Turk J Public Health 2020;18(1):103-112
Available online at: http://dergipark.gov.tr/tjph
doi: 10.20518/tjph.602400

Teknik Not

Sağlık araştırmalarında güç analizinin önemi ve temel prensipleri: Tıbbi çalışmalar üzerinde uygulamalı örnekler

Oya Kalaycıoğlu[a], Serhat Emre Akhanlı[b]

[a] *Dr. Öğr. Üyesi, Bolu Abant İzzet Baysal Ü., Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim AD., Bolu, Türkiye*

[b] *Dr., İstatistik ve Veri Analizi Uzmanı, Public Health England, Londra, Birleşik Krallık*

Geliş tarihi: 06.06.2019, Kabul tarihi:23.12.2019

Öz

Amaç: Güç analizi ile örneklem büyüklüğü tahmini tıbbi araştırmacılar ve etik kurulu üyeleri tarafından çoğu zaman yanlış yorumlanan bir konudur. Bu çalışmada, istatistik alanı dışından olan araştırmacıların, örneklem büyüklüğü, etki büyüklüğü ve istatistiksel güç ile ilgili sorularının cevaplanması amaçlanmıştır. **Yöntem:** Bu kapsamda, farklı tıbbi araştırma düzenlerinde güç analizi gerçekleştirmek için kullanılacak etki büyüklüklerinin ne anlama geldiği ve nasıl hesaplandığı anlatılmıştır. Farklı etki büyüklüklerinde, farklı istatistiksel güç düzeylerinde ve %5 istatistiksel anlamlılık seviyesinde bağımsız gruplar için t-testi, tek yönlü ANOVA ve Ki-kare testleri için gerekli olan örneklem büyüklükleri GPower 3.1 programı kullanılarak hesaplanmıştır. **Bulgular:** Farklı istatistiksel testler için gerçekleştirilen güç analizleri, etki büyüklüğünün örneklem sayısı belirlemede ne derecede önemli rol oynadığını göstermiştir. Gereğinden az sayıda birim üzerinde deney yapılması gerçekte var olan önemli bir etkinin tespit edilememesine neden olabilirken, çok fazla sayıda örnek üzerinde test yapılması istatistiksel olarak anlamlı fakat gerçekte klinik olarak önemsiz bir etkinin tespit edilmesine yol açabilir. **Sonuç:** Sonuç olarak, bir araştırma kapsamında beklenen klinik anlamlılığı ortaya çıkarmak için alınması gereken etki büyüklüğü, araştırma hipotezi ile uyumlu belirlenmelidir. Çünkü araştırma sonucunda verilecek kararların niteliği ve klinik anlamlılığı örneklemin hangi etki büyüklüğüne dayanarak seçildiğine bağlıdır.
Anahtar kelimeler: Güç analizi, örnekleme, örneklem büyüklüğü, etki büyüklüğü, klinik anlamlılık

Sorumlu Yazar: Oya Kalaycıoğlu, Bolu Abant İzzet Baysal Üniversitesi Tıp Fakültesi Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Bolu, Türkiye. Telefon: +90 (374) 254 10 00. E-posta: oyakalaycioglu@ibu.edu.tr

Copyright holder Turkish Journal of Public Health

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. *This is an open Access article which can be used if cited properly.*

The importance and main principles of power analysis in health research: Application examples on medical case studies

Abstract

Objective: Estimation of sample size by power analysis is an oft-misinterpreted topic by medical researchers and ethical committee members. This study aims to answer the questions of non-statisticians regarding sample size, effect size and statistical power. **Material and Method:** In this context, the meaning of the effect size used in power analysis for different medical research designs and how they were calculated are explained. Sample sizes required for independent sample t-test, one-way ANOVA and Chi-Square tests were calculated using the GPower 3.1 software at different effect sizes, different statistical power levels and at 5% statistical significance level. **Results:** Power analyses which were performed for different statistical tests showed the importance of effect size in determining the sample size. Conducting an experiment on an insufficient number of units may result in the undetectability of a significant effect which exists. On the other hand, applying a test on too large a sample may result in the detection of a statistically significant, but in fact, clinically insignificant effect. **Conclusion:** In conclusion, the effect size that should be taken in order to reveal the expected clinical significance in a study should be determined in accordance with the research hypothesis. The quality and clinical significance of the decisions to be made at the end of the research depends on the sample size, which is dependent on the effect size.

Keywords: Power analysis, sampling, sample size, effect size, clinical significance.

Giriş

Bir sağlık araştırmasında kaç birim veya denekten veri toplanacağının belirlenmesi, başka bir deyişle örneklem büyüklüğünün hesaplanması araştırma henüz tasarım aşamasındayken gerçekleştirilmelidir. Bunun bilimsel, etik ve ekonomik açıdan birkaç nedeni vardır.[1] Bilimsel açıdan bakıldığında, çok az sayıda denek üzerinde test yapılması gerçekte var olan önemli bir klinik etkinin tespit edilememesine neden olabilirken, çok fazla sayıda denek üzerinde test yapılması istatistiksel olarak anlamlı, fakat gerçekte klinik olarak önemsiz bir etkinin tespit edilmesine yol açabilir. Bu noktada, araştırmacıların klinik anlamlılık ile istatistiksel anlamlılığı ayırt etmeleri oldukça önemlidir. Etik açıdan incelendiğinde, özellikle insanlar veya hayvanlar üzerinde girişimsel bir deney yapılıyorsa, gereğinden fazla deneği çalışmaya dahil etmek gereksiz zararlara neden olabilir. Bu durumun aksine, gereğinden az sayıda denek üzerinde çalışma yapmak da etik dışıdır, çünkü istatistiksel olarak güçlü olmayan bir çalışma ilgili alana bilimsel olarak nitelikli bir katkı sağlamayabilir ve bunun sonuncunda deneklere boşu boşuna zarar verilmiş olabilir. Son olarak, ekonomik açıdan bakıldığında, gereğinden fazla denek üzerinde test yapılması gereksiz maliyetlere yol açacak; az sayıda test yapılması ise eldeki kaynakların gereksiz tüketilmesine ve israfa neden olacaktır. Bu nedenlerle, birçok bilimsel dergi editörleri ve etik kurullar çalışma protokolüne dahil olan örneklem büyüklüğünün güç analizi ile hesaplanmasını talep etmektedir. Güç analizi, tıbbi araştırmalarda klinik olarak önemli bir etkiyi, belirli bir istatistiksel anlamlılık düzeyinde tespit etmek için gerekli olan en küçük örneklem

büyüklüğünün belirlenmesi için kullanılmaktadır.

Bu çalışmada, güç analizi ile örneklem genişliği belirlemenin temel prensiplerinden bahsedilmiştir. Ayrıca, bağımsız gruplar için t-testi, tek-yönlü ANOVA ve Ki-kare testleri için etki büyüklüğü kavramı anlatılmış ve farklı etki büyüklükleri için gerekli olan örneklem sayıları elde edilmiştir. Üç ayrı hipotetik örnek ile farklı istatistiksel testler için gerekli olan en küçük örneklem büyüklüğü belirlenmiş ve güç analizinin bir araştırmada nasıl raporlanacağı örnekler ile açıklanmıştır.

Yöntem

Güç analizinde belirlenmesi gereken faktörler

Güç analizi ile uygun örneklem büyüklüğünün hesaplanması, bazı faktörlerin seçimine dayanır. Bu faktörler Tablo 1'de özetlenmiştir:

Tablo 1. Güç analizi uygulaması için belirlenmesi gereken faktörler.

Faktör	Nedir?	Nasıl belirlenir?	Örneklem büyüklüğü ile ilişkisi
Anlamlılık düzeyi (α)	Anlamlı bir etki yokken, anlamlı bir etki bulma olasılığıdır.	Genellikle α=0.05 ve α=0.01 kullanılır.	Anlamlılık seviyesinin küçülmesi, örneklem büyüklüğünü arttırır.
Güç (1- β)	Gerçekte var olan anlamlı bir fark veya etkiyi bulma (önemli bir sonuç elde etme) olasılığıdır.	Genellikle 1-β=0.80 veya 1-β=0.90 kullanılır.	Gücün artması, örneklem büyüklüğünü arttırır.
Etki büyüklüğü	Gruplar arasında klinik olarak anlamlı olan en küçük farktır.	Varsa önceki yayınlardan, yoksa araştırmacının geçmiş bilimsel deneyimlerinden veya pilot çalışma yapılarak belirlenebilir.	Gruplar arasındaki küçük bir farkın veya etkinin tespit edilmesi için, büyük örneklem gerekir.
Varyans	Birimler arasındaki değişkenliktir. Nümerik verilerde standart sapmanın karesi olarak hesaplanır.	Varsa önceki yayınlardan, yoksa araştırmacının geçmiş bilimsel deneyimlerinden veya pilot çalışma yapılarak belirlenebilir.	Varyans büyüdükçe, verideki değişkenlik arttırdığından daha büyük bir örneklem boyutuna ihtiyaç duyulur.

Güç analizi gerçekleştirmeden önce, etki büyüklüğü, varsa önceki yayınlardan elde edilir. Daha önce yapılmış bir çalışmanın olmadığı durumlarda, etki büyüklüğü araştırmacının geçmiş bilimsel deneyimlerinden veya pilot çalışma yapılarak belirlenebilir.[2] Etki büyüklüğü iki veya daha fazla grup, örneğin deney ve kontrol grupları arasındaki farkı tespit etmenin bir ölçüsüdür. Bir sonucun sadece istatistiksel olarak anlamlı bulunması (yani p-değerinin 0.05'ten küçük bulunması), bu sonucun klinik olarak da önemli olduğu anlamına gelmez. Örneğin, standart tedavi

ile yeni bir tedavi yönteminin karşılaştırıldığı bir randomize klinik çalışmada, örneklem sayısı çok büyük seçilmiş ise iki tedavi yöntemi arasında oluşan çok küçük bir fark, α=0.05 anlamlılık düzeyinde istatistiksel olarak anlamlı bulunabilir (yani p<0.05 olarak hesaplanabilir). Ancak tek başına bu sonuç yeni tedavinin klinik etkisinin yeterli derecede önemli olduğunun bir göstergesi değildir. Bu nedenle araştırma sonuçlarında p değerleri ile beraber etki büyüklüğünün de bildirilmesi araştırmanın bilimsel niteliği açısından önem taşımaktadır.

Beklenen etki büyüklüğü küçük, orta veya yüksek olmak üzere araştırmacı tarafından ayarlanabilir. Ancak problem burada başlamaktadır. Zira azımsanmayacak sayıdaki araştırmada, ekonomik maliyet, zaman kısıtlılığı ve çok sayıda katılımcıya ulaşmadaki güçlükler nedeni ile küçük örneklem elde etmek için güç analizi yapılırken, etki büyüklüğü manipüle edilmektedir. Öte yandan, etki büyüklüğünün gerektiğinden yüksek olması istatistiksel olarak Tip II hata olasılığını (β), yani gerçekte yanlış olan bir sonucun doğru kabul edilme olasılığını arttırmaktadır. Bu nedenle, araştırma sonucunda elde edilebilecek klinik anlamlılığın ne olduğunu geçmiş çalışmalardan ve deneyimlerden yararlanarak irdelemek oldukça önemlidir.

Cohen (1988) farklı istatistiksel testler için etki büyüklüklerini küçük, orta ve büyük olarak sınıflandırmıştır.[3] En sık kullanılan testler için güç analizinde kullanılacak farklı etki büyüklükleri ve güç düzeylerinde, α=0.05 istatistiksel anlamlılık seviyesinde gerekli olan örneklem büyüklükleri aşağıda ayrı ayrı hesaplanmıştır. Örneklem hesaplamaları için GPower 3.1 programı kullanılmıştır.[4]

Yukarıdaki formülde k çapraz tablodaki toplam hücre sayısını, P_{0i} i numaralı hücrenin sıfır hipotezi altında popülasyon oranının beklenen değerini, P_{1i} ise araştırmacının bu oranlar ile ilgili beklentilerini temsil etmektedir. P_{0i} satır ve sütun toplamlarının çarpımının toplam hücre sayısına bölümü ile hesaplanır.

Etki büyüklüğünün hesaplanması

Bağımsız gruplar için t-testi

Bağımsız gruplar için t-testinde kullanılan etki büyüklüğü ölçütü Cohen'in d katsayısıdır.[3] Bu katsayı, iki bağımsız grup ortalaması arasındaki farkın, iki grubun varyans ortalamalarının karekköküne bölümü ile hesaplanabilir. Cohen'in d katsayısı formülü şu şekilde verilmiştir:

$$d = \frac{\overline{X}_{Grup1} - \overline{X}_{Grup2}}{\sqrt{\left(sd^2_{Grup1} + sd^2_{Grup2}\right)/2}}$$

Tek yönlü ANOVA

Bir araştırmada üç veya daha fazla grup ortalaması arasındaki fark karşılaştırılıyor ise, bu gruplar arasında en az birinin diğerlerinden farklı olduğunu saptamak için tek-yönlü ANOVA testi kullanılır. ANOVA testi için etki büyüklüğü Cohen'in f katsayısı ile hesaplanır[3]:

$$f = \sqrt{\frac{\sum_{j=1}^{k}\left(\mu_j - \mu\right)^2}{k\sigma^2_{hata}}}$$

Bu formülde μ_j her bir grubun popülasyon ortalamasını, μ genel ortalamayı, k toplam grup sayısını ve σ^2_{hata} da grup içi varyansı temsil etmektedir.

Ki-kare testi

İki veya daha fazla grupta, kategorik değişkenlerin gözlenme sıklığı arasında istatistiksel anlamlılık testi olarak ki-kare testi kullanılabilir. Ki-kare testi için etki büyüklüğü Cohen'in w katsayısı olarak tanımlanmıştır[3]:

$$w = \sqrt{\frac{\sum_{i=1}^{k}(P_{1i} - P_{0i})^2}{P_{0i}}}$$

Bulgular

Bağımsız gruplar için t-testi

Farklı etki büyüklüklerinde, farklı güç ve α=0.05 istatistiksel anlamlılık düzeylerinde iki bağımsız grubun ortalamalarının t-testi ile karşılaştırılması için gerekli olan örneklem büyüklükleri Tablo 2'de hesaplanmıştır.

Uygulama: Bağımsız gruplar için t-testi

Bir ilaç çalışmasında, kolesterol düşürücü iki ilacın etkisi karşılaştırılacaktır. Bunun için iki grup hasta randomize olarak iki ayrı ilaç grubuna atanacak, çalışma sonunda her hastanın LDL Kolesterol seviyesi ölçülecek ve gruplar arasındaki farklar karşılaştırılacaktır. Araştırmacılar, daha önceki deneyimlerinden iki grup ortalamaları arasında oluşabilecek 10 birimlik farkın klinik olarak anlamlı olduğu düşünmektedir. Ayrıca daha önceki benzer çalışmalarda birinci grubun standart sapması 18, ikinci grubun standart sapması 22 birim olarak bulunmuştur.

Bu bilgiler doğrultusunda Cohen'in d etki büyüklüğü şu şekilde hesaplanır:

$$d = \frac{10}{\sqrt{(18^2 + 22^2)/2}} = 0.5$$

Tablo 2'de de gösterildiği üzere 0.5 etki büyüklüğü t-testi için orta büyüklükte bir etki büyüklüğüdür. Bu durumda Tablo 2'deki değerler kullanılarak örneklem büyüklüğü hakkında şu yorum yapılabilir: "%80 istatistiksel güç ve α=0.05 anlamlılık düzeyinde, bağımsız gruplar için t-testine göre d=0.50 etki büyüklüğünü tespit etmek için gereken en küçük örneklem genişliği 128 (her grup için 64) olarak hesaplanmıştır". Aynı etki büyüklüğü ve anlamlılık düzeyinde %90 güç için bu sayı 170 olarak bulunmuştur.

Tek yönlü ANOVA

Farklı grup sayıları, güç düzeyleri ve etki büyüklükleri için gerekli olan minimum örneklem büyüklükleri Tablo 3'de verilmiştir:

Tablo 2. Bağımsız gruplar için t-testine göre α=0.05 istatistiksel anlamlılık düzeyinde farklı etki büyüklükleri için gerekli örneklem sayıları.

1-β	**Etki büyüklüğü (Cohen's d)**						
	Küçük ---------- Orta ---------- Büyük						
	0.2	**0.3**	**0.4**	**0.5**	**0.6**	**0.7**	**0.8**
0.80	787	351	198	128	89	66	51
0.90	1053	469	265	170	119	88	68

Tablo 3. Tek yönlü ANOVA testine göre α=0.05 istatistiksel anlamlılık düzeyinde farklı grup sayıları ve etki büyüklükleri için gerekli toplam örneklem sayıları.

Grup sayısı	1-β	Etki büyüklüğü (Cohen's f)						
		Küçük ---------- Orta ---------- Büyük						
		0.1	**0.15**	**0.2**	**0.25**	**0.3**	**0.35**	**0.4**
3	**0.80**	966	431	244	157	110	82	63
	0.90	1268	565	319	205	144	106	82
4	**0.80**	1094	488	277	178	125	93	72
	0.90	1421	634	358	231	161	120	93
5	**0.80**	1198	535	303	196	137	102	79
	0.90	1545	689	390	251	176	131	101

Uygulama: Tek yönlü ANOVA

Bir Faz 3 klinik ilaç araştırmasının bir parçası olarak, geliştirilen bir ilacın üç ayrı dozunun (düşük, orta, yüksek) karaciğer fonksiyonları üzerindeki etkisi incelenmektedir. Bu amaçla, randomize olarak ilaç verilmeyen bir kontrol grubu ve ilaç verilen üç ayrı grup olmak üzere toplam dört gruptaki deneklerin total bilirubin düzeyleri ortalamaları araştırma sonucunda karşılaştırılacaktır. Araştırmacı, deneyimlerine göre dört grubun total bilirubin düzeyleri ortalamalarının sırasıyla 9, 10, 11, 12 birim olacağını ve genel ortalamanın da 11 birim olacağını düşünmektedir. Grup içi standart sapmanın da 4 birim olduğu tahmin edilmektedir. Bu bilgiler doğrultusunda Cohen'in f etki büyüklüğü şu şekilde hesaplanır:

$$f = \sqrt{\frac{(9-11)^2 + (10-11)^2 + (11-11)^2 + (12-11)^2}{3(4)^2}} = \sqrt{\frac{4+1+0+1}{48}} = 0.35$$

Bu durumda Tablo 3'deki değerler kullanılarak örneklem büyüklüğü hakkında şu yorum yapılabilir: "%80 istatistiksel güç ve α=0.05 anlamlılık düzeyinde, tek yönlü ANOVA ile 0.35 Cohen'in f etki büyüklüğünü tespit etmek için gereken en küçük örneklem genişliği toplam 93 olarak hesaplanmıştır". Aynı etki büyüklüğü ve anlamlılık düzeyinde %90 güç için bu sayı 120 olarak hesaplanmıştır.

Ki-kare testi

Farklı serbestlik dereceleri, güç düzeyleri ve etki büyüklükleri için gerekli olan en küçük örneklem büyüklükleri Tablo 4'te verilmiştir:

Tablo 4. Ki-kare testine göre α=0.05 istatistiksel anlamlılık düzeyinde farklı serbestlik dereceleri ve etki büyüklükleri için gerekli örneklem sayıları.

Serbestlik derecesi[a]	**1-β**	**Etki büyüklüğü (Cohen's w)**				
		Küçük ---------- Orta ----------Büyük				
		0.1	**0.2**	**0.3**	**0.4**	**0.5**
1	0.8	785	196	87	49	31
	0.9	1051	263	117	66	42
2	0.8	963	241	107	60	39
	0.9	1265	316	141	79	51
3	0.8	1090	273	121	68	44
	0.9	1417	354	157	89	57
4	0.8	1194	298	133	75	48
	0.9	1541	385	171	96	62
5	0.8	1283	321	143	80	51
	0.9	1647	412	183	103	66

[a] Serbestlik derecesi(degrees of freedom, df) $=$ (satır sayısı -1) $\times$ (sütun sayısı -1)

Uygulama: Ki-kare testi

Bir araştırmada, iki farklı terapi yönteminin anksiyeteyi azaltıp azaltmadığı araştırılmaktadır. Araştırmacı önceki çalışmalardan standart terapi (kontrol grubu) yönteminin hastaların %40'ında anksiyeteyi azalttığını saptamıştır. Yeni terapi yönteminin ise (deney grubu) bu oranı %60'a çıkaracağı düşünülmektedir. Bu bilgilere dayanarak Tablo 5'teki kontenjans tablosu elde edilir.

Tablo 5'te toplam olasılıklar 1.0, kontrol ve deney grubu olasılıkları da ayrı ayrı 0.5 olarak verilmiştir. Bunun nedeni kontrol ve deney grubundaki örnek sayısının birbirine eşit olduğunun düşünülmesidir. Bu nedenle kontrol grubunda tedavinin etkili olduğu örneklerin oranı 0.5x0.4=0.2, deney grubunda tedavinin etkili olduğu örneklerin oranı ise 0.5x0.6=0.3 olarak hesaplanmıştır.

Tablo 5. Uygulama verisi içim kontenjans tablosu

	Kontrol grubu	Deney grubu	Sütun Toplamı
Terapi Etkili	0.2	0.3	0.5
Terapi Etkisiz	0.3	0.2	0.5
Satır Toplamı	0.5	0.5	1.0

Bu tablodaki değerlerle Cohen'in w etki büyüklüğü formülle şu şekilde hesaplanır:

$$w = \sqrt{\left(\frac{0.2 - 0.25}{0.25}\right)^2 + \left(\frac{0.3 - 0.25}{0.25}\right)^2 + \left(\frac{0.3 - 0.25}{0.25}\right)^2 + \left(\frac{0.2 - 0.25}{0.25}\right)^2} = 0.2$$

Bu hesaplamaya göre %80 istatistiksel güç ve α=0.05 anlamlılık düzeyinde, 1 serbestlik dereceli ki-kare testi kullanılarak gerekli en küçük örneklem genişliği toplam 196 (her grupta 88 olmak üzere) olarak hesaplanmıştır (Tablo 4). İstatistiksel güç %90 olarak alındığında bu sayı 263'e çıkmaktadır.

Örneklem büyüklüğü hesaplanırken dikkat edilmesi gereken diğer hususlar

Örneklem büyüklüğü tahmini her zaman yukarıda anlatıldığı kadar basit olmamaktadır. Araştırmalarda sıklıkla farklı türde birden çok değişken ayrı ayrı istatistiksel test ile incelenmektedir. Bu durumda, her bir değişkenin ölçüm türüne uygun olarak güç analizlerinin gerçekleştirilmesi ve elde edilen en büyük örneklem büyüklüğünün kullanılması önerilmektedir.[5]

Eğer araştırma gözlemsel bir çalışma ise, deneysel çalışmalara göre yaklaşık %20 daha fazla örneklem büyüklüğüne ihtiyaç duyulur.[6] Bunun nedeni, randomize olmayan çalışmalarda etki karıştırıcı faktörlerin (confounders) analize dahil edilmesi gerekliliğidir.

Klinik deneylerde, çalışmadan ayrılanlar veya kayıp veriler gibi faktörlerin istatistiksel gücü azaltmaması için güç analizinde hesaplanan örneklem büyüklüğünün %10-20 gibi bir oranda daha fazlası alınmalıdır. Eğer araştırmada tek gruptan toplanan pre-post ölçümler karşılaştırılacaksa (bağımlı gruplar t-testi gibi), kontrol grubu olan benzer bir çalışmaya kıyasla örneklem sayısı %50 azaltılabilir.[6]

İki veya daha fazla grup arasında yapılan karşılaştırmalarda, her gruptaki örnek sayısının eşit alınmasının maksimum istatistiksel gücü ürettiği bilinmektedir.[7] Ancak, deney grubu açısından kaynakların yetersiz olduğu ve deneyin kontrol grubuna göre daha maliyetli olduğu durumlarda, istatistiksel gücü koruyarak kontrol grubuna daha fazla katılımcı tahsis edilebilir.[8] Her gruptaki örneklem sayısının farklı olması durumunda parametrik testler tarafından ihtiyaç duyulan varyans homojenliği varsayımı olumsuz etkilenmektedir.[9]

Yukarıdaki uygulamalı örnekler ve açıklamalar sürekli değişkenlerin normal dağıldığı varsayımına dayanmaktadır. Bu noktada araştırmacılar "Normal dağılmayan değişkenler için non-parametrik testler kullanırsak örneklem büyüklüğü ne olmalıdır?" sorusunu akıllarına getirebilirler. Parametrik olmayan testler, örneklem veri sayısının 30'dan küçük olduğu veya veri dağılımının normal dağılım göstermediği sürekli değişkenler için uygulanan istatistiksel yöntemlerdir. Güç analizi ise ideal olarak veri toplama sürecinden önce yapılmalıdır. Bu durumda, araştırmacının veriyi toplamadan önce incelenecek değişkenlerin istatistiksel olarak nasıl bir dağılım göstereceği konusunda bir varsayımda bulunması çoğu zaman mümkün değildir. Bu nedenle sürekli değişkenler için güç analizinde çoğu zaman parametrik test kullanılacağı varsayılmaktadır. Ancak toplanacak verilerin normal dağılıma uyum sağlamayacağı daha önceki deneyimlere veya çalışmalara dayanarak öngörülmekteyse, ilgili parametrik test ile bulunan örneklem büyüklüğünün %15 fazlasının alınması önerilmektedir.[10]

Güç analizi ile hesaplanan örneklem sayısı, basit rastgele örnekleme yöntemi kullanılan araştırmalar için geçerlidir. Oysa ki, herhangi bir sağlık olayı hakkında gerçekleştirilen saha araştırmaları ve toplum taramalarında, basit rasgele örnekleme yöntemi ile tüm toplumu temsil etmek nerdeyse imkansızdır. Bu nedenle, bu tip araştırmalarda tabakalı örnekleme ve/veya küme örneklemesi sıklıkla kullanılmaktadır. Eğer araştırma kapsamında tabakalı örnekleme veya küme örneklemesi kullanılacak ise, araştırmaya dahil edilmesi gereken örneklem sayısı desen etkisi nedeni ile büyümektedir.[11] Çünkü, tabakalı veya küme örneklem ile elde edilecek verilerin varyansının, basit rasgele örnekleme ile elde edilebilecek verilerin varyansından daha fazla olması beklenmektedir. Literatürde farklı örnekleme yöntemleri kullanıldığında desen etkisinin 1 ila 3 arasında değiştiği[12], küme örneklemesi yöntemi kullanıldığında ise desen etkisinin 1.5 – 2 kat arasında olacağı bildirilmiştir.[6] Desen etkisinin büyüklüğü çalışmadan çalışmaya değişmektedir, çünkü desen etkisi küme içi korelasyon katsayısına (δ) ve her kümedeki örneklem sayısına (n) bağlıdır:

$$\text{Desen etkisi} = 1 + \delta\,(n - 1).$$

Küme içi benzerlik ve dolayısı ile kümeler arasındaki fark arttıkça desen etkisi artmaktadır. Bu durum araştırmaya dahil edilmesi gereken örneklem sayısını arttırmaktadır.

Tartışma

Araştırmacılar ve etik kurul üyeleri örneklem, araştırmanın gücü ve etik konular üzerinde bilinçli yargılarda bulunacaklar ise örneklem büyüklüğü ile ilgili teknik konulara hakim olmalıdırlar. Randomize klinik deneylerin kalitesini denetlemek için geliştirilen ve 25 maddeden oluşan CONSORT (Consolidated Standarts for Reporting Trials)[13] kontrol listesinde yer alan "Örneklem sayısı nasıl belirlendi?" sorusu da klinik deneylerde örneklem sayısının ve güç analizinin önemini göstermektedir.[14]

Örneklem seçiminde ana konu, etki büyüklüğünü araştırma hipotezi ile uyumlu olarak belirlemektir, çünkü araştırma sonucunda verilecek kararların niteliği ve klinik anlamlılığı örneklemin hangi etki büyüklüğüne dayanarak seçildiğine bağlıdır. Gruplar arasında büyük bir farkı, yani büyük etki büyüklüğünü tespit etmek için küçük bir örneklem büyüklüğü yeterli iken, gruplar arasındaki küçük bir farkın anlamlılığı ancak büyük bir örneklem ile tespit edilebilir.

Gereğinden az sayıda birim üzerinde deney yapılması gerçekte var olan önemli bir etkinin tespit edilememesine neden olabilirken, çok fazla sayıda örnek üzerinde test yapılması istatistiksel olarak anlamlı fakat gerçekte klinik olarak önemsiz bir etkinin tespit edilmesine yol açabilir. Bilimsel olarak nitelikli olamayan bu sonuçlar ise hem etik hem de ekonomik açıdan olumsuz sonuçlar doğurmaktadır. Bu nedenle, en uygun örneklem büyüklüğünü hesaplarken etki büyüklüğünü manipüle etmek yerine, geçmiş çalışmalardan ve deneyimlerden klinik anlamlılığı irdeleyerek, etki büyüklüğünün hesaplanması tavsiye edilmektedir. Ayrıca, araştırma sonuçları raporlanırken, p değerleri ile birlikte elde edilen etki büyüklüklerinin de sunulması çalışmanın bilimsel geçerliliğini ve niteliğini arttırmak açısından oldukça önemlidir.

Yazar katkısı: Yazarların makaleye katkısı eşit düzeyde olmuştur.

Çıkar çatışması: Yazarların herhangi bir çıkar çatışması ya da anlaşmazlık durumu bulunmamaktadır. Çıkar çakışması oluşturacak herhangi bir dış katkı veya parasal destek alınmamıştır.

Finansal destek: Çalışma için herhangi bir maddi destek alınmamıştır.

Kaynaklar

1. Hickey GL, Grant WG, Dunning J, Siepe M. Statistical primer: sample size and power calculations—why, when and how?. European Journal of Cardio-Thoracic Surgery 2018; 54(1):4-9.
2. Kul S. Klinik Araştırmalarda Örnek Genişliği Belirleme. Ekstraplevral 2011;11:129-132.
3. Cohen, J. Statistical power analysis for the behavioural sciences. New York: Academic Press, 2013.
4. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. Behaviour research methods 2009; 41(4): 1149-1160.
5. Guo Y, Logan HL, Glueck DH, Muller KE. Selecting a sample size for studies with repeated measures. BMC Medical Research Methodology 2013; 13(1):100.
6. Suresh KP, Chandrashekara S. Sample size estimation and power analysis for clinical research studies. Journal of Human Reproductive Sciences 2012; 5(1):7.
7. Liu X. Statistical power and optimum sample allocation ratio for treatment and control having unequal costs per unit of randomization. Journal of Educational and Behavioral Statistics 2003; 28(3): 231-248.
8. Torgerson D. Designing randomised trials in health, education and the social sciences: an introduction. New York: Palgrave Macmillan, 2012.
9. Rusticus S, Lovato C. Impact of Sample Size and variability on the Power and Type I Error Rates of Equivalence Tests: A Simulation Study. Practical Assessment, Research & Evaluation 2014; 19(11)
10. Lehmann EL. Nonparametrics: Statistical Methods Based on Ranks. Springer, 2006.
11. James EB, Joe WK, Chadwick CH. Organizational Research: Determining appropriate sample size in survey research. Inf Technol Learn Performance J. 2001;19: 43–50.
12. Gene S. Sample size and design effect, NYS DOH, Presented at Albany Chapter of American Statistical Association, 2001.
13. Moher D. CONSORT: an evolving tool to help improve the quality of reports of randomized control trials. Consolidated standards of reporting trials. JAMA 1998; 279:1489-91.
14. Süt N. Klinik araştırmalarda örneklem sayısının belirlenmesi ve güç (power) analizi. RAED Dergisi 2011; 3(1): 29-33.