

# İki Oran Arasındaki Fark İçin Güven Aralıklarının Karşılaştırılması: Bir Simülasyon Çalışması

## Comparison of Confidence Intervals for the Difference of Two Proportions: A Simulation Study

İsmet DOĞAN<sup>a</sup>, Nurhan DOĞAN<sup>a</sup>

<sup>a</sup>Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Afyonkarahisar, Türkiye

**ÖZET Amaç:** Bu çalışmanın amacı, klinik araştırmaların değerlendirilmesinde kullanılan 2 oran arasındaki fark için literatürde mevcut 15 farklı yöntemin %95 güvenirlilikle güven aralıklarını hesaplayarak karşılaştırmaktır. **Gereç ve Yöntemler:** Literatürde, 2 oran arasındaki farkın güven aralığının belirlenmesi amacıyla çok sayıda önerilmiş yöntem söz konusudur. Dolayısıyla hangi durumda hangi yöntemin kullanılacağı belirlenmesi gerekmektedir. Çalışmada, Python-random kütüphanesi kullanılarak  $10 \leq n \leq 1000$  aralığında yer alan 35 farklı  $n$  değeri için veri türetilmiştir. Verilerin türetilmesinde önce  $a$ ,  $b$ ,  $c$  ve  $d$  ile gösterilen gözlemlerden hangisine değer atanacağı, sonra da ilgili göze atanacak değer belirlenmiştir.  $n=10$  için 286,  $n=15$  için 815 ve  $n \geq 20$  için 1.000'er farklı veri seti çalışmada kullanılmıştır. **Bulgular:** Güven aralığının büyüklüğü bakımından yöntemler sıralanacak olursa en dar aralıktan en geniş aralığa doğru sıralama; Agresti-Caffo  $\leq$  Wald = Basit Asimptotik  $\leq$  Beal-Haldane  $\leq$  Anbar  $\leq$  Beal-Jeffreys-Perks  $\leq$  Newcombe Hibrid Skor  $\leq$  Brown-Li  $\leq$  Mee  $\leq$  Uyarlanmış Yule  $\leq$  Miittinen-Nurminen  $\leq$  Süreklilik Düzeltmeli Wald = Süreklilik Düzeltmeli Basit Asimptotik  $\leq$  Yule  $\leq$  Hauck-Anderson şeklindedir. Beal-Haldane yöntemine göre oranlar arası farkların %98,8'i, Beal-Jeffreys-Perks yöntemine göre %99,6'sı, diğer yöntemlerde ise oranlar arası farkların tamamı alt ve üst sınırlar arasında kalmaktadır. Tüm yöntemler için örnek büyüklüğü arttıkça güven aralıklarının daraldığı belirlenmiştir. Güven aralığının simetrikliği bakımından yöntemlere bakıldığında, hemen tüm yöntemlerden elde edilen güven aralıklarının simetrik olduğu ve bunun tüm örneklem büyüklükleri için geçerli olduğu söylenebilir. **Sonuç:** Güven aralıkları, klinik araştırma sonuçlarının yorumlanmasında son derece önemlidir. Çalışmadan elde edilen sonuçlar dikkate alındığında,  $n \leq 30$  için Agresti-Caffo yönteminin,  $30 \leq n \leq 90$  için Anbar ya da Wald yönteminin,  $n \geq 95$  için ise Anbar yönteminin en dar güven aralıklarına sahip olduğu belirlenmiştir.

**ABSTRACT Objective:** The aim of this article is to compare the 15 different methods available in the literature for the difference between the 2 ratios used in the evaluation of clinical studies, by calculating the confidence intervals with 95% confidence. **Material and Methods:** In the literature, there are many proposed methods for determining the confidence interval of the difference between 2 ratios. Therefore, it is necessary to determine which method will be used for which sample size. In the study, data were derived for 35 different  $n$  values in the range of  $10 \leq n \leq 1000$  using the Python-random library. In the derivation of the data, firstly, which cell shown with  $a$ ,  $b$ ,  $c$  and  $d$  will be assigned value, then the value to be assigned to the relevant cell was determined. 286 for  $n=10$ , 815 for  $n=15$  and 1,000 different data sets for  $n \geq 20$  were used in the study. **Results:** If the methods are ranked in terms of the size of the confidence interval, the order from the narrowest to the widest range is Agresti-Caffo  $\leq$  Wald = Simple asymptotic  $\leq$  Beal-Haldane  $\leq$  Anbar  $\leq$  Beal-Jeffreys-Perks  $\leq$  Newcombe Hybrid Score  $\leq$  Brown-Li  $\leq$  Mee  $\leq$  Adjusted Yule  $\leq$  Miittinen-Nurminen  $\leq$  Süreklilik Düzeltmeli Wald = Simple asymptotic with continuity correction  $\leq$  Yule  $\leq$  Hauck-Anderson. According to the Beal-Haldane method, 98.8% of the differences between the rates, 99.6% according to the Beal-Jeffreys-Perks method, and all the differences between the rates in other methods are between the lower and upper limits. For all methods, it was determined that the confidence intervals narrowed as the sample size increased. When we look at the methods in terms of the symmetry of the confidence interval, it can be said that the confidence intervals obtained from almost all methods are symmetrical and this is valid for all sample sizes. **Conclusion:** Confidence intervals are extremely important in interpreting clinical trial results. Considering the results obtained from the study, it was determined that Agresti-Caffo method for  $n \leq 30$ , Anbar or Wald method for  $30 \leq n \leq 90$ , and Anbar method for  $n \geq 95$  had the narrowest confidence intervals.

**Anahtar kelimeler:** Risk farkı; eşit etkinlik deneyleri; güven aralığı; iki sonuçlu veri

**Keywords:** Risk difference; non-inferiority trials; confidence interval; binary data

**Correspondence:** Nurhan DOĞAN

Afyonkarahisar Sağlık Bilimleri Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim ABD, Afyonkarahisar, Türkiye  
**E-mail:** nurhandogan@hotmail.com



Peer review under responsibility of Türkiye Klinikleri Journal of Biostatistics.

**Received:** 19 Sep 2022 **Received in revised form:** 17 Oct 2022 **Accepted:** 17 Oct 2022 **Available online:** 12 Dec 2022

2146-8877 / Copyright © 2023 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/>).

Klinik arařtırmalarda amaç, bir müdahalenin veya tedavinin etkisinin yönünü ve büyüklüğünü; gözlemsel çalışmalarda ise maruziyet ile sonuç arasındaki ilişkiyi tahmin etmektir. Bu etki veya ilişki, ölçülen sonuca bağılı olarak çeşitli şekillerde sunulabilir. İki sonuçlu kategorik verilerin söz konusu olduğu çalışmalarda, etki veya ilişkinin gösterilmesinde oranlar arası fark kullanılabilir.<sup>1</sup> İki oranın fark alınarak karşılaştırılması istatistikte temel bir problemdir ve birçok alanda uygulamaları bulunmaktadır. Klinik deneylerde çoğunlukla yeni tedavinin plasebodan veya standart tedaviden daha üstün (superiority) olduğu gösterilmek istenmesine rağmen yeni tedavinin standart tedaviden kötü olmadığını (non-inferiority=eşit etkinlik) gösteren deneylerin sayısı da giderek artmaktadır. İki oran arasındaki fark veya risk farkı ( $\delta$ ), genellikle kontrollü klinik çalışmalarda ve 2 sonuçlu prospektif çalışmalarda üzerinde durulan önemli konulardan biridir.  $\delta$  için güven aralığı tahmini Agresti (2003) tarafından yapılan çalışma ile literatürde oldukça dikkat çekmiştir.<sup>2</sup>  $\delta$  için elde edilen güven aralığı tahmininin, tedavi edilmesi gereken hasta sayısı (number needed to treat) ile yakın ilişkisinin olması ise klinisyenler tarafından daha fazla ilgi gösterilmesine sebep olmuştur.<sup>3</sup> Eşit etkinlik deneyleri, yeni deneysel tedavinin hâlen kullanılan kontrol tedavisinden kabul edilemez derecede kötü olup olmadığını değerlendirmek için yaygın olarak kullanılmaktadır. Eşit etkinlik ile ilgili hipotez testlerinde, risk farkının, rölatif risk ve odds oranına kıyasla Tip I hatayı iyi kontrol ettiği ve en yüksek istatistiksel gücü elde etmekle kalmayıp en küçük örneklem büyüklüğünü gerektirdiği gösterilmiştir. Bununla birlikte uygulamada, risk farkı, rölatif risk ve odds oranı arasındaki seçimin klinik yorumlara ve düzenleyici bakış açılarına dayanması gerektiği fikrinden dolayı eşit etkinlik ile ilgili klinik çalışmalarda etkinliği değerlendirmek için hangisinin kullanılması gerektiği hâlâ tartışmalıdır.<sup>4</sup> Klinik deneyler, genellikle yan etkileri ve maliyetleri azaltabilecek ya da olağan standart tedaviye göre başka bir avantaja sahip olabilecek yeni deneysel tedaviler veya terapilerle yürütülür. Bu durumlarda genellikle terapötik olarak eş değer veya daha kötü olup olmadığını belirlenmesi amacıyla yeni tedavi standart tedavi ile karşılaştırılır. Yeni tedaviden elde edilen sonuç standart tedaviden elde edilen sonuçtan, önceden belirlenmiş klinik olarak anlamlı bir miktardan farklılık göstermiyorsa yeni tedaviye “eş değer” denir. Ancak sonucu önceden belirlenmiş bir miktarda standart tedaviden daha kötü değilse “eşit etkin” olarak adlandırılır.<sup>5</sup> Deneysel verileri analiz etmek için basit bir yaklaşım, yeni tedavi yöntemi ile hâlen uygulanmakta olan tedavi yöntemini karşılaştırmak ve herhangi bir fark tespit edilmezse tedavileri eş değer ilan etmektir. Bu yaklaşımla ilgili sorun, örneklem büyüklüğünün çok küçük olmasından dolayı güven aralığının çok geniş olması durumunda eş değerliliğin uygun olmayan bir şekilde sonuçlandırılabilir olmasıdır. Bu nedenle amaç, gerçek “denkliği” değerlendirmekse o zaman sıfır hipotezi, tedavilerin farklı olduğu şeklinde olmalıdır. Tedavilerin denkliklerini beyan etmek ve sıfır hipotezini reddetmek için yeterli kanıt olup olmadığı araştırılmalıdır. İki tedavinin tam olarak eş değer etkilere sahip olması pratik olarak imkânsız olduğundan, klinik değerlendirme açısından “denklik”, 2 tedavinin etkilerinin eş değerlik tolerans değeri ( $\Delta$ ) olarak bilinen tolere edilebilir bir miktardan daha fazla farklılık göstermediği anlamına gelir. Bir denklik deneyinde, 2 tedavinin etkileri her iki yönde de  $\Delta$ 'dan daha fazla farklılık gösteriyorsa denklik geçerli değildir. Eşit etkinlik (daha kötü olmama) deneyleri, klinik olarak bir tedavinin bir başka tedaviden önemli ölçüde daha az etkili olmadığını göstermeyi amaçlar. Kontrol ve deney grupları arasında belirli bir miktardan daha az bir fark olduğunu göstermeyi amaçlayan bir deneyde, çalışmadan elde edilecek fark değerinin  $\Delta$  değerine göre daha düşük değere sahip olduğu sıfır hipotezi test edilir. Tedavi oranları arası fark için %95 güven aralığının üst sınırı  $\Delta$  değerinden küçükse o zaman yeni tedavinin standart tedaviden kötü olduğu biçimindeki sıfır hipotezi %5 anlamlılık seviyesinde reddedilir.<sup>6</sup> Eşit etkinlik için kullanılacak olan  $\Delta$ , klinik yargıya dayanmalıdır. Yeni tedavinin eşit etkin olarak kabul edilebilmesi için kontrole ne kadar yakın olması gerektiğine dair yoruma sahiptir.  $\Delta$ 'nın değerinin belirlenmesi kritik bir konu ve zor bir iştir.  $\Delta$ 'nın değerinin belirlenmesinde kontrol olarak kullanılan tedaviye ait başarı oranı dikkate alınabilir. Bu durumda  $\Delta$ 'nın değeri bu orandan daha küçük olmalıdır.<sup>5</sup> Birçok klinik çalışmada, 2 sonuçlu yanıt değişkenlerinin dağılımlarını tanımlamak için oranlar kullanılır. Eşit etkinlik ile ilgili çalışmalarda amaç, önceden belirlenmiş  $\Delta$  için tedavi (test) grubunun kontrol grubundan daha kötü olmadığını göstermektir. Eşit etkinlik değerlendirmesi genellikle tek yanlı bir istatistiksel test kullanılarak yapılır. Test ve kontrol gruplarına ait

başarı oranlarındaki farka odaklanıldığında, güven aralığının alt sınırı hesaplanır. Literatürde, oranlar arasındaki farkın güven aralığını hesaplamak için birçok yöntem vardır ve bunların alt sınırları, tek yanlı eşit etkinlik hipotezinin test edilmesi için potansiyel olarak uygulanabilir.<sup>7</sup> Oranlar arası fark, randomize kontrollü araştırmalar ve kohort çalışmaları için önemli bir etki ölçüsüdür. Etki ölçüsü olarak doğal değerine ek olarak, tahmini ve güven aralığı, tedavi edilmesi gereken sayının tahminini ve güven aralığını elde etmek için kullanılır. Bu fark, söz konusu olay zararlı ise risk farkı, epidemiyolojide ise atfedilebilir risk olarak adlandırılır.<sup>8</sup> Üstün olma, eşit etkinlik veya eş değerlik deneyleri için %5 veya daha düşük bir  $\alpha$  hata oranının sağlanması, genellikle %95lik bir güven aralığı olan  $(1 - \alpha)$  güven aralığının belirli sınırlar içinde kalmasını gerektirir.  $x$  tedavisi ile  $y$  tedavisi arasındaki etki farkı için üstünlük, %95 güven aralığının alt sınırının 0'dan büyük olduğu anlamına gelir. Eşit etkinlikte ise fark  $\Delta$  değerinden büyük olduğu sürece %95 güven aralığı 0'dan küçük değer içerebilir. %95 güven aralığının üst sınırının,  $\Delta$  değerinden küçük değer alması, yeni tedavinin eşit etkin olmadığını gösterir. %95 güven aralığının  $\Delta$  değerini içermesi durumunda eşit etkinlik konusunda bir şey söylenemez. Eşit etkinlik için %95 güven aralığının üst ve alt sınırlarının  $\Delta$  ve  $-\Delta$  arasında olması gerekir.<sup>9</sup> Bu çalışmada amaç, farklı örneklem büyüklükleri dikkate alınarak 2 oran arasındaki fark için önerilen güven aralıkları ile ilgili yöntemleri, güven aralıklarının genişliği ve farkları kapsama kriterleri bakımından karşılaştırmaktır.

## GEREÇ VE YÖNTEMLER

Oranlar arası farkın hesaplanmasında kullanılan formüller dikkate alındığında, klasik 2x2 tablo örneği Tablo 1'de verildiği şekilde oluşturulmaktadır.

**TABLO 1:** Oranlar arası fark hesaplamaları için 2x2 tablo örneği.

	İyileşti	İyileşmedi	Toplam
Yeni tedavi	$n_{11}$	$n_{12}$	$n_{1+}$
Standart tedavi	$n_{21}$	$n_{22}$	$n_{2+}$
Toplam	$n_{+1}$	$n_{+2}$	$n$

Tabloda yer alan  $n_{11}$ , yeni tedavi yöntemi ile iyileşen hasta sayısını,  $n_{12}$  iyileşmeyen hasta sayısını,  $n_{21}$  standart tedavi yöntemi ile iyileşen hasta sayısını,  $n_{22}$  ise iyileşmeyen hasta sayısını göstermektedir. Örneklemeden elde edilen oranlar kullanılarak  $\delta$ 'nın tahmini değeri;

$$\hat{\delta} = \hat{p}_1 - \hat{p}_2 = \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}} \quad (1)$$

eşitliği yardımıyla hesaplanır.<sup>8</sup> Randomize kontrollü çalışmalarda, sonuçların yorumlanması için güven aralığı gerekli olsa da hangi yöntemin en iyi olduğu konusunda bir fikir birliği yoktur. Eşit etkinlik ile ilgili çalışmalarda, her bir grupta denek sayısının 30 ve üzerinde olması durumunda yöntemlerin kapsama olasılıklarının birbirine daha benzer olduğu gösterilmiştir. Ancak bir deneyden elde edilen güven aralıkları, eşit etkinlik olmadığına dair çalışmanın sonucunu değiştirecek kadar büyük olsa bile farklı yöntemlerden elde edilen güven aralıkları arasında önemli farklılıklar olup olmadığı hâlâ belirsizliğini korumaktadır.<sup>10</sup> Çalışmada, 2 (değerlendirici) x 2 (kategori) tablolardan yararlanarak  $\hat{\delta}$  için geliştirilen güven aralıkları ile ilgili yöntemleri karşılaştırmak için bir simülasyon çalışması yapılmıştır. Çalışmada, Python-random kütüphanesi kullanılarak  $10 \leq n \leq 1000$  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 üretilmiştir.  $n = 10$  için 286,  $n = 15$  için 815 ve  $n \geq 20$  için 1.000'er farklı

veri seti çalışmada kullanılmıştır. Veri türetimine ilişkin Python 3.9.10 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 seçimi yapılır.

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

Adım 4. Tüm kategoriler için 2 ve 3. adımlar tekrarlanır ( $k'$ ncü kategori için bir tam sayı seçildikten sonra seçilen sayıların toplamına bakılır.). Eğer bu toplam, istenilen toplamdan fazla ise bu geçersiz bir veri olduğundan 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 1. 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 toplamı 10 veya 15 olan ve tekrar etmeyen 1.000 tane veri bulmak imkânsız olduğundan  $n = 10$  için 286 ve  $n = 15$  için 815 adet veri türetilmiştir.).

Adım 7. Son olarak, türetilen verilerden çalışmada dikkate alınan yöntemlere ait formüller kullanılarak alt ve üst sınır değerleri hesaplanır.

$n = 10$  için 286,  $n = 15$  için 815 ve  $n \geq 20$  için 1.000'er farklı veri seti çalışmada kullanılmıştır. Ancak güven aralığı hesaplamalarında bazı durumlarda hesaplama yapılamadığı için veri seti sayıları her bir örneklem büyüklüğü ve yöntem için farklılık göstermiştir. Farklı örneklem büyüklükleri ve yöntemler için hesaplamalarda kullanılan simülasyon sayıları Tablo 2'de verilmiştir.

Otuz beş farklı örneklem büyüklüğü için 15 farklı yöntemden elde edilen güven aralıkları, güven aralıklarının genişliği dikkate alınarak karşılaştırılmıştır. Literatürde yer alan ve bu çalışmada da dikkate alınan yöntemlerin çoğu birbirinden bağımsız ortaya konmuş yöntemler olmasına rağmen bazı yöntemler birbirinden türetilmiş yöntemlerdir. Örneğin Agresti ve Caffo, Wald yönteminin uygulanmasından önce dikkate alınan her iki gruptaki gözlem sayılarına 2 eklenmesi (1 başarılı, 1 başarısız) şeklinde bir değişiklik önermişlerdir. Newcombe, grupları ayrı ayrı dikkate almış, her bir oran için Wilson skor yöntemini kullanarak alt ve üst sınırları elde etmiş ve elde edilen alt ve üst sınırları birleştirerek güven aralığını hesaplamıştır. Diğer bir yöntem ise kısıtlı en çok olabilirlik tahminine dayanan Miettinen ve Nurminen yöntemidir.<sup>10</sup> Çalışmada dikkate alınan yöntemlere ait güven aralıklarının elde edilmesinde kullanılan eşitlikler aşağıda verilmiştir. Alt ve üst sınırların hesaplaması Excel uygulaması ile gerçekleştirilmiştir. Eşitliklerde yer alan  $c(\alpha)$  ve  $z_{\alpha/2}$ 'nin değeri %95 güven aralığı için 1,96 olarak alınmıştır.

Basit Asimptotik yöntem için güven aralığı<sup>11</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\frac{n_{11}n_{12}}{n_{1+}^3} + \frac{n_{21}n_{22}}{n_{2+}^3}} \quad (2)$$

**TABLO 2:** Hesaplamalarda kullanılan simülasyon sayıları.

n	Yöntem grubu I	Yöntem grubu II	Yöntem grubu III	Yöntem grubu IV
10	286	264	243	224
15	815	784	752	724
20	1.000	977	958	931
25	1.000	985	965	949
30	1.000	991	977	966
35	1.000	993	982	967
40	1.000	993	987	978
45	1.000	996	992	987
50	1.000	998	995	991
55	1.000	999	990	992
60	1.000	1.000	992	997
65	1.000	999	994	996
70	1.000	998	996	993
75	1.000	998	997	994
80	1.000	999	996	997
85	1.000	999	999	997
90	1.000	999	998	997
95	1.000	998	994	996
100	1.000	998	996	996
125	1.000	1.000	998	1.000
150	1.000	1.000	999	1.000
175	1.000	1.000	1.000	1.000
200	1.000	1.000	999	999
225	1.000	1.000	1.000	1.000
250	1.000	1.000	1.000	999
300	1.000	1.000	1.000	1.000
350	1.000	1.000	999	1.000
400	1.000	1.000	1.000	1.000
450	1.000	1.000	999	1.000
500	1.000	1.000	1.000	1.000
600	1.000	1.000	1.000	1.000
700	1.000	1.000	1.000	1.000
800	1.000	1.000	1.000	1.000
900	1.000	1.000	1.000	1.000
1.000	1.000	1.000	1.000	1.000
Genel	34.101	33.968	33.797	33.670

Yöntem Grubu I: Agresti-Caffo;

Yöntem Grubu II: Basit Asimptotik, Süreklilik Düzeltmeli Basit Asimptotik, Wald, Süreklilik Düzeltmeli Wald, Brown-Li, Yule, Uyarlanmış Yule, Newcombe Hibrid Skor, Anbar, Beal-Haldane, Beal-Jeffreys-Perks;

Yöntem Grubu III: Mee, Miettinen-Nurminen;

Yöntem Grubu IV: Hauck-Anderson.

Süreklilik Düzeltmeli Basit Asimptotik yöntem için güven aralığı<sup>11</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\frac{n_{11}n_{12}}{n_{1+}^3} + \frac{n_{21}n_{22}}{n_{2+}^3} + \frac{1}{2} \left( \frac{1}{n_{1+}} + \frac{1}{n_{2+}} \right)} \quad (3)$$

Hauck-Anderson yöntemi için güven aralığı<sup>12</sup>

$$\hat{\delta} \pm \left( z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_{1+}-1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_{2+}-1}} + \frac{1}{2 \min(n_{1+}, n_{2+})} \right) \quad (4)$$

Wald yöntemi için güven aralığı<sup>8</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_{1+}} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_{2+}}} \quad (5)$$

Wald güven aralığı  $n_{11}, n_{11}, n_{11}$  ve  $n_{11}$  değerlerinden etkilenmektedir. Buna göre;  $n_{11} = n_{21} = 0$  veya  $n_{12} = n_{22} = 0$  ise güven aralığı  $(0; 0)$ ,  $n_{11} = n_{22} = 0$  ise güven aralığı  $(-1; -1)$   $n_{12} = n_{21} = 0$  ise güven aralığı  $(1; 1)$  olur.

Süreklilik Düzeltmeli Wald yöntemi için güven aralığı<sup>8</sup>

$$\hat{\delta} \pm \left[ z_{\alpha/2} \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_{1+}} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_{2+}}} + \frac{1}{2} \left( \frac{1}{n_{1+}} + \frac{1}{n_{2+}} \right) \right] \quad (6)$$

Agresti-Caffo yöntemi için güven aralığı<sup>8,13</sup>

$$\check{p}_1 - \check{p}_2 \pm z_{\alpha/2} \sqrt{\frac{\check{p}_1(1-\check{p}_1)}{\check{n}_{1+}} + \frac{\check{p}_2(1-\check{p}_2)}{\check{n}_{2+}}} \quad (7)$$

$$\check{n}_{1+} = n_{1+} + 2$$

$$\check{n}_{2+} = n_{2+} + 2$$

$$\check{p}_1 = \frac{n_{11} + 1}{\check{n}_{1+}}$$

$$\check{p}_2 = \frac{n_{21} + 1}{\check{n}_{2+}}$$

Brown-Li yöntemi için güven aralığı<sup>14,15</sup>

$$\check{p}'_1 - \check{p}'_2 \pm z_{\alpha/2} \sqrt{\frac{\check{p}'_1(1-\check{p}'_1)}{n_{1+}} + \frac{\check{p}'_2(1-\check{p}'_2)}{n_{2+}}} \quad (8)$$

$$\check{p}'_1 = \frac{n_{11} + 0.5}{n_{1+} + 1}$$

$$\check{p}_2 = \frac{n_{21} + 0.5}{n_{2+} + 1}$$

Yule yöntemi için güven aralığı<sup>14</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\left(\frac{1}{n_{1+}} + \frac{1}{n_{2+}}\right) \bar{p}\bar{q}} \quad (9)$$

$$\bar{p} = \frac{n_{11} + n_{21}}{n_{1+} + n_{2+}}$$

$$\bar{q} = 1 - \bar{p}$$

Uyarlanmış Yule yöntemi için güven aralığı<sup>14</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\left(\frac{1}{n_{1+}} + \frac{1}{n_{2+}}\right) \check{p}\check{q}} \quad (10)$$

$$\check{p} = \frac{\frac{n_{2+}n_{11}}{n_{1+}} + \frac{n_{1+}n_{21}}{n_{2+}}}{n_{1+} + n_{2+}}$$

$$\check{q} = 1 - \check{p}$$

Newcombe Hibrid Skor yöntemi için güven aralığı<sup>8,11</sup>

$$\hat{p}_i \left( \frac{n_{i+}}{n_{i+} + z_{\alpha/2}^2} \right) + \frac{1}{2} \left( \frac{z_{\alpha/2}^2}{n_{i+} + z_{\alpha/2}^2} \right) \pm z_{\alpha/2}^2 \sqrt{\frac{1}{n_{i+} + z_{\alpha/2}^2} \left[ \hat{p}_i (1 - \hat{p}_i) \left( \frac{n_{i+}}{n_{i+} + z_{\alpha/2}^2} \right) + \frac{1}{4} \left( \frac{z_{\alpha/2}^2}{n_{i+} + z_{\alpha/2}^2} \right) \right]} \quad i = 1, 2$$

$l_1: i = 1$  için alt smür,

$l_2: i = 2$  için alt smür,

$u_1: i = 1$  için üst smür,

$u_2: i = 2$  için üst smür,

$$\begin{cases} \text{Alt Smür} = \hat{\delta} - \sqrt{(\hat{p}_1 - l_1)^2 + (u_2 - \hat{p}_{12})^2} \\ \text{Üst Smür} = \hat{\delta} + \sqrt{(\hat{p}_2 - l_2)^2 + (u_1 - \hat{p}_1)^2} \end{cases} \quad (11)$$

Anbar yöntemi için güven aralığı<sup>16</sup>

$$\bar{\delta} = \frac{\hat{\delta}}{1 + c(\alpha)/n_{2+}} - \frac{\left(\hat{p}_1 - \frac{1}{2}\right) c(\alpha)/n_{2+}}{1 + c(\alpha)/n_{2+}}$$

$$L = \sqrt{\frac{c(\alpha)/n_{2+}}{[1 + c(\alpha)/n_{2+}]^2} \left[ (1 + c(\alpha)/n_{2+}) \left\{ \frac{1}{4} \left( 1 + \frac{n_{2+}}{n_{1+}} \right) - \frac{n_{2+}}{n_{1+}} \left( \hat{p}_1 - \frac{1}{2} \right)^2 \right\} - \left( \frac{1}{2} - \hat{p}_2 \right)^2 \right]}$$

$$\begin{cases} \text{Alt Smir} = \bar{\delta} - L \\ \text{Üst Smir} = \bar{\delta} + L \end{cases} \quad (12)$$

Beal-Haldane yöntemi için güven aralığı<sup>11,17</sup>

$$\theta^* = \frac{\hat{\theta} + z_{\alpha/2}^2 v(1 - \varphi)}{1 + z_{\alpha/2}^2 u}$$

$$w = \frac{z}{1 + z_{\alpha/2}^2 u} \sqrt{[u\{4\varphi(1 - \varphi) - \hat{\theta}^2\} + 2v(1 - 2\varphi)\hat{\theta} + 4z_{\alpha/2}^2 u^2(1 - \varphi)\varphi + z_{\alpha/2}^2 v^2(1 - 2\varphi)^2]}$$

$$\varphi = \frac{1}{2} \left( \frac{n_{11}}{n_{1+}} + \frac{n_{21}}{n_{2+}} \right)$$

$$\hat{\theta} = \left( \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}} \right)$$

$$u = \frac{1}{4} \left( \frac{1}{n_{1+}} + \frac{1}{n_{2+}} \right)$$

$$v = \frac{1}{4} \left( \frac{1}{n_{1+}} - \frac{1}{n_{2+}} \right)$$

$$\begin{cases} \text{Alt Smir} = \theta^* - w \\ \text{Üst Smir} = \theta^* + w \end{cases} \quad (13)$$

Beal-Jeffreys-Perks yöntemi için güven aralığı<sup>11,17</sup>

Beal-Haldane güven aralığı ile benzerdir. Ancak,  $\varphi$ 'nin değeri

$$\varphi = \frac{1}{2} \left( \frac{n_{11} + 0.5}{n_{1+} + 1} + \frac{n_{21} + 0.5}{n_{2+} + 1} \right)$$

olarak hesaplanmaktadır.



Mee yöntemi için güven aralığı<sup>15,18,19</sup>

$$\hat{\delta} \pm z_{\alpha/2} \sqrt{\lambda \left( \frac{\tilde{p}_1(1-\tilde{p}_1)}{n_{1+}} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_{2+}} \right)} \quad (14)$$

$$\lambda = 1$$

$$\tilde{p}_1 = 2p \cos(a) - \frac{L_2}{3L_3}$$

$$\tilde{p}_2 = \tilde{p}_1 + \hat{\delta}$$

$$a = \frac{\pi + \cos^{-1}(q/p^3)}{3}$$

$$q = \frac{L_2^3}{(3L_3)^3} - \frac{L_1 L_2}{6L_3^2} + \frac{L_0}{2L_3}$$

$$p = \operatorname{sgn}(q) \sqrt{\frac{L_2^2}{(3L_3)^2} - \frac{L_1}{3L_3}}$$

$$L_3 = n$$

$$L_2 = \hat{\delta}(n_{1+} + 2n_{2+}) - n - (n_{11} + n_{21})$$

$$L_1 = \hat{\delta}(\hat{\delta}n_{2+} - n - 2n_{21}) + (n_{11} + n_{21})$$

$$L_0 = \hat{\delta}n_{21}(1 - \hat{\delta})$$

Miettinen-Nurminen yöntemi için güven aralığı<sup>20</sup>

Mee güven aralığı ile benzerdir. Ancak,  $\lambda$ 'nın değeri,

$$\lambda = \frac{n}{n-1}$$

olarak hesaplanmalıdır.

## BULGULAR

Çalışmada dikkate alınan yöntemlere ait her bir örneklem büyüklüğü için elde edilen oranlar arası farklar Tablo 3'te verilmiştir. Çalışmada dikkate alınan yöntemlerin Tablo 2 ve Tablo 3'te farklı gruplarda yer aldığı görülmektedir. Bu farklılık, güven aralığı hesaplamalarında 0/0 belirsizliğine sahip durum sayısının yöntemden yöntem farklılık göstermesi ve Tablo 3'ün 2. sütununda yer alan yöntemlerin tamamında aynı sonuçların elde edilmesinden kaynaklanmaktadır.

**TABLO 3:** Oranlar arası farklara ait ortalama ve standart sapma değerleri.

n	Yöntem grubu I	Yöntem grubu II	Yöntem grubu III
10	0,0000±0,33068	0,0000±0,51399	0,0000±0,39047
15	-0,0005±0,35023	0,0000±0,48375	-0,0006±0,39842
20	0,0046±0,35323	0,0046±0,45726	0,0051±0,39262
25	0,0113±0,36592	0,0131±0,45288	0,0122±0,39954
30	-0,0221±0,36873	-0,0282±0,43910	-0,0239±0,39714
35	-0,0069±0,37851	-0,0088±0,44446	-0,0078±0,40560
40	0,0016±0,37443	0,0019±0,42984	0,0018±0,39757
45	0,0009±0,38199	0,0026±0,42963	0,0011±0,40283
50	-0,0069±0,39480	-0,0058±0,44097	-0,0069±0,41482
55	-0,0024±0,40105	-0,0017±0,44441	-0,0022±0,41965
60	0,0054±0,38380	0,0043±0,42011	0,0046±0,40014
65	-0,0089±0,38394	-0,0084±0,41935	-0,0086±0,39958
70	-0,0204±0,39420	-0,0219±0,42781	-0,0212±0,40885
75	0,0053±0,40207	0,0055±0,43491	0,0052±0,41655
80	-0,0075±0,39613	-0,0087±0,42447	-0,0077±0,40887
85	-0,0025±0,39352	-0,0014±0,42194	-0,0022±0,40611
90	0,0041±0,41082	0,0032±0,43919	0,0037±0,42365
95	-0,0020±0,39971	-0,0034±0,42439	-0,0027±0,41112
100	-0,0182±0,40547	-0,0198±0,42957	-0,0187±0,41626
125	0,0009±0,40160	-0,0002±0,42057	0,0004±0,41045
150	-0,0014±0,40133	-0,0007±0,41915	-0,0013±0,40952
175	-0,0181±0,40413	-0,0194±0,41771	-0,0188±0,41058
200	0,0203±0,39948	0,0201±0,41201	0,0202±0,40551
225	-0,0176±0,41089	-0,0178±0,42112	-0,0175±0,41592
250	-0,0115±0,40468	-0,0117±0,41466	-0,0117±0,40932
300	-0,0128±0,41325	-0,0133±0,42144	-0,0129±0,41725
350	-0,0125±0,41299	-0,0126±0,41977	-0,0125±0,41623
400	-0,0081±0,40850	-0,0084±0,41507	-0,0083±0,41160
450	0,0208±0,40584	0,0208±0,41110	0,0207±0,40839
500	-0,0088±0,39144	-0,0086±0,39620	-0,0089±0,39382
600	0,0118±0,41194	0,0115±0,41607	0,0116±0,41413
700	-0,0006±0,40998	-0,0007±0,41386	-0,0006±0,41189
800	0,0168±0,40762	0,0170±0,41085	0,0169±0,40906
900	0,0035±0,39164	0,0034±0,39428	0,0034±0,39298
1.000	-0,0141±0,40740	-0,0140±0,40999	-0,0140±0,40877
Genel	-0,0028±0,39478	-0,0032±0,42643	-0,0030±0,40772

Yöntem Grubu I: Agresti-Caffo;

Yöntem Grubu II: Anbar, Basit Asimptotik, Süreklilik Düzeltmeli Basit Asimptotik, Wald, Süreklilik Düzeltmeli Wald, Yule, Uyarlanmış Yule, Newcombe Hibrid Skor, Beal-Haldane, Mee, Miettinen-Nurminen, Beal-Jeffreys-Perks, Hauck-Anderson;

Yöntem Grubu III: Brown-Li.

Tablo 3'ten de görüldüğü üzere, oranlar arası en küçük farklar Agresti-Caffo yönteminden, en büyük farklar ise Brown-Li yönteminden elde edilmiştir. Çalışmada dikkate alınan yöntemlere ait her bir örneklem büyüklüğü için elde edilen güven aralıklarının en küçük ve en büyük değerleri Tablo 4'te verilmiştir.

**TABLO 4a:** Güven aralıklarına ait alt/üst sınır değerleri.

n	Agresti-Caffo yöntemi	Anbar yöntemi	Newcombe yöntemi	Wald yöntemi	Süreklilik Düzeltmeli Wald yöntemi
10	-1,14/1,14	-1,24/1,24	-1,41/1,41	-1,20/1,20	-1,65/1,65
15	-1,15/1,15	-1,24/1,24	-1,39/1,39	-1,20/1,20	-1,60/1,60
20	-1,15/1,16	-1,24/1,19	-1,36/1,39	-1,20/1,20	-1,57/1,57
25	-1,17/1,17	-1,24/1,12	-1,39/1,39	-1,19/1,19	-1,53/1,53
30	-1,12/1,15	-1,12/1,13	-1,32/1,32	-1,17/1,14	-1,52/1,52
35	-1,18/1,18	-1,13/1,24	-1,38/1,38	-1,17/1,12	-1,51/1,51
40	-1,13/1,15	-1,10/1,14	-1,30/1,29	-1,17/1,14	-1,43/1,49
45	-1,17/1,05	-1,01/1,09	-1,27/1,23	-1,13/1,13	-1,32/1,37
50	-1,16/1,18	-1,10/1,12	-1,27/1,38	-1,11/1,10	-1,37/1,51
55	-1,10/1,17	-1,23/1,08	-1,31/1,37	-1,18/1,15	-1,49/1,53
60	-1,14/1,14	-1,19/1,06	-1,34/1,34	-1,14/1,09	-1,51/1,51
65	-1,09/1,12	-1,06/1,07	-1,23/1,26	-1,05/1,15	-1,42/1,42
70	-1,10/1,12	-1,04/1,08	-1,25/1,28	-1,12/1,17	-1,39/1,36
75	-1,14/1,09	-1,12/1,19	-1,34/1,28	-1,13/1,14	-1,50/1,46
80	-1,09/1,09	-1,07/1,14	-1,30/1,21	-1,19/1,11	-1,47/1,29
85	-1,09/1,07	-1,05/1,00	-1,28/1,21	-1,18/1,11	-1,44/1,28
90	-1,18/1,11	-1,06/1,23	-1,37/1,25	-1,03/1,15	-1,52/1,40
95	-1,02/1,04	-1,03/1,01	-1,16/1,18	-1,05/1,03	-1,19/1,20
100	-1,05/1,12	-1,01/1,02	-1,26/1,17	-1,17/1,02	-1,34/1,14
125	-1,05/1,02	-1,02/1,01	-1,19/1,20	-1,01/1,11	-1,17/1,20
150	-1,00/1,07	-1,05/0,99	-1,13/1,23	-1,04/1,14	-1,19/1,31
175	-1,07/1,01	-1,00/1,00	-1,20/1,16	-1,09/1,02	-1,26/1,09
200	-1,03/1,00	-1,00/1,01	-1,18/1,06	-1,01/1,01	-1,16/1,02
225	-1,04/0,99	-0,98/0,99	-1,20/1,05	-1,11/1,00	-1,18/1,02
250	-0,99/1,01	-1,00/0,98	-1,19/1,16	-1,10/1,01	-1,26/1,32
300	-1,00/1,00	-0,99/0,98	-1,06/1,06	-1,02/1,01	-1,04/1,03
350	-1,01/0,98	-0,97/1,00	-1,12/1,02	-1,00/0,98	-1,06/0,99
400	-1,01/0,99	-0,99/1,00	-1,06/1,03	-1,02/1,00	-1,04/1,01
450	-1,00/0,99	-0,97/1,00	-1,09/1,02	-1,02/0,99	-1,04/1,00
500	-1,00/0,99	-0,98/0,99	-1,04/1,05	-1,00/1,00	-1,01/1,02
600	-0,98/1,04	-1,00/0,97	-1,04/1,08	-0,98/1,02	-0,99/1,03
700	-1,01/0,96	-0,95/0,98	-1,06/0,99	-1,02/0,96	-1,03/0,97
800	-1,08/0,97	-0,96/1,00	-1,15/0,99	-1,00/0,97	-1,06/0,98
900	-0,95/0,97	-0,96/0,95	-1,01/0,99	-0,96/0,97	-0,96/0,98
1.000	-0,99/0,91	-0,91/0,99	-1,00/1,10	-0,99/1,00	-1,00/1,10
Genel	-1,18/1,18	-1,24/1,24	-1,41/1,41	-1,20/1,20	-1,65/1,65

**TABLO 4b:** Güven aralıklarına ait alt/üst sınır değerleri (devamı).

n	Basit Asimptotik yöntem	Süreklilik Düzeltmeli Basit Asimptotik yöntem	Beal-Haldane yöntemi	Beal-Jeffreys-Perks yöntemi	Brown-Li yöntemi
10	-1,20/1,20	-1,65/1,65	-1,21/1,21	-1,31/1,31	-1,56/1,56
15	-1,20/1,20	-1,60/1,60	-1,22/1,22	-1,34/1,34	-1,57/1,57
20	-1,20/1,20	-1,57/1,57	-1,19/1,23	-1,32/1,36	-1,53/1,58
25	-1,19/1,19	-1,53/1,53	-1,14/1,14	-1,28/1,28	-1,58/1,58
30	-1,17/1,14	-1,52/1,52	-1,13/1,16	-1,28/1,30	-1,49/1,49
35	-1,17/1,12	-1,51/1,51	-1,23/1,23	-1,38/1,38	-1,59/1,59
40	-1,17/1,14	-1,43/1,49	-1,13/1,14	-1,23/1,29	-1,32/1,47
45	-1,13/1,13	-1,32/1,37	-1,10/1,06	-1,20/1,07	-1,28/1,22
50	-1,11/1,10	-1,37/1,51	-1,11/1,24	-1,20/1,39	-1,34/1,59
55	-1,18/1,15	-1,49/1,53	-1,07/1,20	-1,13/1,35	-1,50/1,57
60	-1,14/1,09	-1,51/1,51	-1,05/1,04	-1,08/1,05	-1,54/1,54
65	-1,05/1,15	-1,42/1,42	-1,06/1,10	-1,10/1,18	-1,41/1,41
70	-1,12/1,17	-1,39/1,36	-1,08/1,12	-1,16/1,23	-1,38/1,35
75	-1,13/1,14	-1,50/1,46	-1,20/1,11	-1,35/1,27	-1,54/1,48
80	-1,19/1,11	-1,47/1,29	-1,17/1,06	-1,32/1,07	-1,50/1,13
85	-1,18/1,11	-1,44/1,28	-1,13/1,04	-1,13/1,11	-1,21/1,18
90	-1,03/1,15	-1,52/1,40	-1,23/1,10	-1,39/1,15	-1,58/1,19
95	-1,05/1,03	-1,19/1,20	-1,00/1,03	-1,09/1,14	-1,19/1,19
100	-1,17/1,02	-1,34/1,14	-1,10/1,01	-1,11/1,01	-1,14 / 1,16
125	-1,01/1,11	-1,17/1,20	-1,02/1,01	-1,11/1,02	-1,20/1,09
150	-1,04/1,14	-1,19/1,31	-0,99/1,00	-1,00/1,00	-1,02/1,17
175	-1,09/1,02	-1,26/1,09	-1,02/0,99	-1,03/0,99	-1,17/1,01
200	-1,01/1,01	-1,16/1,02	-1,01/1,00	-1,02/1,01	-1,20/1,02
225	-1,11/1,00	-1,18/1,02	-1,00/0,98	-1,00/0,99	-1,09/0,99
250	-1,10/1,01	-1,26/1,32	-1,04/1,07	-1,05/1,23	-1,07/1,37
300	-1,02/1,01	-1,04/1,03	-0,98/1,00	-0,98/1,00	-1,01/1,01
350	-1,00/0,98	-1,06/0,99	-0,99/0,98	-1,00/0,98	-1,01/0,98
400	-1,02/1,00	-1,04/1,01	-1,01/0,99	-1,02/0,99	-1,02/1,00
450	-1,02/0,99	-1,04/1,00	-1,00/0,99	-1,00/0,99	-1,01/0,99
500	-1,00/1,00	-1,01/1,02	-0,99/0,99	-1,00/0,99	-1,00/0,99
600	-0,98/1,02	-0,99/1,03	-0,98/1,02	-0,98/1,04	-0,98/1,05
700	-1,02/0,96	-1,03/0,97	-1,01/0,96	-1,01/0,96	-1,01/0,96
800	-1,00/0,97	-1,06/0,98	-0,95/0,97	-0,99/0,97	-1,10/0,97
900	-0,96/0,97	-0,96/0,98	-0,95/0,97	-0,95/0,97	-0,96/0,97
1.000	-0,99/1,00	-1,00/1,10	-0,99/0,91	-0,99/0,91	-0,99/0,98
Genel	-1,20/1,20	-1,65/1,65	-1,23/1,24	-1,39/1,39	-1,59/1,59

**TABLO 4c:** Güven aralıklarına ait alt/üst sınır değerleri (devamı).

n	Hauck-Anderson yöntemi	Mee yöntemi	Miettinen-Nurminen yöntemi	Yule yöntemi	Uyarlanmış Yule yöntemi
10	-1,73/1,73	-1,59/1,59	-1,64/1,64	-1,72/1,72	-1,62/1,62
15	-1,73/1,73	-1,60/1,60	-1,63/1,63	-1,68/1,68	-1,51/1,51
20	-1,68/1,73	-1,60/1,60	-1,62/1,62	-1,66/1,66	-1,44/1,44
25	-1,73/1,73	-1,60/1,60	-1,62/1,62	-1,65/1,65	-1,39/1,39
30	-1,49/1,63	-1,60/1,57	-1,61/1,58	-1,64/1,62	-1,36/1,36
35	-1,59/1,59	-1,60/1,57	-1,61/1,58	-1,63/1,62	-1,33/1,33
40	-1,60/1,68	-1,60/1,59	-1,61/1,61	-1,63/1,63	-1,31/1,31
45	-1,60/1,64	-1,59/1,57	-1,60/1,58	-1,61/1,58	-1,26/1,29
50	-1,41/1,39	-1,60/1,54	-1,61/1,55	-1,62/1,58	-1,28/1,28
55	-1,68/1,43	-1,59/1,56	-1,59/1,57	-1,61/1,60	-1,26/1,26
60	-1,51/1,55	-1,38/1,38	-1,38/1,38	-1,44/1,44	-1,25/1,25
65	-1,30/1,53	-1,58/1,58	-1,58/1,58	-1,60/1,60	-1,24/1,24
70	-1,23/1,43	-1,60/1,60	-1,60/1,60	-1,61/1,62	-1,22/1,23
75	-1,58/1,65	-1,39/1,57	-1,39/1,58	-1,44/1,59	-1,23/1,23
80	-1,31/1,57	-1,49/1,17	-1,49/1,17	-1,52/1,22	-1,22/1,22
85	-1,47/1,39	-1,35/1,53	-1,36/1,54	-1,38/1,54	-1,19/1,20
90	-1,20/1,69	-1,25/1,21	-1,26/1,21	-1,29/1,27	-1,21/1,19
95	-1,14/1,32	-1,55/1,35	-1,56/1,35	-1,56/1,38	-1,20/1,20
100	-1,41/1,21	-1,33/1,35	-1,33/1,36	-1,34/1,38	-1,20/1,20
125	-1,17/1,22	-1,34/1,32	-1,35/1,32	-1,37/1,33	-1,15/1,18
150	-1,47/1,43	-1,16/1,24	-1,16/1,24	-1,18/1,27	-1,13/1,16
175	-1,38/1,14	-1,23/1,07	-1,23/1,07	-1,26/1,14	-1,12/1,13
200	-1,15/1,02	-1,34/1,38	-1,34/1,38	-1,36/1,38	-1,14/1,13
225	-1,20/1,02	-1,14/1,03	-1,14/1,03	-1,14/1,09	-1,13/1,09
250	-1,22/1,03	-1,03/1,60	-1,04/1,60	-1,10/1,60	-1,10/1,11
300	-1,04/1,03	-1,14/1,02	-1,14/1,02	-1,15/1,08	-1,08/1,08
350	-1,13/0,99	-0,99/1,02	-0,99/1,02	-1,10/1,08	-1,10/1,06
400	-1,02/1,00	-1,31/1,01	-1,31/1,01	-1,31/1,08	-1,08/1,08
450	-1,04/1,00	-1,00/0,98	-1,00/0,98	-1,09/1,05	-1,09/1,05
500	-1,00/1,02	-1,03/1,00	-1,03/1,00	-1,08/1,05	-1,06/1,06
600	-0,98/1,03	-1,02/1,02	-1,02/1,02	-1,04/1,08	-1,04/1,08
700	-1,02/0,97	-0,99/1,20	-0,99/1,20	-1,03/1,20	-1,05/1,01
800	-1,06/0,97	-1,02/0,99	-1,02/0,99	-1,07/1,02	-1,07/1,02
900	-0,96/0,97	-1,04/0,98	-1,04/0,98	-1,05/1,02	-1,00/1,02
1.000	-1,00/1,14	-0,99/0,96	-0,99/0,96	-1,04/0,97	-1,04/1,00
Genel	-1,73/1,73	-1,60/1,60	-1,64/1,64	-1,72/1,72	-1,62/1,62

Tablo 4'ten de görüldüğü üzere, güven aralığının büyüklüğü bakımından yöntemler sıralanacak olursa en dar aralıktan en geniş aralığa doğru sıralama; Agresti-Caffo  $\leq$  Wald = Basit Asimptotik  $\leq$  Beal-Haldane  $\leq$  Anbar  $\leq$  Beal-Jeffreys-Perks  $\leq$  Newcombe Hibrid Skor  $\leq$  Brown-Li  $\leq$  Mee  $\leq$  Uyarlanmış Yule  $\leq$  Miettinen-Nurminen  $\leq$  Süreklilik Düzeltmeli Wald = Süreklilik Düzeltmeli Basit Asimptotik  $\leq$  Yule  $\leq$  Hauck-Anderson şeklindedir. Beal-Haldane yöntemine göre oranlar arası farkların %98,8'i, Beal-Jeffreys-Perks yöntemine göre %99,6'sı, diğer yöntemlerde ise oranlar arası farkların tamamı alt ve üst sınırlar arasında kalmaktadır. Tüm yöntemler için örnek büyüklüğü arttıkça güven aralıklarının daraldığı belirlenmiştir. Güven aralığının simetrikliği bakımından yöntemlere bakıldığında, hemen tüm yöntemlerden elde edilen güven aralıklarının simetrik olduğu ve bunun tüm örneklem büyüklükleri için geçerli olduğu söylenebilir.

## TARTIŞMA

Eşit etkinlik deneyleri için hipotez, tasarım ve yorum üstün olma deneylerinden farklıdır. Eşit etkinlik deneyleri, yeni tedavinin bazı yönlerden standart tedaviye göre avantajlar sunduğu fikriyle tasarlanmaktadır. Yeni tedavinin standart tedavi ile benzer veya standart tedaviden daha kötü olmadığını kabul edilmesi için tolerans olarak kullanılacak fark ( $\Delta$ ) değerinin deneyden önce belirlenmesi gerekir.  $\Delta$ 'nın değerinin belirlenmesinde, hem göreceli fark hem de mutlak fark yöntemleri kullanılabilir. Eşit etkinlik ve üstün olma için sıklıkla ardışık testler kullanılmaktadır. Plaseboya karşı herhangi bir tedaviyi test etmenin etik olmadığı durumlarda eşit etkinlik deneyleri gerekli olabilir. Eşit etkinlik deneylerinde doğru olmayabilecek ve test edilmeyen doğal varsayımlar vardır. Ardışık eşit etkinlik deneyleri, yeni tedavilerin eşit etkin olmadığı gösterilmiş olsa bile daha az etkili tedaviler sunabilir. Ayrıca düşük kaliteli deneyler eşit etkin sonuçları desteklemektedir. Klinisyenler,  $p$  değerleri yerine etki büyüklüğü ve güven sınırları üzerinde odaklanmalıdır.<sup>21</sup> Bir çalışmanın amacı bir tedavinin kontrolden daha iyi olduğunu göstermekse tek yanlı güven aralıkları ilgi çekicidir. Pratikte, mevcut aralıkların çoğu yaklaşık değerlerdir. Yani aralığın 2 oranını içermesi olasılığı, aralık onu hedeflese de her zaman en azından nominal düzey değildir. Güven aralıklarının çoğu, nominal seviyeden çok daha düşük asimptotik infimum kapsama olasılığına sahip yaklaşık aralıklardır. Ayrıca büyük örneklem büyüklüğünde bile, infimum kapsama olasılığı hâlâ nominal seviyeden çok daha az olabilir ve bu miktara yakınsamamaktadır. Büyük örneklem üzerinde çalışmak uygulamada maliyetli olabilir. Bu nedenle kesin optimal güven aralıkları, doğruluk ve kesinlik ile geçerli istatistiksel çıkarımlar yapmak için kritik hâle gelir. Aslında Wald yöntemine dayalı güven aralığı, asimptotik normalite varsayımına sahip olmasına rağmen herhangi bir örneklem büyüklüğü ve herhangi bir nominal  $1 - \alpha$  seviyesi için 0 olan bir infimum kapsama olasılığına sahiptir. Bu nedenle doğru kapsamı garanti edemedikleri için bu tür yaklaşık aralıklar kullanıldığında örneklem büyüklüğü sorgulanabilir. En az  $1 - \alpha$ 'lık bir infimum kapsama olasılığını garanti eden tam aralıklarda bu sorun yoktur.<sup>22</sup> Kapsam olasılıkları bakımından güven aralığı yöntemleri karşılaştırılmış, Wald yönteminin Agresti-Caffo, Newcombe Hibrid Skor yöntemi ve Miettinen-Nurminen yöntemlerine kıyasla daha düşük performans gösterdiği belirlenmiştir. Wald ve Süreklilik Düzeltmeli Wald yöntemlerinden elde edilen güven aralıklarının, nominal kapsama olasılıkları düşük olmasına rağmen kullanılmaya devam etmektedir. Newcombe'un Hibrid Skor yöntemi ve Agresti-Caffo yöntemleri, deney ve kontrol grubuna ait örneklem büyüklüklerinin eşit olup olmamasına bağlı olmaksızın, nominal yakın kapsama olasılıklarına sahiptir.<sup>23</sup> Wald yöntemi, deney grubunun örneklem büyüklüğünün kontrol grubunun örneklem büyüklüğünden göreceli olarak daha fazla olması durumunda nominal düzeyde Tip I hatanın kontrolü açısından uygun performans gösterir. Ancak Agresti-Caffo yöntemi bu durumlarda daha yüksek güce sahiptir. Newcombe Hibrid Skor yöntemi, hemen tüm örneklem büyüklüklerinde Tip I hata kontrolü sağlar, ancak bu genel kontrol gücü kaybına neden olur. Bu sonuçlar, uygun yöntem seçiminin örneklem büyüklüğü dağılımına bağlı olduğunu göstermektedir. Tüm örneklem büyüklüğü senaryolarında yeterli gücü ve Tip I hatanın uygun kontrolünü sağlayan yöntemlere karar vermek önemlidir.<sup>7</sup>

## SONUÇ

Güven aralıkları, klinik araştırma sonuçlarının yorumlanmasında son derece önemlidir. Çalışmadan elde edilen sonuçlara göre güven aralıklarının genişliği dikkate alındığında  $n \leq 30$  için Agresti-Caffo yönteminin,  $30 \leq n \leq 90$  için Anbar ya da Wald yönteminin,  $n \geq 95$  için ise Anbar yönteminin en dar güven aralıklarına sahip olduğu belirlenmiştir. Oranlar arası fark değerini kapsama bakımından yöntemler karşılaştırıldığında, Beal-Haldane ve Beal-Jeffreys-Perks yöntemleri dışındaki tüm yöntemlerde farkların tamamının alt ve üst sınırlar arasında kaldığı görülmüştür. Dolayısıyla bu 2 yöntem diğer yöntemlere göre düşük performans göstermiştir. Bu çalışmada dikkate alınan yöntemler, güven aralıklarının genişliği ve güven aralıklarının farkları kapsama kriterleri dikkate alınarak karşılaştırılmıştır. Uygun yöntem karar vermek için bu kriterler ile birlikte güç ve Tip I hata kriterlerinin de dikkate alındığı çalışmaların yapılması gerekmektedir.

### 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

- Holmberg MJ, Andersen LW. Estimating risk ratios and risk differences: alternatives to odds ratios. JAMA. 2020;324(11):1098-9. PMID: 32930746
- Agresti A. Dealing with discreteness: making 'exact' confidence intervals for proportions, differences of proportions, and odds ratios more exact. Stat Methods Med Res. 2003;12(1):3-21. PMID: 12617505
- Zou G, Donner A. A simple alternative confidence interval for the difference between two proportions. Control Clin Trials. 2004;25(1):3-12. PMID: 14980745
- Ni S, Yu Q, Zhong Z, Yang M, Zhao Y, Wu J, et al. Risk difference, relative risk, and odds ratio for non-inferiority clinical trials with risk rate endpoint. J Biopharm Stat. 2022;1-16. PMID: 35791856
- Tunes da Silva G, Logan BR, Klein JP. Methods for equivalence and noninferiority testing. Biol Blood Marrow Transplant. 2009;15(1 Suppl):120-7. PMID: 19147090; PMCID: PMC2701110
- Hahn S. Understanding noninferiority trials. Korean J Pediatr. 2012;55(11):403-7. PMID: 23227058; PMCID: PMC3510268
- Dann RS, Koch GG. Methods for one-sided testing of the difference between proportions and sample size considerations related to non-inferiority clinical trials. Pharm Stat. 2008;7(2):130-41. PMID: 17526060
- Fagerland MW, Lydersen S, Laake P. Recommended confidence intervals for two independent binomial proportions. Stat Methods Med Res. 2015;24(2):224-54. PMID: 21996567
- Oczkowski SJ. A clinician's guide to the assessment and interpretation of noninferiority trials for novel therapies. Open Med. 2014;8(2):e67-72. PMID: 25009686; PMCID: PMC4085087
- Bai AD, Komorowski AS, Lo CKL, Tandon P, Li XX, Mokashi V, et al; McMaster Infectious Diseases Fellow Research Group. Confidence interval of risk difference by different statistical methods and its impact on the study conclusion in antibiotic non-inferiority trials. Trials. 2021;22(1):708. PMID: 34656155; PMCID: PMC8520289
- Newcombe RG. Interval estimation for the difference between independent proportions: comparison of eleven methods. Stat Med. 1998;17(8):873-90. PMID: 9595617
- Hauck WW, Anderson S. A comparison of large-sample confidence interval methods for the difference of two binomial probabilities. Am Stat. 1986;40:318-22. doi: 10.2307/2684618
- Agresti A, Caffo B. Simple and effective confidence intervals for proportions and differences of proportions result from adding two successes and two failures. Am Stat. 2000;54(4):280-8. doi: https://users.stat.ufl.edu/~aa/articles/agresti\_caffo\_2000.pdf
- Brown L, Li X. Confidence intervals for two sample binomial distribution. J Stat Plan Inference. 2005;130(1-2):359-75. doi: 10.1016/j.jspi.2003.09.039

15. Laud PJ, Dane A. Confidence intervals for the difference between independent binomial proportions: comparison using a graphical approach and moving averages. *Pharm Stat.* 2014;13(5):294-308. PMID: 25163425
16. Anbar D. On estimating the difference between two probabilities, with special reference to clinical trials. *Biometrics.* 1983;39(1):257-62. PMID: 6871355
17. Beal SL. Asymptotic confidence intervals for the difference between two binomial parameters for use with small samples. *Biometrics.* 1987;43(4):941-50. PMID: 3427177
18. Mee RW. Confidence bounds for the difference between two probabilities. *Biometrics.* 1984;40(4):1175-6. <https://www.jstor.org/stable/2531174>
19. Farrington CP, Manning G. Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Stat Med.* 1990;9(12):1447-54. PMID: 2281232
20. Miettinen O, Nurminen M. Comparative analysis of two rates. *Stat Med.* 1985;4(2):213-26. PMID: 4023479
21. Leung JT, Barnes SL, Lo ST, Leung DY. Non-inferiority trials in cardiology: what clinicians need to know. *Heart.* 2020;106(2):99-104. PMID: 31672779; PMCID: PMC6993027
22. Shan G, Wang W. ExactCldiff: An R Package for computing exact confidence intervals for the difference of two proportions. *The R Journal.* 2013;5(2):62-70. doi: 10.32614/RJ-2013-026
23. Reed JF. Improved confidence intervals for the difference between two proportions. *J Mod Appl Stat Methods.* 2009;8(1):208-14. doi: 10.22237/jmasm/1241137020