

Klinisyenlerin ve Araştırmacıların Bilimsel Araştırmalarda Sorgulamaları Gereken İki Temel Kavram: Güvenilirlik ve Geçerlilik

TWO BASIC CONCEPTS THAT CLINICIANS AND RESEARCHERS SHOULD THINK OVER CAREFULLY IN INVESTIGATIONS: RELIABILITY AND VALIDITY

Dr. Zeliha Aslı ÖCEK,^a Dr. Şafak Taner GÜRSOY^a

^aHalk Sağlığı ABD, Ege Üniversitesi Tıp Fakültesi, İZMİR

Özet

Bu yazının amacı, klinisyenlerin araştırma makalelerini okurken sorgulamak, araştırmacıların ise araştırmalarının tüm aşamalarında kaçınmak zorunda oldukları hata kaynaklarına yönelik temel düzeyde bir rehber oluşturmaktır.

Hata, araştırmada ölçülen değer ile gerçek değer arasındaki farktır. Rastgele (random) ve sistematik olmak üzere iki tipi vardır. Belli bir yönde olmayan hata rastgele, geliş güzel dağılmayan, belli bir yöndeki hata ise sistematik hata olarak adlandırılır. Her iki hatanın da bulunmaması araştırma sonuçlarının doğruluğunu güvence altına alır.

Doğruluğun iki temel aşaması vardır: güvenilirlik ve geçerlilik. Güvenilirlik bilginin tutarlılığıdır, birden fazla kez uygulanan bir ölçümden her defasında benzer bilgiler elde edebilme düzeyini yansıtır. Geçerliliğin sağlanması için temel koşul rastgele hatanın bulunmamasıdır. Rastgele hata örnek grubun özelliklerinden ve/veya ölçümlerdeki tutarsızlıklardan kaynaklanabilir.

Geçerlilik ölçüm amacına uygunluk ve içinden örnek alınan evrene genelleme yapabilme anlamına gelir. Geçerliliğinin iç ve dış olmak üzere iki bileşeni vardır. Bir araştırmanın hipotezini uygun şekilde test edebilme yeteneğini yansıtan iç geçerliliğin sağlanabilmesi için birinci koşul rastgele hata olasılığının ortadan kaldırılması, ikinci koşul ise tüm araştırma sürecinin her türlü sistematik hata olasılığından arındırılmasıdır. Sistematik hata araştırmacıların bilinçli ya da bilinçsiz olarak yan tutmaları sonucunda gerçekleşir. Dış geçerlilik ise araştırma grubu için elde edilen sonuçların tüm topluma ve araştırılan konuya ne düzeyde genellenebilir olduğunu tanımlar.

Anahtar Kelimeler: Sonuçların güvenilirliği; yan tutma; kafa karıştırıcı faktörler; epidemiyoloji

Türkiye Klinikleri J Gynecol Obst 2007, 17:310-320

Abstract

The purpose of this article is to develop a basic guide focused on sources of error that the clinicians should think over carefully while reading an article and the researches should avoid at every stage of their researches.

Error is the difference between measurement value and true value. There are two types of error; random and systematic. Random errors are due to chance and can affect the score in an unpredictable way. Systematic errors are predictable errors of measurement. The accuracy of the study result guaranteed by being free from both random and systematic errors.

Accuracy has two components; reliability and validity. Reliability is the consistency of data and having similar scores on multiple trials. The main requirement of validity is the absence of random error. Random error should occur due to the characteristic of the sample and/or inconsistency of measurement.

Validity means being appropriate for the aims of measurement and to generalize it to the target population. The two types of validity are; external and internal validity. The first condition to establish internal validity, which reflects the power of testing the hypothesis, is inexistence of random error where the second is being free from the systematic error in each stage of research. Systematic error occurs due to the researches bias conscious or unconscious. External validity determines the level of generalization of the data to the community and the research subject.

Key Words: Reproducibility of results; bias (epidemiology); confounding factors (epidemiology); epidemiology

Bir klinisyen okuduğu her araştırma makalesinin ardından çok önemli iki soru ile karşılaşır; Bu araştırmanın sonuçlarına

inanabilir miyim ve eğer ilk soruya evet yanıtını veriyorsa, bu sonuçlar benim hastalarım ile ilgili mi? Klinisyenin bu iki soruyu yanlış yanıtlaması kaynakları boşa harcamasından hastalarının yaşamını tehdit edebilecek çok ciddi hatalar yapmasına kadar varan sonuçlara neden olabilir.¹ Sürekli mesleki gelişimini sağlamak için her klinisyen bilimsel araştırmaları izlemek zorundadır. Bununla birlikte okuduğu araştırmaların sonuçlarını eleştiri

Geliş Tarihi/Received: 20.12.2006 Kabul Tarihi/Accepted: 12.02.2007

Yazışma Adresi/Correspondence: Dr. Zeliha Aslı ÖCEK
Ege Üniversitesi Tıp Fakültesi
Halk Sağlığı ABD, İZMİR
zeliha.ocek@ege.edu.tr

Copyright © 2007 by Türkiye Klinikleri

süzgecinden geçirmek klinisyenlerin hastalarına karşı olan temel sorumlulukları arasındadır.

Araştırmayı gerçekleştiren ve sunan taraf açısından baktığımızda sorumluluklar çok daha genişdir ve bu sorumlulukları gerçekleştiren araştırmacı pek çok güçlük ile karşılaşır. Karşılaştığı ilk güçlük toplum sağlığını etkileyen ve/veya sağlık alanındaki bilimsel teorilere katkıda bulunabilen bir araştırma konusu bulmaktır. Bu da çok yoğun ve uzun bir döneme yayılan bir bilimsel okuma süreci anlamına gelir. Araştırma konusuna karar veren araştırmacı tasarlama ve yürütme aşamalarında hem etik ilkeler çerçevesinde kalmayı, hem mümkün olan en az kaynağı harcamayı hem de araştırdığı ilişkiyi mümkün olan en doğru şekilde ölçmeyi amaçlar. Araştırma sonuçlarının yorumlanması ve bilimsel ortam ile paylaşılması aşamasına gelindiğinde işler hala kolaylaşmamıştır. Çünkü araştırmanın doğruluğunu tehdit eden hata olasılıkları bu aşamalarda da araştırmacının peşini bırakmamaktadır.

Bu yazının amacı, klinisyenlerin araştırma makalelerini okurken sorgulamak, araştırmacıların ise araştırmalarının tüm aşamalarında kaçınmak zorunda oldukları hata kaynaklarına yönelik temel düzeyde bir rehber oluşturmaktır.

Hata

Hata, araştırmada ölçülen değer ile gerçek değer arasındaki farktır. Rastgele (random) ve sistematik olmak üzere iki tipi vardır. Rastgele hata belli bir yönde olmayan hatadır. Bir araştırmada öğrencilerin boylarının ölçüldüğünü varsayalım, tamamen geliş güzel olarak bazı ölçümlerin gerçekte olduğundan daha yüksek, bazılarının daha düşük olması rastgele hataya örnek olarak gösterilebilir. Buna karşın çocukların boyları ayakkabıları çıkartılmadan ölçülürse, tüm ölçümler gerçek değerlerden belli bir yönde farklı olacak, yani daha yüksek bulunacaktır. Gelişi güzel dağılmayan, belli bir yöndeki hata sistematik hata olarak adlandırılır.² Her iki hatanın da bulunmaması araştırma sonuçlarının doğruluğunu belirler (Tablo 1).

Doğruluk

Doğruluğun iki temel aşaması vardır: güvenilirlik ve geçerlilik.³ İngilizce kaynaklarda reliability,

Tablo 1. Hata tanımı ve tipleri.

Hata

- Araştırmada ölçülen değer ile gerçek değer arasındaki fark
- Rastgele (Random) Hata:** Belli bir yönde olmayan hata
Sistematik Hata: Gelişi güzel dağılmayan, belli bir yöndeki hata

precision, agreement, consistency, repeatability, reproducibility gibi çok sayıda farklı isimlerle geçen güvenilirlik dilimize kesinlik ya da tutarlılık olarak da çevrilmektedir. Güvenilirlik bilginin tutarlılığıdır, birden fazla kez uygulanan bir ölçümden her defasında benzer bilgiler elde edebilme düzeyini yansıtır.⁴ Geçerlilik (validity) ise ölçüm amacına uygunluk ve içinden örnek alınan evrene genelleme yapabilme anlamına gelir.⁵

I. Güvenilirlik

Güvenilirliğin sağlanması için temel koşul rastgele hatanın bulunmamasıdır. Rastgele hata örnek grubun özelliklerinden ve/veya maruziyet ve etkinin ölçümündeki tutarsızlıklardan kaynaklanabilir.⁶

I. A. Örnek seçimine bağlı hata

Rastgele hatanın ortaya çıkmasının birinci nedeni gözlem sayısının yetersizliğidir. Bu açıdan değerlendirildiğinde güvenilirlik ölçülen değer istatistiksel stabilitesi olarak tanımlanabilir. Örnek olarak gebelik döneminde ultrason ile muayenenin yeni doğanlarda malformasyon riski üzerindeki etkisini değerlendiren bir çalışma uygulandığını ve aşağıdaki verilere dayanarak, rölatif riskin “2” olarak hesaplandığını varsayalım (Tablo 2).⁷

Tablo 2. Gebelik döneminde ultrasonografi yapıma durumuna göre bebeklerde malformasyon varlığı (hipotetik veriler üzerinden).

Ultrasonografi	Malformasyon var		Sağlıklı		Toplam
	Sayı	%	Sayı	%	
Var	2	0.5	398	99.5	400
Yok	1	0.2	399	99.8	400
Toplam	3	0.4	797	99.6	800

Rölatif Risk= Ultrason uygulananlarda malformasyon insidansı/Ultrason uygulanmayanlarda malformasyon insidansı= (2/400)/(1/400)= 2.

Peki bu verilerden yola çıkıp, ultrasonografinin malformasyon riskini iki katına çıkarttığı sonucuna varabilir ve bu bilgiyi toplumla paylaşabilir miyiz? Bu çalışmada sistematik hata ne kadar etkin bir şekilde önlenmeye çalışılmış olursa olsun, olgu sayısı çok azdır. Maruz kalanlar ve kalmayanlar arasında gözlenen fark şansa bağlı olarak ortaya çıkmıştır.⁷

Örnekleme hatasını azaltmanın ya da güvenilirliği arttırmanın birinci yolu araştırmaya alınan kişi sayısını arttırmaktır. Araştırma grubunun büyüklüğü istatistiksel örnek büyüklüğü belirleme formülleri aracılığıyla saptanır. Çalışma tasarımı, evren büyüklüğü ve çalışmanın istenen gücü ile ilişkili olan bu formüller sadece teknik özelliktedir ve bir çalışmadan elde edilecek olan bilginin değerini dikkate almazlar. Çalışma büyüklüğünün planlanmasındaki ve buna bağlı olarak güvenilirliğin belirlenmesindeki temel karar çalışma sonuçlarının daha güvenilir olması ile maliyet arasındaki dengedir. Burada yapılması gereken bir maliyet etkinlik analizidir. Fakat çalışmanın sonuçlarından yararlanacak kişi sayısını tam olarak bilmek hiçbir zaman mümkün değildir. Üstelik bir araştırmacının yararları genellikle sayılarla ölçülemeyen sosyal, politik ve biyolojik faktörleri içerir. Bu nedenlerle maliyet etkinlik analizinin uygulanması pek çok zorluğu ve etik tartışmayı beraberinde getirir.⁷

Günlük aspirin kullanımının kardiyovasküler mortalite hızı üzerindeki etkisini değerlendiren ve 100 000 erkeğin dahil edildiği bir kohort çalışması yapıldığını varsayalım. Bu çalışma büyüklüğünün çok iyi bir güvenilirlik sağlayacağı düşünülebilir. Fakat sadece 100 erkeğin her gün aspirin kullandığını öğrenirsek bu görüşümüzü değiştiririz. 99 900 karşılaştırma bireyi olmasına rağmen, aspirin kullanan 100 erkekte kardiyovasküler mortalite olguları çok az sayıda olacağından etkinin doğru olarak tahmin edilmesi beklenemez. Her iki grupta da 50 000'er erkeğin yer aldığını, fakat bütün erkeklerin 30-39 yaşları arasında olduğunu varsayalım. Dengeli dağılımın güvenilirliği arttırmasına rağmen, bu yaş grubundaki erkekler nadiren kardiyovasküler

hastalık nedeniyle ölümler ve her iki gruptaki olgu sayısı azlığına bağlı olarak güvenilir bir risk oranı tahmini yapılamaz. Bu örnekler güvenilirliğin araştırma büyüklüğünün yanı sıra maruziyet sıklığı ve olgu sıklığı gibi bir dizi faktörle ilişkili olduğunu göstermektedir. Bu ilişkiler araştırmacılara güvenilirliği arttırmak için uygulanabilecek bazı maliyet etkin stratejiler tanımlamaktadır. Örneğin, eğer maruziyete toplumda nadir rastlanıyorsa, çalışma popülasyonu maruz kalma oranının yüksek olduğu bir alt gruptan seçilebilir.⁷ Pestisit maruziyetinin etkileri tüm toplumda değil, ziraat işçilerinden alınan bir örnek grupta araştırılabilir. Farklı maruziyet düzeyleri söz konusuysa ve maruziyet düzeyinin artması ile birlikte etkinin artması bekleniyor, fakat farklı maruziyet düzeylerinin etkilerinin değerlendirilmesi amaçlanmıyorsa çalışma maruz kalmayan ve yüksek düzeyde maruz kalan bireylerle sınırlı tutulabilir.⁷ Sigara ve akciğer kanseri ilişkisinin günde bir paketin üzerinde sigara içenler ile hiç sigara içmeyenleri karşılaştırarak sorgulanması bu stratejiye örnek olarak gösterilebilir.

I. B. Maruziyet ve etkinin ölçümüne bağlı hata

Araştırma sonuçlarının güvenilirliğini tehdit eden ikinci neden bir araştırmacının farklı ölçümleri ya da farklı araştırmacıların farklı ölçümleri arasındaki dikkatsizlik, bilgisizlik ve standardizasyon yetersizliği gibi durumlara bağlı tutarsızlıklardır. Örneğin pek çok çalışmada çocukların sabah yapılan boy ölçümlerinde öğleden sonraları yapılan ölçümlere göre yaklaşık 0.5 cm daha kısa bulunduğu bildirilmiştir.⁴ Bir çalışmada çocukların bazılarının sabah, bazılarının da öğleden sonra ölçülmesi rastgele hataya neden olacaktır. Güvenilirliğin artırılması için, araştırma protokollerinin hazırlanması, araştırmadaki işlemleri ve ölçüm yöntemlerini açıklayan rehberler geliştirilmesi, araştırma ekibine eğitim verilmesi, pilot çalışma uygulayarak standardizasyonu tehdit eden durumların belirlenmesi ve ölçüm sonuçlarının tutarlılığının değerlendirilmesi, ölçüm yöntemlerini daha az karmaşık hale getirerek hata olasılığının azaltılması, aynı değişkenin iki kez ölçülmesi ve ölçüm ortalamasının alınması gibi stratejiler uygulanmaktadır.⁶

Tablo 3. Geçerlilik tanımı ve tipleri.

Geçerlilik
Güvenilirlik
<ul style="list-style-type: none"> • Rasgele hatanın olmaması • Bilginin tutarlılığı • Araştırma yönteminden her defasında benzer bilgiler elde edilmesi
İç Geçerlilik
<ul style="list-style-type: none"> • Sistematik hatanın olmaması • Araştırmanın hipotezini test edebilme yeteneği
Dış Geçerlilik
<ul style="list-style-type: none"> • Araştırma sonuçlarını tüm topluma ve koşullara genellenebilirliği

II. Geçerlilik

Geçerliliğini iç ve dış olmak üzere iki temel bileşeni vardır. Birinci bileşen elde edilen sonuçların araştırmaya dahil edilen grup açısından, ikinci bileşen ise tüm toplum açısından geçerliliğidir³ (Tablo 3).

II. A. İç geçerlilik

Çalışmanın iç yapısı ile ilişkili olan iç geçerlilik bir araştırmanın hipotezini uygun şekilde test edebilme yeteneğini, bir diğer ifade ile araştırmaya katılan kişiler üzerinde yapılan gözlemlerde elde edilen sonuçların doğruluğunu yansıtır.

İç geçerliliğin sağlanması için birinci koşul rastgele hata olasılığının ortadan kaldırılmış olması, yani araştırma sonuçlarının güvenilirliğidir. İkinci koşul araştırmanın tasarımı ile başlayıp yayınlanması ile sonlanan sürecin her türlü sistematik hata olasılığından arındırılmış olmasıdır.³

Sistematik hata araştırmacıların bilinçli ya da bilinçsiz olarak yan tutmaları sonucunda gerçekleşir. Aksakoğlu⁸ bir araştırmada sistematik hataların sıklığı ve yan tutmayı dürüst araştırmacı için korkulu bir tuzak, dürüst olmayan araştırmacı için ise tehlikeli bir silah olarak tanımlamıştır.

Yan tutma İngilizce terminolojide “bias” olarak adlandırılır ve dilimize “taraf tutma”, “yanlılık” gibi ifadelerle de çevrilmiştir. Sackett ve Choi yan tutmayı araştırma süreci boyunca gerçekleşebileceği aşamalara göre sınıflandırmıştır: araş-

tırmanın örnek grubunun özelliklerinin belirlenmesi ve seçimi, deneysel girişimin uygulanması, maruziyet ve sonuçların ölçümü, veri analizi, sonuçların yorumlanması ve yayınlanma.⁸ Daha sık kullanılan bir başka yaklaşım ise yan tutmayı üç gruba ayırır: seçmede yan tutma, bilgide yan tutma, kafa karıştırıcı yan tutma (Tablo 4).^{1,3,9} Bu yazıda da yan tutma bu sınıflandırmanın önerdiği başlıklar altında tartışılacaktır.

II. A. 1. Seçmede yan tutma

Seçmede yan tutma, araştırma grubuna girecek bireyleri belirleme ve/veya bu bireylerin araştırma boyunca gerçekleşen izlemlere katılımını sağlama sürecinde gerçekleşen sistematik hata olarak tanımlanabilir. Bu yan tutmanın en önemli özelliği maruziyet ve hastalık arasındaki ilişkinin araştırmaya katılan ve araştırmaya katılma kriterlerini taşıyan ancak araştırmaya katılmayan kişiler açısından farklı oluşudur.¹⁰

Servikal kanser ve human papilloma virüsü arasındaki ilişkiyi değerlendiren bir kohort araştırması yapıldığını, çalışmanın uygulandığı şehrindeki tüm kadınları temsil eden bir örnek grup seçildiğini, bu örneğe giren, kendisine jinekolojik muayene uygulanmasını kabul eden ve muayene sonucunda servikal kanser bulgusuna rastlanmayan kadınların araştırma grubunu oluşturduklarını, araştırma grubunun birer yıllık aralarla 6 yıl boyunca servikal kanser ve human papilloma virüsü varlığı açısından tarandığını, kadınların tarama sonuçları hakkında bilgilendirildiklerini ve araştırmanın sonunda papilloma virüsüne rastlanan ve rastlanmayan kadınların servikal kanser insidanslarının karşılaştırıldığını varsayalım. Bu çalışmada seçmede yan tutma olasılığı araştırma grubunun oluşturulmasında başlamaktadır. Kanser oluşumunda rol oynayan çeşitli risk faktörlerini taşıyan, jinekolojik sorunları

Tablo 4. Yan tutma tanımı ve tipleri.

Yan Tutma
<ul style="list-style-type: none"> • Araştırmada sistematik hata varlığı
Seçmede Yan Tutma
Bilgide Yan Tutma
Kafa Karıştırıcı Yan Tutma

olan kadınlar bu araştırmaya katılmaya daha fazla gönüllü olacaklardır. Ayrıca 6 yıl süren izlemlerde mutlaka izlem kayıpları gerçekleşir, fakat human papilloma virüsü gözlenmeyen, risk grubunda yer almayan ve jinekolojik sorunları olmayan kişilerin izlemlere devam etmek için motivasyonu çok düşük olacaktır. Diğer yandan virüsü taşıyan, ailesinde kanser öyküsü olan, çok sayıda cinsel yolla bulaşan enfeksiyon geçirmiş, menstrüel düzensizlikleri veya benzer sorunları olan bir kadının izlemlere devam etme olasılığı çok daha yüksektir. Servikal kanser gözlenme olasılığına göre ayrılan kişilerin araştırmaya farklı oranlarda katılımı virüs ve kanser arasında aslında var olmayan bir ilişkinin belirlenmesine ya da ilişkinin abartılmasına neden olabilir. Bu örnekte görüldüğü gibi kohort araştırmalarında çalışmaya kabul edenlerin ve izlemlere devam edenlerin yaşam tarzı, demografik özellikler, maruziyet ve hastalık sıklığı gibi değişkenler açısından, katılmayı ret edenlerden ve izlemlerden çıkanlardan farklı oluşu epidemiyolojik çalışmaların geçerliliğini tehdit eden önemli bir etmendir.

Seçmede yan tutma farklı araştırma tasarımlarında farklı şekillerde ortaya çıkabilir. Olgu kontrol çalışmalarında olguların ve kontrollerin araştırma sorusu dışındaki önemli bir özellik bakımından farklı olmalarından kaynaklanır.¹ Deneysel araştırmalarda ise araştırmaya dahil edilecek gönüllülerin seçiminde ve/veya bu gönüllülerin müdahale ve kontrol gruplarına yerleştirilmelerinde ortaya çıkar. Müdahale ve kontrol gruplarına ayrılan bireylerin belli bir özellik ya da özellikler bakımından birbirlerinden farklı olmaları randomizasyon aracılığıyla önlenir.⁹ Randomizasyonun amacına ulaşabilmesinde hastaların ve bu hastaları tedavi ve müdahale gruplarına ayıran kişilerin randomizasyon yöntemi konusunda kör tutulmuş olmaları çok kritiktir.^{7,9} Bunu açıklamak için, miyokard enfarktüsü tedavisinde antikoagülan kullanımının etkinliğini gösteren ilk çalışmaların yayınlandığı yıllarda ülkemizde üniversite hastanelerinde bir araştırma yapıldığını, bu araştırma öncesinde çalışmaya katılacak tüm kardiyologlarla bir toplantı yapılarak, son dönemlerdeki yayınların sonuçlarının aktarıldığını, araştırmanın amaçlarının ve nasıl uygulanacağını açıklandığını varsayalım. Uygulanan randomizasyon

yöntemi de araştırma boyunca hastanelere tek günlerde yatan hastalara antikoagülan verilmesi, çift günlerde yatanlara ise verilmemesi şeklinde olsun. Bu araştırmada antikoagülan tedavisinin etkili olduğuna inanan ve randomizasyon yöntemini de bilen doktorlar, durumu çok ağır olmayan hastalarını acil serviste bekletip, yatışı ertesi gün yapabilirler. Ağır durumdaki olguları ise bekletme şansları olmadığı için gününe bakmaksızın kardiyoloji bölümüne kısa zamanda yatırmak zorundadırlar. Bu tip bir yan tutma ağır olguların çift günlerde, hafif olguların tek günlerde birikmesine neden olacaktır. Bu örnekte olduğu gibi hastaların araştırmanın hangi grubunda yer aldıklarının gizli tutulmamasına bağlı ortaya çıkan yan tutmanın genellikle tedavi etkinliğinin gerçekte var olandan daha yüksek düzeyde belirlenmesine neden olduğu saptanmıştır.⁹

Seçmede yan tutmanın iki tipi kendilerini açıklayan araştırmacıların isimleri ile anılır: Berkson ve Neyman tipi yan tutma. Berkson tipi yan tutma (veya paradoksu) olguların ve kontrollerin hastaneye yatış hızlarının farklı olmasına bağlı yan tutmayı tanımlar. Berkson bu fenomenin eş zamanlı bir hastalığın varlığından kaynaklandığını açıklar. Örneğin doktorların rahim içi aracı olan salpingitis hastalarını hastaneye yatırma olasılığı hormonal bir kontrasepsiyon yöntemi uygulayan salpingitis'li hastaları yatırma olasılıklarından çok daha güçlüdür. Hastane tabanlı bir olgu kontrol çalışmasında bu durum olgularda gerçekte var olduğundan daha yüksek bir oranda rahim içi araç kullanma sıklığının belirlenmesine ve rahim içi araca bağlı riskin abartılmasına neden olacaktır.¹

Neyman tipi yan tutma, insidans-prevalansda yan tutma veya seçici sağkalıma bağlı yan tutma olarak da adlandırılmaktadır. Çok hızla ölümle sonuçlanan hastalıklarda olgu grubunun toplumdaki tüm olguları temsil etmemesi nedeni ile ortaya çıkar. Örneğin miyokard enfarktüsü ile güneşin altında ağır egzersiz yapma arasındaki ilişkiyi değerlendiren hastane tabanlı bir çalışmada yolda ölen ve bu nedenle hastaneye ulaşamayan hastalar araştırmaya dahil edilmezler. Bu durumda ağır egzersiz ile enfarktüs arasındaki güçlü ilişki olduğundan daha düşük düzeyde ölçülür.¹

Seçmede yan tutmanın bir diğer tipi maskeleyen meme olarak tanımlanmaktadır (unmasking, detection signal). Bir maruziyetin ya da sonucun varlığının bu sonucun aranmasına neden olduğu durumlarda ortaya çıkar. Örneğin östrojen destek tedavisi daha önce hiç bir semptomun gözlenmediği endometriyum kanseri hastalarda kanamaya ve bu nedenle tanı testlerinin uygulanmasına neden olabilir.¹ Bu durumda herhangi bir uyaranla karşılaşmayan sessiz kanserler atlanır, fakat östrojen kullanan kanserlere daha fazla tanı konur, bu da olgu grubunun maruziyet sıklığı bakımından toplumdaki tüm kanser olgularını temsil etmemesine neden olur.

II. A. 2. Bilgide yan tutma

Bilgide yan tutma, gözlemde, sınıflandırmada, ölçümde yan tutma olarak da adlandırılır ve maruziyet veya sonucun ya da her ikisinin birden ölçümünün ve sınıflandırmasının sistematik hata nedeniyle geçerli olmamasıdır.³

Bilgide yan tutma olasılığını mümkün olan en düşük düzeye indirmenin temel koşulu araştırmada ölçülmek istenen değişkenlerin ve bu değişkenleri yansıtan bilgilerin doğru tanımlanmasıdır. 1981 yılında ABD’de vajinal spermid ve konjenital anomaliler arasındaki ilişkiyi değerlendiren bir kohort çalışması gerçekleştirilmiştir. Bu çalışmada spermid kullanımına yönelik bilgiler eczane kayıtlarından elde edilmiş ve hamileliğin bitişinden önceki son 600 gün içinde herhangi bir spermid kullanmış olan bütün kadınlar spermid kullanıcısı olarak sınıflandırılmıştır. Araştırmada spermid kullanan kadınlar arasında major malformasyon prevalansı iki kat daha yüksek bulunmuştur. Fakat bu çalışmada spermid kullanımının tanımından kaynaklanan önemli yan tutma olasılıkları mevcuttur. İlk hata içeriklerine bakılmaksızın farklı spermid tiplerinin ortak bir başlık altında toplanmasıdır. İkinci hata kaynağı maruziyet bilgisinin eczane kayıtlarından elde edilmiş olmasıdır. Etkenle karşılaşmış olarak sınıflandırılan kadınların bazıları hiç spermid kullanmamış olabilir ya da etkenle karşılaşmamış olarak sınıflandırılanlar başka kaynaklardan elde ettikleri spermidleri kullanmış olabilirler. Üçüncü hata kaynağı maruziyetin çok geniş bir zaman diliminde değerlendirilmesidir: doğum veya düşükten önceki 600 gün içinde bir reçete ya-

zılmış olması. Pek çok konjenital defektin gelişmesi için gerekli süre hamileliğin ilk üç ayını kapsayan organogenesis dönemidir. Her ne kadar geniş maruziyet penceresi bu dönemi içine alsın da spermid maruziyetinin hiçbir öneminin olmadığı dönemleri de (örneğin defektin gelişiminden sonra) kapsamaktadır.¹¹

Araştırmalarda değerlendirilen değişkenlerin büyük bir bölümünün doğrudan ölçülmesi mümkün değildir. Bu tip değişkenler için ampirik tanımlar aracılığıyla temsil eden değişkenler, yani göstergeler belirlenir. Örneğin pasif içiciliğin myokard enfarktüsü üzerindeki etkisini araştıran bir çalışmada, pasif içicilik diğer kişilerin tütün dumanını inhale etme, miyokard enfarktüsü ise kalp kas dokusunun nekrozu olarak tanımlanabilir. Bu teorik tanımlar hangi bilgilerin toplanması gerektiğini belirler. Fakat geniş kapsamlı bir çalışmada tütün inhalasyonunun ve miyokardial nekrozun doğrudan ölçülmesi mümkün değildir. Pasif sigara içme sigara içenlerle geçirilen saat sayısı olarak, miyokardial enfarktüs ise göğüs ağrısı, enzim düzeyi, EKG bulguları vb gibi bazı değişkenler temel alınarak tanımlanabilir. Ampirik tanımların ya da göstergelerin doğru belirlenememesi, teorik tanımları temsil edememesi hatalı sınıflandırmaya neden olur. Bilgide yan tutmanın ikinci ortaya çıkış şekli ölçüm hatasıdır. Enzim düzeyi sistematik olarak hatalı ölçülüyorsa ölçüm hatasından kaynaklı yanlış sınıflandırma söz konusudur. Düzey doğru ölçülüyor, fakat sadece enzim düzeyi enfarktüs varlığını başarıyla temsil edemiyorsa ölçüm hatası olmadan sınıflandırmada yan tutma söz konusudur.⁷

Özetlersek üç farklı hata olasılığı söz konusudur: 1) Doğru göstergelerin seçilmemesi; 2) Göstergelerin doğru ölçülememesi; 3) Göstergelerin hem seçiminde hem de ölçümünde hata yapılması.

Araştırmacılar gözlemsel araştırmalarda maruziyetin, deneysel araştırmalarda müdahalenin ve her iki araştırma tipinde de hastalığın sınıflandırılmasında bu üç farklı hata olasılığı ile karşılaşılırlar.

- Maruziyetin/Müdahalenin Sınıflandırılmasında Yan Tutma

Olgu kontrol çalışmaları geçmiş maruziyetin hatırlanmasına dayanır. Bu da hatırlamada yan

tutmaya neden olabilir. Olgular hastalıklarının nedenini bulmak için belleklerini zorlarlar, ama sağlıklı kontrollerin böyle bir motivasyonu yoktur. Bu nedenle olgular genellikle daha fazla şey hatırlarlar. Örneğin kürtaj ve göğüs kanseri arasındaki olası ilişki çok kritik bir medikal ve politik tartışma konusudur. Çok sayıda olgu kontrol çalışmasında kürtajın ardından kanser riskinin arttığı bildirilmiştir. Fakat araştırmacılar görüşmelerde bildirilen kürtaj öyküsü ile sağlık kayıtlarını karşılaştırdıklarında kontroller arasında sistematik bir şekilde kürtaj sayısının daha düşük bildirildiğini, olguların ise doğru sayılar bildirdiklerini belirlemişler ve kürtaj ile kanser arasında bulunduğu iddia edilen ilişkinin aslında bu sistematik hataya bağlı olduğunu açıklamışlardır. Daha sonra gerçekleştirilmiş ve yan tutma olasılığı en düşük düzeyde tutulmuş olan kohort çalışmalarında kürtaj ile göğüs kanseri arasında hiçbir ilişki bulunmadığı gösterilmiştir.¹ Bu örnekte olduğu gibi maruziyet hakkındaki bilginin olgular meydana geldikten sonra toplanması sistematik hataya neden olabilir. Olgu grubu karşılaştırma grubuna göre geçmiş maruziyeti daha çok hatırlayabilir ve daha fazla bilgi aktarabilir (hatırlamada yanlışlık) ya da sosyal açıdan daha kabul edilebilir yanıtlar verebilir (annelerin çocuklarını dövdüklerini söylememeleri gibi). Bir diğer olasılık da araştırmacının hastalığın varlığı nedeniyle maruziyetin ölçüm ve sınıflandırılmasında önyargılı olmasıdır.¹²

Bilginin araştırmaya katılan tüm bireylerde aynı yolla toplanmaması da önemli bir hata kaynağıdır. Örneğin bir olgu kontrol çalışmasında maruziyet bilgisinin olgularla hastane yataklarının başında yapılan görüşmelerde, toplumdan kaynaklı kontrol grubu ile ise telefon görüşmelerinde elde edilmesi ya da olguların eşleriyle, kontrollerin ise kendileriyle görüşülmesi bilginin ölçümde gruplar arasında sistematik bir farkın doğmasına neden olur.¹

Deneyisel çalışmalarda maruziyetin yerini araştırmacının uyguladığı müdahale alır. Müdahalenin hatalı sınıflandırılmasının en önemli nedeni uyumta yan tutmadır.⁹ Örneğin risk düzeyi yüksek olan hastaların egzersiz programına devam etmemesi ya da tam aksine risk düzeyi düşük olan kişilere göre beslenmelerine daha fazla dikkat etmeleri müdahalenin etkinliğini belirler. Müdahalenin bir

şekilde kontrol grubunu da etkilemesi kontaminasyona bağlı yan tutma olarak adlandırılır. Genellikle toplumsal deneylerde müdahale ve kontrol grupları arasındaki ilişkilere, medya, sağlık çalışanları vb araçlara bağlı olarak ortaya çıkar.⁹ Örneğin sigarayı bıraktırmaya yönelik bir programın etkinliğinin değerlendirildiği bir çalışmada araştırma grubunun iki fabrikanın işçilerinden oluştuğunu, fabrikalardan birindeki işçilerin müdahale diğerindekilerin ise kontrol grubu olduklarını varsayalım. Programa dahil olan bir işçi diğer fabrikada çalışan yakınlarını da etkileyebilir, bu da müdahalenin etkinliğinin olduğundan daha düşük düzeyde belirlenmesine neden olur.

• Hastalığın Sınıflandırılmasında Yan Tutma

Belli bir risk faktörüne maruz kalan kişiler daha fazla hastalık semptomu bildirme veya daha fazla doktora gitme eğiliminde olabilirler. Benzer şekilde muayene ve gözlemleri yapan personel tarafından da eğer bir kişinin potansiyel bir risk faktörüne maruz kaldığı biliniyorsa o kişinin hasta olma olasılığı abartılabilir. Sigara içen ve içmeyenlere amfizem insidans hızlarının karşılaştırıldığı bir kohort çalışması uygulandığını varsayalım. Amfizem özel bir tıbbi kontrol uygulanmadıkça tanısız kalan bir hastalıktır. Fakat sigara içenler bu maruziyetleri nedeniyle daha fazla sağlık riski ile karşı karşıya olduklarını biliyorlarsa ya da daha fazla sağlık sorunları yaşıyorlarsa daha fazla tıbbi bakım alırlar ve amfizemlerin belirlenme olasılığı da diğer gruba göre daha yüksek olur. Benzer şekilde hastalarının sigara içtiği için daha fazla risk altında olduğunu bilen doktorlar sigara içen grupta gerçekte olduğundan daha fazla gruba amfizem tanısı koyabilir.³

Journal of American Medical Association'da toksik şok sendromu etiyolojisinde tampon kullanımını ile ilgili iki farklı görüşün yer aldığı iki makale yayınlanmıştır. Birinci makale toksik şok sendromundaki risk faktörlerini değerlendiren bir olgu kontrol çalışmasıdır. Bu çalışmada olgu grubunu kendi hekimleri tarafından toksik şok tanısı konmuş olan 50 kadın, kontrol grubunu da bu kadınlarla yaşa ve bölgeye göre eşleştirilmiş olan 150 kadın oluşturmuştur. Menstrüasyon döneminde olgu grubunun %100'ünün, kontrol grubunun ise

%45'inin tampon kullandığı belirlenmiş ve tampon kullanımının toksik şok sendromuna yol açtığı sonucuna varılmıştır. İkinci makalede ise toksik şok sendromu ile tampon kullanımını değerlendiren çalışmalarda saptanan ilişkilerin yan tutmadan kaynaklandığı savunulmuştur. Araştırmacılar tanıya yan tutulduğunu çünkü doktorların eğer hastalarının menstrüasyon döneminde olduklarını ve tampon kullandıklarını biliyorlarsa toksik şok sendromu tanısını koyduklarını belirtmişlerdir. Örnek olarak şigella enteriti olduğu halde menstrüasyon döneminde olduğu ve tampon kullandığı için toksik şok sendromu tanısı konan bir kadın gösterilmiştir. Ayrıca yayınlamada yan tutma da olduğu belirtilmiştir. Yazarlara göre toksik şok ve tampon kullanımı arasındaki ilişki o yıllarda çok popüler hale geldiği için bu konuda yapılan yayınlar artmıştır. Bu iki değişken arasında anlamlı bir ilişki saptamayan çalışmalar ise dergilere gönderilmemektedir ya da yayınlanmamaktadır.¹²

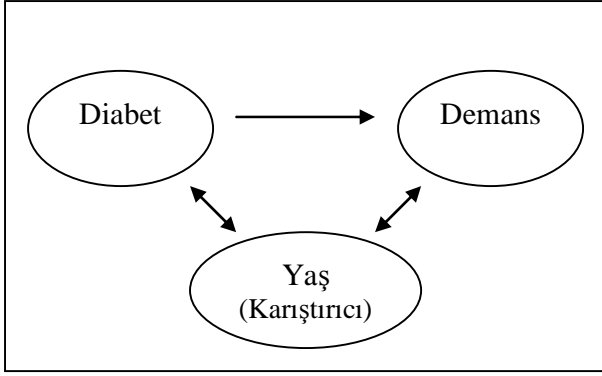
Özellikle deneysel çalışmalarda körlük uygulanmaması hastalığın sınıflandırılmasında yan tutmaya neden olur. İntranasal alpha₂-interferon kullanımının soğuk algınlığına karşı koruyucu bir etkiye sahip olup olmadığı randomize kontrollü çift kör bir çalışmada değerlendirilmiştir. Müdahale grubuna interferon içeren, kontrol grubuna ise plasebo içeren nasal sprey verilmiştir. İzlemler sırasında araştırma grubu üst solunum yollarına ait her türlü semptomlarını kaydettikleri bir günlük tutmuştur. Herhangi bir semptom ortaya çıktığında araştırma hemşiresi ileri muayene ve tanı için telefonla aranmıştır. Araştırma çift kördür, yani bir kişinin interferon mu yoksa plasebo içeren bir sprey mi kullandığını hem kendisi hem de araştırma hemşiresi bilmemektedir. Bu çalışmada körlük uygulanmaması iki yönlü bilgide yan tutmaya neden olabilir. Araştırma grubundaki bireylerin hangi spreyi kullandıklarını bilmeleri yaşam biçimlerini, riskli davranışlarını, örneğin enfekte kişilerden uzak durup durmamalarını etkileyebilir, bu da müdahalenin etkisinin doğru ölçülmesini önleyebilir. Ayrıca plasebo kullanan kişiler daha fazla semptom bildirme, araştırma hemşiresini daha fazla arama, hemşire ise plasebo kullandığını bildiği kişilere daha fazla tanı koyma eğiliminde ola-

bilir, bu da hastalığın sınıflandırılmasında sistematik hataya neden olabilir.¹³

II. A. 3. Kafa karıştırıcı yan tutma

Araştırılan ilişkinin bir dış faktörün karışması nedeniyle doğru yorumlanmaması (kafa) karıştırıcı yan tutma olarak adlandırılır. Kafa karıştırma pek çok yazar tarafından bir yan tutma tipi olarak sınıflandırılırken, bazı kaynaklarda da yan tutmanın bir alt başlığı değil, iç geçerliliği tehdit eden bir diğer etmen olarak adlandırılmaktadır.^{8, 14} Bunun nedeni kafa karıştırmanın araştırma grubuna girecek bireylerin seçimi veya değişkenlerin sınıflandırılması aşamalarında değil, elde edilen sonuçların analiz edilmesinde yorumlanmasında ortaya çıkmasıdır. Bu aşamalarda kafa karıştırıcı etmenlerin uygun bir şekilde kontrol edilmemesi nedeniyle ortaya çıkan yan tutma iyatrojenik yan tutma olarak da adlandırılmaktadır.¹⁵

Kafa karıştırıcı çalışmadaki risk faktörü ve sonuçtan ayrı, fakat bağımsız bir şekilde hem risk faktörü hem de sonuçla ilişkili olan ve böylece gerçekte var olmayan bir ilişkiyi var gibi gösterebilen veya gerçekte var olan bir ilişkiyi maskeleyebilen bir deyişkendir. Örneğin, diyabet ve demans arasındaki ilişkiyi değerlendiren bir çalışma yapıldığını, yeni tanı almış olan diyabetik yetişkinlerin ve diyabeti olmayan yetişkinlerden oluşan kontrol grubunun kümülatif demans insidansının belirlenmesi amacıyla 10 yıl boyunca izlendiğini ve sonuçta diyabeti olan kişilerde demans gelişme olasılığının, diyabeti olmayanlara göre dört kat daha yüksek bulunduğunu varsayalım. Her ne kadar araştırma sonuçları diyabet ve demans arasında güçlü bir ilişki bulunduğunu gösterse de bu ilişki diyabeti olan grubun daha yaşlı olmasına ve bu nedenle demans insidansının daha yüksek olmasına bağlı olabilir. Yaş faktörü kontrol edildiğinde diyabeti olan kişilerde demans olasılığı daha düşük bulunabilir.¹¹ Yani ilk değerlendirmede yaşın karıştırıcı etkisi nedeniyle diyabet ve demans arasındaki ilişki abartılmıştır (Şekil 1). Oral kontraseptif kullanımının miyokard enfarktüsü riskini arttırdığını savunan çalışmalar da kafa karıştırıcı faktörün etkisine yönelik tipik en örnekler arasındadır. Bu çalışmalarda bulunan ilişki kontraseptif kullanan kadınların daha



Şekil 1. Bir kafa karıştırıcı faktör örneği.

fazla sigara içmesine ve bu nedenle enfarktüs risklerinin artmasına bağlıdır.¹ Oral kontraseptif ve enfarktüs ilişkisinde sigara içme karıştırıcı faktördür.

Araştırma planlaması ve analizi aşamalarında karıştırıcı faktörlerin tanımlanması ve etkilerinin randomizasyon, eşleştirme, çok değişkenli analiz gibi teknikler aracılığıyla mümkün olan en düşük düzeye indirilmesi çok önemlidir. Fakat hem maruziyetle hem de hastalıkla ilişkili tüm değişkenler karıştırıcı faktör değildir.² Örneğin, sık bira tüketiminin sık pizza tüketimi ile ilişki olduğunu ve sık bira tüketiminin rektal kanser için bir risk olduğunu düşünelim. Pizza tüketimi bir karıştırıcı faktör müdür? İlk bakışta cevabın evet olduğu düşünülebilir, çünkü pizza tüketimi hem bira içme hem de rektal kanser ile bağlantılıdır. Fakat pizza tüketimi rektal kanserle sadece sekonder olarak bira tüketimi ile ilişkisi nedeniyle ilişkili ise, bir karıştırıcı değildir. Karıştırıcı bir faktör maruz kalmayan bireylerde de hastalık gelişimi için bir risk göstergesi olmalıdır. Eğer pizza tüketimi bira içmeyenler arasında da rektal kanser için bir risk oluştursaydı, bir karıştırıcı olabilirdi. Sonuç olarak karıştırıcı faktör hem maruziyetle hem de hastalıkla dolaylı değil, doğrudan ilişki gösteren bir değişkendir.

Karıştırıcı faktörü diğer değişkenlerden ayırd etmek için bir başka ilişkiyi değerlendirelim. Sigara içme ve kalp hastalığı ilişkisinde artan kan basıncı bir karıştırıcı mıdır? Basıncıdaki artış hastalık için kesinlikle bir risk faktörüdür ve aynı zamanda maruziyetle doğrudan ilişkilidir, çünkü

sigara içmeye bağlı olarak gelişebilmektedir. Maruz kalmayan bireylerdeki hastalık için de bir risk faktörü olabilir, çünkü kan basıncı sigara dışındaki nedenlere bağlı olarak da yükselebilir. Bununla birlikte tam bir risk faktörü olarak değerlendirilemez, çünkü sigara içmenin etkisi kan basıncının etkisi aracılığıyla gerçekleşmiştir. Maruziyet ve hastalık ilişkisinde, nedensellik zincirindeki bir aşamayı temsil eden bir değişken karıştırıcı faktör olarak değil, aracı (intermediate) faktör olarak değerlendirilmelidir.³

II. B. Dış geçerlilik

Dış geçerlilik, araştırma grubu için elde edilen sonuçların tüm topluma ve araştırılan konuya ne düzeyde genellenebilir olduğunu tanımlar. Bu nedenle genellenebilirlik olarak da adlandırılmaktadır. Örneğin, sigaranın akciğer kanserine yol açması, sedanter yaşam tarzının obeziteye neden olması sadece özel bir grup için değil, evrensel olarak geçerlidir.

İstatistiksel bir kavram olmayan dış geçerlilik sadece istatistiksel olarak uygun büyüklükte bir grup seçilmesi ile sağlanamaz. Eğer bu yeterli olsaydı hayvan çalışmalarında elde edilen tüm sonuçlar insanlar için de genellenebilirdi.³

Dış geçerlilik iki başlık altında değerlendirilir. Birinci başlık olan *popülasyon geçerliliği*, örnek gruptan elde edilen sonuçların hedef topluma genellenebilirliğini gösterir. Araştırmanın gerçekleştirildiği koşulların diğer koşullara genellenebilir olması ise *ekolojik geçerlilik* olarak adlandırılır.³ Randomize kontrollü bir klinik deney ile bir aşımın etkinliğinin değerlendirildiğini varsayalım. Bu klinik deneye katılan gönüllülerden elde edilen sonuçların tüm toplum için de geçerli olması popülasyon geçerliliğidir. Aşımın sadece klinik veya laboratuvar ortamlarında sağlanabilecek özel koşullarda değil, sağlık ocaklarında ve alan çalışmalarında da etkili olabilmesi ise ekolojik geçerliliğidir.

II. C. İç geçerlilik ve dış geçerlilik ilişkisi

İç geçerlilik, dış geçerlilik için gereklidir, ancak dış geçerliliği garanti etmez. Elde edilen sonuçlar araştırma grubu için doğru olabilir, ama diğer toplumlara ve koşullara genellenemiyorsa dış geçerlilik sağlanamamıştır.

Geçerliliğin bu iki tipi araştırmaların gerçekleştirildiği ortamla çok yakından ilişkilidir.³ Bunu laboratuvar ortamında ve alanda gerçekleştirilen çalışmalar aracılığıyla açıklayabiliriz. Laboratuvar ortamında gerçekleştirilen çalışmalara örnek olarak simülasyon kullanılan deneysel araştırmaları alalım. Bu araştırmaların iç geçerliliği yüksektir, çünkü dış faktörlerin kontrol edilmesi kolaydır, buna karşın laboratuvar ortamı dışına çıktığında sonuçların genellenebilirliği düşebilir. Alanda gerçekleştirilen çalışmalara örnek olarak ise gözlemsel epidemiyolojik çalışmaları alalım. Bu araştırmalarda da çok sayıda dış faktör işe karışacağı için iç geçerliliğin sağlanması güçtür, fakat genellenebilirliği daha yüksek sonuçlar elde edilebilir.

Açıklayıcı ve Pragmatik Klinik Deneysel Çalışmalar

Klinik deneysel araştırmalar önceliklerinin iç ya da dış geçerlilik olmasına göre açıklayıcı ve pragmatik deneyler olarak ikiye ayrılır.¹⁶ Çok sayıda dışlanma kriteri olan, pek çok özel koşulun varlığını gerektiren ve bu nedenle de iç geçerliliği çok yüksek olan deneysel araştırmalar açıklayıcı (explanatory) deneyler olarak adlandırılır. Açıklayıcı deneylere genellikle müdahaleye uyunc gösterme, izlemlere devam etme olasılığı yüksek olan, araştırılan hastalık dışında ek bir sağlık sorunu bulunmayan kişiler ve homojen hasta gruplarına hizmet veren merkezlerde tedavi gören hastalar dahil edilir.¹⁶ Açıklayıcı deneyler “ilaç veya müdahale ideal koşullarda işe yarıyor mu?” sorusuna yanıt ararlar.¹¹ Eğer bu soruya olumlu bir yanıt verebilirlerse aynı uygulamanın etkisi pragmatik deneylerle gerçek yaşam koşullarında, yani gerçek klinik uygulamalarda değerlendirilir. Yönetim çalışmaları ise bir müdahalenin etkisini gerçek yaşam koşullarında değerlendirir. Pragmatik çalışmalar tedavinin uygulanma olasılığı olan tüm hasta popülasyonunu temsil eden hastalar üzerinde gerçekleştirilir. Bu hastalar farklı uyunc düzeyleri gösterebilirler, çok sayıda eşlik eden hastalıkları bulunabilir ve farklı ilaçlar kullanıyor olabilirler.¹⁶

Açıklayıcı çalışmaların uygulanması zorunludur. Fakat bu çalışmalar dış geçerliliğin iki bileşeni olan popülasyon geçerliliği ve ekolojik geçerlilik

açısından oldukça sınırlıdır ve bu nedenle bir müdahalenin etkinliği hakkında karar verebilmek için yeterli düzeyde bilgi sağlamazlar. Bir tedavinin etkisi en iyi şekilde çok dikkatli tasarlanmış randomize kontrollü klinik pragmatik çalışmalar aracılığıyla ölçülebilir. Bu çalışmalardan olumlu sonuç alınması müdahalenin sadece etkili olduğunu göstermez, aynı zamanda gerçek yaşamda da kullanılabileceğini kanıtlar.¹⁶

Pragmatik klinik çalışmalar klinisyenlere ve sağlık hizmeti yöneticilerine rehberlik yapar. Bununla birlikte pragmatik çalışmalarda iç ve dış geçerlilik arasındaki denge çok kritiktir. Açıklayıcı çalışmalar müdahale dışındaki hemen tüm değişkenler üzerinde çok önemli düzeyde kontrol olanağı sağlarken, pragmatik deneyler araştırma sonuçlarını genelleyebilmek için dış geçerliliği çok yüksek tutmayı amaçlarlar. Bununla birlikte pragmatik deneylerin tehlikesi genellenebilirliği sağlamaya yönelik çabaların iç geçerliliği kabul edilemez bir düzeye indirmesidir.¹⁶

Klinik deneysel çalışmalarda iç geçerliliğin sağlanmasında randomizasyon çok kritiktir ve randomize olmayan klinik çalışmaların çok yanıltıcı sonuçlar verebildikleri gösterilmiştir.¹⁷ Randomizasyonun yanı sıra, mümkün olan durumlarda verilerin toplanmasında ve analizinde mutlaka körlük uygulanmalıdır. Genellenebilirliğin sağlanması için ise dışlanma kriterleri minimum düzeyde tutulmalı, hastalar ve hekimler istedikleri tedavinin seçiminde özgür bırakılarak araştırmanın pragmatik olması ve hastaların zarar görmemesi sağlanmalıdır.¹⁶

KAYNAKLAR

1. Grimes DA, Schulz KF. Bias and causal associations in observational research. *The Lancet* 2002;359:248-52.
2. Katz D. *Epidemiology Biostatistics and Preventive Medicine Review*. 1st ed. Philadelphia: WB Saunders Company; 1997. p.25-6.
3. Rothman KJ, Greenland S. *Modern Epidemiology*. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998. p.135-47.
4. Abramson JH, Abramson ZH. *Survey Methods in Community Medicine. Epidemiological Research, Programme Evaluation, Clinical Trials*. 5th ed. London: Churchill Livingstone; 1999. p.171-7.
5. Şencan H. *Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik*. Ankara: Seçkin; 2005. s.724-6.

6. Higgins PA, Straub AJ. Understanding the error of our ways: Mapping the concepts of validity and reliability. *Nurs Outlook* 2006;54:23-9.
7. Norell SE. *Workbook of Epidemiology*. Oxford: Oxford University Press; 1995. p.16-170.
8. Aksakoğlu G. *Sağlıkta Araştırma Teknikleri ve Analiz Yöntemleri*. İzmir: D.E.Ü Rektörlük Matbaası; 2001.
9. Delgado-Rodriguez M, Javier L. Bias. *J Epidemiol Community Health* 2004;58:635-41.
10. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615-25.
11. Aschengrau A, Seage GR. *Essentials Epidemiology in Public Health*. USA: Jones and Bartlett Publishers; 2003. p.251-98.
12. Gehlbach S. *Interpreting the Medical Literature*. 3rd ed. USA: McGraw-Hill Inc; 1993. p.3-111.
13. Monto AS, Shope TC, Schwartz SA, Albrecht JK. Intranasal interferon-alpha 2b for seasonal prophylaxis of respiratory infection. *J Infect Dis* 1986;154:128-33.
14. Maclure M, Schneeweiss S. Causation of bias: The episcopes. *Epidemiology* 2001;12:114-22.
15. Bulterys M, Morgernstern H. Confounding or intermediate effect? An appraisal of iatrogenic bias in perinatal AIDS research. *Paediatr Perinat Epidemiol* 1993;7:387-94.
16. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: The struggle between external and internal validity. *BMC Medical Research Methodology* 2003;3:28.
17. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating nonrandomized intervention studies. *Health Technology Assessment* 2003; 7 (27) (Executive Summary).