



## Örnek Büyüklüğü, Güç Kavramları ve Örnek Büyüklüğü Hesaplaması

Selim Kılıç<sup>1</sup>

### ÖZET:

Örnek büyüklüğü, güç kavramları ve örnek büyüklüğü hesaplaması

Örnek büyüklüğü, güç kavramları ile örnek büyüklüğü hesaplamaları makalelerde sıklıkla yer alan, yazı içinde belirtilmemesi durumunda gönderilen araştırmaları değerlendiren hakemler ve editörlerce eksikliği vurgulanan konulardandır. Bu yazının amacı verilen örneklerle makale yazacak, okuyacak ve değerlendirecek olanlara bu kavramlar hakkında katkı sağlamaktır.

**Anahtar sözcükler:** örnek büyüklüğü, beta hata, güç

### ABSTRACT:

Sample size, power concepts and sample size calculation

Sample size, power and sample size calculation concepts are commonly stated issues in manuscripts and the lackness of them are indicated by the reviewers and editors if they were not explained in the submitted manuscripts. The aim of this review is to help individuals who will write, read and review manuscripts through endowed examples.

**Key words:** sample size, beta error, power

**Journal of Mood Disorders 2012;2(3):140-2**

<sup>1</sup>MD, 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: drselimkiloc@gmail.com

Kabul tarihi / Date of acceptance: 20 Eylül 2012 / September 20, 2012

### Bağıntı beyanı:

S.K.: Yazarlar bu makale ile ilgili olarak herhangi bir çıkar çatışması bildirmemişlerdir.

### Declaration of interest:

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

Daha önceki sayılarda istatistiksel önemliliği değerlendirmede kullanılan bir ölçüt olan p değeri ve tedavinin etkinliğini değerlendirmede katkı sağlayan epidemiyolojik bir ölçüt olan tedavi için gerekli sayı (TGS) ve zarar için gerekli sayı (ZGS) kavramlarını tartışmıştık (1,2). Bu yazıda da örnek büyüklüğü, güç (power) kavramları ile örnek büyüklüğü hesaplanmasının örneklerle açıklanması amaçlanmıştır.

Planladığımız ve yürüttüğümüz araştırmalarda çoğunlukla örnekleme çalışıyoruz. Evrenin özelliklerini yansıtmaya düşüncesiyle evrenden belirli yöntemlerle seçilmiş birimlerin oluşturduğu topluluk olan örneklem, matematiksel ifadeyle evrenin bir alt kümesidir (3,4). Örneğin Türkiye’de 15 yaş ve üstü grupta depresyon sıklığını-prevalansını saptamak çalışma konumuz olsun. Bu durumda evrenimiz Türkiye’de yaşayan 15 yaş ve üstü nüfustur. Bu da Türkiye İstatistik Kurumu (TÜİK) 2011 yılı sonuçlarına göre 55 milyonun üzerinde bir değere karşılık gelmektedir (5). Bu çalışmayı milyonlarca kişiyi depresyon tanısı yönünden değerlendirerek yürütmemiz olanaklı değildir. Bu durumda evrenden (55 milyon kişiden) temsil özelliği olan bir örneklem seçmemiz gerek-

mektedir. Peki örneklem için en az kaç kişi seçmemiz gerekiyor?

Araştırmalarda örneklem büyüklüğünü saptamak için farklı formüller geliştirilmiştir. Örnekleme temel ilke evreni temsil niteliğine sahip bir örneklem seçilmesidir. Bu özellik örneklemin niteliksel olarak alt kümesi olduğu evreni temsil etmesi yani evrenin özelliklerini taşımasıdır ki çok önemlidir. Niteliksel olarak uygun olmayan yani evreni temsil etmeyen bir örneklemin büyüklüğü ne kadar artırılsa artırılsın bu örneklemden elde edilecek sonuçlara güvenmek sıkıntılıdır zira uygun olmayan örnek seçimine bağlı (selection bias) elde edilecek sonuçlarda bir yanlılık- taraf tutma (bias) söz konusu olacaktır (6). Buna karşılık seçime bağlı bir yanlılık yani örnekleme niteliksel bir sorun olmadığı sürece örnekleme birim sayısının artması yani örnek büyüklüğünün artması durumunda ait olduğu evreni temsil niteliği de artacaktır. Ancak büyüklüğün artması maliyet, zaman ve personel gibi her zaman temini olanaklı olmayan kaynakların artmasına neden olmaktadır. Bu durumda çalışma için gerekli uygun (optimal) örnek büyüklüğünü saptamak gereklidir. Örneklem büyüklüğünü etkileyen önemli

unsurlardan biri "kabul edilen örnekleme hatası"dır ve araştırmacı tarafından belirlenir. Bu değer küçüldükçe, yani gerçeğe, evrene ait değerlere en yakın sonuçlar tahmin edilmek istedikçe gereksinim duyulacak örneklem büyüklüğü artacaktır. Sözgelimi %1 örnekleme hatası öngörülen örneklemin büyüklüğü, %3 örnekleme hatası öngörülen örneklemden daha büyük olacaktır. Bilimsel çalışmalarda kabul edilen anlamlılık düzeyi genellikle %1 veya %5 olarak seçilir. Günümüzde bu hesaplamaları internet ortamında yapabileceğimiz web adresleri bulunmaktadır ve farklı seçenekleri değerlendirme adına büyük kolaylık sağlamaktadır (7-9). Evrenin büyüklüğünün bilindiği durumlarda kullanacağımız formül (10):

$$N = \frac{(t1-a)^2 (p.q)}{S^2(N-1) + S^2(p.q)}$$

N: Evrendeki kişi sayısı. Bizim çalışmamız için bu değer 55 milyon

p: Araştırdığımız durumun sıklığı. Daha önceki çalışmalardan öngörümüz %10 olsun=0.10

q:  $1-p=1-0.10=0.90$  (Yani incelediğimiz durumun görülme sıklığı)

s:  $\alpha=0.01$  (Araştırmacı tarafından kabul edilen örnekleme hatası)

t (1-0.95): Kabul edilen anlamlılık düzeyi yani tip 1 (alfa) hata düzeyi = %5'e karşılık gelen değer 1.96

$$55.000.000 (1.96 \times 1.96) (0.1 \times 0.9)$$

$$n = \frac{55.000.000 (1.96 \times 1.96) (0.1 \times 0.9)}{(0.01)^2 (55.000.000 - 1) + (0.01)^2 (0.1 \times 0.9)} = 3458$$

Sonuç olarak Türkiye'de 15 yaş ve üstünde depresyon sıklığı için bir araştırma planladığımızda daha önceki çalışmalardan yola çıkarak depresyon sıklığı için öngördüğümüz değer %10 olduğunda, örnekleme hatası olarak %1 hatayı kabul ettiğimizde ve kabul edilen istatistiksel anlamlılık düzeyini %5 kabul ettiğimizde en az 3458 kişilik bir grupta bu çalışmayı yürütmemiz gerekmektedir. Çalışmanın niceliksel boyutu yani örnek büyüklüğü için en az gerekli sayı bu değerdir ve bu değeri ne kadar artırırsak tip 1 hata ( $\alpha$  hata) düzeyimiz azalacaktır. Buna karşılık sayıyı ne kadar artırırsak artıralım, seçtiğimiz örneklem evreni temsil etmiyorsa seçime bağlı yanlılık (selecti-on bias) nedeni ile doğru sonuca-gerçeğe ulaşmamız o

derece güç olacaktır. Yani temsiliyet araştırmalarda olmazsa olmaz çok önemli bir durumdur ve örneklemin evreni temsil etmesi olasılığını artırmak için de rastgele örnekleme yöntemleri ile çalışma grubunu belirlemek gerekmektedir.

Buna karşılık bir ilacın diğer bir ilaca/plaseboya veya bir tedavi yönteminin diğer üstünlüğünü değerlendirmek için planladığımız klinik araştırmalardaki örnek büyüklüğü hesaplamasında ise farklı bir yaklaşım ve hesaplama yöntemlerini kullanıyoruz. Günümüzde bu hesaplamalar için internet üzerinden ulaşılabilecek birçok kullanımı kolay web adresi bulunmaktadır ve bir kaçının web adresi verilmiştir (11-13).

Aşağıda bir örnekle -bu web adreslerinden birinde ulaşılan programda- görsel olarak hesaplama açıklanmıştır (11). Klinikte depresyon tedavisinde kullanılan ve tedavi başarısı %50 olan A isminde bir ilacımız var. Yeni kullanıma girmiş ve tedavi etkinliği en az %70 olarak faz 3 çalışmalarında bulunmuş B isminde de bizim çalışmak istediğimiz yeni bir ilaç var. Çalışma sorumuz bu yeni ilacın ilaç A'dan en az %20 üstün olduğunu yani ilaç B'nin %70 ve üzerinde klinik etkinliği olduğunu göstermek için her bir çalışma grubu için en az kaç hastada tedavi sonuçlarını değerlendirmemiz gerektiğidir.

Yeni ilacın etkinliği için (ilaç B) öngördüğümüz %70'i en üst kutucuğa yazıyoruz. Alt kutucuğa çalışmada karşılaştırmada kullanılan ilacın (ilaç A) etkinliği %50'yi yazıyoruz. Bir alt kutucuğa da bu çalışma için kabul ettiğimiz alfa hata (tip 1) düzeyini yazıyoruz. Tip 1 hata gerçekte karşılaştırılan iki ilaç arasında istatistiksel olarak anlamlı fark olmadığı halde bizim çalışmada yanlışlıkla fark var olarak bulma olasılığımızdır yani yanlış pozitif bulma olasılığıdır. Tip 1 hata için kabul edilen sınır değer genellikle %5'tir. Yapılan hatanın ciddi sonuçlar vereceği değerlendirildiğinde bu hata payı küçültülür, %1 hatta %0.1 kabul

edilebilir. En alttaki kutuda yazacağımız değer ise beta hata (tip 2) değeridir. Beta hata gerçekte karşılaştırılan iki grup (ilaç/yöntem) arasında istatistiksel anlamlı fark varken, bu farkın araştırmacı(lar) tarafından bulunamaması durumudur (yanlış negatif). Bu değer de küçük olması tercih edilir ama çalışmalarda %20'ye kadar kabul edilmektedir. Zira bu değeri ne kadar küçültürsek örnek büyüklüğü o kadar artacaktır. Eğer zaman, para, insan gücü gibi kaynaklarımız büyük gruplarda çalışmaya olanak sağlıyorsa ve örnek büyüklüğünün artması etik anlamda çıkabilecek sorun olasılığını artırmayacaksa kabul edilen beta hata değeri de küçültülebilir. Buna karşılık araştırmacının gücü (power) dediğimiz kavram ise 1-beta hataya karşılık gelen değerdir. Yani beta hatanın %20 kabul edildiği bir çalışmanın gücü (1-0.20) 0.80, beta hatanın %5 kabul edildiği bir çalışmanın gücü (1-0.05) 0.95 olarak kabul edilir. Çalışmalarda araştırmacının gücü 0.80 ve üzerinde olması istenmektedir (14,15).

Yukarıdaki örnekteki sonuçları yorumlarsak; klinikte kullanılan ve %50 tedavi başarısı olan A ilacına karşılık, en az %70 tedavi başarısı beklediğimiz yeni ilacımız ilaç B'nin istatistiksel olarak üstünlüğünü göstermek için

-çalışmadaki alfa hatayı %5, beta hatayı %20 kabul ettiğimizde- her bir grupta en az 73 hastayı çalışmaya almamız gerekecektir. Eğer yeni ilaçtaki (ilaç B) tedavi başarısı %70'ten fazlaysa çalışma için gerekli olan örnek büyüklüğü azalacaktır. Örneğin ilaç B için %80 tedavi başarısı durumunda örnek büyüklüğü her bir grup için 30 hastaya, %90 tedavi başarısında ise 15 hastaya düşecektir. Buna karşılık ilaç B ve A için %70 ve %50 tedavi başarı hızlarını değiştirmedeğimizde, alfa hatayı yine %5 kabul edip beta hatayı %10 veya %5'e düşürdüğümüzde her bir grup için en az gerekli hasta sayısı sırası ile 101 ve 107 olmaktadır.

Sonuç olarak; günümüzde örnek büyüklüğü ve güç kavramları ile örnek büyüklüğü hesaplamaları araştırmalarda sıklıkla yer alan, yazı içinde belirtilmemesi durumunda gönderilen araştırmaları değerlendiren hakemler ve editörlerce eksikliği vurgulanan ve yazı içinde hesaplamaların nasıl yapıldığının ayrıntılı açıklanması istenebilen kavramlardır. Bu yazının -ilgilenenlerin yararlanabileceği bazı kaynaklar da verilerek- makale yazacak ve değerlendirecek olanlara katkı sağlayacağı değerlendirilmektedir.

## Kaynaklar:

1. Kılıç S. Neyin Peşindeyiz? Kutsal p değerinin mi (istatistiksel önemlilik) yoksa klinik önemliliğin mi? JMOOD. 2011;1:46-48
2. Kılıç S. Tedavi için gerekli sayı (TGS) ve zarar için gerekli sayı (ZGS) kavramları klinisyene ne mesajlar veriyor? JMOOD. 2012;2:91-94
3. Özmen Ahmet. Örnekleme. Sosyal Bilimlerde Araştırma Yöntemleri. Anadolu Üniversitesi Yayınları. Eskişehir, 1999, 27.
4. Dawson B, Trap RG. Basic and Clinical Biostatistics, Lange Medical Books/McGraw-Hill, 2004; 69.
5. Adrese Dayalı Nüfus Kayıt Sistemleri Sonuçları 2011, Türkiye İstatistik Kurumu (TÜİK) Matbaası, 2012, 1.
6. Morton RF, Hebel JR, McCarter RJ. A Study Guide to Epidemiology and Biostatistics, Aspern Publisher Inc., 1996,74.
7. <http://www.raosoft.com/samplesize.html>
8. <http://www.macorr.com/sample-size-calculator.htm>
9. <http://www.surveysystem.com/sscalc.htm>
10. Naing L, Winn T, Rusli BN. Practical issues in calculating the sample size for prevalence studies. Archives of Orofacial Sciences 2006;1:9-14
11. <http://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/samplesizecalculators.aspx>
12. <http://www.stat.ubc.ca/~rollin/stats/ssize/>
13. <http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize>
14. Pagano M, Gauvreau K.. Hypothesis Testing. Principles of Biostatistics, Duxbury Press Wadsworth Inc., 1993, 222.
15. Dawson B, Trap RG. Basic and Clinical Biostatistics, Lange Medical Books/McGraw-Hill, 2004; 107.