



Etki Büyüklüğü

Selim Kılıç¹

ÖZET:

Etki büyüklüğü

Makaleleri değerlendirirken istatistiksel önemlilik yani p değeri dikkate alınan bir ölçüttür. Buna karşılık en az istatistiksel önemlilik kadar önemli olan klinik önemlilik yazarlarca veya okurlarca her zaman dikkate alınmayan bir durumdur. Bu makalede klinik önemi gösteren önemli bir ölçüt etki büyüklüğü kavramı, hesaplanması ve yorumlanması ile ilgili bazı bilgiler okuyucuya sunulmuştur.

Anahtar sözcükler: etki büyüklüğü, klinik önemlilik

ABSTRACT:

Effect size

Statistical significance (p value) is a considerable measure when the manuscripts were reading or interpreting. On the other hand clinical importance, which is important at least as statistical significance, may not consider every time by authors or readers. In this manuscript some information related to an important measure, which indicates clinical importance "effect size concept (ES)", ES calculation and ES interpretation are presented.

Key words: effect size, clinical significance

Journal of Mood Disorders 2014;4(1):44-6



¹MD, Gülhane Askeri Tıp Fakültesi, Ankara-Türkiye

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

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

Kabul tarihi / Date of acceptance: 28 Şubat 2014 / February 28, 2014

Bağıntı beyanı:

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

Declaration of interest:

S.K.: The author declare that they have no conflict of interests regarding the content of this article.

Çalışmalarda karşılaştırılan gruplar arası farkın önemliliği-anlamlılığını değerlendirmede hemen her zaman dikkate alınan husus istatistiksel önemlilik olup olmadığıdır. Bir başka ifade ile p değerinin 0.05'ten küçük olmasıdır. Çoğu araştırmacı veya okur çalışmada bulunan p değeri <0.05 olduğunda, bulunan sonucun önemli olduğunu, hatta bu değer sifıra çok yaklaştığında sonucun çok ama çok önemli olduğunu düşünür veya yorumlar. Oysaki p değeri gruplar arası bulduğumuz farkın veya -yapılan analiz korelasyon analizi ise- incelenen değişkenler arasındaki doğrusal ilişkinin şansa bağlı (tesadüfi) ortaya çıkıp çıkmadığını gösteren bir sonuçtur. Bu değer sifıra ne kadar yakın olursa, bulduğumuz farkın veya doğrusal ilişkinin şansa bağlı ortaya çıkma olasılığının çok düşük olduğunu gösterir fakat bulduğumuz sonucun klinik olarak önemli bir sonuç olduğunu kanıtlamaz. "p" değerinin çalışmada incelenen örnek büyüklüğünden etkilendiğini, büyük örneklerde gerçekleştirilen çalışmalarda gruplara ait ortalamalar-ortancalar arası klinik olarak önemi olmayan çok küçük farklar olsa dahi, p değerinin 0.05'ten

küçük olabileceğini daha önceki bir yazımızda tartışmıştık (1).

Çalışmadaki gruplara ait sonuçlar arası farkın önemli olup olmadığını gösteren bir başka ölçüt ise etki büyüklüğüdür. Etki büyüklüğü (EB), klinik olarak anlamlı farklılığın ortaya konması için ilgilenilen sonuç değişkenine göre iki ortalama ya da iki oran arasındaki beklenen farklılık olarak ifade edilebilir (2,3). Etki büyüklüğü, biraz basitleştirerek tanımlarsak, yeni denenen bir yöntemin, eskisine kıyasla ne kadar fark oluşturduğu kavramı olup farklı şekillerde hesaplanabilmektedir. Herhangi bir çalışma planlarken en az gerekli örnek büyüklüğü hesaplamasında etki büyüklüğü mutlaka dikkate alınması gereken bir kavramdır. Etki büyüklüğüne öncelikle konu ile ilgili literatür tarayarak, bu konuda yapılmış veya ulaşılmış çalışma yoksa araştırmacının geçmiş deneyimleri ile karar vermek mümkün olabilir (4,5). Herhangi bir şekilde örneği olmayan bir çalışma için ise bir pilot-ön çalışma düzenlenerek etki büyüklüğü belirlenebilir. Gruplar arasındaki beklenen farklılık yani etki büyüklüğü ne kadar

büyük olursa, o farkı istatistiksel olarak ortaya koymak daha kolay olacağı için daha küçük örnek büyüklüğü ile çalışmak yeterli olacaktır (4).

Etki büyüklüğünün hesaplanmasında en yaygın kullanılanı Cohen tarafından geliştirilen hesaplama (d) olmakla birlikte, Hedge's d, Glass's Δ gibi hesaplamalara da literatürde rastlanabilir (6). Cohen genel bir öneri olmak üzere, d değerinin 0,2'den küçük olması durumunda, etki büyüklüğünün zayıf, 0,5 olması durumunda orta ve 0,8'den büyük olması durumunda ise kuvvetli olarak tanımlanabileceğini söylemektedir. Ancak, 0,2'lik bir d değerinin bile kuvvetli bir etki olarak ele alınabileceği özel durumların da olabileceği unutulmamalıdır (6,7).

Klinik çalışmalarda etki büyüklüğü değerinin ≥ 0.5 olması önerilir. Gruplar arası fark (d) sembolü ile gösterilmektedir ve klinik araştırmalarda karşılaştırılacak gruplar arasındaki mutlak farkı belirtir. Örneğin çalışmada ilgilenilen sonuç değişkeni (primary outcome-dependent variable) bir oran ise, tedavi grubu ile kontrol grubu arasındaki ilgilenilen olayın gözlenme oranı farkı, başka bir ifade ile karşılaştırılan gruplar arasındaki mutlak değişim farkıdır. Örneğin ilgilendiğimiz olay panik atak hastalarında farklı tedavi seçenekleri ile ikinci bir atağın önlenmesi olsun. Tekrarlayan atak A tedavi grubunda %14, diğer grupta ise %8 olarak gözleniyorsa mutlak fark $d = \%6$ birimdir. Buna karşılık çalışmada incelenen olay ölçümsel (sürekli) bir değişken ise gruplara ait ortalamalar arası farkın grupların standart sapmalardan elde edilecek harmanlanmış (pooled) standart sapmaya bölümü EB'yi verir (5,8).

Bir örnekle etki büyüklüğünü (EB) hesaplayacak olursak karşılaştırdığımız iki gruba ait ortalamalar 24 ve 20 olsun. Buna karşılık gruplara ait standart sapmalarda sırasıyla 5 ve 4 olsun.

$$EB = \frac{A \text{ grubunun ortalaması} - B \text{ grubunun ortalaması}}{\text{Harmanlanmış (pooled) standart sapma}}$$

$$SS_{\text{pooled}} = \sqrt{(SS^2_{\text{grupA}} + SS^2_{\text{grupB}}) / 2}$$

$$SS_{\text{pooled}} = \sqrt{(5^2 + 4^2) / 2} = 4.53$$

$$EB = \frac{24 - 20}{4.53} = 0.88$$

Yorumumuz: Bu sonuç bize iki grup arasındaki farkın klinik olarak önemli kabul edilecek büyük bir fark olduğunu göstermektedir.

Çalışmada birden fazla değişken, sonuç değişkeni olarak incelenecekse, bu durumda her bir değişken için fark değerleri (d) ve etki büyüklükleri (EB) hesaplanır. Bu d ve EB değerleri için belirli bir alfa hata ve güç (1-beta hata) değerleriyle örneklem sayıları saptanır. Saptanan en yüksek örneklem sayısı çalışmada alınması gerekli örneklem sayısını gösterir.

"d" etki büyüklüğüyle ilgili gözden kaçmaması gereken bir diğer önemli nokta ise, bunun da bir istatistik olduğudur. Yani, bir başka örneklem seçilip çalışma yinlendiğinde farklı bir d değeri elde edilecektir. Dolayısıyla, d değeri için de güven aralığı hesaplanmalıdır. Böylelikle, güven aralığı başlığı altında verilen avantajlar etki büyüklüğü için de sağlanmış olacaktır (6).

Günümüzde psikiyatri alanında yapılan çalışmalarda da etki büyüklüğü kavramı yaygın olarak kullanılmakta ve tartışılmaktadır. İki uçlu bozukluklu hastalarda, depresyon hastalarında, anksiyete hastalarında, uyku bozukluğu hastalarında yapılan çalışmalara ait örnekler literatürde yer almaktadır ve bu çalışmalardan bazılarında ayrıntılı sonuçlar ilgili kaynaklara ulaşılarak incelenebilir (9-13).

Sonuç olarak çalışmalar planlanırken, en az gerekli örnek büyüklüğü hesaplanırken tip 1 (alfa) ve tip 2 (beta) hatanın yanı sıra klinik önemi ortaya koymada bir kriter olan etki büyüklüğü de mutlaka dikkate alınmalı, çalışma sonuçları bulunan etki büyüklüğü değeri göz önüne alınarak yorumlanmalıdır.

Kaynaklar:

1. Kılıç S. Neyin Peşindeyiz? Kutsal p Değerinin mi (İstatistiksel Önemlilik) Yoksa Klinik Önemliliğin mi? *Journal of Mood Disorders*. 2011;1:46-48.
2. Dawson B, Trapp RG. Reading the medical literature: In *Basic&Clinical Biostatistics*. Third Ed. New York. Lange Medical Books/McGraw Hill. 2001:250-1.
3. Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Getting ready to estimate sample size: Hypotheses and Underlying Principles. In *Designing Clinical Research*, 2nd Ed. Lippincott Williams&Wilkins, 2001, 55-56.
4. Kul S. Klinik araştırmalarda örnek genişliği belirleme. *Ekstraplevral*. 2011, 129-32.

5. Kirby A, Gebiski V, Keech AC. Determining the sample size in a clinical trial. *MJA*. 2002;177:256-7.
6. Yıldırım HH, Yıldırım S. Hipotez Testi, Güven Aralığı, Etki Büyüklüğü ve Merkezi Olmayan Olasılık Dağılımları Üzerine. *İlköğretim Online*. 2011;10:1112-23.
7. Cohen J. The analysis of variance. In *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates. 1988, 274-87.
8. Süt N. Klinik araştırmalarda örneklem sayının belirlenmesi ve güç (power) analizi. *RAED Dergisi*. 2011;3:29-33.
9. Chengappa KN, Bowie CR, Schlicht PJ, Fleet D, Brar JS, Jindal R. Randomized placebo-controlled adjunctive study of an extract of withania somnifera for cognitive dysfunction in bipolar disorder. *J Clin Psychiatry*. 2013;74:1076-83.
10. Smith EG, Deligiannidis KM, Ulbricht CM, Landolin CS, Patel JK, Rothschild AJ. Antidepressant augmentation using the N-methyl-D-aspartate antagonist memantine: a randomized, double-blind, placebo-controlled trial. *J Clin Psychiatry*. 2013;74:966-73.
11. Sarris J, Stough C, Bousman CA, Wahid ZT, Murray G, Teschke R, Savage KM, Dowell A, Ng C, Schweitzer I. Kava in the treatment of generalized anxiety disorder: a double-blind, randomized, placebo-controlled study. *J Clin Psychopharmacol*. 2013; 33:643-8.
12. Ishak WW, Christensen S, Sayer G, Ha K, Li N, Miller J, Nguyen JM, Cohen RM. Sexual satisfaction and quality of life in major depressive disorder before and after treatment with citalopram in the STAR*D study. *J Clin Psychiatry*. 2013;74:256-61.
13. Krystal AD, McCall WV, Fava M, Joffe H, Soares CN, Huang H, Grinell T, Zummo J, Spalding W, Marshall R. Eszopiclone treatment for insomnia: effect size comparisons in patients with primary insomnia and insomnia with medical and psychiatric comorbidity. *Prim Care Companion CNS Disord*. 2012;14:PCC.11m01296. doi: 10.4088/PCC.11m01296.