,01	0,43±0,21	0,56±0,21
250	0,00-1,00	0,43±0,21	0,56±0,21	0,00-1,01	0,43±0,21	0,56±0,21
300	0,00-1,00	0,43±0,21	0,55±0,21	0,00-1,00	0,43±0,21	0,55±0,21
350	0,00-1,00	0,44±0,21	0,55±0,21	0,00-1,00	0,44±0,21	0,55±0,21
400	0,01-1,00	0,45±0,20	0,55±0,21	0,01-1,00	0,45±0,20	0,55±0,20
450	0,00-1,00	0,46±0,21	0,56±0,20	0,00-1,00	0,46±0,21	0,56±0,20
500	0,00-1,00	0,45±0,20	0,54±0,20	0,00-1,00	0,45±0,20	0,54±0,20
600	0,00-0,99	0,47±0,21	0,55±0,21	0,00-0,99	0,47±0,21	0,55±0,21
700	0,00-0,99	0,46±0,21	0,54±0,21	0,00-0,99	0,46±0,21	0,54±0,21
800	0,02-0,99	0,47±0,21	0,54±0,21	0,02-0,99	0,47±0,21	0,54±0,21
900	0,01-0,99	0,47±0,20	0,54±0,20	0,01-0,99	0,47±0,20	0,54±0,20
1.000	0,00-0,97	0,46±0,20	0,52±0,20	0,00-0,97	0,46±0,20	0,52±0,20
Genel	-0,13-1,13	0,39±0,22	0,60±0,22	-0,08-1,18	0,40±0,22	0,61±0,22

Sınırlar: Alt sınır için Min, üst sınır için Max değerlerini göstermektedir, Alt/Üst sınırları için değerler $\bar{x} \pm \sigma$ olarak verilmiştir,

TABLO 4d: Yöntemlere göre \widehat{AUC} sınır değerleri için tanımlayıcı istatistikler.

N	Agresti-Coull Yöntemi			Frekansçı Yaklaşım		
	Sınırlar	Alt sınır	Üst sınır	Sınırlar	Alt sınır	Üst sınır
10	-0,04-1,04	0,26±0,18	0,74±0,18	0,00-1,00	0,27±0,19	0,73±0,19
15	-0,03-1,03	0,29±0,19	0,71±0,19	0,00-1,00	0,30±0,20	0,70±0,20
20	-0,03-1,03	0,32±0,19	0,68±0,19	0,00-1,00	0,32±0,20	0,68±0,19
25	-0,02-1,02	0,34±0,20	0,67±0,20	0,00-1,00	0,34±0,20	0,67±0,20
30	-0,02-1,02	0,33±0,19	0,64±0,20	0,00-1,00	0,34±0,19	0,64±0,20
35	-0,02-1,02	0,35±0,20	0,64±0,20	0,00-1,00	0,36±0,20	0,64±0,20
40	-0,02-1,02	0,37±0,19	0,64±0,19	0,00-1,00	0,37±0,20	0,63±0,20
45	-0,01-1,01	0,37±0,20	0,63±0,20	0,00-1,00	0,37±0,20	0,63±0,20
50	-0,01-1,01	0,38±0,20	0,62±0,20	0,00-1,00	0,38±0,21	0,62±0,20
55	-0,01-1,01	0,39±0,21	0,62±0,21	0,00-1,00	0,39±0,21	0,62±0,21
60	-0,01-1,01	0,39±0,20	0,61±0,20	0,00-1,00	0,39±0,20	0,61±0,20
65	-0,01-1,01	0,39±0,20	0,60±0,20	0,00-1,00	0,39±0,20	0,60±0,20
70	-0,01-1,01	0,39±0,20	0,59±0,21	0,00-1,00	0,39±0,20	0,59±0,21
75	-0,01-1,01	0,40±0,21	0,60±0,21	0,00-1,00	0,41±0,21	0,60±0,21
80	-0,01-1,01	0,40±0,20	0,59±0,20	0,00-1,00	0,40±0,20	0,59±0,20
85	-0,01-1,00	0,40±0,20	0,59±0,20	0,00-0,99	0,40±0,20	0,59±0,20
90	-0,01-1,00	0,41±0,20	0,59±0,21	0,00-1,00	0,41±0,20	0,59±0,21
95	-0,01-1,01	0,41±0,21	0,59±0,21	0,00-1,00	0,41±0,21	0,59±0,21
100	-0,01-1,01	0,41±0,20	0,58±0,20	0,00-1,00	0,41±0,21	0,58±0,21
125	0,00-1,01	0,42±0,20	0,58±0,20	0,00-1,00	0,42±0,21	0,58±0,20
150	0,00-1,00	0,43±0,20	0,57±0,20	0,00-1,00	0,43±0,20	0,57±0,20
175	0,00-0,99	0,43±0,21	0,56±0,21	0,00-0,99	0,43±0,21	0,56±0,21
200	0,00-1,00	0,45±0,20	0,57±0,20	0,00-1,00	0,45±0,20	0,57±0,20
225	0,00-1,00	0,43±0,21	0,55±0,21	0,01-1,00	0,43±0,21	0,55±0,21
250	0,01-1,00	0,44±0,20	0,55±0,20	0,01-1,00	0,44±0,21	0,55±0,20
300	0,01-1,00	0,44±0,21	0,54±0,21	0,01-1,00	0,44±0,21	0,54±0,21
350	0,00-0,99	0,44±0,21	0,54±0,21	0,00-0,99	0,44±0,21	0,54±0,21
400	0,02-1,00	0,45±0,20	0,54±0,20	0,02-1,00	0,45±0,20	0,54±0,20
450	0,01-0,98	0,47±0,20	0,55±0,20	0,00-1,00	0,47±0,20	0,55±0,20
500	0,00-1,00	0,45±0,20	0,53±0,20	0,00-1,00	0,45±0,20	0,54±0,20
600	0,00-0,99	0,47±0,21	0,55±0,21	0,00-0,99	0,47±0,21	0,55±0,21
700	0,00-0,99	0,47±0,21	0,53±0,21	0,00-0,98	0,47±0,21	0,53±0,21
800	0,02-0,98	0,48±0,21	0,54±0,21	0,02-0,98	0,48±0,21	0,54±0,21
900	0,02-0,99	0,47±0,20	0,53±0,20	0,02-0,99	0,47±0,20	0,53±0,20
1.000	0,00-0,96	0,46±0,20	0,52±0,20	0,00-0,96	0,46±0,20	0,52±0,20
Genel	-0,04-1,04	0,41±0,21	0,58±0,21	0,00-1,00	0,41±0,21	0,59±0,21

Sınırlar: Alt sınır için Min, üst sınır için Max değerlerini göstermektedir, Alt/Üst sınırları için değerler $\bar{x} \pm s$ olarak verilmiştir,

Tablo 4'ten de görüldüğü üzere, güven aralıklarının kapsama eğilimi dikkate alındığında çalışmada dikkate alınan yöntemler iyiden kötüye doğru logit dönüşüm tabanlı Mann-Whitney, Mann-Whitney, Agresti-Coull, Wilson, Frekansçı Yaklaşım, Süreklilik Düzeltmeli Wilson, Uyarlanmış Wald ve Süreklilik Düzeltmeli Uyarlanmış Wald yöntemi şeklinde sıralanabilir. Alt/üst sınırlara ait minimum-maksimum değerlerinin 0 – 1 aralığında kalması bakımından ise bu sıralama Logit Dönüşüm Tabanlı Mann-Whitney, Wilson, Frekansçı Yaklaşım, Süreklilik Düzeltmeli Wilson, Agresti-Coull, Uyarlanmış Wald, Süreklilik Düzeltmeli Uyarlanmış Wald ve Mann-Whitney şeklindedir.

TARTIŞMA

ROC eğrisi, tanı testlerinde sürekli belirteçlerin doğruluğunun bir ölçüsü olarak sıklıkla kullanılır. AUC, ROC eğrisi için tartışmasız en yaygın kullanılan özet indekstir. Biyomedikal ve farmasötik bilimlerde örneklem büyüklüğünün çok küçük olduğu önemli uygulamalar vardır. Asimptotik sonuçlara dayalı bazı yöntemlerin örneklem büyüklüğü küçük olduğunda ve/veya gözlemlenen AUC değeri bire yakın olduğunda bozulabileceği iyi bilinmesine rağmen bu sorunların ciddiyeti ve hangi yöntemlerin iyi performans gösterebileceği konusunda literatürde sınırlı sayıda tartışma bulunmaktadır ve bu çalışmalar parametrik, parametrik olmayan ve Bayes yöntemleri arasındaki kapsamlı karşılaştırmalardan yoksundur. Üstelik, çalışmalarda ampirik kapsam olasılıkları dikkate alınmamış ve bu nedenle AUC için güven aralıklarının performanslarının karşılaştırılması göz ardı edilmiştir.¹⁵ Bu çalışmada, farklı örneklem büyüklüklerinin varlığında AUC'nin güven aralığını oluşturmak için benimsenebilecek 8 farklı yaklaşım karşılaştırılmıştır. Göz önünde bulundurulan yöntemler yaygın olarak kullanılan yöntemlerdir. Farklı yöntemleri karşılaştırmak için çeşitli senaryolardan üretilen verilerle bir simülasyon çalışması gerçekleştirilmiştir. Tıbbi testlerde küçük örneklem büyüklüğü senaryosu yaygın olmasına rağmen AUC için güven aralığının oluşturulmasına yönelik çeşitli yöntemlerin küçük örneklem büyüklüğü özelliklerine ilişkin kapsamlı bir çalışma, literatürde büyük ölçüde eksiktir. Feng ve ark. tarafından yapılan 29 farklı yöntemin karşılaştırıldığı çalışmada, AUC değeri ne kadar büyükse ve örneklem büyüklüğü ne kadar küçükse, farklı yaklaşımların sonuçları arasındaki tutarsızlığın da o kadar büyük olduğu, parametrik yaklaşımların parametrik olmayan yaklaşımlardan daha iyi performans gösterme eğiliminde olduğu belirtilse de çalışmada dikkate alınan yöntemler arasında Mann-Whitney istatistiğine dayalı yöntemlerin genel olarak diğerlerinden üstün olduğu, diğer yöntemlere göre daha iyi kapsama eğiliminde olduğu sonucuna ulaşılmıştır.¹⁵ Çalışmadan elde edilen bulgular, Feng ve ark. tarafından yapılan çalışma bulguları ile benzerlik göstermektedir. Buna göre güven aralıklarının \widehat{AUC} değerlerini kapsama eğilimi bakımından $n \geq 175$ için yöntemlerin çoğu benzer güven aralıkları sağlar.¹⁵ Bu nedenle tercih edilen bir yaklaşım yoktur. $n < 175$ durumunda ise \widehat{AUC} güven aralıklarının belirlenmesinde logit dönüşüm tabanlı Mann-Whitney ve dönüşümsüz Mann-Whitney yöntemlerinin kullanılması gerektiği sonucuna ulaşılmıştır. Alt ve üst sınır değerlerinin < 0 ya da > 1 olmasına göre ise logit dönüşüm tabanlı Mann-Whitney, Wilson ve Frekansçı Yaklaşımın diğer yöntemlere göre $0 \leq alt, üst sınır \leq 1$ aralığında değerler aldığı belirlenmiştir. Güven aralıklarının \widehat{AUC} değerlerini kapsama eğilimi bakımından $n \geq 175$ için yöntemlerin çoğu benzer güven aralıkları sağlar. Bu nedenle tercih edilen bir yaklaşım yoktur. $n < 175$ durumunda ise \widehat{AUC} güven aralıklarının belirlenmesinde logit dönüşüm tabanlı Mann-Whitney ve dönüşümsüz Mann-Whitney yöntemlerinin kullanılması gerektiği sonucuna ulaşılmıştır. Alt ve üst sınır değerlerinin < 0 ya da > 1 olmasına göre ise logit dönüşüm tabanlı Mann-Whitney, Wilson ve Frekansçı Yaklaşımın diğer yöntemlere göre $0 \leq alt, üst sınır \leq 1$ aralığında değerler aldığı belirlenmiştir. Gengsheng Qin ve Hotilovac tarafından yapılan çalışmada, 9 farklı yöntem karşılaştırılmış, ampirik olabilirliğe dayalı yöntem ile logit dönüşüm tabanlı Mann-Whitney yöntemine ait güven aralıklarının iyi bir kapsama doğruluğuna sahip oldukları belirlenmiştir. Dönüşümsüz Mann-Whitney aralığı ise yazarlar tarafından önerilmemiştir.¹⁶ Logit dönüşüm tabanlı Mann-Whitney yöntemine ilişkin çalışmadan elde edilen bulgular Gengsheng Qin ve Hotilovac tarafından elde edilen bulgular ile benzerlik göstermektedir. Ancak Gengsheng Qin ve Hotilovac tarafından önerilmemesine rağmen dönüşümsüz

Mann-Whitney yönteminin, çalışmadan elde edilen bulgular dikkate alındığında diğer yöntemlere göre önerilebilir bir yöntem olduğu belirlenmiştir.¹⁶

SONUÇ

Örnekleme büyüklüğünün etkisinin araştırılması ve simülasyon verilerine uygulanması, bazı sonuçları ve önerileri ortaya çıkarmıştır. Karşılaştırılan yöntemler arasında gerek kapsama eğilimi gerekse sınır değerlerin 0 – 1 aralığında değerler alması kriterleri dikkate alındığında logit dönüşüm tabanlı Mann-Whitney yönteminin en iyi yöntem olduğu sonucuna ulaşılmıştır.

Finansal Kaynak

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Çıkar Çatışması

Bu çalışma ile ilgili olarak yazarların ve/veya aile bireylerinin çıkar çatışması potansiyeli olabilecek bilimsel ve tıbbi komite üyeliği veya üyeleri ile ilişkisi, danışmanlık, bilirkişilik, herhangi bir firmada çalışma durumu, hissedarlık ve benzer durumları yoktur.

Yazar Katkıları

Bu çalışma hazırlanırken tüm yazarlar eşit katkı sağlamıştır.

KAYNAKLAR

1. Gonçalves L, Subtil A, Rosario Oliveira M, de Zea Bermudez P. ROC curve estimation: an overview. REVSTAT-Statistical Journal. 2014;12(1):1-20. [\[Crossref\]](#)
2. Kumar R, Indrayan A. Receiver operating characteristic (ROC) curve for medical researchers. Indian Pediatr. 2011;48(4):277-87. [\[Crossref\]](#) [\[PubMed\]](#)
3. Obuchowski NA. ROC analysis. AJR Am J Roentgenol. 2005;184(2):364-72. [\[Crossref\]](#) [\[PubMed\]](#)
4. Halligan S, Altman DG, Mallett S. Disadvantages of using the area under the receiver operating characteristic curve to assess imaging tests: a discussion and proposal for an alternative approach. Eur Radiol. 2015;25(4):932-9. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
5. Cho H, Matthews GJ, Harel O. Confidence intervals for the area under the receiver operating characteristic curve in the presence of ignorable missing data. Int Stat Rev. 2019;87(1):152-77. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
6. Pardo MC, Franco-Pereira AM. Nonparametric ROC summary statistics. REVSTAT- Statistical Journal. 2017;15(4):583-600. [\[Link\]](#)
7. du Prel JB, Hommel G, Röhrig B, Blettner M. Confidence interval or p-value?: Part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int. 2009;106(19):335-9. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
8. Bamber D. The area above the ordinal dominance graph and the area below receiver operating characteristic graph. J Math Psychol. 1975;12(4):387-415. [\[Crossref\]](#)
9. Kottas M, Kuss O, Zapf A. A modified Wald interval for the area under the ROC curve (AUC) in diagnostic case-control studies. BMC Med Res Methodol. 2014;14:26. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
10. Mbizvo GK, Lamer AJ. Receiver operating characteristic plot and area under the curve with binary classifiers: pragmatic analysis of cognitive screening instruments. Neurodegener Dis Manag. 2021;11(5):353-60. [\[Crossref\]](#) [\[PubMed\]](#)
11. Muschelli J. ROC and AUC with a binary predictor: a potentially misleading metric. J Classif. 2020;37(3):696-708. [\[Crossref\]](#) [\[PubMed\]](#) [\[PMC\]](#)
12. Wilson E. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc. 1927;22:209-12. [\[Crossref\]](#)
13. Agresti A, Coull B. Approximate is better than "exact" for interval estimations of binomial proportions. Am Stat. 1998;52(2):119-26. [\[Crossref\]](#)
14. Stalbovskaya V, Hamadicharef B, Ifeachor EC. Sample size determination using ROC analysis. Plymouth, United Kingdom: 3rd International Conference on Computational Intelligence in Medicine and Healthcare (CIMED2007); 2007. [\[Link\]](#)
15. Feng D, Cortese G, Baumgartner R. A comparison of confidence/credible interval methods for the area under the ROC curve for continuous diagnostic tests with small sample size. Stat Methods Med Res. 2017;26(6):2603-21. [\[Crossref\]](#) [\[PubMed\]](#)
16. Gengsheng Qin, Hotilovac L. Comparison of non-parametric confidence intervals for the area under the ROC curve of a continuous-scale diagnostic test. Stat Methods Med Res. 2008;17(2):207-21. [\[Crossref\]](#) [\[PubMed\]](#)