

Kanıtın yokluğu, yokluğun kanıtı mıdır?

Prof. Dr. Fahri Ovalı



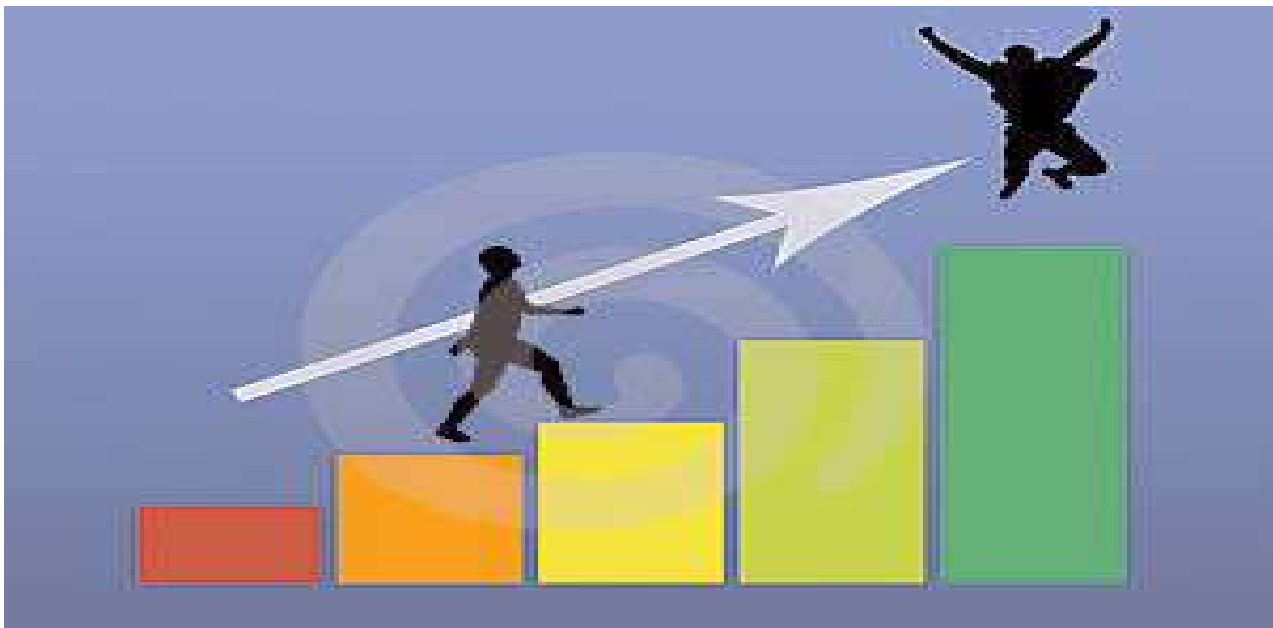
İstanbul'da doğdu. 1979 yılında Darüşşafaka Lisesi'ni, 1985 yılında İstanbul Tıp Fakültesi'ni bitirdi. 1991 yılında İstanbul Tıp Fakültesi'nde çocuk sağlığı ve hastalıkları uzmanlık eğitimini tamamladı. 1993-2003 yılları arasında İstanbul Tıp Fakültesi Yenidoğan Yoğun Bakım Ünitesi'nde çalıştı. 1996 yılında doçent, 2003 yılında profesör oldu. 2003-2005 yılları arasında Afyon Kocatepe Üniversitesi Tıp Fakültesi Çocuk Sağlığı ve Hastalıkları Anabilim Dalı Başkanlığı ve Dahili Tıp Bilimleri Bölüm Başkanlığı görevlerinde bulundu. 2005 yılından itibaren Zeynep Kamil Kadın ve Çocuk Hastalıkları Hastanesi Yenidoğan Yoğun Bakım Ünitesi Klinik Şefliği görevini yürütmektedir. Editörlüğünü yaptığı 6 kitabı, yurtiçi ve yurtdışında yayınlanmış 150'den fazla makalesi bulunmaktadır.

Bilimsel çalışmalar ve araştırmaların sonuçları değerlendirilirken sıklıkla istatistiksel yöntemler kullanılır ve sonuçta, elde edilen verilerin “anlamlı” veya “anlamsız” olduğu yargısına varılır. Hatta birçok yazar ve editör, “istatistiksel anlamlılık”, yani “ $p < 0.05$ ” için her şeyini vermeye hazırdır. Ancak, anlamlı kabul edilen sonuçların her zaman klinik açıdan önemli olmadığı da bilinmektedir. İstatistiksel olarak ileri derecede anlamlı olan bir sonuç, klinik açıdan anlamsız olabilir ve hekimlere

hiçbir “anlam” ifade etmeyebilir. Fakat istatistiksel olarak “anlamsız” olanların dikkate alınmaması veya yanlış yorumlanması daha da vahimdir. Bir diğer yanlış yorumlanan durum ise, “etkisiz olduğuna dair kanıt bulunmasının”, “etkili olduğuna dair kanıt bulunmaması” ile karıştırılması veya “kanıt yokluğunun” “yok olduğuna dair kanıt bulunmaması” ile karıştırılmasıdır. Durumu daha iyi anlayabilmek için, bazı temel kavramları tekrar irdelemekte yarar vardır.

Klinik çalışmalarda genellikle farklılığın < 0.05 veya < 0.01 olduğu durumlar, “anlamlı” olarak kabul edilir ve bu

farkın, karşılaştırılan iki durumdan birinin daha iyi veya etkili olduğunu “gösterdiği” ifade edilir. Ancak, bu farklılık meselesini dikkatli ele almak gerekir. Eğer, $p < 0.05$ düzeyini kullanıyorsak, % 5 olasılıkla, tip I hata yapıyoruz demektir. Ancak bu % 5, her zaman gerçeği göstermez. Eğer fark tam olarak 0.049 ise, 0.051 olduğunda aniden buharlaşıp gidecek midir? Bilindiği gibi, tip I hata demek, gerçekte farklılık olmamasına rağmen, sonuçların fark olduğunu gösterdiği durumdur. Tip II hata ise, gerçekte farklılık olmasına rağmen, sonuçların fark olmadığını gösterdiği duruma verilen addır. İstatistiksel testlerin yanlış olarak kullanıl-







İstatistiksel olarak ileri derecede anlamlı olan bir sonuç, klinik açıdan anlamsız olabilir ve hekimlere hiçbir "anlam" ifade etmeyebilir. Fakat istatistiksel olarak "anlamsız" olanların dikkate alınmaması veya yanlış yorumlanması vahimdir.

diği durumlarda ortaya çıkan hataları ise adlandırmak bile mümkün değildir. Ancak okuyucuların, araştırma sonuçlarını değerlendirebilmek için yine de hangi testin hangi durumlarda kullanılması gerektiğine ait temel bilgileri sahip olmaları gerekir.

Birçok klinisyen, istatistiksel olarak anlamlı kabul edilen farklılıkların, klinik öneme sahip olduğuna dair bir önyargıya sahiptir. Ancak bazı durumlarda, $p < 0.0001$ bile olsa, bu durumun klinik bir önemi olmayabilir. Bazıları, $p < 0.048$ gördüğünde, artık bahsi geçen tedaviyi hastalarına rahatlıkla uygulayabileceği düşüncesine kapılır veya yıllardır uyguladığı bir tedavinin farkının yeni bir çalışmada 0.06 olduğunu gördüğünde, aniden tedaviden vazgeçer. Böyle çalışmaları değerlendirirken dikkatli olunmasında fayda vardır. Anlamın, anlamlı olmadığı durumlar şu şekilde sıralanabilir:

- Örneklem sayısı büyüktür. İstatistiksel anlamlılık gösterme ihtimali yüksektir. Klinisyen şu soruyu sormalıdır: Bu fark, klinik olarak önemli midir?
- $p < 0.05$ olmasına rağmen, yetersiz sayıda denekte incelenen parametrede küçük değişiklikleri belirtmektedir. İstatistiksel olarak anlamlı olmasına rağmen, klinik anlamı çok azdır veya yoktur. Klinisyen bu durumda, verileri şüpheyle karşılamalı ve uygulamalarını hemen değiştirmemelidir.
- Birçok değişik ve fazla sayıda denekle yapılan çalışmalarda anlamlı fark görülmüştür. Gerçek fark küçük ve belki de önemsiz olabilir. Klinisyen, bu meta-analizi de şüpheyle karşılamalı ve uy-

gulamalarını hemen değiştirmemelidir.

Diğer yandan, bazı durumlarda ise anlamlı bulunmayan istatistiksel sonuçların klinik anlamı fazla olabilir. Burada da yine aynı soru sorulmalıdır: Bu anlamsız istatistik, klinik olarak anlamlı olabilir mi? Böyle çalışmaları değerlendirirken, sorulması gereken sorular şunlar olmalıdır:

- İncelenen sonuç parametreleri nelerdir ve nasıl tanımlanmıştır?
- Çalışma dışı bırakılanlar kimlerdir?
- Örneklem büyüklüğü fazla olsa da, istenen sonuç parametrelerine uygun mudur?
- Kontrol grubundaki insidans nedir?
- Bu sonuçlara ulaşabilmek için daha iyi bir örneklem yapılabilir miydi?
- Böyle bir çalışmadaki sonuçlar da anlamlı olur muydu?
- Bir girişim sonucunda "pozitif" bir etki elde edilmişse, bu etki önemli midir?
- Yan etkiler incelenmiş midir?
- Eğer "anlamlı farklar bulunmamışsa", bu durum, güvenliği de garanti eder mi? Şu ifadelere dikkat edin: "Bu uygulama güvenli bulunmuştur ve uygulanmalıdır".
- Yayında "anlamlı" kabul edilen sonuçla, sizin ve hastalarınızın "anlamlı" kabul ettiği sonuç aynı mıdır?

Tüm bu çalışmalarda dikkat edilmesi gereken bir diğer önemli konu da, "payda" dır.

Hastalıkların insidansları, görülme hızları, risk faktörleri gibi durumlarda, payda önemli hale gelir. Örneğin, prostat kanseri insidansından bahsediliyorsa, ele alınan toplum, tüm toplum mudur, yalnız erkekler midir, yoksa yalnız belirli bir yaşın üzerindeki erkekler midir? Eğer tüm toplum ele alınmışsa, kızlar ve çocuklar da hesaba katılmış demektir! Ya da, prematüre retinopatisi insidansı bildiriliyorsa, 1.500 gramın altında doğan tüm bebekler mi hesaba katılmıştır? Oysa prematüre retinopatisi, 4. haftadan sonra ortaya çıkar. Doğan tüm bebeklerin hesaba katılmasıyla, doğumdan sonra 4. haftasını doldurmadan ölen birçok bebek de insidans hesabına katılmış olacak ve insidans daha düşük gözükülecektir. Dolayısıyla, çalışmalar değerlendirilirken sorulması gereken sorulardan birkaçı şöyle olmalıdır: Payda nedir? Ne olmalıdır? Riskli popülasyon nedir?

Tedavi edilmesi gereken hasta sayısı da, istatistiksel yöntemlerde sık kulla-

nılır. Bu sayı, mutlak risk azalmasının tersi olarak hesaplanır. Örneğin, bir tedaviyle kontrol grubuna göre % 5 daha fazla hasta tedavi edilmişse, tedavi edilmesi gereken hasta sayısı 20 olarak bulunur. Yani, bir hastada başarılı olabilmek için aynı tedaviyi 20 hastaya uygulamanız gerekmektedir. Bu rakam, tedavinin etkinliğini anlamada daha faydalı olabilir. İdeal olanı, bu rakamın 1 olmasıdır. Ancak birçok durumda bu rakam 10'dan büyüktür ve elinizdeki hastanın bu tedaviden faydalanma olasılığı da düşüktür. Eğer, tedavi edilmesi gereken hasta sayısı nisbeten düşükse (6 - 10 gibi), o zaman sorulması gereken soru ise "neyin karşılığında?" olmalıdır. Eğer elde edilecek sonuç klinik olarak önemli ise, o zaman tedavi edilecek hasta sayısının 10 olması, 100 olmasından iyidir.

Cochrane, herhangi bir sağlık uygulaması için şu soruları sorar: "Bu iş uygun mu?", "Pratikte uygulanabilir mi?" ve "Yapmaya değer mi?". Bu soruların yanıtları, "evet", "belki" ve "hayır" olarak verilebilir. Geniş kapsamlı yapılan çalışmaların hepsi de söz edilen girişimin sonuç vermediğini gösteriyorsa, bu tedavinin etkili olmadığını kanıtı olarak ele alınabilir. Tek bir randomize kontrollü çalışmada elde edilen olumsuz sonuçlar hemen genellenmemelidir. Karl Popper'in klasik örneğinde olduğu gibi, dünyada hiç siyah kuğu yoktur diyebilmek için, dünyadaki tüm kuğuları görmek gerekir. Bazı tedavilerin etkili olmadığı iddiaları, bazı doktorların hastalarını tedavi etmelerini engelleyebildiği gibi, diğer başka tedavilerin yan etkilerine de açık hale getirebilir. Editörler ve okuyucular, "azaltmadı", "etkisiz", "etkili değil" gibi ifadelerden kaçınmalı ve dikkat etmelidirler. Çünkü aslında gösterilen şey, "etkiye ait kanıt bulunmadığı"dır yoksa "etkisiz olduğunun kanıtlandığı" değildir. Bu nedenle, kullanılması gereken ifadeler "anlamlı farklılıklar bulunamadı," veya "destekleyici veya reddedici kanıtlar yetersizdir," şeklinde olmalıdır. Genel olarak $p > 0.05$ bulunan çalışmalar, yalnızca "fark bulunduğu"na kanıtın yokluğu"nu veya "istenmeyen etkileri olduğuna dair kanıtın yokluğu"nu ortaya koyar. Bu çalışmaları, etkisizliğin kanıtı olarak göstermek yanıltır ve aldatmacadır. Böyle çalışmalarda bazen, geniş güvenlik sınırları, gerçeği ortaya koyar ve kanıt yokluğuna açıklık getirir. Randomize kontrollü çalışmalarda kanıt yok deniyorsa, güvenlik sınırları mutlaka bilinmelidir.

"Önemli sonuçlar" genellikle hastalar ve aileler için önemli kabul edilen sonuçlardır. Ancak ölüm, felç, görme ve işitme kaybı gibi bu sonuçlar çalışmalarda "sıklıkla nadirdir". Dolayısıyla, is-



tatistiksel bir farklılığın gösterilebilmesi için, örneklem büyüklüğünün oldukça fazla olması ve risk altındaki popülasyonun iyi belirlenmesi (paydanın dikkatli seçilmesi) gerekir. Diğer bir deyişle, çalışma için hedeflenen sonuçların, klinik açıdan önemli olup olmadığına dikkat edilmelidir.

Kanıtla dayalı tıbbin temelini, randomize kontrollü çalışmalar (RKÇ) olduğu söylenir. Ancak, altın standart da denilen bu çalışmalar her zaman fazla parlak olmayabilir. Bazı şeyleri bildiğimizi biliriz ve bunun için de RKÇ bulunmasına gerek yoktur. Örneğin, saatte 200 km hızla giden bir arabadan atlayan bir kişinin, sağlıklı kalıp kalmayacağı hakkında yapılmış bir RKÇ var mıdır? Kanıtla dayalı tıp açısından bunun yanıtı verilemez ama yanıtı herkes bilir! Zaten etik açıdan bazı çalışmaların yapılması olanaksız olduğu için, yalnızca elde edilen veriler veya tahminler ile klinik uygulamalar yapılabilmektedir. Randomize kontrollü çalışmalar, klinik girişim ile sonuç arasındaki ilişki basit ve kısa ise ve sonuçlar başka durumlara da rahatlıkla uygulanabilir ise kıymetlidir. Ancak RKÇ'ler genellikle pahalıdır, daima yapaydır, seçilmiş ve kısıtlı bir grupta yapılmışlardır ve bazı kişiler dışarıda bırakılmıştır. Halbuki gerçek hastaların durumları ve sorunları, çok daha farklı ve karmaşıktır. Diğer bir deyişle, bu çalışmalar, biraz "hayattan kopuk"lardır. Yol gösterici olabilirler ama gerçekleri tam olarak yansıtamayabilirler. Bu nedenle RKÇ'lerin, gerçek hayattaki gözlemsel çalışmalar, yani pratik klinik çalışmalar ile desteklenmesi gerekir. Yani, kanıtla ulaşmak, yalnızca RKÇ ile mümkün ve

doğru değildir, non-randomize çalışmaların da bu bağlamda önemi büyüktür. Zaten gerçek hayatta da RKÇ'ler her zaman doktorların klinik soru(n)larına tam olarak yanıt verememektedir. Hastaların sağlığından ve tedavisinden birebir sorumlu olanlar doktorlardır, RKÇ'ler değil. Fakat bu durum, herkesin her istediğini yaptığı kaotik bir ortama da yol açmamalıdır. Günlük hayatta sıklıkla kullanılan bazı ifadelerin gerçekte ne anlamlara geldiğine dair örnekler verilebilir.

Görüldüğü gibi, kanıtla dayalı tıp çalışmalarının dikkatle değerlendirilmemesi birçok hatayı beraberinde getirmektedir. Kanıtlar, gerçek kanıt olmadıkça, uygulamada sorunlarla karşılaşmak kaçınılmazdır. Bir tedavi veya girişimin gerçekten daha iyi veya daha etkili olduğu ve yalnızca istenen etkileri meydana getirdiğini kanıtlamak kolay değildir. Bir tedavi veya uygulamanın her hastada gerekli olmadığı, etkili olmadığı ve yapılmadığı takdirde herhangi bir önemli istenmeyen etkiye yol açmayacağını kanıtlamak ise çok daha zordur. Randomize kontrollü çalışmalardan elde edilen sonuçlar ne kadar etkileyici olursa olsun iyi klinik uygulamalar için tek başına yeterli değildir. Biz, hastalarımızın her birinden tek tek sorumluyuz; randomize kontrollü çalışmalar, büyük adamlar, rehberler veya uzmanlar değil. Bizi, neyi uyguladığımız kadar, neyi uygulamadığımız da sorumlu kılar. Umulur ki, bu yazıyı buraya kadar okuduktan sonra, artık sizin de bazı uzman görüşleri (uzmanların görüşleri değil) olur. Unutmayın; hastalar sizin hastanız.

Kaynaklar

Sinclair JC. Weighing risks and benefits in treating the individual patient. *Clin Perinatol* 2003; 30: 251-68

Joffe M. Evidence of absence can be important. *BMJ* 2003; 326: 1267

Alderson P, Groves T. What doesn't work and how to show it. *BMJ* 2004; 328: 473

Alderson P, Chalmers I. Survey of claims of no effect in abstracts of Cochrane reviews. *BMJ* 2003; 326: 475

Sinclair JC, Haughton DE, Bracken MB, Horbar JD, Soll RF. Cochrane neonatal systematic reviews: a survey of the evidence for neonatal therapies. *Clin Perinatol* 2003; 30: 285-304

Sackett DL. Patients and therapies: getting the two together. *N Engl J Med* 1978; 298: 278-0

Chalmers I. Well informed uncertainties about the effects of treatment. *BMJ* 2004; 328: 475-6

Fretheim A, Schunemann HJ, Oxman AD. Improving the use of research evidence in guideline development: 15. Disseminating and implementing guidelines. *Health Res Policy Syst* 2006; 4: 27

Mainland D. Medical statistics-thinking vs. arithmetic. *J Chronic Dis* 1982; 35: 413-7

İfade	Gerçek anlamı
"Benim deneyimime göre, bu tedavi işe yarar"	Ben aynı hatayı tekrarlarım
"Bu tedaviyle ilgili bizim bir sorunumuz olmadı"	Verilerini dikkatli değerlendirmiyorum
"Ben bu işi birçok kez yaptım ve çok başarılı oldum"	Yalnızca 2-3 kez yaptı veya gözlem kapasitesi yetersiz
"Bu iş bize doğru geliyor"	Başkalarının (örneğin hastaların) ne çektiği durumumuzda değil
"Ben şimdiye kadar öyle bir sorun görmedim"	Ya çalışmıyor ya dikkat etmiyor ya riskli hastaları takip etmiyor veya sorun olduğunda tatildeydi
"Böyle yapıyoruz çünkü falan ünlü (doktor) böyle dedi"	Ben kendim düşünmüyorum ve hep başkalarını takip ederim
"Böyle yapıyoruz çünkü falan ünlü üniversite (veya hastane) böyle yapıyor"	Doğru olmayan birçok işi, birçok yer yapabilir
"Biz bu işi 10 yıldan beri böyle yapıyoruz ve hiçbir şey olmadı"	Benim kafamı karıştırma, ben ne yapacağıma önceden karar vermişim
"Çalışmalar net değil"	Çalışmaların yalnızca özetini okumuş veya çok eskiden okuduğu çalışmaları hatırlıyor
"Baktığımız hastaların en iyi ve en etkili tedavi alması için elimizden geleni yapıyoruz"	Güzel. "Kanıtları" dikkatle değerlendir, yan etkileri gözle, zarar vermemeye çalış