

Klinik arařtırmalarda örneklem sayısının belirlenmesi ve güç (power) analizi

Sample size determination and power analysis in clinical trials

Necdet Süt

Trakya Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, Edirne

Özet

Derlemenin amacı klinik arařtırmalarda, i) örneklem sayısı ve gücünü belirlemenin arařtırma sonuçları üzerine etkisini ve önemini vurgulamak, ii) örneklem sayısını belirlerken ne tür bileşenlere dikkat edilmesi gerektiği konusunda arařtırmacılara yol gösterici bilgiler aktarmaktır. Klinik arařtırmalarda güç (power) analiziyle örneklem sayısının belirlenmesini etkileyen bileşenler olarak, I. tip hata (α) düzeyi, II. tip hata (β) düzeyi, gruplar arası fark ya da etki büyüklüğü değeri, gruplara dağıtım oranı ve arařtırma hipotezinin yönü sayılabilir. Klinik çalışmalarda çalışmanın başlangıcında güç analiziyle gerekli minimum örneklem sayısının belirlenmesiyle çalışma sonuçlarının güvenilirliğinin sağlanmasının yanı sıra kaynak israfının da önüne geçilir.

Anahtar sözcükler: Güç analizi, örneklem sayısı, I. tip hata, II. tip hata, etki büyüklüğü

Summary

The aim of this review is to: i) stress the effect of sample size and power analysis on study results, ii) construct information to guide researchers on the components that should be taken into consideration in order to calculate the sample size. The components which effect sample size calculation with power analysis in clinical trials are type I error (α), type II error (β), difference between the groups or effect size, allocation ratio and direction of study hypothesis (one/two sided). Sample size calculation with power analysis in the planning stage of clinical trials guarantees the reliability of the study results, in addition to preventing source extravagance.

Key words: Power analysis, sample size, type I error, type II error, effect size

Klinik arařtırmalar hipotezin oluşturulması, planlama, finansman sağlanması, etik kurul onayı alınması, hastalardan yazılı/sözlü onay alınması, verilerin biyosız elde si ve elektronik ortama aktarılması, verilerin doğru istatistiksel yöntemlerle analizi ve sonuçların uygun raporlanması gibi uzun süreç, emek ve maliyet gerektiren arařtırmalardır.

Klinik arařtırmalar karşılařtırmaya dayanan analitik arařtırmalardır. Çalışma sonucunda karşılařtırılan yöntemler arasında ister “fark var ($p<0.05$)” bulunsun, ister “fark yok ($p>0.05$)” bulunsun bu durum gerçeği yansıtmalıdır. Bunun için örneklem sayısı ve güç (power) analizi klinik arařtırmalarda önemli rol oynamaktadır. Randomize klinik arařtırmaların kalitesini yükseltmek amacıyla geliř-

tirilen ve NEJM, Lancet, Rheumatology gibi çok sayıda saygın tıp dergilerince desteklenen CONSORT (*Consolidated Standards for Reporting Trials*) listesinin bir maddesinin “Örneklem sayısı nasıl belirlendi?”^[1] sorusundan oluşması klinik çalışmalarda örneklem sayısının ve gücün belirlenmesinin önemini göstermektedir.

Örneklem sayısı çalışmanın başlangıcında çalışmanın gücü dikkate alınarak belirlenmelidir. Aksi halde çalışmanın gücünde örneklem sayısının yetersizliğinden kaynaklanacak düşüşler, yanlış olabilecek sonuçlara, arařtırmacıların-hastaların emek/zamanlarının boşa geçmesine ve kaynak israfına neden olacaktır.

Klinik çalışmada örneklemin, sayısal açıdan yeterli ve yapısal açıdan toplumun tümünü temsil etme gücüne sa-

İletişim / Correspondence:

Doç. Dr. Necdet Süt. Trakya Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, 22030, Edirne
Tel: 0284 235 76 41 • e-posta: necdetsut@yahoo.com

Çıkar çakışması / Conflicts of interest: Çıkar çakışması bulunmadığı belirtilmiştir. / No conflicts declared.

hip olması gereklidir.^[2] Örneklemden doğru sonuçlar elde edebilmek ve elde edilen sonuçları topluma genelleştirebilmek için örneklem seçiminde toplumun bazen tabaka (stratum), blok gibi alt katmanlara ayrılması gerekebilir.

Her bir klinik çalışma tasarımı için güç ve örneklem sayısı hesaplamasında kullanılan formüller farklı ve oldukça karmaşık olduklarından çözümlerinin bilgisayar programları yardımıyla yapılması araştırmacılar için daha uygun olacaktır. Güç analizi ve örneklem sayısı hesaplamak için Statistica, NCSS ve G-Power gibi istatistiksel paket programlar kullanılabilir.

Güç Analizi

Güç (*power*) kavramı; iki ya da daha fazla ilaç/tedavi yönteminin etkinliklerini karşılaştırırken gerçekte fark olduğunda kullanılan istatistiksel testin bunu saptayabilme olasılığıdır. 1-β olarak hesaplanan güç klinik çalışmalarda minimum %80 olmalıdır. Güç analiziyle örneklem sayısını belirlerken aşağıda bahsedilen bileşenlerin belirlenmesi ve bu bileşenler doğrultusunda hesaplamalara gidilmesi gereklidir.^[3,4]

Güç (*power*) analiziyle örneklem sayısını hesaplamak için gerekli bileşenler

1. I. tip hata (α) düzeyi
2. II. tip hata (β) düzeyi
3. Gruplar arası fark (*d-difference*) ya da etki büyüklüğü *EB-effect size*)
4. Gruplara dağıtım oranı
5. Araştırma hipotezinin (H_1) yönü

1. I. tip hata (α)

Herhangi bir ilaç/yöntemin etkinliğinin araştırıldığı kontrollü bir çalışmada, etkinlik açısından gerçekte “fark olmadığı” araştırma sonucu yanlışlıkla “fark var” bulunması olasılığıdır. I. tip hatanın (α) klinik araştırmalarda maksimum %5 olarak seçilmesi uygundur.

2. II. tip hata (β)

Herhangi bir ilaç/yöntemin etkinliğinin araştırıldığı kontrollü bir çalışmada, etkinlik açısından gerçekte “fark var” iken araştırma sonucu yanlışlıkla “fark yok” bulunması olasılığıdır. 1-β değeri çalışmanın gücünü (*power*) gösterir ve klinik çalışmalarda en düşük *power* değeri %80, dolayısıyla β (II. tip hata) da maksimum 0.2 olarak seçilmelidir.

3. Gruplar arası fark (*Difference*) ya da etki büyüklüğü (*Effect Size*)

İstatistiksel anlamlı bir farkın aynı zamanda klinik olarak da bir anlam ifade edip etmediğini gösteren ölçüt “Etki Büyüklüğü (EB)”dir. Klinik çalışmalarda etki büyüklüğü değerinin ≥ 0.5 olması önerilir.

Gruplar arası fark (*d*) ise klinik araştırmalarda karşılaştırılacak gruplar arası mutlak farkı ifade etmektedir.

Örneklem sayısının belirlenmesinde kullanılacak *d* ya da EB değeri, araştırmada incelenecek ana sonuç (*primary outcome*) değişkeninin toplumsal değeri hakkında elde edilen ön bilgilere (literatür bilgisi) dayalı olarak hesaplanır.

Kontrollü bir çalışmada *d-EB* aşağıdaki gibi hesaplanır:

- İlgilenilen ana değişken (*primary outcome*) bir oran ise, tedavi grubu ile kontrol grubu arasındaki ilgilenilen olayın gözlenme oranı farkı kabaca *d*'yi verir. Diğer bir ifadeyle tedavi grubu ile kontrol grubu arasındaki mutlak değişim farkıdır. Grup sayısı ikiden fazla ise en yüksek oran ile en düşük oran arasındaki farktır. Örneğin ilgilenilen bir semptom (olay) tedavi grubunda %20, kontrol grubunda %12 oranında gözleniyorsa mutlak fark $d = \%8$ birimdir.
- İlgilenilen olay ölçümsel (nicel) bir değer ise, tedavi grubu ortalaması ile kontrol grubu ortalaması arasındaki fark *d*'yi verir. Bu farkın kontrol grubunun standart sapmasına bölümü ise EB'yi verir.

$$d = \text{Tedavi grubunun ortalaması} - \text{Kontrol grubunun ortalaması}$$

$$EB = \frac{\text{Tedavi grubunun ortalaması} - \text{Kontrol grubunun ortalaması}}{\text{Kontrol grubunun standart sapması}}$$

Uygulamada etki büyüklüğünün değerleri aşağıdaki gibi yorumlanır:

Etki Büyüklüğü	Yorum
$EB < 0.1$	Önemsiz etki
$0.1 \leq EB < 0.3$	Düşük etki
$0.3 \leq EB < 0.5$	Orta derecede etki
$EB \geq 0.5$	Büyük etki

Grup sayısı ikiden fazlaysa *d* ve EB aşağıdaki formüllerle hesaplanır:

$$d = \text{En büyük grup ortalaması} - \text{En küçük grup ortalaması}$$

$$EB = \frac{\text{En büyük grup ortalaması} - \text{En küçük grup ortalaması}}{\sqrt{\text{Hata kareler ortalaması}}}$$

4. Gruplara dağıtım oranı (*Allocation Ratio*)

Gruplara dağıtım oranı örneklem sayısını etkiler.^[4] Bu nedenle örneklem sayısı olabileceği en düşük düzeyde belirlenmek isteniyorsa plasebo kontrollü bir çalışmada 1:1 (tedavi/ilaç grubuna alınan her 1 hastaya karşılık plasebo grubuna da 1 hasta alınması) kuralına göre dağıtım yapılmalıdır. Bazen gruplara dağıtım oranı 1:2 ya da 1:3 gibi daha büyük seçilebilir. Bu gibi durumlarda 1:1 oranıyla aynı güce ulaşabilmek için daha büyük örneklemle çalışılması gereklidir.

5. Araştırma hipotezinin (H₁) yönü

Klinik çalışmalarda araştırma hipotezinin yönü tek yönlü ya da iki yönlü olabilir. Tek yönlü hipotezler büyüklük (>) ya da küçüklük (<) yönünde farklılığı test ederken iki yönlü hipotezler yön belirtmeksizin farklılığı (≠) test eder. İki yönlü oluşturulan hipotez testlerinde testin gücü tek yönlülere göre daha düşüktür.^[3]

Güç Analizi Türleri

Önsel (*priori*), deneysel (*post hoc*) ve uzlaşmacı (*compromise*) olmak üzere 3 tür güç analizi vardır.^[3,5]

Önsel (*Priori*) güç analizi

Çalışma başlamadan önce güç analizi yapılmasıdır. Bunun için örneklem sayısı, d ya da EB değeri, α değeri, gruplara dağıtım oranı ve araştırma hipotezinin yönünün bilinmesi gereklidir.

Deneysel (*Post hoc*) güç analizi

Önsel güç analizi çalışma başlangıcında yapılırken, deneysel güç analizi çalışmanın sonunda yapılır ve çalışmanın gücü hesaplanarak sonuçların anlamlı etkiye sahip olup olmadığı belirlenir. Deneysel gücü hesaplayabilmek için her bir gruptaki örneklem sayılarının, d ya da EB değerinin, α değerinin, gruplara dağıtım oranının ve araştırma hipotezinin yönünün bilinmesi gereklidir.

Uzlaşmacı (*Compromise*) güç analizi

Yaygın kullanım alanı olmamakla birlikte $q=\beta/\alpha$ denmesini temel alan güç analizidir. Etki büyüklüğü düşük (EB=0.2) olduğunda hesaplanan örneklem sayısı yüksek olacaktır, fakat çok sayıda örnekleme incelemek mümkün değilse bu durumda güç değerinin minimum %80 olması yerine $q=\beta/\alpha$ formülüyle α ve β arasında bir denge oluşturulur. Örneğin plasebo kontrollü bir çalışmada EB=0.50 iken α=0.05 ve hedeflenen güç=0.80 temel alınarak her bir grup için n=64'er birim örneklem sayısı hesaplanır. Fakat EB=0.25 ise α=0.05 ve hedeflenen güç=0.80 temel alınarak her bir grup için n=253'er birim örneklem sayısı hesaplanır. Küçük etki büyüklüğünde (0.25) çalışılırken bu kadar sayıda örnekleme çalışmaya dahil etmek mümkün değilse bu durumda uzlaşmacı güç analizi uygulanır. Bunun için $q=\beta/\alpha$ formülünden $q=4$ olarak hesaplanır. Eğer risk dengesi olarak $q=4$ almaya karar verilirse ve her bir gruptan n=80'er kişi alarak çalışma yürütülürse daha sonra bu uzlaşma değeri için tekrar güç analizi uygulanır. Bunun sonunda çalışmanın gücü %51 olarak hesaplanır. Uzlaşmacı güç analiziyle çalışmanın gücüne ilişkin orta bir yol bulunmuş olur.

Örnek Uygulama

Romatoid artrit tedavisinde kullanılan bir ilacın etkinliğinin plasebo kontrollü araştırıldığını varsayalım. Birincil sonlanım noktası (*primary outcome*) olarak DAS28 skorunu ele alalım. İlacın DAS28 skorunda 0.83 (6.59 vs 5.76) birimlik bir düşüşünün anlamlı olduğunu ve kontrol grubunun standart sapmasının 1.35 olduğunu önceki kaynak çalışmalardan (eğer böyle bir çalışma yoksa kendimizin yapacağı pilot çalışmadan) belirlediğimizi varsayalım.

a. I. tip hata α=0.05, II. tip hata β=0.2 (power=%80) ve gruplara dağıtım oranı 1:1 olacak şekilde her bir grup için kaç hasta almalıyız?

Adım 1: I. tip hata α düzeyinin belirlenmesi: α=0.05 olarak seçildi.

Adım 2: II. tip hata β düzeylerinin belirlenmesi: β=0.2 (power=%80) olarak seçildi.

Adım 3: Gruplar arası fark (d) ve etki büyüklüğünün (EB) belirlenmesi:

$$d = 6.59 - 5.76 = 0.83 \quad EB = \frac{6.59 - 5.76}{1.35} = 0.615$$

Adım 4: Gruplara dağıtım oranının belirlenmesi: Gruplara dağıtım oranı olarak 1:1 seçildi.

Adım 5: Hipotezin yönünün belirlenmesi: Araştırma hipotezi 2 yönlü seçildi.

Örneklem sayısı aşağıdaki formül kullanılarak hesaplanır:^[2,6]

$$n = \frac{2\alpha^2(Z_{\alpha/2} + Z_{\beta})^2}{d^2} = \frac{2*(1.35)^2*(1.96 + 0.84)^2}{0.83^2} = 42$$

Her bir grup için n=42'şer hastanın çalışmaya alınması gerekmektedir. Statistica 7.0 programıyla yukarıdaki veriler baz alınarak hesaplanan örneklem sayıları ve buna karşılık gelen güç değerleri **Şekil 1**'de görülmektedir.

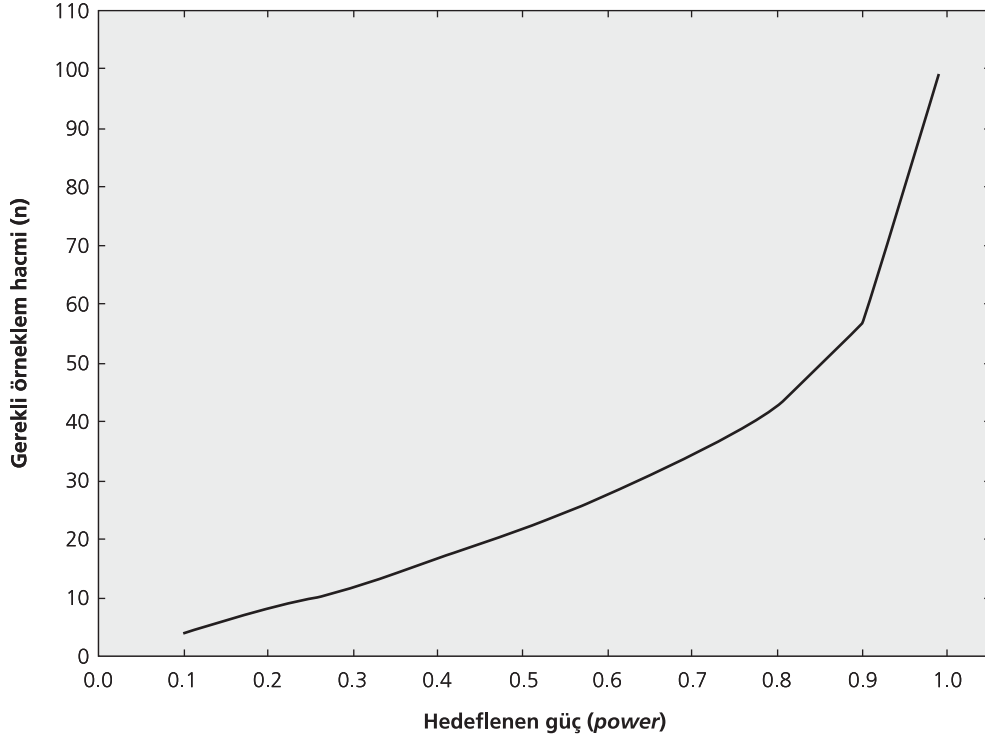
b. Çalışmanın gücünü %80'den %90'a çıkarmak istersek yukarıdaki diğer kriterler aynı kalmak koşuluyla her bir grup için kaç hasta almalıyız?

$$n = \frac{2\alpha^2(Z_{\alpha/2} + Z_{\beta})^2}{d^2} = \frac{2*(1.35)^2*(1.96 + 1.28)^2}{0.83^2} = 56$$

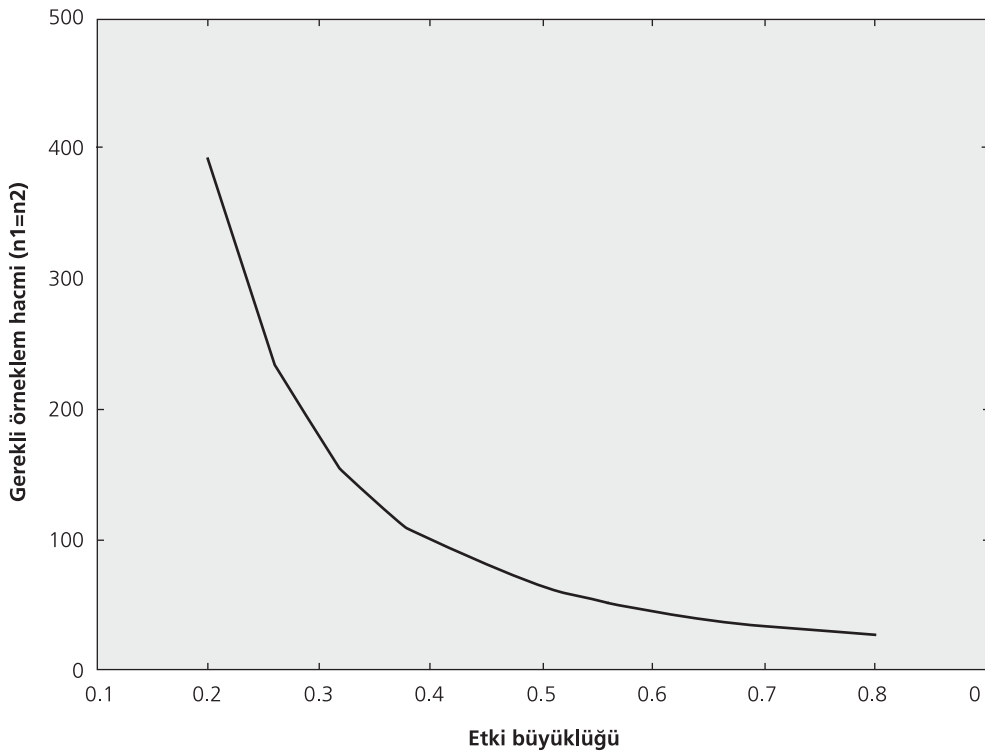
Çalışmanın gücünü %80'den %90'a çıkardığımızda her bir grup için n=56'şer hastanın alınması gerekecektir (**Şekil 1**).

c. I. tip hata α=0.05, II. tip hata β=0.2 (power=%80) ve gruplara dağıtım oranı 1:2 (1 tedaviye karşılık 2 kontrol) olacak şekilde her bir grup için kaç hasta almalıyız?

Gruplara dağıtım oranını 1:1'den 1:2'ye çıkardığımızda tedavi grubu için n=32, kontrol grubu için n=64 olmak



Şekil 1. Güç değerlerine karşılık gelen örneklem sayıları.



Şekil 2. Güç değerlerine karşılık gelen örneklem sayıları.

üzere toplam 96 hastanın alınması gerekecektir (Bu sayıların hesaplanmasında karmaşık olduğundan dolayı formüller verilmemiş olup hesaplama G-power adlı istatistiksel program ile yapılmıştır).

Etki büyüklüğü değerlerine karşılık gelen örneklem sayıları **Şekil 2**'de gösterilmiştir.

Sonuç

Klinik araştırmalarda güç (power) analiziyle örneklem sayısı belirlenmesinde, I. tip hata (α) düzeyi, II. tip hata (β) düzeyi, d ya da etki büyüklüğü, gruplara dağıtım oranı ve araştırma hipotezinin yönü etki eden ve mutlaka bilinmesi gereken bileşenlerdir.

Şekil 2'de görüldüğü gibi etki büyüklüğü ile örneklem sayısı arasında ters yönlü bir ilişki vardır. Belirli bir güç değerinde (örn. %80) etki büyüklüğü arttıkça çalışmaya dahil edilmesi gereken örneklem sayısı azalacaktır. Bu nedenle etki büyüklüğünün uygun değişken seçilerek belirlenmesi gerekir. Uygulamada 2 grup karşılaştırmalarında etki büyüklüğünün ≥ 0.50 olması gerekirken 3 ya da daha fazla grubun karşılaştırıldığı çalışmalarda ≥ 0.45 olmalıdır.

Çalışmada birden fazla değişken bağımsız olarak incelenecekse bu durumda her bir değişken için fark değerleri (d) ve etki büyüklükleri hesaplanır. Bu d ve EB değerleri için belirli bir α ve güç değeriyle örneklem sayıları saptanır. Saptanan en yüksek örneklem sayısı çalışmanın örneklem sayısını gösterir.

- α küçüldükçe güç azalır.
- Gruplar arası fark (d) ve etki büyüklüğü (EB) azalırsa güç azalır.

- Örneklem sayısı azalırsa güç azalır.
- Standart sapma büyüdükçe güç azalır.
- İki yönlü hipotezlerde güç tek yönlülere göre daha düşüktür.

Hayvan deneylerinde güç değerini çok yüksek ya da çok düşük tutarak çalışma planlamak gereksiz yere hayvan itlafına neden olur. Bu nedenle güç değerini optimum (%80) tutarak örneklem sayısını belirlemek etik açıdan oluşabilecek sakıncaları ortadan kaldırır.

Sonuç olarak, bilimsel çalışmalarda gereğinden az sayıda örneklem incelemek çalışma sonuçlarının gücünü düşürürken gereğinden fazla örneklem incelemek boş yere emek ve kaynak israfına neden olur. Bu nedenle çalışmanın başlangıcında çalışma hipotezine uygun olarak gerekli minimum örneklem sayısının belirlenmesiyle çalışma sonuçlarının güvenilirliğinin sağlanmasının yanı sıra kaynak israfının da önüne geçilir.

Kaynaklar

1. The CONSORT Statement. URL:<http://www.consort-statement.org>, Erişim Tarihi: 04 Ocak 2011.
2. Şenocak M. Klinik Biyoistatistik. Nobel Tıp Kitabevi, İstanbul, 2009.
3. Özdamar K. Modern Bilimsel Araştırma Yöntemleri, Kaan Kitabevi, Eskişehir, 2003.
4. Kirby A, Gebisi V, Keech AC Determining the sample size in a clinical trial, MJA, 177:256-257, 2002.
5. Hinton PR. Statistics explained. Taylor & Francis; 2 edition, NY, USA, 2007.
6. Sumbüloğlu V, Sumbüloğlu K. Klinik ve Saha Araştırmalarında Örneklem Yöntemleri ve Örneklem Büyüklüğü. Alp Ofset Matbaacılık Ltd. Şti, Ankara, 2005.