



Neyin Peşindeyiz? Kutsal p Değerinin mi (İstatistiksel Önemlilik) Yoksa Klinik Önemliliğinin mi?

Selim Kılıç¹

ÖZET:

Neyin peşindeyiz? Kutsal p değerinin mi (istatistiksel önemlilik) yoksa klinik önemliliğinin mi?

Makaleler okunurken okuyucuların çoğunluğu makalede verilen p değerlerine bakarak sonuçlar ve önemlilik konusunda karar verirler. Oysa ki istatistiksel önemliliği gösteren p değeri, klinik önemliliği desteklediğinde önemlidir. Bu yazıda örnekler üzerinde klinik ve istatistiksel önemlilik tartışılmıştır.

Anahtar sözcükler: İstatistiksel önemlilik, klinik önemlilik

ABSTRACT:

What do we be after? Holy p value (statistical significance) or clinical significance?

Most of the readers decide on results and significance according to the p values which are given in the manuscript when they are reading. In fact, the p value which shows the statistical significance is important if it supports the clinical significance. In this article clinical and statistical significance are discussed on examples.

Key words: Statistical significance, clinical significance

Journal of Mood Disorders 2011;1:46-48

¹MD, Gülhane Askeri Tıp Fakültesi, Epidemiyoloji BD, Ankara-Turkey

Yazışma Adresi / Address reprint requests to: Selim Kılıç, Gülhane Askeri Tıp Fakültesi, Epidemiyoloji BD, Ankara-Turkey

Telefon / Phone: +90-312-304-2000

Elektronik posta adresi / E-mail address: drselimkili@gmail.com

Kabul tarihi / Date of acceptance: 10 Mart 2011 / March 10, 2011

Bağıntı beyanı:

S.K.: Yazar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemiştir.

Declaration of interest:

S.K.: The author reported no conflict of interest related to this article.

Yapılan çalışmalarda bir makale okurken en çok okunan kısmının başlık ve sonra da özet kısmı olduğu belirtilmiştir. Çoğu okuyucu makale ilgili kararını makalenin tamamını okumadan özet kısmında ya da bulgularda verilen p değer(ler)ine bakarak vermektedir. $p > 0.05$, istatistiksel olarak önemli değil. Karar: Çalışmada elde edilen

sonuç önemli değildir, makalenin tamamını okumaya değmez. Acaba böyle bir yaklaşım doğru mudur? Gelin aşağıdaki örnek üzerinden tartışalım.

Bu örnekte etken (E) olarak tanımladığımız steroid tedavisi görmek olsun. Olgular dediklerimiz ise iki uçlu bozukluk gelişenler olsun. Bu durumda; etken (+) olanlar

SONUÇ SETİ 1

	Olgular	Sağlamalar	Toplam	χ^2 , RR, p ve %95 GA değerleri
E (+)	9	51	60	$\chi^2 = 1.3$, $p = 0.13$ RR = 1.8
E (-)	5	55	60	
Toplam	14	106	120	%95 GA = [0.6 - 4.9]

SONUÇ SETİ 2

	Olgular	Sağlamalar	Toplam	χ^2 , RR, p ve %95 GA değerleri
E (+)	90	510	600	$\chi^2 = 12$, $p = 0.0002$ RR = 1.8
E (-)	50	550	600	
Toplam	140	1060	1200	%95 GA = [1.3- 2.5]

SONUÇ SETİ 3

	Olgular	Sağlamalar	Toplam	χ^2 , RR, p ve %95 GA değerleri
E (+)	600	1400	2000	$\chi^2 = 12$, $p = 0.0002$ RR = 1.2
E (-)	500	1500	2000	
Toplam	1100	2900	4000	%95 GA = [1.1- 1.3]

steroid tedavisi alan kişiler, etken (-) olanlar steroid tedavisi almayanlar, olgular iki uçlu bozukluk gelişenler, sağlamlar ise iki uçlu bozukluk gelişmeyen kişilerdir. Farklı durumları değerlendireceğimiz üç çalışmanın verileri ve p değerleri aşağıda üç ayrı SONUÇ SETİ olarak tablolatılmıştır.

İlk sonuç setine göre; steroid tedavisi alan ve almayan 60 kişilik 2 ayrı grubun izlendiği bir kohort çalışmasıdır. Buna göre izlemin sonunda steroid tedavisi alanlarda 9, almayanlarda ise 5 iki uçlu bozukluk olgusu gelişmiştir. Paydalar her ikisinde de eşit (60) olduğundan, payların yani 9 ve 5'in bölümünden elde ettiğimiz 1.8 değeri bize her 2 gruptaki insidans hızlarının bölümünden elde ettiğimiz rölatif riski (risk oranı=risk ratio) verir. Bu sonuçlara göre yorumumuz: Steroid tedavisi alanlarda almayanlara göre iki uçlu bozukluk gelişimi 1.8 kat yüksektir. Peki ilk sonuç setinde verilen p değeri yani 0.13 nedir? p değeri her 2 gruba ait oranların farkının önemli olup olmadığını değerlendiren (9/51 ile 6/54) ki-kare testine ait sonuçtur. $p=0.13$ sonucu da bu oranların dağılımının istatistiksel olarak önemli olmadığını ifade etmektedir. Peki altta yatan %95 GA (güven aralığı) değeri neyi ifade etmektedir? Bu değer aynı çalışma benzer gruplarda 100 kez yapılırsa 95 çalışmada 0.6 ile 4.9 arasında değişen bir rölatif risk değerinin bulunacağını ifade etmektedir. Yani bu çalışmada RR değerini 1.8 bulduk ama benzer gruplarda 100 kez çalışmayı tekrarlasaydık bu değerler arasında bir değer bulacaktık demektir. Peki 0.6 ile 4.9 arasında 1 değeri var mıdır? Vardır. Bu ne demektir? Yapacağımız bu 100 çalışmadan 95'ine ait sonuçlarda bulacağımız RR değerlerinden biri de 1 olabilir. Yani "steroid tedavisi alan ve almayanlar arasında iki uçlu bozukluk gelişme riski yönünden fark olmazdı, yani bulduğumuz sonuç istatistiksel olarak önemli değildir" demektir. Evet görüldüğü gibi %95 güven aralığı 1 değerini içerdiği için, p değerimiz de istatistiksel olarak önemli değil yani 0.05'ten büyük ($p=0.13$) olarak bulunmuştur. **Sonuç steroid tedavisi almayanlara göre alanlarda %80'lik bir risk artışı olmasına rağmen bu artış istatistiksel olarak önemli değildir.**

Gelelim ikinci sonuç setine. Dikkat ederseniz tüm değerler 10 ile çarpılmıştır. Steroid tedavisi alanlar ve almayanlar bu kez 60 değil 600 kişidir. Her 2 grupta iki uçlu bozukluk gelişenlerin sayısı da 10 kat artmış, 9 yerine 90, 5 yerine 50 olmuştur. Bu durumda RR değeri değişmiştir. Payda her 2 grupta da eşit (600) olduğundan payların bölümü yani 90/50'den yine 1.8 değerine ulaşmıştır.

Fakat bu sefer ki-kare testine ait p değeri değişmiştir. $p=0.002$ bulunmuştur. Yani ki-kare testi sonucu istatistiksel olarak önemlidir. Bu durumda yorumumuz "steroid tedavisi alanlarda ve almayanlarda iki uçlu bozukluk gelişme riskinde istatistiksel olarak önemli bir farklılık vardır veya steroid tedavisi alanlarda almayanlara göre iki uçlu bozukluk gelişme riski istatistiksel olarak önemli oranda artmıştır" şeklinde olacaktır. %95 GA değerine baktığımızda da (1.3-2.5) bulunmuştur. Yani 1 değerini içermemektedir, yani p değeri bu durumda 0.05'ten küçük olacaktır ki 0.002 olarak bulunmuştur. **Peki niye RR değişmediği halde p değeri küçülmüştür? Yanıt örnek büyüklüğümüz toplam 120'den 1200'e çıkmıştır.**

Üçüncü sonuç setimize baktığımızda steroid tedavisi alanlarda almayanlara göre iki uçlu bozukluk gelişme riski sadece %20 artmış olarak bulunmuştur. Yani rölatif risk değeri 1.2'dir ve ilk 2 sonuç setine göre risk artışı azalmıştır. Buna karşılık p değeri üstteki sette olduğu gibi 0.002 olarak bulunmuştur. **Risk artışı azalmıştır ama p değeri değişmemiştir. Bu nasıl olmuştur?** Şimdi grup büyüklüklerine bir bakalım. İkinci sette steroid tedavisi alan ve almayan 600 kişiden oluşan toplam 1200 kişide çalışma gerçekleştirilmişti. Üçüncü sette ise bu gruplar her biri 2000 kişiden oluşmaktadır. Yani toplam katılımcı sayısı 4000'e yükselmiştir.

Gelelim bu sonuç setlerinin bize ne anlattığına. Bu 3 grupta da bulduğumuz RR değerleri "klinik önemliliği" gösteren ölçütlerdir. Yani ilk 2 grupta %80'lik bir risk artışı bulunmuşken, üçüncü sonuç setine göre risk artışı %20'dir yani klinik olarak daha az önemlidir. **p değerleri ise "istatistiksel önemliliği" göstermektedir.** Bu durumda ilk grupta üçüncü gruba göre risk artışı daha fazla olsa da (%80'e karşılık %20) p değeri ilk grupta çalışma daha küçük bir grupta (toplam 120 kişi) gerçekleştirildiği için istatistiksel olarak önemli bulunmamıştır. Buna karşılık üçüncü sonuç setinde risk artışı sadece %20'de kalsa da (steroid tedavisi alanın iki uçlu bozukluk riskini artırdığına yönelik daha az bir klinik önem bulunmuş olsa da) çalışma toplam 4000 kişide gerçekleştirildiği için yani örnek büyüklüğü daha fazla olduğu için istatistiksel olarak önemli bir fark bulunmuştur.

Sonuç olarak

1. İstatistiksel önemlilik yani p değeri örnek büyüklüğünden etkilenir. Yani klinik önemliliği fazla olmayan bir durumu örnek büyüklüğünüz arttıkça istatistiksel

olarak önemli bulabilirsiniz.

2. Klinik önemlilik ise örnek büyüklüğünden etkilenmez.
3. Peki biz neyin peşinde olmalıyız? Klinik önemliliği destekleyecek istatistiksel önemliliğin. Bunun içinde çalışmanın başında öngördüğümüz klinik önemliliği destekleyecek istatistiksel önemliliği ($p < 0.05$) bulmamızı sağlayacak en az gerekli örnek büyüklüğünü hesaplayarak çalışmamızı planlayıp yürütmeliyiz. Böyle yapmazsak; ilk sonuç setinde olduğu gibi klinik olarak önemli bir risk artışını istatistiksel olarak önemli bulamayabiliriz. Diğer tarafta da, daha az klinik önemliliği olan bir çalışma sonucunu da sırf daha

büyük örnekleme gerçekleştirildiği için ve bu nedenle p değeri istatistiksel olarak önemli olduğu için daha önemli bir sonuçmuş gibi yorumlayabiliriz.

Özetle; biz bir çalışma planlarken klinik önemi destekleyen bir istatistiksel önemliliğin peşinde olmalıyız ve okuduğumuz yazılarda da bu hususa dikkat ederek sonuçları dikkatli yorumlamalıyız.

Not: Bu yazıda bildirilen iki uçlu bozukluk ve steroid kullanımı ilişkisi örneği ve ilgili sayısal veriler istatistik yorumlama eğitimi için tamamen kurgulanmış sanal verilerdir, klinik amaçla kullanılamaz.

Kaynaklar:

1. Dawson B, Trapp RG. Reading the medical literature: In Basic&Clinical Biostatistics. Third Ed. New York. Lange Medical Books/McGraw Hill. 2001:304-305.
2. Pagano M, Gauvreau K. Confidence Intervals. In Principles of Biostatistics. Belmont. Duxbury Press. 1992: 195-209.
3. Greenberg RS, Daniels SR, Flanders WD, Eley JW, Boring JR. Cohort Studies. In Medical Epidemiology. New York. Lange Medical Books/McGraw Hill. 2001:121.