

Hayvan Deneylerinde Verimliliği Artıracak Deneysel Tasarımları ve Denek Sayısı

Experimental Designs and Sample Sizes for Increased Efficiency in Animal Experimentation

Öz

Hayvanlar üzerinde yürütülen deneysel araştırmalar bilimin en temel aşamasıdır. İnsanlar ve hayvanlar (özellikle memeliler) arasındaki anatomik ve fizyolojik benzerlikler araştırmacıları yeni tedavileri önce hayvan modelleri üzerinde değerlendirmeye sevk ettiği için, bu tip deneyler insan sağlığı ile ilgili bilimsel araştırmalarda da hayati önem arz etmektedir. Buna karşın hayvanların yanlış uygulama veya eksik planlama gibi nedenlerle gereksiz yere telef edilmesi şiddetle karşı çıkılan bir durumdur. Deneysel araştırmalardan etkin bir şekilde yararlanabilmek ve etik ihlallerin önüne geçebilmek için planlama aşamasında dikkat edilmesi gereken birçok husus mevcuttur. Bu çalışmada, bu gibi hususlar ana hatlarıyla ortaya konmuş ve uygun denek sayısını belirlemede araştırmacılara yardımcı olabilecek çeşitli metodolojilerden bahsedilmiştir. Ayrıca yaygınca kullanılmayan, ancak hayvan deneylerinde denek sayısını ve maliyeti en aza indirebilecek ve iç ve dış geçerliliği artıracak deneysel tasarımlarına değinilmiştir. Son olarak, hayvan deneylerinin aynı materyal ile tekrarlanması ve metaanalizlerle değerlendirilmesinin önemi vurgulanmıştır.

Anahtar Sözcükler: deneysel tasarımları; hayvan deneyleri; hayvan modelleri; örneklem genişliği; randomizasyon

Abstract

Animal experimentation is the most fundamental stage of science. Such trials are also crucial in the scientific research for human health as the anatomical and physiological similarities between humans and animals (particularly mammals) lead researchers to try new treatments first on animal models. However, wasteful use of animals due to reasons like malpractice and poor planning has led to heavy criticism. There are many issues that should be taken into consideration during the planning phase in order to make effective use of experimental research and prevent ethical violations. This study outlines such issues and discusses various methodologies that can help researchers in determining the appropriate sample size. It also touches on uncommon experimental designs that are capable of minimizing costs and sample sizes and increasing internal and external validity in animal experimentation. Finally, importance of repeating animal tests with the same materials and their evaluation by meta-analysis is highlighted.

Keywords: animal experimentation; animal models; experimental design; randomization; sample size

Handan Ankaralı¹, Seyit Ankaralı²

¹ İstanbul Medeniyet Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

² İstanbul Medeniyet Üniversitesi, Tıp Fakültesi, Fizyoloji Anabilim Dalı

Geliş/Received : 22.04.2019
Kabul/Accepted: 22.05.2019

DOI: 10.21673/anoloklin.556640

Yazışma yazarı/Corresponding author

Handan Ankaralı
İstanbul Medeniyet Üniversitesi, Tıp Fakültesi,
Biyoistatistik ve Tıp Bilişimi Anabilim Dalı,
İstanbul, Türkiye
E-posta: handanankarali@gmail.com

ORCID

Handan Ankaralı: 0000-0002-3613-0523,
Seyit Ankaralı: 0000-0003-3752-0846

GİRİŞ

Deney (laboratuvar) hayvanlarının kullanımı bilimin gelişiminde hayati bir rol oynamaktadır. Bu hayvanlar, çeşitli biyolojik ve genetik araştırmalarda, tıbbi müdahale araştırmalarında, ilaç araştırmalarında, davranış modellerinde ve eğitim alanlarında sonuçları insanlara genellemek amacıyla kullanılmaktadır. Geçmişten günümüze bu araştırmalar, yeni ilaç ve tıbbi araçların geliştirilmesi için kilit çalışmalar olarak önemini korumaktadır (1). Böylece insana özgü sorunların çözümünde, ortaya çıkabilecek risk minimize edilmektedir. Ancak insan sağlığını iyileştirirken ve hayat kalitesini artırırken hayvan haklarını gözetmeye ve verilecek zararın en az olmasına azami ölçüde dikkat etmek gerekir. Bu nedenle hayvan deneylerinin planlama aşamasında standartlar geliştirilmeli, tekrarlanabilirliği sağlayacak şeffaflıkta denemeler düzenlenmeli ve bilim dünyasında yaygın olarak kullanılan kılavuzlar takip edilmelidir (*ARRIVE—Animals in Research: Reporting In Vivo Experiments* ve *GSPC—Gold Standard Publication Checklist*).

DENEYSEL ARAŞTIRMALARDA PLANLAMA

Bilimsel araştırmalarda planlama, en önemli aşamadır. Planlama sürecinin titizlikte ele alınması ve çalışmayı yapıyor muşcasına detaylı olarak düşünülmesi ve gerektiğinde B, C planlarının oluşturulması büyük önem taşır. Deney hayvanları üzerinde yapılan çalışmaların standardize edilebilmesi için daha planlama aşamasında, çalışmanın amacına göre seçilecek biyolojik hayvan modeline, kullanılacak hayvan türüne, yöntemlere, denek sayısına ve en az denekle en etkili sonuca ulaşmayı sağlayacak deney tasarımına karar verilmelidir. Ayrıca uyulması gereken etik ilkeler ile (istenmeyen etkilere göre standardizasyonu sağlamak için) körleme veya randomizasyon prensipleri de bu aşamada belirlenmelidir (2–4).

1. Hayvan deneylerinde uyulması gereken kurallar

Hayvan deneyleri planlanırken yaygın olarak 3R kuralına uyulması tavsiye edilir. Ancak son zamanlarda farklı farklı çalışmalarda adı geçen kurallar bir araya getirildiğinde toplam 7 adet alt kuraldan oluşan 7R kuralı önerilmektedir (5,6). Deney planlamasında dikkate alınan bu kurallar kısaca aşağıdaki gibi tanımlanır.

Birinci R (Replace—yerine geçirme): Mümkünse deney planlanırken hayvan modeli yerine kullanılabilen *in vitro* bir model veya aynı hedefe ulaştıracaksa filogenetik skalada daha yüksekte yer alan hayvan yerine daha aşağıda bulunan bir hayvan kullanılmalıdır.

İkinci R (Reduce—azaltma): Deneyin en az sayıda hayvan ile yapılması gerektiğini ifade eder. Bunun için istatistiksel yöntemlerle denek sayısına karar verilmeli ve deney süresince hayvan sayısında azalma sağlayabilecek tüm olası çevresel unsurlar gözden geçirilmelidir.

Üçüncü R (Refine—arındırma): Hayvanların doğumlarından deneylerde kullanımlarının sonlandırılmasına kadar geçen süreçte rahatlarını (*animal welfare*) ve mümkün olduğu kadar az acı duymalarını sağlamak için gerekli önlemlerin alınmasını ifade eder.

Dördüncü R (Responsibility—sorumluluk): Araştırmacının sorumluluklarının farkında olmasını ifade eder.

Beşinci R (Rehabilitate—iyileştirme): Deney sonunda eğer mümkün ise hayvanın yeniden sağlığına kavuşturulmasıdır.

Altıncı R (Reproducibility—tekrar üretilebilirlik): Farklı laboratuvarlarda benzer deney maddeleriyle aynı metot uygulanarak aynı deney sonuçlarının elde edilebilmesidir.

Yedinci R (Rigor—titizlik): Sağlam ve tarafsız deneysel tasarım, yöntem seçimi ve raporlamada titizlik, deney ve sonuçlarında tam şeffaflık gibi bilimsel konulara azami dikkat edilmesidir.

Yedinci R olan *rigor* kavramında, *uygun deney tasarımının seçimi* diğer konular arasında kaldığı için yeterince vurgulanmamaktadır. Uygun deney tasarımının belirlenmesi ve bu tasarıma göre veri analizinin yapılması, hem denek sayısını belirleme aşamasında hem de mevcut varyasyon kaynaklarını doğru analiz etmek açısından deneyin en önemli iki bileşenidir. Bu nedenle bu yazıda, *rigor* ile genel bilimsel titizlik vurgulanmış, doğru tasarım ve veri analizi için yeni bir R (*right analysis—doğru istatistiksel analiz*) tanımlanarak literatüre bir R kuralı daha eklenmiştir.

Sekizinci R (Right analysis—doğru istatistiksel analiz): Bir deneyin planlama aşamasından veri analizine kadar olan süreçte aynı istatistik grubu veya danışmanı ile çalışılmalıdır. Uygun deney tasarımının seçiminde

veya başka bir deyişle doğru istatistik modelinin belirlenmesinde mutlaka istatistik danışmanının bilgisine başvurulmalıdır. Deney koşullarında mevcut olan varyasyonların nasıl analiz edileceğine dair tanımlanmış çok sayıda deney tasarımı mevcuttur. Ayrıca istatistiksel analiz yapacak danışmana, ölçülebilecek ve ölçülemeyecek varyasyon kaynakları açıklanarak deney tasarımının doğru seçiminde yardımcı olmak gerekir.

Deney tasarımının seçimi ve buna bağlı olarak yapılacak veri analizi, sonuçlarla doğrudan ilişkili olduğu için özel bir önem verilmesi gerekir. Günümüzde yapılmış ve yayımlanmış birçok deneysel araştırmada “tesadüf parselleri deney tasarımı” (*randomized plot design*) veya “eşleştirilmiş gruplar düzeni” (*paired or matched design*) kullanılmıştır. Bu deney tasarımlarından elde edilen veriler, *independent samples t-test*, *Mann-Whitney U test*, *one-way ANOVA*, *Kruskal-Wallis test*, *paired samples t-test* veya *Wilcoxon signed-rank test* ile analiz edilmektedir. Ancak konuyla ilgili literatür taramalarında deney tasarımının hatalı seçildiği çok sayıda yayımlanmış araştırma olduğu ve verilerin basit yöntemlerle analiz edildiği görülmüştür. Bu ise tekrar üretilebilirliği olumsuz etkileyen ve hatta yanıltıcı sonuçlara neden olabilen ciddi durumların başında gelir.

Bu koşullar dikkate alındığında, istatistiksel metodoloji ve planlamanın iyi bir istatistik uzmanı tarafından ele alınmasının büyük bir önem arz ettiği söylenebilir.

2. Biyolojik hayvan modelleri

Literatürde tanımlanmış çeşitli deneysel hayvan modelleri bulunmaktadır. Bu modeller biyolojik hayvan modelleri olarak da adlandırılabilir. Uygun model seçiminde, deneyin primer amacı rol oynar. Ayrıca seçilen hayvan modelinde hangi hayvan türü ile çalışılacağı da önemlidir. Bu seçim, elde edilecek sonucun insanlarda kliniğe en iyi şekilde uyarlanabilmesi için büyük önem taşır. Ayrıca tür seçiminde maliyet, nesiller arası süre ve daha alt sınıf hayvan kullanılmasına dikkat edilmelidir.

Genel olarak literatürde aşağıda açıklanan dört tip biyolojik hayvan modeli ile karşılaşılır:

İndüklenmiş veya transgenik modeller: Çeşitli müdahale veya maddelerle hayvanlarda deneysel olarak oluşturulan hastalıkların hedef organizmada görülen-

lere benzemesine dayanır. Tür seçimi gerekir.

Spontan modeller: Genetik olarak kendiliğinden gelişen ve insanlardaki hastalığa benzer patoloji ve semptomlar gösteren hayvanlarda geliştirilen modellerdir. Burada söz konusu hayvanlar deney materyali olarak kullanılır.

Negatif modeller: Spontan modellerin tersi olup bazı hastalıkların hiç görülmediği türler içeren modellerdir.

Olası (*orphan*) modeller: Bazı hastalıklar önce hayvanlarda gelişebilir. Bu hastalıkların insanda gelişmesi durumunda kurulacak modellerdir.

Ayrıca insan sağlığını koruma ve iyileştirme ile dolaylı ilişki içinde bulunan ve kendisinden çeşitli ürünler elde edilen hayvanlar üzerinde yapılan “seleksiyon çalışmaları” da ayrı bir model olarak düşünülebilir.

3. Kontrol grubu

Bilimsel deneylerde, deneysel hataları ve önyargıları ortadan kaldırmak için kontrol grubu kullanılmalıdır. Böylece yabancı değişkenlerin etkisini en aza indirmek veya istenmeyen değişkenlerin etkisini gözlemlemek mümkün olacaktır. Çalışmalarda genellikle bir veya birkaç çeşit kontrol grubuna ihtiyaç duyulur. Amaca göre, aşağıda tanımlanan kontrol gruplarından biri veya birkaçı seçilebilir (7).

Pozitif kontrol: Tedavinin beklenen etkisini gösterir. Deney grubunda etkisi araştırılan muamele (*treatment*—işlem, metot, ilaç vd.) sonuçlarda düzelmeye neden olurken, kontrol grubunda da aynı yönde bir değişiklik olması durumudur. Örneğin, klasik tedavi verilen grup.

Negatif kontrol: Deneyin istenen sonucunu vermeyecek şekilde tasarlanmış grubudur. Örneğin, saf su veya serum fizyolojik verilen grup.

Sahte (*sham*) kontrol: Bir prosedür veya test maddesinin, sahtesi ile taklit edildiği gruptur. Plasebo kullanımını farmasötik çalışmalarda sağlanan sahte kontrole bir örnektir.

Araç: Test edilecek madde serum fizyolojik dışında bir çözücüde çözülüyor ise bu çözücünün tek başına etkisini incelemek amacıyla yeni bir grup oluşturulmalıdır. Örneğin bazı maddeler sadece alkol ya da DMSO gibi kimyasallarda çözülürler.

Kontrol grupları iç (*internal*) veya dış (*external*) grup olarak temin edilebildiği gibi eş zamanlı (*concurrent*) ya da eski çalışmalardan temin edilen (*historical*) gruplar da olabilir. Hangi kontrol grubunun kullanılması gerektiği sorusuna “araştırma için tanımlanan hipotezler” cevap verir.

Yapılacak deney için kontrol gruplarının gerekliliği ve çeşidi planlama aşamasında netleştirilmelidir. Kontrol gruplarının dış kaynaklı olması veya eski çalışmalardan temin edilmesi mümkün ise, denek israfının ve gereksiz hayvan kullanımının önüne geçilmiş olacağı için, bu seçenek etik açıdan daha uygun olacaktır.

4. Pilot çalışma

Hayvan deneyleri genel olarak, (i) pilot ve açıklayıcı deneyler ve (ii) bir hipotezi test etmek için kurulan deneyler olmak üzere, iki ana başlık altında toplanabilir (8).

Deneye başlamadan önce, bir ya da bir dizi pilot çalışma yapılmalıdır. Bu çalışma, ana deneyin nasıl sonuçlanacağını öngörmek, ölçülecek olan değişkenlerin ölçümünün kesinliğini bilmek ve önerilen deneylerin lojistiğini kontrol etmek için yapılır. Ayrıca pilot çalışmalar, araştırmacıların el becerisi ve bilgi düzeyini artırabilir ve de standart sapma ve etki büyüklüğü gibi örneklem büyüklüğünün hesaplanması için gerekli bilgileri sağlayabilir (9). Yanı sıra, yeni hipotezler üretmek için keşif yapma imkanı sunar. Pilot çalışmalarındaki örneklem büyüklüğü, araştırmacının önceki deneyimlerine ya da (önceden veri mevcut olmadığında) çalışma tahminine dayanır. Hipotez testi yapılmaz. Geçerli ve güvenilir olmaları halinde pilot çalışmalardan elde edilen veriler ana çalışmaya dahil edilebilir (10).

5. Örneklem büyüklüğü

Örneklem büyüklüğü; deney tasarımına, hayvan modeline, hayvan türüne, çalışılan sonuç değişkenlerine, grup sayısına, primer hipoteze (veya en önemli hipotezlere) göre değişkenlik gösterebilir. Ayrıca deney hayvanları ile ilgili çalışmalarda etik hususlar da dikkate alınarak gereksiz hayvan kullanımından kaçınılmalıdır. Her araştırma tasarımı (veya istatistik modeli) için ayrı bir *power* (güç) analizi yapılmakta, hatta aynı tasarımda amaç ve ölçüm skalası değiştiğinde gereken örneklem genişliği de değişebilmektedir. Örneklem genişliği ge-

reğinden az alındığında sonuçların gücü düşmekte ve var olan farklılıkları anlamlı bulmak zorlaşmaktadır. Buna karşın büyük örneklemle çalışıldığında ise etik sorunlarla karşılaşmaktadır. Ayrıca klinik çalışmalar için örneklem genişliği hesabı aydınlatılmış bir konu iken, hayvan denekler üzerindeki çalışmalar için bu konuda aynı netlik sağlanamamıştır. Deney hayvanları üzerinde yapılan araştırmalar incelendiğinde, ağırlıklı olarak örneklem genişliği hesabında iki farklı yöntem kullanıldığı görülmüştür. Birinci yöntem, *power* analizi olup daha bilimsel ve daha çok tercih edilen bir yöntemdir (9). Bu nedenle etik kurullar, dergi editör ve hakemleri hayvan deneylerinde denek sayısının nasıl belirlendiğini ve *power* analizi yapıp yapılmadığını sorgulamaktadır. *Power* analizi neticesinde elde edilen denek sayısı bilimsel açıdan gerekli olan minimum sayıdır. Bu sayı, bir dizi bilimsel bilgi yardımıyla hesaplandığı için, deneyden elde edilecek sonuçların genellebilirlik özelliğini doğrudan etkiler.

5.1. Güç (*power*) analizi

Etki büyüklüğü ve sinyal: gürültü (*effect size & signal: noise*): Çalışmalarda etki büyüklüğü, “klinik olarak önemli veya bilimsel açıdan ilginç olan fark veya değişim” olarak tanımlanır. Burada tanımlanan etki büyüklüğü, öznel bir tahmin değeridir ve klinik veya biyolojik olarak önem taşıyan değişimi ifade eder (9,10). Bu nedenle çalışma hipotezini ortaya koyan araştırmacı(lar) tarafından bu değer ne olması ve ne kadar büyüklükte olması gerektiği belirlenmelidir.

Gürültü ise primer hipotezde yer alan sonuç (*outcome*) veya sonlanım değişkenindeki (*end-point variable*) varyasyonu ifade eder ve genellikle standart sapma ile ölçülür.

Etki büyüklüğü, sinyal ve gürültü kavramları aşağıdaki gibi formüle edilir:

$$\text{standart dize edilmiş etki büyüklüğü} = \frac{\text{grup ortalamaları (veya diğer istatistikler) arası fark}}{\text{standart sapma}} = \frac{\text{sinyal (etki)}}{\text{gürültü (varyasyon)}}$$

Bu oran $\geq 1,5$ olduğunda gruplardaki denek sayısının ≤ 10 olması yeterli görülmektedir. Mesela bu oran 2’ye çıktığında her grupta 6 denek olması yeterli bulunmuştur. Bu sayılar güç eğrileri yardımıyla elde edilmektedir (10).

Uygun grup örneklem genişlikleri *nQuery Advisor*, *G*Power* veya *Minitab* programları ile hesaplanabilir. Ayrıca www.biomath.info gibi, çok sayıda hesaplama sitesi mevcuttur. Ancak bunların büyük bir bölümü veri analizinde basit istatistiksel yöntemler kullanılacağı zaman örneklem genişliği için kullanılabilir. Eşitlikteki standart sapma, üzerinde çalışılan değişkenin (özellik) örneklem içindeki varyasyonu gösterir.

Sonuç değişkenleri (outcomes): Deneylede üç tip sonuç değişkeni tanımlanır:

- Dikotomus (kategorik veya nominal değişken)
 - yaşıyor/öldü, kan grupları gibi
- Sürekli değişken (*ratio, interval, ordinal*)
 - ağırlık, hacim = ratio
 - sıcaklık = interval
 - herhangi bir ölçek (*ordinal*)
- Zamana bağlı bir olayın görülme durumu
 - 6 ay sonra ölüm oranı, 1 yıl sonra tekrarlamaya sıklığı gibi

Dikotomus değişkenler hız veya oran olarak ifade edilir. Bu tip ölçümlerdeki amaç grupların oranları arasındaki değişimleri incelemektir. Sonuç değişkeni dikotomus olduğu zaman örneklem genişliğini hesaplamak için aşağıdaki bilgilere ihtiyaç duyulur.

- Çalışmanın gücü (1-β) (β, ikinci tip hata yapma olasılığı)

Genellikle prior güç >%80 kabul edilir

- Anlamlılık düzeyi (α) (α, birinci tip hata yapma olasılığı)

Genellikle %5 kabul edilir, yani p<0,05

- Araştırmacının ortaya koymak istediği biyolojik/klinik olarak önem taşıyan gruplar arası fark

Bu fark pilot çalışmalardan, önceki çalışmalardan veya araştırmacının tecrübe ve bilgilerinden elde edilebilir.

Bir araştırmada birden çok sonuç değişkeni varsa, son maddede gerekli olan bilgi en önemli sonuç değişkeni için elde edilir. Birden fazla sonuç değişkeninin hepsi araştırmacı için aynı önemi taşıyor ise, hepsi için ayrı ayrı örneklem genişliği hesaplanır ve çıkan en büyük sayı deneyde kullanılır. Bu tip çalışmalarda bir gruptaki örneklem genişliği aşağıdaki formül ile hesaplanır:

$$n=C \frac{p_c q_c + p_e q_e}{d^2} + \frac{2}{d} + 2$$

p_c : kontrol grubunda olayın görülme sıklığı,
 q_c : kontrol grubunda olayın görülmemeye sıklığı,
 p_e : deney grubunda olayın görülme sıklığı,
 q_e : deney grubunda olayın görülmemeye sıklığı,
 d : iki grup arasında araştırmacının incelemek istediği fark.

C, sabit terim olup aşağıdaki tablodan hesaplanır.

α→ 1-β↘	0,05	0,01
0,8	7,85	11,68
0,9	10,51	14,88

Sonuç değişkeni sürekli ve gruplar bağımsız iken (iki grupta farklı denekler yer aldığı zaman) örneklem genişliğinin hesabında aşağıdaki adımlar izlenir. Standart sapma (*standard deviation—SD*) ve standardize edilmemiş etki büyüklüğü, iki grup arasında araştırmacının incelemek istediği fark pilot veya önceki çalışmalardan elde edilir. *Prior* güç ve birinci tip hata kararlaştırılır. Bu tip çalışmalarda bir gruptaki örneklem genişliği aşağıdaki formülle hesaplanır.

$$n=1 + 2xC \left(\frac{SD}{d} \right)^2$$

Sonuç değişkeni sürekli ve gruplar bağımlı iken (tek gruptan iki farklı zamanda veya iki ayrı muamele altında ölçüm alındığı durumlar) örneklem genişliğinin hesabında aşağıdaki formül kullanılır.

$$n=2 + C \left(\frac{SD}{d} \right)^2$$

İki sürekli değişken arasındaki korelasyon çalışmalarda örneklem genişliği formülü, izleyen eşitlikte tanımlanmıştır.

$$n=2 + \frac{4C}{\left[\ln \left(\frac{1+r}{1-r} \times \frac{1-r_0}{1+r_0} \right) \right]^2}$$

r : başlangıçta öngörülen veya test edilmek istenen ilişki miktarı

r_0 : belirlenen korelasyon katsayısı

Eğer deneydeki amaç bir olayın bir hayvan kolonisinde veya grubunda görülme sıklığının anlamlı olup olmadığını incelemek ise örneklem genişliği aşağıdaki eşitlikle hesaplanır.

$$n = \frac{\log a}{\log p}$$

Burada p etkilenmeme veya görülmeme sıklığıdır ve genellikle %5 alınır.

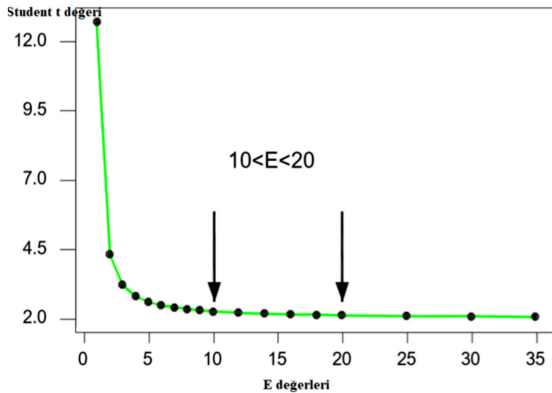
5.2. Örneklem büyüklüğünü belirlemenin bir başka yöntemi: kaynak denklemi (resource equation)

Kaynak denklemi az sayıda denek kullanmak zorunluluğu olduğu durumlarda, uygun örneklem genişliğini belirlemek için tercih edilen bir yaklaşımdır (9). Hesaplaması kolay ancak $power$ analizine göre daha kaba bir yöntemdir. Etki büyüklüğü veya standart sapma hakkında bir öngörü veya bilgi yoksa çok sayıda sonuç değişkeni üzerinde çalışılacaksa veya karmaşık deney tasarımları kullanılacaksa bu yöntem tercih edilmelidir. Kaynak denklemi:

Hata serbestlik derecesi = $E = (\text{Toplam hayvan sayısı}) - (\text{Grup sayısı})$

$10 < E < 20$ (Görsel 1)

E değerinin 10 ile 20 arasında olması yeterli görül-



Görsel 1. Kaynak denklemi

mektedir. Eğer $E < 10$ çıkarsa deneye hayvan eklemek daha önemli sonuç alma olasılığını artıracaktır. Belirlenen E değerine karşılık gelen toplam denek sayısı araştırmada kullanılır. Denekler ucuz veya nesiller arası süre kısa ise E değeri daha büyük alınabilir.

6. Geçerlilik

Denemelerin iç geçerliliği (internal validity): Denek sayısına bilimsel yöntemlerle karar verildikten sonra, yanlışlık kaynaklarını elemine etmek için randomizasyon ve körleme prensiplerine başvurulduğunda ve materyalin homojen olmadığı durumlarda homojen hale

getirecek istatistiksel araçlardan (bloklama, tabakalama, kovaryans analizi gibi) yararlandığında deneyin iç geçerliliği artacaktır. Ayrıca deneyin yanlışlıklardan uzak olması iç geçerliliğinin yüksek olduğu yönünde bir güvencedir.

Denemelerin dış geçerliliği (external validity): Deneylerin tekrarlanabilirlik özelliğini destekleyen her türlü girişim denemelerin dış geçerliliğini artırır. Örneğin, faktöriyel dizayn, bloklama gibi analizlerle diğer faktörler üzerinde (cinsiyet, tür, yaş, kafes, zaman, yer etkileri gibi) tekrarlanabilirlik sağlanarak deney sonuçlarının genellenebilirliği genişletilmiş olur.

7. Randomizasyon

Seçim yanlışlığını gidermeye yönelik randomizasyon ilkesi, deneklere uygulanacak işlemlerin rastgele dağıtılmasını ya da bir işlemin uygulanması için deneklerin popülasyondan rastgele seçilmesini gerektirir. Böylece zaman içinde veya muamele sırasında meydana gelebilecek değişiklikler kontrol altına alındığı gibi denekler arası farklılığa neden olan ancak etkisi tam olarak bilinmeyen veya ölçülmeyen/ölçülemeyen özelliklerin etkisi de gruplara rastgele dağıtılmış olur. Randomizasyon, kartlar üzerine denek numaraları yazıldıktan sonra rastgele numaralar arasından seçim yaparak, rastgele sayılar tablosu kullanılarak veya *Excel*, *Minitab* gibi programların random fonksiyonu kullanılarak gerçekleştirilebilir. Randomizasyon ve körleme kullanılmayan hayvan çalışmalarında gruplar arası farkların anlamlı bulunma olasılığı bu yöntemlerin kullanıldığı çalışmalara kıyasla daha yüksektir (8). Randomizasyon deneklerin popülasyondan rastgele seçiminde uygulanacaksa, söz konusu popülasyon, üzerinde durulan özellikler bakımından benzer bireylerden oluşmalı ve varyasyonun kaynağı genetik farklar veya ölçülemeyen biyolojik farklar olmalıdır. Burada tanımlanan popülasyon, çalışılacak konu ile ilgili bütün bireyleri içeren veya başka bir deyişle araştırma bulgularının genelleneceği hedef kitleyi tanımlamaktadır. Buna karşın, deneklere uygulanacak işlemlerin rastgele dağıtılması daha sık uygulanan bir ilkedir.

Veri analizi aşamasında, uygun istatistiksel modele karar vermek için deneklerin seçim şekli büyük önem taşır. Deneklerin popülasyondan tamamen rastgele yöntemlerle seçilerek muamele gruplarına atanması

durumunda deney tasarımı, “tamamen rastgele deney tasarımı” olarak adlandırılır. Ancak popülasyonda homojenliği bozan etkenler söz konusu olduğunda ise bu etkenlere göre bloklayarak yapılarak muamele etkileri “blok deney tasarımları”na göre analiz edilir.

8. Performans ve saptama yanlılığı ve körleme (blinding/masking)

Hayvan deneylerinde performans ve saptama yanlılığı, deneyi yürüten araştırmacıların ve/veya veri analizi yapan istatistik uzmanının bilerek ya da bilmeyerek çalışma sonucunu etkilemesi olarak tanımlanabilir. Ortak bir ifadeyle “tespit yanlılığı” (*ascertainment bias*) olarak da adlandırılabilir (12). Bu tür yanlılıkların önlenmesinde körleme tekniği kullanılır ve yanlılığa sebep olabilecek araştırmacıların, çalışma gruplarından hangisinin deney hangisinin kontrol grubu olduğunu bilmemesi sağlanır. Böylece deneyin iç geçerliliğinin artması sağlanmış olur.

9. Deney tasarımları (experimental designs)

Eksik veya hatalı tasarlanmış deneyler, gereksiz hayvan kullanımı nedeniyle etik ihlallere, zaman kaybına ve ekonomik kayba yol açar. Hatta hipotezlerle ilgili yanlış sonuçlar ortaya çıkmasına neden olabilir. Yayımlanan birçok hayvan deneyinde planlama ve veri analizi aşamalarında ciddi eksiklikler veya hatalar mevcuttur (13). Çeşitli varyasyon kaynaklarını modele alarak analiz eden, etkisi incelenen muamele hakkında daha güçlü sonuçlar veren ve denek sayısını daha verimli kullanmayı sağlayan deney tasarımlarının genel özellikleri aşağıda açıklanmıştır. Bu tasarımların tarım ve endüstri alanında yaygınca, günümüze kadarki hayvan deneylerinde ise nadiren kullanıldığı gözlenmiştir. Hem etik hem de ekonomik açıdan birçok kısıtlılık altında planlanan deney hayvanı içerikli araştırmaların başta insan sağlığına hizmet amaçlı yürütüldüğü göz önüne alındığında, söz konusu sorunları minimize edecek bu tasarımların kullanılması tavsiye edilmektedir (14).

9.1. Tamamen rastgele deney tasarımları (tesadüf parselleri deney tasarımı—completely randomized designs)

Planlaması ve sonuçların analizi kolay olduğu için literatürde en sık kullanılan deney tasarımıdır. Denekler

muamele gruplarına tamamen rastgele atanır. Gruplardaki denek sayısı iki kata kadar farklı olabilir ve bu koşulda sonuçlar olumsuz etkilenmez. Denekler veya çevre koşulları heterojen yapıdaysa sonuçların gücü düşük çıkar. İyi planlandığında iç geçerliliği yüksek olur, ancak dış geçerliliği düşüktür. İki veya daha çok grubun karşılaştırılmasında *one-way ANOVA* modeli veya Kruskal–Wallis testi kullanılmaktadır (iki grup olduğunda *independent samples t-test* veya Mann–Whitney U testi özel bir durum olarak geliştirilmiştir) (15).

9.2. Blok deney tasarımları (block designs)

Blok kavramı, etkisi araştırılan muamele dışında, etkisi olduğundan şüphelenilen veya etkisi bilinen karıştırıcı (*confounding*) faktörleri ifade eder. Bu kavramı barındıran çok sayıda deney tasarımı mevcuttur. Bu tasarımlar tarımsal ve endüstriyel alanda yaygınca kullanılmasına karşın etik hususların büyük önem taşıdığı laboratuvar hayvan deneylerinde maalesef asgari düzeyde kullanılmıştır (16).

9.2.1. Tek blok içeren denemeler

Blok içindeki denek sayısının etkisi incelenen muamele sayısına eşit olma durumu, farklı olma durumu ve blok ile muamele etkileşiminin önem taşıdığı durumlara göre farklı isimler altında adlandırılır. Tasarımın en basit şeklinde, blok faktörünün her seviyesinde muamele sayısı kadar denek bulunur. Ayrıca materyalin homojenliğini bozan faktör sayısı (blok) bir tanedir. Mesela çalışılan özellik bakımından farklı batında doğan yavrular arasında anlamlı fark olması bekleniyorsa batın, blok faktörü olarak adlandırılır ve birinci batın, blok faktörünün birinci seviyesini, ikinci batın ikinci seviyesini, ..., gösterir. Çalışmada bir negatif kontrol grubuna ilaveten iki ilaç dozu daha denenecek ise etkisi araştırılan faktörün adı “ilaç” olur ve bu faktörün kontrol, birinci doz, ikinci doz gibi 3 ayrı seviyesi vardır. Ayrıca her bir batında 3'er adet doğan deney hayvanına 3 muamele rastgele atanır ve birinci batında doğanlar (birinci blok) üzerinde 3 muamele etkisi denenmiş olunur. Benzer şekilde diğer batınlar (blok seviyeleri) için de benzeri şekilde işlem yapılır. Tasarımın planlaması bu şekilde olup her muamele her blok seviyesinde sadece bir deneye uygulanır. Bu deney tasarımı tamamen rastgele deney tasarımı olarak düşünüldüğü zaman, batın etkisi incelenemez veya göz ardı

edilmiş olur. Bu ise deneyin dış geçerliliğini düşürür ve tekrarlanmasını engelleyebilir. Blok denemelerinde az sayıda denek kullanılarak çok sayıda etkenin etkisi daha doğru değerlendirilmiş olur.

Bazı durumlarda her blok seviyesinde muamele sayısından daha fazla denek yer alabilir. Bu durumda blok ile muamele etkileşimi merak ediliyor ise veri analizine dahil edilir, merak edilmiyor veya anlamlı bulunmuyor ise etkileşimsiz model kullanılır. Örneğin bir hiperlipidemi ilacının iki farklı dozunun tavşanlarda kolesterol üzerindeki etkisi incelenecek olsun. Kolesterol bakımından cinsiyetler arasında fark olması beklenen bir durum olduğu için cinsiyet blok faktörü, ilaç ise etkisi araştırılan faktör (muamele) olarak adlandırılır. Bu deneyde araştırmacı, ilaç dozları arasındaki kolesterol farkının cinsiyete göre değişmeyeceğini öngörüyor veya değişip değişmeyeceğini merak etmiyor veya değişmediğini test ederek görüyor ise, blok ile muamele etkileşimi modele dahil edilmez, sadece blok faktörünün etkisi ortadan kaldırılır. Ancak eğer etkileşimi merak ediyor veya modele aldığı anda etkisini anlamlı buluyorsa, “ilaç dozları arasındaki kolesterol farkının cinsiyete göre değiştiğini” göstermiş olur.

Bazı durumlarda aynı ebeveynlerden birden fazla bütünde doğan veya farklı ebeveynlerden doğan denekler aynı deneme içinde yer alabilir. Bazen de aynı deneklerin sağ ve sol taraflarına, ya da çift organlarına farklı muameleler uygulanmak durumunda kalınabilir. Birinci durumda ebeveynler, ikinci durumda ise sağ ve sol ölçümler bağımlı olup “blok” olarak modele alınır (17).

Bu açıklamalardan görüleceği üzere, sonuç üzerine etkisi araştırılan faktör dışında etkisi bilinen, incelenmek istenen veya ölçülmediği için değerlendirilemeyen faktörlerin planlama ve analiz aşamalarında çok iyi düşünülmesi gerekir.

9.2.2. En az iki blok içeren denemeler (Latin square designs)

Latin kare tasarımlarında, deney materyalinin homojenliğini bozan iki etken vardır ve bunlar 1. blok ve 2. blok olarak adlandırılır. Bu blokların düzeyleri aynı sayıda veya farklı olabilir. Klasik Latin kare tasarımında iki blok faktörüne ait seviye sayısı ile muamele faktörünün seviye sayısı aynı olmalıdır. Denemedeki amaç, etkisi merak edilmeyen veya bilinen bu iki bloktan kaynaklanan varyasyonu elemine etmektir. Böylece

etkisi incelenecek olan muamele hakkında daha güçlü sonuçlar elde edilir. Veri analizi açısından bir bloklu tasarımlara benzer. Muamelelerin sadece iki düzeyi karşılaştırılmak isteniyorsa veya sekizden fazla düzeyi var ise deney Latin kare düzeninde planlanamaz.

Etkisi aynı anda incelenecek muamele sayısı ve blok sayısı iki ise veya üç blok faktörüne ilaveten bir muamele var ise Latin kare düzeni, Greko-Latin kare ve dikdörtgen yapıdaki düzen Youden kare düzeni olarak adlandırılır (16). Mesela depresyon tedavisinde bir antidepresan ilacın 4 farklı dozunun etkisi 4 farklı hayvan üzerinde denemek istensin. Bu denemede hayvanlar blok olarak kabul edilir. Belirli aralıklarla ilk 3 doz uygulandıktan sonra en az bir hayvanda yorgunluk veya başka yan etki ortaya çıkarsa deney sonlandırılır. Bu durumda her hayvana 3 doz verilmiş olur. Dört hayvan, 3 periyot ve 3 doz olduğu için tasarım dikdörtgen yapıda olup Youden kare deney tasarımına ait model ile analiz edilir.

9.2.3. Eksik blok tasarımları (incomplete block designs)

Faktör sayısı arttıkça deneylerde yeterince homojen yapıda muamele kombinasyonları bulmak zorlaşır. Bu durumda homojen yapıda olan muamele kombinasyonlarından sadece bir kısmı incelenebilir. Buna “eksik bloklar” adı verilir. Bu deneyde, etkisi incelenecek muamele düzeylerinden bir kısmı bir blokta, diğer kısmı başka bir blokta denir (14).

9.2.4. Dört-köşe deney tasarımları (lattice designs)

Faktör sayısının ve doğal olarak faktör (muamele) kombinasyonlarının çok olduğu ve her faktör (veya kombinasyonları) için aynı derecede bilgi edinmek istendiği durumlarda eksik bloklar düzeninden olan dört-köşe tasarımlar kullanılmalıdır. Yukarıda anlatılan eksik bloklarda veya bölünmüş parseller deney tasarımlarında bazı faktörler (veya kombinasyonlar) hakkında elde edilen bilgiler daha az güvenilir olmaktadır. Dengeli veya dengesiz yapıda olabilen dört-köşe tasarımlar mevcuttur. Yeterince homojen materyal bulunmadığı takdirde dengesiz dört-köşe deney tasarımlarından yararlanılır (14). Örneğin, depresyon modeli çalışmasında sabah, öğle ve akşam (bloklar) üç farklı antidepresan ilacın üç farklı sıcaklık ortamında verilmesi amaçlanmış olsun. Denemede (3 ilaç x 3 sıcaklık =) 9 muamele kombinasyonu bulunmaktadır. Bu kom-

binasyonlar en az bir defa denenecek şekilde ratlara sabah, öğle ve akşam uygulanır.

9.2.5. Change-over denemeleri

Change-over denemeleri, Latin kare denemelerinin tekrarlı ölçüm içeren formudur. Bu denemelerde her deneye birden fazla muamele çeşidi belli bir sırayla ileriye dönük olarak uygulanır. Deneyde bireyler arası farklar (biyolojik/genetik) blok etkisi olarak alınır ve bu farkların etkisi giderilir. Change-over deney tasarımları içinde yaygın kullanım alanı bulan 2x2 cross-over denemesidir. Ayrıca bu denemeler tekrarlı ölçümlerde Latin kare tasarımları, switch-over denemeleri veya time-series denemeleri olarak da adlandırılmaktadır. En önemli avantajları; ölçülmeyen veya diğer denemelerde hata varyansı olarak kabul edilen denekler arası farklılığın etkisinin giderilmesi, az sayıda denekle yürütülebilmesi ve maliyeti düşürmesidir. Bu denemelerde, etkisi incelenecek muamelenin çeşitleri her birey üzerinde denenir ve muamele çeşitlerinin dağıtımı Latin kare düzeninde yapılır. Muamele çeşitlerinin verilmiş sırasının etkisi de dikkate alınarak giderilir. En yaygın kullanıldığı alanlar, tarımsal araştırmalar, psikolojik deneyler, biyoşeyerlik deneyleri, diyet uygulamaları ve kronik hastalıklarla ilgili klinik araştırmalardır. Ayrıca muamelelerin birbiri ile etkileşime geçmemesi için bekleme süresinin (washout period) iyi belirlenmesi gerekir ve kısa olması istenir (15).

9.3. Faktöriyel deney tasarımları

Faktöriyel deney tasarımlarında, etkisi merak edilen en az iki muamelenin (işlem, tedavi, ilaç gibi) sonucu üzerine direkt etkileri ve etkileşim (interaksiyon) etkileri birlikte araştırılır. Bu amaçla bir muamelenin her bir seviyesi, diğer muamele ya da muamelelerin tüm seviyelerinde tekrar edilerek muamele kombinasyonları oluşturulur. Bu kombinasyonlarda en az iki denek yer almalıdır. Etkisi incelenecek muameleler dışında modelde kovaryant ve/veya blok faktörleri de yer alabilir. Bu deney tasarımları ise “tesadüf parsellerinde faktöriyel denemeler,” “tesadüf bloklarında faktöriyel denemeler,” “faktöriyel kovaryans analizi” gibi isimlerle anılır. Ayrıca faktöriyel denemeler Latin kare düzeninde de planlanabilir. Faktöriyel denemelerin en basit formu iki faktörlü varyans analizi modelleridir. Bu tasarımlarda daha az sayıda denekle daha fazla bilgi elde edilir. Faktöriyel tasarımlarla elde edilmiş sonuçlarda

iç geçerlilik düşerken dış geçerlilik artar. Bu modeller, çevre ve muamele, genetik ve çevre gibi önemli etkileşimleri de analiz etmeye olanak sağlar (16).

9.4. Bölünmüş parseller deney tasarımları (split-plot designs)

En az iki faktörün etkisinin incelendiği ve görünüş itibarıyla faktöriyel deney tasarımına benzeyen bölünmüş parseller deney tasarımında, faktörlerden biri ve faktör etkileşimleri hakkında daha hassas ve güçlü bilgiler elde etmek amaçlanır. Etkisi daha az merak edilen faktörün her bir seviyesine, etkisi daha çok merak edilen ikinci faktörün seviyeleri atanır. Üst seviyede yer alan faktöre ana parsel, alt seviyede yer alan faktöre ise alt parsel adı verilir. Ayrıca faktörlerin uygulama kolaylığı açısından bir farklılık varsa da bu düzen tercih edilebilmektedir. Uygulaması zor olan faktör ana parsel, kolay olan ise alt parsel atanır. Buradaki parsel kavramı deney materyali anlamında kullanılır. Bölünmüş parseller düzeni de blok düzeninde veya Latin kare düzeninde tertiplenebilir veya kovaryant değişkeni içerebilir (14,16). Mesela ortam sıcaklığı ile üç farklı besin çeşidinin kilo artışı üzerindeki etkisi incelenmek istendiğinde, ana parsellere farklı sıcaklık düzeyleri, alt parsellere ise her bir sıcaklık için üç ayrı besin çeşidi atanır. Bu işlem, tersi durumu gerçekleştirmekten daha kolaydır. Bu deney sonucunda sıcaklıklar arası fark, besin çeşitleri arası fark ve besin çeşitleri arasındaki farkın sıcaklığa göre değişme durumu (veya tersi) incelenmiş olur.

9.5. Tekrarlanan deney tasarımları

Bir veya daha fazla faktör hakkında daha fazla genellenen bilgi edinmek amacıyla deneyin farklı yerlerde, zamanlarda, materyallerle veya laboratuvarlarda tekrarlanması durumudur. Bu düzen deneylerin tekrarlanabilirlik özelliğini incelemek amacıyla yapıldığı için, etkileşim teriminin anlamlı bulunmaması incelenen muamelenin farklı koşullarda aynı sonucu vereceğini (tekrarlanabilirlik özelliğinin var olduğunu) gösterir (16). Bu deney tasarımı daha çok yeni bir madde/metot/ilâç deneneceği zaman tercih edilmelidir. Tekrarlanan deneyler daha çok bölünmüş parseller düzenine uyarlar. Tesadüf parselleri, blok düzenleri veya Latin kare düzeninde gerçekleştirilebilirler. Mesela, dört farklı antiepileptik ilacın etkisi üç ayrı laboratuvarda denenmiş olsun. Burada ilk adım her bir laboratuvarı

kendi içinde değerlendirmektir. Daha sonra laboratuvarlar arası fark ve ilaçlar arası fark ile laboratuvar etkileşimi hesaplanır. Deney tasarımında, laboratuvarlar ana parsel, ilaçlar alt parsel konumundadır. Bu deneyde en önemli konu, ilaçlar arası farkın laboratuvarlara göre değişip değişmediğini incelemek olduğu için laboratuvarların rastgele seçilmesi gerekir.

9.6. İç-içe deney tasarımları (nested designs)

İç-içe deney tasarımında, faktöriyel tasarımlardaki gibi etki incelenen en az iki faktör vardır. Ancak faktöriyel düzende bir faktörün seviyeleri diğer faktörün tüm seviyelerinde kullanılırken, iç-içe tasarımlarda, içe yerleştirilen faktörün seviyeleri dışa yerleştirilen faktör seviyelerinde tekrarlanmaz (14,16). Mesela bir deneyde, üç farklı erkek rat ile çiftleşen dört dişi rattan doğan yavruların doğum ağırlıkları arasında farklılık olup olmadığı merak edilmiş olsun. Burada her erkek rata ait yavruların ağırlıkları arasında erkek farklılığının etkisi yoktur. Bu deneyde de aslında bir tekrarlanma vardır. Tekrarlama erkek ratlar arasındadır. Ancak bölünmüş parsellerde farklı laboratuvarlarda aynı ilaçlar denenirken (*crossed factor*), iç-içe tasarımlarda farklı erkek ratlarla çiftleşen dişi ratlar farklıdır (*nested factor*). Bu deney tasarımındaki temel amaç denek sayısını artırmak amacıyla işe karışan dış faktörlerin (burada erkek ratlar) etkilerini elemine etmek ve böylece iç faktörün (dişi rat faktörü) etkisini daha güvenli olarak test edebilmektir.

9.7. Ardışık deney tasarımları (sequential designs)

Örneklem genişliğinin kaç olması gerektiği konusunda bir ön bilgi veya hesaplama olanağı olmadığı durumlarda tercih edilen bir deney tasarımıdır. Ardışık olarak örneklem genişliği değişebilir ve her değerlendirme aşamasında bir analiz yapılarak öngörülen hipotezler test edilir. Önceden belirlenen sonlandırma kriterine ulaşıldığı an deneme sonlandırılır. Bu kriter maksimum örneklem genişliğine erişilmeden primer hipotezin reddedilmesi şeklinde tanımlanabilir. Maksimum örneklem genişliğine ulaşılsa dahi hipotez reddedilmemiş ise çalışma sonlandırılır ve muamelenin etkisi olmadığı sonucuna varılır.

10. Kovaryans analizi

Kovaryans analizi faktöriyel deney tasarımları, blok denemeleri gibi genel bir tanımlamadır. Kovaryant,

etkisi olduğundan şüphe edilen veya etkisi bilinen sürekli tipteki değişkene verilen isimdir. Sağlık araştırmalarında genellikle yaş, vücut kitle indeksi gibi özellikler birçok biyolojik ölçümü etkiler. Bu nedenle bu tip değişkenler kovaryant olarak alınır. Blok denemelerinde "blok" faktörü de kovaryant gibidir. Ancak blok, kategorik yapıda bir değişken veya faktöre verilen bir isimdir. Kovaryant içeren modellerin adlandırılmasında mutlaka "kovaryans" kavramı yer alır. Bu denemelerdeki amaç, blok denemelerinde olduğu gibi kovaryant değişkeninin etkisini incelemek değil, kovaryantın etki karıştırıcı bir rolü olduğu için bu etkiyi gidermektir. Kurulan istatistiksel modellerde kovaryant/bloklara göre düzeltme yapılarak asıl etkisi incelenecek muamelenin düzeltilmiş etkileri elde edilir. Modellerde kovaryant sayısı birden fazla olabilir.

11. a ve p değerleri

Hipotez testine başlamadan önce belirlenen "a değeri" ve hipotez testi sonucunda hesaplanan "p değeri" birinci tip hata veya yanlış olma olasılığı olarak tanımlanır. Ancak a nominal, yani önceden kabul edilen yanlış olma olasılığı, p ise hesaplanan yanlış olma olasılığı olarak adlandırılır. p değeri testler sonucunda elde edildikten sonra başlangıçta kabul edilen yanlış olma olasılığı (a) ile karşılaştırılır. $p < a$ ise araştırılan ilişki veya fark istatistiksel olarak anlamlıdır; bu durumda bu karara varılırken %5'ten az yanlış söz konusudur. $p > a$ ise araştırılan ilişkinin anlamlı olmadığı sonucuna varılır. Bütün hipotez testlerinde p değerine bakılarak sonucun yorumlanışında bu adımlar izlenir (18).

12. İstatistiksel anlamlılık ve biyolojik anlamlılık

İstatistiksel olarak anlamlı bulunan sonuçlar her zaman biyolojik anlam taşımayabilir. İstatistiksel olarak anlamlı bulunmayan bir sonuç da biyolojik olarak anlamlı olabilir. Örneklem genişliği arttıkça farkları anlamlı bulma ihtimali artacağı için, büyük örneklerde sonuçların istatistiksel olarak anlamlı bulunma ihtimali artar. Ancak istatistiksel olarak anlamlı her sonucun biyolojik bir önemi olmayabilir. Biyolojik önemi olmadığı araştırmacılar tarafından ifade edilen farkların/ilişkilerin istatistiksel olarak anlamlı çıkması, kullanılacak bir bilgi değildir. Öte yandan araştırmacının önemseydiği ancak istatistiksel değerlendirme sonucunda anlamlı çıkmayan farklar veya ilişkiler söz

konusu olduğunda ise, test sonuçlarında hesaplanan anlamlılık düzeyi (p değeri) gözden geçirilmeli ve bu değer 0,05 ile 0,20 arasında olması durumunda ya deney yeniden planlanmalı ya da deneyin eksiklikleri “tartışma ve öneriler” bölümünde vurgulanarak elde edilen sonuçlar kısıtlı bir duruma genellenmelidir.

Biyolojik anlamlılığın olduğu ancak istatistiksel anlamlılığın olmadığı durumlarda, şu hususlar tekrar gözden geçirilmeli ya da deney tekrarlanmalıdır: Deney tasarımının uygunluğu, örneklem genişliğinin doğru tespit edilip edilmediği, örnekleme yönteminin uygunluğu, doğru istatistiksel testin kullanılıp kullanılmadığı.

SONUÇ VE ÖNERİLER

Bilim için yapılan deneylerde kullanılan hayvanların telef olmasını önlemek ve elde edilecek sonuçların güvenilir olmasını sağlamak amacıyla daha planlama aşamasında uygun denek sayısına ve deney tasarımına bilimsel yöntemlerle karar verilmesi gerekmektedir. Deney tasarımını yanlış seçmek gereken denek sayısının da yanlış belirlenmesine neden olacağı için, önce tasarıma karar vermek, sonra da uygun örneklem genişliğini hesaplamak gerekir.

Bu yazıda değinilen deney tasarımları kendi içinde de oldukça geniş varyasyonlara sahiptir. Her deneyde tasarım planlama ve veri analizi aşamalarında istatistiksel danışmanlık alınması şarttır. Deney tasarımlarına dair ayrıntıların ve hesaplama adımlarının verilmemiş nedeni, söz konusu tasarımlar hakkında sadece farkındalık uyandırmaktır. Deney tasarımları hakkında ayrıntılı bilgi için Hinkelmann ve Kempthorne’un eseri (15) gibi kaynakların iyice incelenmesi önerilir. Yukarıda değinilen deney tasarımlarından elde edilen verilerin analizinde SAS, Minitab, Stata, NCSS gibi paket programlar kullanılmakta ve SPSS programında da birçok tasarım modeli Syntax editörüne yazılarak analiz edilebilmektedir.

Örneklem genişliği *power* analizi ile belirlendiğinde, birçok hayvan deneyi için denek sayısı genellikle etik açıdan kabul edilemeyecek kadar büyük çıkar. Bu durumda araştırmacılar, örneklem genişliğini kaynak denklemi veya ardışık deney tasarımları ile belirleyebilir. Ayrıca etkisi incelenecek muamele sayısı azaltılabilir, 10 farklı doz yerine 5 farklı doz incelemek

gibi. Küçük örneklemle çalışıldığında, karşılaştırılan gruplar arası farkların istatistiksel olarak anlamlı çıkma olasılığı düşeceği için, grup farklarının biyolojik/klinik anlamlılığının da dikkate alınması gerekir. Özellikle p değeri 0,05 ile 0,20 arasında ise farkın biyolojik olarak önem taşıyıp taşımadığı tartışılmalıdır. Bu öneriler dışında, hayvan deneylerinde veya az denekle çalışmak zorunda kalınan diğer denemelerde, aynı hipotezin farklı zamanlarda veya farklı araştırmacılar tarafından tekrarlanması ve araştırma sayısının 4’ün üzerine çıkması durumunda metaanaliz ile sonuçlar yeniden değerlendirilebilir. İlaç araştırmaları gibi bir muamelenin çok sayıda çeşidi (mesela aynı ilacın farklı dozları veya aynı amaçla kullanılan farklı ilaçlar gibi) mevcut olduğunda, birbirinden farklı çeşitleri karşılaştırmış araştırmalar bir araya getirilerek *network* metaanalizi ile mevcut olmayan karşılaştırmalar yapılabilir. Yanı sıra metaanalizler yardımıyla birçok ölçüme dair tanımlayıcı istatistikler çıkarılabilir ve bu değerler yeni araştırmaların örneklem genişliği hesabında veya karşılaştırmalarında kullanılabilir.

Örneklem büyüklüğü, deney tasarımları ile direkt ilişkilidir ve kullanılacak deney tasarımına göre değişiklik gösterir. Ayrıca grup sayısı, çalışılan değişkenlerin tipi ve primer hipotezler örneklem büyüklüğünün belirlenmesinde etkili faktörlerdir. Bu nedenle konu ile ilgili tek bir formül veya yaklaşım bulunmamaktadır. Araştırmacılar örneklem büyüklüğünü belirleme aşamasında mutlaka uzman desteği almalıdır. Ancak araştırmacıların bu konuda bilmeleri veya sunmaları gereken en önemli bilgi, etki büyüklüğüdür. Bu değer, araştırmacının önemseddiği biyolojik/klinik fark veya ilişkidir. Bu bilgi istatistik uzmanı tarafından bilinemez.

Ayrıca yayımlanmış araştırmalardaki deneysel koşulların ayrıntılı olarak tanımlandığı ve ham verilerin aktarıldığı çeşitli veri tabanları geliştirilerek bunlar diğer araştırmacıların kullanımına sunulabilir. Böylece pozitif, negatif ve *sham* kontrol gruplarına dair veriler rahatlıkla temin edilebilir.

Etik gerekçeler nedeniyle az denekle çalışma zorunluluğu olan hayvan deneylerinden elde edilecek bulguların genellenmesi konusunda da titizlik gösterilmelidir. Bu amaçla, başlangıçta hedef popülasyon sınırları net olarak çizilmeli, deneyin tüm adım ve koşulları çeşitli kılavuzlar eşliğinde ayrıntılı olarak rapor edilmelidir.

KAYNAKLAR

1. de Aguilar-Nascimento JE. Fundamental steps in experimental design for animal studies. *Acta Cir Bras.* 2005;20(1):2-7.
2. Naycı S. Hayvan çalışmalarında planlama ve uygun model ve denek seçimi nasıl yapılmalı? *Türk Toraks Derg.* 2013;14(Ek 2):10-4.
3. Kumar AHS. Effectively communicating the 5R's (replace, reduce, refine, reuse, and rehabilitate) of research ethics, biomedical waste, personalized medicines and the rest. *J Nat Sci Biol Med.* 2012;3(1):1-2.
4. Demirkılıç U. Deneysel modelleme, araştırmanın planlanması, değerlendirme ve yayın etiği. *Türkiye Klin Kalp Damar Cerrahisi—Özel Konular.* 2013;5(3):14-7.
5. Aske KC, Waugh CA. Expanding the 3R principles: more rigour and transparency in research using animals. *EMBO Rep.* 2017;18(9):1490-2.
6. Tüfek H, Özkan Ö. 4R rule in laboratory animal science. *Commagene J Biol.* 2018;21(1):55-60.
7. Johnson PD, Besselsen DG. Practical aspects of experimental design in animal research. *ILAR J.* 2002;43(4):202-6.
8. Bebar V, Luyten D, Heard K. Emergency medicine animal research: does use of randomization and blinding affect the results? *Acad Emerg Med.* 2003;10(6):684-7.
9. Jaykaran C, Kantharia ND. How to calculate sample size in animal studies? *J Pharmacol Pharmacother.* 2013;4(4):303-6.
10. Bek Y, Güneren E, Hökelek M. Etkili örnek büyüklüğünün belirlenmesi ve uygun istatistik yöntemin seçimi. *Türk Plast Rekonstr Est Cer Derg.* 2002;10:98-105.
11. In J. Introduction of a pilot study. *Korean J Anesthesiol.* 2017;70(6):601-5.
12. Akın B, Koçoğlu D. Randomize kontrollü deneyler. *HU-HEMFAD.* 2017;4(1):73-92.
13. Festing, MFW. Reduction of animal use: experimental design and quality of experiments. *Lab Anim.* 1994;28:212-21.
14. Düzgüneş O, Kesici T, Kavuncu O, Gürbüz F. Araştırma ve Deneme Metotları (İstatistik Metotları II). Ankara: Ankara Üniversitesi Ziraat Fakültesi Yayınları, 1021; 1987.
15. Hinkelmann K, Kempthorne O. Design and Analysis of Experiments: Advanced Experimental Design. New Jersey: John Wiley & Sons; 2005.
16. Neter J, Wasserman W, Kutner MH. Applied Linear Statistical Models: Regression, Analysis of Variance, and Experimental Designs, 3. ed. Boston: Irwin; 1990.
17. Festing, MFW. Randomized block experimental designs can increase the power and reproducibility of laboratory animal experiments. *ILAR J.* 2014;55(3):472-6.
18. Ankaralı H, Cangür Ş, Sungur MA. Formülsüz Biyoistatistik. İstanbul: BETİM; 2015.