

## Temel Klinik Araştırma Düzenleri: Paralel Gruplu Araştırmalar ve Çapraz Gruplu Araştırmalar

Dr. Oktay ÖZDEMİR

Omega Araştırma Organizasyon Eğitim Danışmanlık Ltd. Şti. Ankara

✓ Bu yazıda temel klinik araştırma yöntemleri özetlenmiştir.  
**Anahtar kelime:** Klinik araştırma yöntemleri

✓ **Basic Clinical Research Systems: Parallel Group Studies and Cross-Over Group Studies**  
Basic clinical research methods are summarized in this paper.  
**Key word:** Clinical research methods.

İki ya da daha fazla hasta grubunun bir çalışmanın parçası olarak, fakat farklı gruplarda yer aldığı çalışmalara bağımsız kontrollü araştırmalar (paralel gruplu araştırmalar) denir. Örneğin iki farklı tedavinin etkilerinin karşılaştırılması amacıyla yapılan bir çalışmaya alınan 60 hastadan 30'una A ilacı, 30'una B ilacının verilmesi durumunda bu çalışma, paralel gruplu çalışma olarak adlandırılır (Şekil 1).

Öte yandan çalışmaya alınan tüm hastalara farklı zamanlarda tüm tedavilerin verilmesi de mümkündür. Yukarıdaki örnekte çalışmaya alınan hastalardan 30'una birer ay süreyle önce A, sonra B ve diğer 30 kişiye önce B, sonra A tedavisinin verildiğini düşünün. Bu durumda bu çalışma çapraz gruplu (cross-over) çalışma olarak adlandırılır (Şekil 2).

Paralel ve çapraz çalışma düzeni, tıpta özellikle ilaç etki ve yan etki çalışmalarında ve tanısal işlemlerin karşılaştırılmalarında en sık kullanılan araştırma yöntemleridir. İkisinin kendine özgü üstünlük ve sakıncaları vardır.

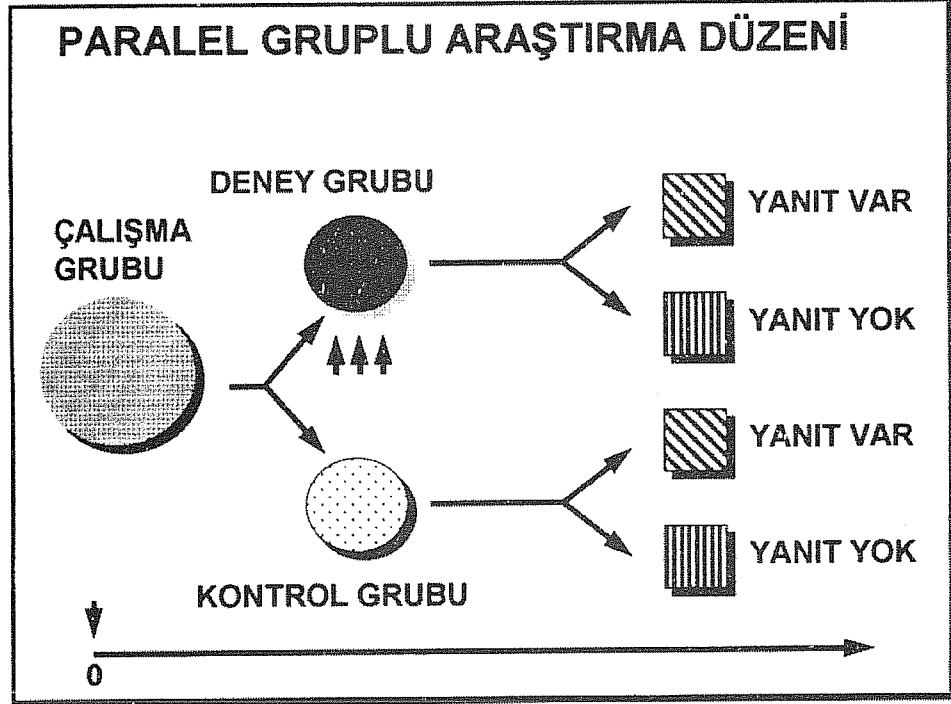
### Paralel Gruplu Araştırmalar

Paralel çalışmalarda farklı kişilere farklı

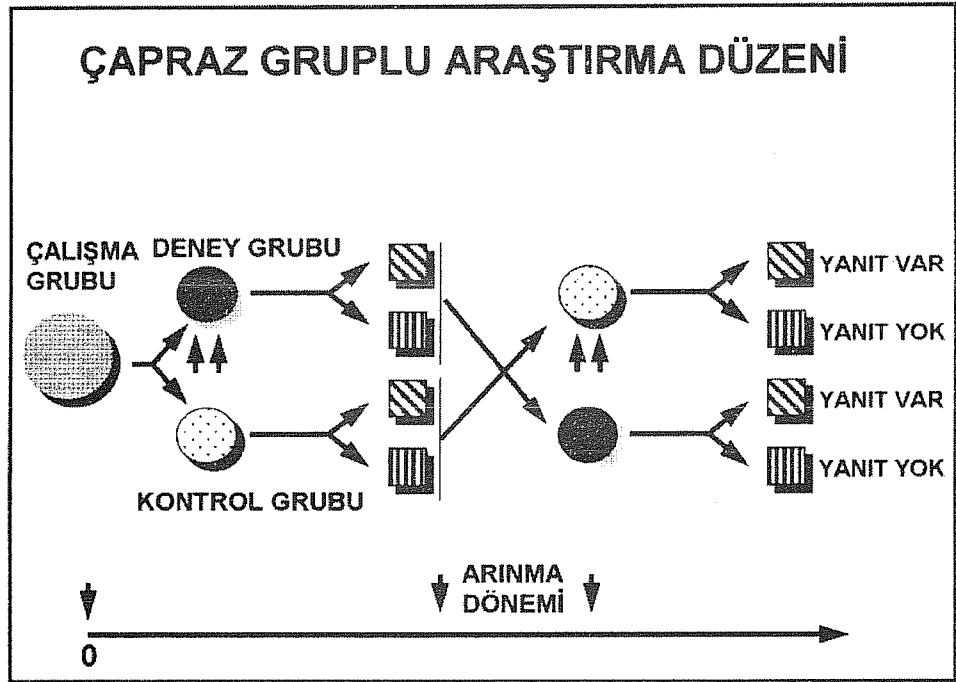
ilaçlar verilip, gruplar arasında fark olup olmadığı incelenir ve eğer fark varsa, bunun nedeninin ilaçlar arasındaki farklılardan kaynaklandığı sonucuna varılmaya çalışılır. Bu sonucun doğru ve güvenilir olabilmesi için farklı ilaç gruplarına alınacak kişilerin ilaca yanıtı etkileyecek diğer etkenler açısından farklı olmamasının sağlanması gerekir.

Farklı tedavi gruplarına alınan hastalar, tedavi verilmeden önce yaş, hastalığın ağırlığı, eşlik eden diğer medikal sorunlar gibi prognoza etki eden diğer etkenler açısından farklılık gösterirse, tedavi sonuçlarının değerlendirilmesinde taraf tutma ya da önyargı (*bias*) olur. Bu etkenlere "prognostik etkenler" ya da "kontrol değişkenleri" (*covariate, confounding variable*) adı verilir. Prognostik etkenlerin gruplarda farklı dağılım göstermesi ise dengesizlik (*imbalance*) olarak tanımlanır.

Grupların prognostik etkenler açısından istatistiksel olarak farklı olmadığını gösterilmesi, grupların dengeli olduğunu kanıtlamaz. Tedavi grupları arasında istatistiksel olarak önemli fark olmamasına dayanarak, grupların gerçekten benzer olduğunu kabul



Şekil 1. Paralel gruplu araştırma düzeni. Aşağıdaki kalın ok çalışmanın yönünü ve "0" çalışmanın başladığı anı temsil etmektedir.



Şekil 2. Çapraz gruplu araştırma düzeni. Aşağıdaki kalın ok çalışmanın yönünü ve "0" çalışmanın başladığı anı temsil etmektedir.

etmek doğru değildir. Örneğin iki gruba 25'lerden 50 hastanın randomize edilmesi sırasında cinsiyet açısından bir gruba 10 erkek, 15 kadın, diğer gruba ise bunun tam tersi sayılarda denek seçilmiş olduğu takdirde, hesaplanan p değeri 0.20 olacaktır. Yani iki grup arasında cinsiyet açısından istatistiksel olarak önemsiz düzeyde fark olduğu halde, iki grubun cinsiyet açısından dengesiz olacağı söylenebilir. Önemli olan, gruplar arası dengesizliğin istatistiksel önemi değil, klinik önemidir.

Prognostik etken dengesizliğinin üç çözümü vardır:

1. Randomizasyon.
2. Tabakalı analiz.
3. İstatistik analizde asıl etki (yani tedavi) dışındaki etkenleri de (yani diğer prognostik etkenleri de) dikkate alan bir model seçilmesi.

Bu üç yöntem ayrı ayrı ya da daha sık olarak birlikte kullanılır.

### Randomizasyon

Paralel düzende yapılan bir çalışmada önce hastalar çalışmaya alınma ölçütleri açısından değerlendirilir. Ölçütlere göre uygun olmayan hastalar çalışma dışı bırakılır. Uygun hastalar ise tedavi gruplarından birine alınır. Hastalara verilecek tedavinin seçilme yöntemi yapılacak çalışmanın sonuçlarının güvenilirliği açısından çok önemlidir. Tedavi seçiminin araştırmacının insiyatifinde olmasının önemli sakıncaları vardır. Örneğin yeni ve etkinliği pek de bilinmeyen bir antidepresif ilacın plaseboyla karşılaştırıldığı bir çalışmada araştırmacılar bilinçli ya da bilinçsiz olarak, depresyon bulgularının ağır olduğu hastalarda ilacı, hafif olduğu hastalarda plaseboyu seçebilir. Böylece ilaç grubunda ağır depresyonlu, plasebo grubunda hafif depresyonlu hastalar yer alacağı için elde edilecek sonuçlar ilaç aleyhine olabilir.

Bu nedenle hastaların çalışma gruplarına seçiminin rastgele olmasının sağlanması gerekir. Bunu sağlayan yöntemlere randomizasyon adı verilir. Randomizasyon, hastaların tedavi gruplarına seçilmesinde bilinçli ya da bilinçsiz olarak yapılan "manipülasyon"ları engeller.

En kolay randomizasyon yöntemi basit randomizasyondur. Örneğin hastalara verilecek tedavinin yazı-tura ile belirlenmesi bir basit randomizasyon yöntemidir. Bu yöntemle hastaların prognostik etkenler açısından durumları dikkate alınmadığı halde, özellikle denek sayısı fazla tutulduğunda genellikle dengeli dağılım sağlanabilir. Denek sayısı az olduğu zaman basit randomizasyonla bile gruplar arasında dengesizlik olabilir. Denek sayısı arttıkça, gruplar arasında dengesizlik olma olasılığı azalır.

Denek sayısının az olduğu bir çalışmada basit randomizasyon tek başına dengesizliğin azaltılmasında yeterli olmayabilir. Bu durumda randomizasyona şu iki yöntemden biri ya da her ikisi birlikte eklenmelidir:

1. Tabakalandırma,
2. Bloklama.

Araştırmacı, önce hastaları bir ya da daha çok prognostik etkene göre sınıflandırır (tabakalandırır), daha sonra, her tabakaya baştan belirlenmiş (ve genellikle eşit) sayıda hasta girecek şekilde hastaları seçer (bloklar). Örneğin her iki erkek hastadan birini bir gruba, diğerini diğer gruba seçecek şekilde bir randomizasyon şeması kurar ve kadınları da aynı şekilde seçerse, çalışma grupları cinsiyete göre tabakalandırılmış ve ikili gruplar halinde bloklanmış olur. Örneğin ilk erkek hastayı yazı-tura ile A grubuna seçtikten sonra gelen erkek hasta zorunlu olarak B grubuna alınır. Üçüncü erkek hastada yine yazı-tura ile ilaç seçimi belirlenir. Aynı işlem kadınlar için de benzer şekilde yapıldığında, tabakaların birleştirilmesi ile gruplar ken-

diliğinden dengeli olur. Yazı-tura yerine istatistik kaynak kitaplar ya da bilgisayarlardan elde edilebilecek rastgele sayılar tablosundan yararlanarak da randomizasyon yapılabilir. Tablo I'de cinsiyetin tabaka değişkeni olduğu ve blokların ikişer kişiden oluştuğu bir randomizasyon şeması görüyorsunuz. Blok I'de ve II'de ikişer erkek, blok III ve IV'de ikişer kadın yer almaktadır. Her bloktaki hastaların biri ilaç grubuna, diğeri ise plasebo grubuna seçilmiştir. Böylece iki grubun cinsiyet açısından benzer olması sağlanmıştır. Tablo I'deki randomizasyon şemasına göre ilk erkek hastaya ilaç, ikinci ve üçüncü erkek hastaya plasebo, dördüncü erkek hastaya ilaç verilecektir.

**Tablo I.** Cinsiyet Açısından Tabakalandırılmış ve Her Bloкта İki Hasta Olmak Üzere Bloklama Yapılmış Bir Çalışmadaki Randomizasyon Şeması.

Tabaka	Blok	İlaç	Plasebo
Erkek	I	+	+
	II	+	+
Kadın	III	+	+
	IV	+	+
		4	4

Tabakalandırmaya mutlaka bloklama da eşlik etmelidir. Aksi takdirde basit randomizasyon ile bloklama yapılmadan yapılan tabakalandırma arasında, gruplar arası dengenin sağlanması açısından fark yoktur. Ama tabakalandırma olmaksızın bloklama yapılabilir. Örneğin tek amaç gruplarda denek sayısının eşit olması ise, tek başına bloklama yeterlidir.

Birden çok özellik açısından da tabakalandırma yapılabilir. Kaç adet tabakalandırma değişkeni kullanılacağına seçiminde tedaviye yanıtı etkileyen prognostik etkenlerin (ya da kontrol değişkenlerinin) araştırmacı tarafından, henüz araştırma başlamadan belirlenmesi önemlidir. Prognostik etkenlerin neler olduğu ya literatürdeki bilgilere dayanarak, ya da araştırmacının deneyimlerine dayanarak seçilmelidir. Örneğin skolyoz cerrahisi sırasında desmopressin kullanılması perioperatif kanamayı azaltacağı hipoteziyle yapılan plasebo-kontrollü bir çalışmada kanama miktarını etkileme olasılığı olan özellikler açısından desmopressin ve plasebo gruplarının benzer olması gereklidir. Bu nedenle hastanın yaşı, operasyona alınacak bölgedeki vertebra sayısı ve skolyoz nedeni gibi etkenler açısından tabakalandırma gerekebilir. Tabakalandırmada özen gösterilmesi gereken bir konu, her tabaka değişkeninin nasıl kategorize edileceğidir. Tablo I'deki örnekte "cinsiyet" in iki adet kategoriden oluşacağına kuşku yoktur. Ancak yaş, vertebra sayısı gibi sayısal değişkenler için tabakaların kategorilerinin araştırmacı tarafından belirlenmesi gerekecektir.

Eğer gereğinden fazla kategori kullanılırsa, araştırmaya alınacak hasta sayısının çok artması ve buna rağmen gruplar arası dengenin tam olarak sağlanamaması gibi sakıncalar ortaya çıkabilir. Örneğin "yaş" tabakasının kategorileri "0-4, 5-9, 10-14, 15-19, ..." gibi beşer yıllık ve "vertebra sayısı" tabakasının kategorileri "1-3, 4-6, 7-9, ..." gibi üçer vertebra olarak belirlendiğini ve "skolyoz nedeni" nin ise "doğumsal, edinsel" olarak iki kategoriden oluştuğunu varsayalım. Operasyon bölgesindeki vertebra sayısı "7-9" grubuna giren 6 yaşında (yani 5-9 yaş grubuna giren) doğumsal skolyozu olan bir hasta desmopressin grubuna alındığında, plasebo grubuna da vertebra sayısı ve yaşı benzer olan

doğumsal skolyozlu bir hastanın alınması gerekecektir. Aksi takdirde gruplar, prognostik etkenler açısından dengeli olmayacak ve tabakalandırma işe yaramamış olacaktır.

Öte yandan tabakaların kategorilerinin fazla geniş olması da sakıncalıdır. Örneğin "vertebra sayısı" tabakasının kategorileri "1-10 ve >10 vertebra" olarak belirlenirse, 2 vertebraya müdahale yapılacak bir hastayla 10 vertebraya müdahale yapılacak bir hasta aynı kategoride ele alınacak ve kanama miktarına vertebra sayısının etkisi açısından bu iki hasta arasında fark olmadığı varsayılacaktır.

Hem tabaka sayısı, hem de tabakalardaki kategori sayısı arttıkça araştırmanın yürütülmesi ve grupların dengesinin sağlanması güçleşir. Bu nedenle prognostik etkisi belirgin olmayan değişkenlere göre tabakalandırma yapılmasından kaçınılması uygun olur. Aynı şekilde tabakaların kategorilerinin alt ya da üst sınırlarının belirlenmesinde de prognostik açıdan aralarında fark olduğu varsayılan sınırlar seçilmelidir. Örneğin alt solumun yolu infeksiyonu gelişmesinde yaşın etkisine göre grupların benzerliğini sağlamak için hastaların "45'in altında - üstünde" olarak iki kategoriye ayrılması yeterli olursa "20-24, 25-29, 30-34, ..." gibi dar kategoriler oluşturmaya gerek yoktur.

Tabakaların kategorilerinin eşit genişliklerde olması gerekmez. Örneğin özellikle ilk yaşta ve 60 yaşından sonra daha ağır seyreden ve bu nedenle ilaca yanıt açısından bu yaş gruplarında farklı seyir gösteren bir hastalık için yaş açısından tabakalandırma yapılacağı zaman kategoriler "0-1 yaş, 2-20 yaş, 21-60 yaş ve >60 yaş" olarak belirlenebilir. Önemli olan kategorilerin alt ya da üst sınırlarının, hastaları prognoz açısından farklı olan gruplara ayırmasıdır.

Tekrar skolyoz cerrahisi örneğine dönersek, araştırmacı tarafından yaşın "<20, 20-49 ve >49" olarak üç kategoriye, vertebra sayısı

tabakasının "1-5, 6-10 ve >10" olarak üç kategoriye ve "skolyoz nedeni"nin ise "doğumsal, edinsel" olarak iki kategoriye ayrıldığını düşünelim. Bu durumda bu üç prognostik etken açısından  $3 \times 3 \times 2 = 18$  kombinasyon olacaktır (Tablo II).

Çalışma süresince bu 18 kombinasyondan bazılarında uygun hastaların bulunması tabakalandırma ve bloklama için bir engel değildir. Örneğin hem desmopressin, hem de plasebo grubunda 20 yaşından küçük, 10'dan fazla vertebraya müdahale yapılacak doğumsal skolyozu olan hasta olmaması sorun oluşturmaz. Önemli olan toplam 18 altgruptaki hasta sayılarının desmopressin-plasebo gruplarında eşit olmasıdır. Ancak her şeye rağmen tüm prognostik etkenler açısından gruplar arasında denge tam olarak sağlanamayabilir.

#### **Tabakalı analiz**

Prognostik etkenler açısından grupların dengeli olmasını sağlamanın en güvenilir yolu tabakalı randomizasyondur. Ancak kimi zaman araştırmaya başlarken prognostik önemi olmadığı sanılan bir etkenin araştırma sırasında ya da araştırmanın yürütülmesi bittikten sonra prognostik öneminin olduğu ve bu özellik açısından grupların dengeli olmadığı anlaşılabilir. Örneğin araştırmanın başlangıcında henüz bilinmeyen bir etkenin prognostik önemi olduğu, araştırma henüz devam ederken ya da bittikten sonra başka araştırma grupları tarafından gösterilebilir, ya da daha sıklıkla araştırmaya başlamadan önce yapılan literatür taramasının yetersiz olmasına bağlı olarak, başlangıçta yapılan randomizasyonda bu etkene göre tabakalandırma yapılmamış olabilir.

Bu durumda prognostik etkenlerin dengesizliğinin çözümünde kullanılacak bir yöntem, analizi tabakalandırmaktır. Tabakalı analiz, çalışma grupları içindeki prognostik

**Tablo II.** Yaş, Vertebra Sayısı ve Skolyoz Nedeni Açısından Tabakalandırma Yapılan Bir Çalışmada Desmopressin ve Plasebo Grubuna Alınan Hasta Sayıları. Görüldüğü Gibi "<20 Yaş, 6-10 Vertebra Sayısı ve Doğumsal Skolyoz" ve "20-49 Yaş, 6-10 Vertebra Sayısı ve Doğumsal Skolyoz" Gruplarında Hasta Sayıları Eşit Değildir. Buna Bağlı Olarak Grupların Tam Dengeli Olmadığı Söylenebilir. Bazı Gruplarda Hem Desmopressin, Hem de Plasebo Gruplarında Hasta Olmaması Grupların Dengesi Açısından Sakınca Yaratmaz.

Yaş	Vertebra sayısı	Skolyoz nedeni	Desmopressin	Plasebo
<20	1-5	Doğumsal	3	3
		Edinsel	-	-
	6-10	Doğumsal	4	5
		Edinsel	1	1
	>10	Doğumsal	2	2
		Edinsel	3	3
20-49	1-5	Doğumsal	1	1
		Edinsel	2	2
	6-10	Doğumsal	3	2
		Edinsel	1	1
	>10	Doğumsal	2	2
		Edinsel	4	4
49	1-5	Doğumsal	-	-
		Edinsel	2	2
	6-10	Doğumsal	-	-
		Edinsel	1	1
	>10	Doğumsal	-	-
		Edinsel	3	3
<b>Toplam</b>			<b>32</b>	<b>32</b>

etkenler açısından homojen altgrupların analizinin ayrı ayrı yapılmasıdır. Bu yöntemle çalışmanın gücü azalır, ama ciddi dengesizlik nedeniyle işe yaramayacak bir çalışmada yapılması gereken budur.

Tabakalı analizin nasıl yapıldığını şu örnekte inceleyelim:

$\beta$ -bloker ve kalsiyum antagonistlerinin anstabil anjina pektoriste miyokard infarktüs (MI) gelişmesini önlemek üzerindeki etkilerinin karşılaştırılması amacıyla düzenle-

nen bir çalışmaya alınan 80 hasta iki gruba ayrılıyor ve bir gruba  $\beta$ -bloker, diğer gruba kalsiyum antagonisti veriliyor. Tedaviye yanıtın farklı cinsiyetlerde farklı olmayacağı varsayıldığı için, hastaların ilaç gruplarına ayrılmasında cinsiyet, tabaka değişkeni olarak kullanılmıyor. Tedaviye yanıt için ilaç başlandıktan sonraki 6 ay içinde hastaların ne kadarının MI geçirdiği gözleniyor.

Tablo III'ten Tablo VIII'e kadar yer alan tablolarda bu çalışmadaki verilere ilişkin çapraz

**Tablo III.** İki Tedavi Grubunda MI Geçiren Hasta Sayılarının Gösterildiği Çapraz Tablo.

	Miyokard infarktüsü		Toplam
	+	-	
β-bloker	21	19	40
Kalsiyum antagonist	30	10	40
<b>Toplam</b>	<b>51</b>	<b>29</b>	<b>80</b>

**Tablo IV.** İki Tedavi Grubundaki Deneklerin Cinsiyete Göre Dağılımının Gösterildiği Çapraz Tablo.

	Erkek	Kadın	Toplam
β-bloker	10	30	40
Kalsiyum antagonist	30	10	40
<b>Toplam</b>	<b>40</b>	<b>40</b>	<b>80</b>

tablolar görülmektedir. β-bloker grubunda 21, kalsiyum antagonist grubunda 30 hastanın MI geçirdiği gözleniyor ve ki-kare analizi ile β-bloker grubundaki MI oranının kalsiyum antagonist grubundan istatistiksel olarak önemli derecede az olduğu görülüyor (p=0.03) (Tablo III).

Ancak gruplardaki deneklerin cinsiyetlere göre dağılımının farklı olduğu görülüyor. β-

bloker grubunda 10 erkek, 30 kadın, kalsiyum antagonist grubunda 30 erkek, 10 kadın var (p<0.001) (Tablo IV). Bu nedenle farklı cinslerde ilaçların etkilerinin farklı olabileceği, ya da verilen ilaç ne olursa olsun farklı cinslerde hastalığın seyrinin farklı olabileceği gibi sorular akla geliyor. Gerçekten de iki farklı tedavi grubundaki denekler cinsiyet ve prognoz arasındaki ilişki açısından incelendiği zaman her iki ilaç grubunda da erkeklerde MI oranının daha yüksek olduğu saptanıyor (erkeklerde p=0.04, kadınlarda p<0.001) (Tablo V ve VI).

Acaba β-bloker grubunda infarktüs oranının düşük olmasının nedeni ilacın daha etkili olması mıdır, yoksa bu grupta hastalığın daha kötü seyrettiği anlaşılan cinsten, yani erkeklerden daha az denek bulunması mıdır? Acaba her iki tedavi grubundaki deneklerin cinsiyet açısından dağılımı benzer olsaydı, sonuç yine β-bloker lehine mi bulunacaktı? Bu sorunun çözümü için tabakalı analiz yöntemi kullanılabilir. İnfarktüs oranlarının, tedaviye yanıt ya da hastalığın seyri açısından farklı olan alt-gruplarda, yani erkeklerde ve kadınlarda tedavi gruplarındaki ayrı ayrı incelenmesi grupların cinsiyet açısından dengesiz olmasının neden olduğu yorumlama gücünü ortadan kaldıracaktır. Gerçekten de hem erkeklerde (p=0.41) (Tablo VII), hem de kadınlarda (p=0.46) (Tablo VIII) hangi ilaç

**Tablo V.** β-Bloker Grubunda Farklı Cinslerde MI Geçiren Denek Sayılarının Gösterildiği Çapraz Tablo.

	Miyokard infarktüsü		Toplam
	+	-	
Erkek	8	2	10
Kadın	13	17	30
<b>Toplam</b>	<b>21</b>	<b>19</b>	<b>40</b>

**Tablo VI.** Kalsiyum Antagonisti Grubunda Farklı Cinslerde MI Geçiren Denek Sayılarının Gösterildiği Çapraz Tablo.

	Miyokard infarktüsü		Toplam
	+	-	
Erkek	27	3	30
Kadın	3	7	10
<b>Toplam</b>	<b>30</b>	<b>10</b>	<b>40</b>

**Tablo VII.** İki Tedavi Grubunda MI Geçiren Erkek Denek Sayılarının Gösterildiği Çapraz Tablo.

	Miyokard infarktüsü		Toplam
	+	-	
β-bloker	8	2	10
Kalsiyum antagonistisi	27	3	30
<b>Toplam</b>	<b>35</b>	<b>5</b>	<b>40</b>

kullanılırsa kullanılsın, prognozun farklı olmadığı görülmektedir.

Bu örnekte olduğu gibi çalışma bittikten sonra altgrupların belirlenmesine "sonradan-tabakalandırma" (post-stratification) adı verilir. Çeşitli etkenlerin prognostik öneminin derecesinin belirlenmesinde sonradan-tabakalandırma yararlıdır, ancak bu yöntemde iki sorun vardır:

1. Çalışmanın gücü azalır ve bazı tabakalarda denek sayısı çok az olacağı için istatistik analiz sonuçları güvenilir olmayabilir.
2. Gruplar arası karşılaştırma sayısı artacağı için "istatistiksel önemlilik enflasyonu" riski vardır.

Yeri gelmişken çalışmanın gücü ya da istatistik analizin gücünü ne anlam geldiğini açıklamakta yarar var. Örneğin iki ilaca yanıtın incelendiği bir çalışmada, iki grupta tedaviye yanıt oranı arasında fark olduğu varsayımı çalışmada toplanan verilere dayanarak test edilir ve analiz sonucunda genellikle herkesin çok aşına olduğu bir p değeri elde edilir. P değeri, kabaca iki grup arasında gerçekte fark olmadığı halde, şans eseri olarak fark varmış gibi sonuç elde edilmiş olması olasılığıdır. Örneğin bir ilaca yanıt oranı %45, diğer ilaca yanıt oranı %60 ve analiz sonucunda hesaplanan p değeri 0.12 ise, iki ilaç arasında gözlenen bu farkın tamamen tesadüfi olma olasılığı %12'dir. Ya da

**Tablo VIII.** İki Tedavi Grubunda MI Geçiren Kadın Denek Sayılarının Gösterildiği Çapraz Tablo.

	Miyokard infarktüsü		Toplam
	+	-	
β-bloker	13	17	30
Kalsiyum antagonistisi	3	7	10
<b>Toplam</b>	<b>16</b>	<b>24</b>	<b>40</b>

daha doğru bir anlatımla, bu araştırma aynı sayıda hasta ile ve aynı koşullarda 100 kez yapılırsa, iki ilaç arasında 12 çalışmada fark bulunamayacak, 88'inde fark bulunacaktır.

Öte yandan gerçekte fark olduğu halde, yapılan çalışmadan elde edilen verilerin analizi ile fark olmadığı gibi bir hata da yapılabilir. Bir çalışmanın gücü, gerçekte fark olduğu zaman bunun istatistik analizle de gösterilebilmesi, yani saptanabilmesi olasılığı olarak tanımlanabilir. Örneğin bir çalışmanın gücünün 0.85 olması, bu çalışma ile eğer gerçekte de fark varsa bunun % 85 olasılıkla saptanabileceği anlamına gelir.

Bu çalışmanın analizini 6 tane çapraz tablo ile yapabildik (Tablo III-VIII). MI geçirmeyi önleme etkisi açısından β-bloker ve kalsiyum antagonistisi arasında fark olup olmadığı amacıyla yapılan çalışmaya 80 hasta alındığı halde erkekler (n=40) ve kadınlar (n=40) olmak üzere altgruplarda analiz yapıldığı için (yani daha az sayıda hasta ile analiz yapıldığı için) çalışmanın gücü önemli derecede azaldı. Eğer çalışmanın başlangıcında cinsiyetin prognoz üzerine etkisi olabileceği öngörülse ve buna göre hastalar tabakalandırılarak randomize edilseydi, 80 hastayla yapılmış olan bu çalışmadaki istatistik analizin gücüne, aşağı yukarı 40 hasta ile yapılacak bir çalışmayla ulaşılabilecekti.



Tabakalı analizin tek sakıncası çalışmanın gücünün azalması değildir. İstatistikte genel ilke olarak mümkün olan en az sayıda analiz ile inceleme yapılmalıdır. Yukarıdaki örnekte hastalara verilen ilaç, cinsiyet ve prognoz olmak üzere üç değişken arasındaki ilişkiyi 6 farklı tabloda analiz etmek yalnızca zahmetli olduğu için değil, aynı zamanda istatistiksel önemlilik enflasyonuna neden olacağı için de sakıncalıdır. İstatistiksel önemlilik enflasyonu, çok sayıda analiz yapınca gerçekte fark olmadığı halde fark varmış gibi sonuç elde edilmesidir. Örneğin Tablo III'ün analizi ile hesaplanan p değerinin 0.03 olması, iki ilaç arasında prognoz açısından tamamen şansa bağlı olarak fark olması olasılığının %3 olduğunu göstermektedir. Bilindiği gibi çoğu kez p değerinin 0.05'ten daha küçük olması halinde istatistiksel olarak önemli fark olduğu söylenir. Ancak bu araştırmada çok sayıda analiz yapıldığı için her analizle hesaplanan p değerlerinin ne kadar güvenilir olduğu tartışılır hale gelecektir. Bu durumda Tablo III'ün analizi ile 0.03 olarak hesaplanan p değeri gerçekte olduğundan daha küçük olarak hesaplanmış olacaktır.

#### **Sonuçların Prognostik Etkene Göre Standardizasyonu (Adjustment)**

Adjustment ya da doğrudan standardizasyon, tabakalardaki sonuçların ağırlıklı ortalamalarının kullanılmasıdır. Ağırlıkların

seçimi, analizin amacına göre değişir.

Standardizasyonun nasıl yapıldığını şu örnekte inceleyelim:

Hipertansiyonda A ve B ilaçlarının plasebo ile karşılaştırıldığı bir çalışmada gruplarda cinsiyet açısından ciddi dengesizlikler olduğu görülmüştür (Tablo IX). Bu nedenle kan basıncındaki düşme miktarı olarak tanımlanan ilaca yanıt kriteri erkek ve kadınlar için ayrı ayrı analiz edilmiştir (Tablo X). Erkek ve kadınlarda kan basıncındaki azalma miktarı arasında farklılık olduğu görülmektedir. Bu nedenle erkekleri ve kadınları içeren tüm grubun sonuçları güvenilir olmayacaktır.

Eğer, gruplar cinsiyet açısından dengeli olsaydı, tedaviye yanıtın nasıl olacağını kestirmek için sonuçlar standardize edilebilirdi. A ilacı alan hastalarda gözlenen ortalama kan basıncı düşmesi 23.9 mmHg'dır. Çalışmaya alınan 150 kişiden yarısı erkek, yarısı kadın olduğuna göre, bu 150 kişiye A ilacı verilseydi, kan basıncındaki ortalama düşme  $(25.4 \times 0.50) + (22.4 \times 0.50) = 23.9$  mmHg olacaktır. Hesaplanan bu değer standardize edilmiş ortalama kan basıncı azalma miktarıdır. Bu gruptaki hastaların cinsiyete göre dağılımı tüm gruptaki dağılımın aynısı olduğu için, yani yarı yarıya erkek ve kadın olduğu için standardize ortalama, gözlenen ortalamadan farklı değildir. B ilacı verilen grupta gözlenen ortalama kan basıncı düşmesi 34.0 mmHg'dır. Benzer şekilde he-

**Tablo IX.** Gruplardaki Hastaların Cinsiyete Göre Dağılımı.

	A ilacı	B ilacı	Plasebo	Toplam
Erkek	20	15	40	75
Kadın	20	35	20	75
Toplam	40	50	60	150

**Tablo X.** Kan Basıncında Ortalama Düşme Miktarı (mmHg).

	A ilacı	B ilacı	Toplam
Erkek	25.4	22.4	1.2
Kadın	22.4	38.9	8.9
Genel ortalama	23.9	34.0	4.5
Standardize ortalama	23.9	30.6	5.1

saplandığında standardize ortalama,  $(22.4 \times 0.50) + (38.9 \times 0.50) = 30.6$  mmHg olarak hesaplanmıştır. Yani bu grupta 15 erkek, 35 kadın değil de, 25 erkek, 25 kadın olsaydı, ortalama kan basıncı düşmesi 30.6 mmHg olacaktı. Plasebo grubunda ise gözlenen ortalama 4.5 mmHg, hesaplanan standardize ortalama ise  $(1.2 \times 0.50) + (8.9 \times 0.50) = 5.1$  mmHg'dır. Bu üç ortalama, artık cinsiyetin etkisinden arıtılmıştır, çünkü her üçü de % 50'si erkek, % 50'si kadın olan hipotetik popülasyonlarda gözlenmesi beklenen değerlerdir.

#### **Kontrol Değişkenlerinin Etkilerinin Arıtıldığı İstatistik Modeller**

Gruplar arasında dengesizlik olduğu zaman kullanılabilir bir yöntemin tabakalı analiz olduğundan ve bu yöntemin sakıncalarından yukarıda söz edildi. Dikkat ederken, tabakalı analizde prognostik etkenlere göre her alt grubun analizini ayrı ayrı yaparken, analize aldığımız hasta sayısı çalışmaya aldığımız hasta sayısından daha az oluyor ve buna bağlı olarak çalışmanın gücü azalıyor. Tabakalı analizin bu sakıncalarını ortadan kaldıran ve sonuç üzerine prognostik etkenlerin (kontrol değişkenlerinin) etkilerinin arıtıldığı ve tedavinin "saf" etkisinin belirlenebildiği matematik modeller vardır. Bu yöntemlerle sonuç üzerine etkisi olan etkenler (verilen ilaç, cinsiyet, hastalığın şiddeti, yaş vb.) parçalara ayrılarak incelenir. Buna göre yukarıdaki örnekte kan basıncındaki düşme miktarına tedavinin katkısı ve cinsiyetin katkısı ayrı ayrı hesaplanabilir.

Anstabil anjina pektorisinde  $\beta$ -bloker ve kalsiyum antagonistlerinin karşılaştırıldığı çalışmada tabakalı analizle iki ilaç arasında fark olmadığı, ilk bakışta gözlenen farkın iki gruptaki hastaların cinsiyete göre dağılımları arasındaki farktan kaynaklandığı 6 farklı çapraz tablonun istatistiksel analizi ile gösterilmişti. Ancak bu sonuca ulaşabilmek

için çalışmanın gücünün azalması ve istatistiksel önemlilik enflasyonu riskini gözlemledik. Bu çalışmada alternatif olarak log-lineer analiz yapıldığında iki tedavi grubundaki hastaların cinsiyet dağılımının benzer olmadığı ( $p < 0.0001$ ), prognozun iki cinside benzer olmadığı ( $p = 0.0001$ ), ancak tedavi ile prognoz arasında ilişki olmadığı ( $p = 0.88$ ) saptanmıştır. Burada üzerinde önemle durulması gereken nokta tedavi ile prognoz arasında ilişki olmadığını söylerken, cinsiyetin etkisinin tamamen arıtılmış olmasıdır.

Prognostik etkenlerin etkilerinin arıtılmasında hangi durumlarda hangi istatistik modellerin uygulanabileceği Tablo XI'de özetlenmiştir.

Sonradan-tabakalandırma ile grup karşılaştırmaları sayısındaki artmaya bağlı "istatistiksel önemlilik enflasyonu", bu istatistik modellerle önlenemez. Bu yöntemlerle, hem tedavinin etkisi prognostik etkenlerden arıtılmış, hem de prognostik etkenlerin önemi test edilmiş olur.

Gerek tabakalı analiz, gerekse kontrol değişkenlerinin etkilerinin arıtıldığı modellerin amacı, gruplar arası dengesizliğin neden olduğu taraf tutmayı, ya da önyargıyı ortadan kaldırmaktır. Ancak iki yöntem arasında önemli bir fark vardır. Tabakalı analizde, tabakalardaki denek sayısı azalacağı için, çalışmanın gücünün azalması pahasına, taraf tutma ortadan kaldırılabılır. Kontrol değişkenlerinin etkilerinin arıtıldığı modeller kullanıldığında ise, (eğer doğru model kullanılırsa) hem taraf tutma azalır, hem de sonuçların doğruluk derecesi artar. Böylece istatistiksel önemliliği sınırda olan yargıların doğruluğu artabilir, ya da bazen tersine farklı tedavi grupları arasında istatistiksel olarak önemli görünen farklılıkların tedavinin kendisinden değil, gruplar arası prognostik etken dengesizliğinden kaynaklandığı ortaya konabilir. Model kullanılmasının bir diğer yararı

**Tablo XI.** Kontrol Değişkenlerinin Etkilerinin Artıldığı İstatistik Modeller.

Grup sayısı	İzlenen değişken	Kontrol değişkeni	Temel analiz	Tabakalı analiz	Alternatif model
2	Sayısal	Sayısal	T testi	-	Kovaryans analizi
2	Sayısal	Kategorik	T testi	T testi	İki yönlü ANOVA
2	Dikotom*	Sayısal	X2 testi	-	Lojistik regresyon analizi
2	Dikotom	Kategorik	X2 testi	X2 testi	Log-lineer analiz
2	>2 düzeyli kategorik**	Sayısal	X2 testi	-	Diskriminant analiz
2	>2 düzeyli kategorik	Kategorik	X2 testi	X2 testi	Log-lineer analiz
>2	Sayısal	Sayısal	Tek yönlü ANOVA	-	Kovaryans analizi
>2	Sayısal	Kategorik	Tek yönlü ANOVA	Tek yönlü ANOVA	İki yönlü ANOVA
>2	Dikotom	Sayısal	X2 testi	-	Lojistik regresyon analizi
>2	Dikotom	Kategorik	X2 testi	X2 testi	Log-lineer analiz
>2	>2 düzeyli kategorik	Sayısal	X2 testi	-	Diskriminant analiz
>2	>2 düzeyli kategorik	Kategorik	X2 testi	X2 testi	Log-lineer analiz

\*: Yanıt kriteri "yanıt var-yok" ya da "öldü-yaşiyor" gibi iki sonuçtan ibaretse.

\*\* : Yanıt kriteri "yanıt tam-kısmi-yok" ya da "hafif-orta-ağır" gibi ikiden daha çok sonuç içeriyorsa.

ise prognostik etkenlerin önemlilik derecesinin belirlenmesi ve daha sonra yapılacak çalışmaların planlanmasında yararlı olacak veriler elde edilmesidir.

### Prognostik Etken Denetim

#### Yöntemlerinin Karşılaştırılması

Çalışma grubu (örneklem) yeterince genişse, basit randomizasyon, ölçülen ve ölçülmeyen prognostik etkenler açısından yeterli dengeyi genellikle sağlar. Randomizasyona bir-iki önemli prognostik etkene göre tabakalandırma ve bloklama eklenirse, gruplar daha da dengeli olur. Herşeye karşın, çalışmaya başlamadan önce denetlenmeyen bazı özellikler açısından gruplarda dengesizlik olabilir. Bu durumda denek sayısının izin verdiği ölçüde sonradan-tabakalandırma yapılarak tabakalar ayrı ayrı analiz edilebilir.

Grupların dengesi mükemmel olsa bile, analiz sırasında prognostik etkenlere göre standardizasyon yapılması ile gruplarda tedaviye yanıtın karşılaştırılması daha da saflaştırılmış olur.

Sonradan-tabakalandırma ve prognostik etkenlere göre standardizasyon birbirini tamamlayan iki yöntemdir. Sonradan-tabakalandırma ile prognostik etkenler açısından farklı olmalarına karşın, grupların karşılaştırılmasının ne kadar geçerli olduğu test edilmiş olur. Standardizasyon ile sonuçların doğruluğu daha da belirginleştirilip, çalışmadan daha fazla bilgi edinilmiş olur.

Araştırmanın amacı, çalışmaya alınan hastalardaki sonuçların tüm hastalara genelleştirilebilmesidir. Bu yazıda anlatılan randomizasyon ve prognostik etken denetim çabalarının nedeni budur.

### Çapraz Gruplu Araştırmalar

Paralel çalışmalarda çalışma gruplarından birine alınan hastaya tek bir tedavi uygulanır. Çapraz çalışmalarda ise aynı hastaya iki ya da daha fazla tedavi, belli bir plan çerçevesinde arka arkaya uygulanır.

İki tedavili çapraz çalışmada her hastanın A tedavisine yanıtı ile B tedavisine yanıtı karşılaştırılır. Böylece tedaviye yanıtı etkileyen hasta özelliklerinin etkileri artırılmış olur. Çapraz çalışma ile hastalar arası değişkenliğin etkisi artılmakla birlikte, biyolojik özelliklerin zamanla değişimi ya da ölçümler arasındaki değişkenliğin etkileri kontrol edilemez. Eğer bu etkiler hastalar arası değişkenlikten daha azsa ya da daha önemsizse, çapraz çalışma düzeni ile, paralel çalışma düzenine göre daha az sayıda denekle aynı doğrulukta bir çalışma yapılabilir. Yani çapraz çalışmalar, paralel çalışmalara göre daha güçlüdür.

İki antihipertansif ilacın sistolik kan basıncı üzerine etkileri dört hipertansif hastada çapraz çalışma düzeni ile araştırılıyor. Hastalara önce A, sonra B ilacı veriliyor. Tablo XII'de bu hastaların A ilacı ve B ilacı alırken sistolik kan basıncındaki düşme miktarları görülmektedir. Dört hastada da sistolik kan basıncındaki düşmenin B ilacı ile daha fazla olduğu ve ortalama sistolik kan basıncı düşmesi arasındaki farkın 9 mmHg ve farkın standart hatasının\* 1.2 mmHg olduğu görülmektedir.

Eğer Tablo XII'deki 8 ölçüm, her iki grupta 4'er hasta bulunan toplam 8 denekli bir paralel çalışmadan elde edilmiş olsaydı, sis-

tolik kan basıncındaki düşme miktarları arasındaki farkın standart hatası 1.20 değil 8.28 olacak ve iki tedavi grubu arasındaki fark istatistiksel olarak önemli olmayacaktı. (Hesaplarsanız, standart hata 8.28 olduğunda farkın %95 olasılıkla alabileceği değerlerin arasında 0, hatta negatif değerler de olduğunu görürsünüz; yani B ilacı lehine görülen 9 mmHg fark gerçek durumu çok iyi temsil etmemektedir. Oysa, çapraz çalışma düzenine göre hesaplanmış olan standart hata 1.20'dir ve A ile B tedavileri arasındaki fark istatistiksel olarak önemli bulunmuştur.

Ancak bu sonucun yorumlanmasında dikkat edilmesi gereken bir konu tüm hastalar önce A, sonra B tedavisinin verilmiş olmasıdır. Yani tedaviler arasında bulunan fark, "A ile B tedavileri arasında fark olduğu" şeklinde yorumlanabileceği gibi, hangi tedavi olursa olsun, "birinci tedavi ile ikinci tedavi arasında fark olduğu" şeklinde de yorumlanabilir. Burada olduğu gibi tedavilerin hangi sırayla verildiğinin önemini ekarte etmek için birinci tedavinin bitimi ile ikinci te-

**Tablo XII.** A Ve B İlaçlarını Alırken Gözlenen Sistolik Kan Basıncındaki Düşme Miktarı (mmHg Olarak).

Hasta No	Tedavi		Fark
	A	B	
1	42	51	9
2	18	24	6
3	21	33	12
4	18	27	9
Ortalama	24.75	33.75	9.00
Standart hata	5.70	6.00	1.20

\* : Bir değer in standart hatası ne kadar küçükse, o değer in o kadar iyi kestirildiği anlaşılmalıdır. Daha iyi akıld a kalabilecek bir örnek vermek gerekirse, bir değ er in gerçek değ eri, % 95 olasılıkla çalışmanın analizi ile hesaplanan değ erden standart hatanın iki katı kadar küçük ya da iki katı kadar büyük değ erler arasında olabileceği söylenebilir. Örneğ imizdeki çalışmada iki grup arasındaki fark 9 mmHg ve farkın standart hatası 1.2 mmHg olduğ una göre, 9 mmHg olarak kestirdiğ imiz farkın gerçek değ erinin % 95 olasılıkla 9 ± 2.4, yani 6.6 ile 11.4 mmHg arasında yer alabileceğ ini, 6.6'dan küçük olma olasılığ ının % 2.5 ve 11.4'ten büyük olma olasılığ ının yine % 2.5 olduğ unu ö ne sürebiliriz.

davinin başlangıcı arasında iki ilacın da verilmemesi bir dönem (arınma dönemi) olması ya da daha da iyisi hastaların yarısına A, diğer yarısına B tedavisini önce verip, arınma döneminden sonra diğer tedaviye geçmek daha doğru olacaktır.

Tedavilerin verilme sırasının önemi olmaması ve denekler arası değişkenliğin bu çalışmadakinin aynısı olması koşulları ile paralel çalışma düzeni ile aynı düzeyde istatistiksel önemlilik saptayabilmek için, 14 kat daha fazla deneye gerek olacaktır. Yani her iki tedavi grubunda 28'lerden toplam 56 hasta gerekecektir. Değişkenliğin azalmaması halinde bile, çapraz çalışma düzeni ile paralel çalışma düzeninin yarısı kadar denek ile aynı istatistiksel sonuca ulaşılır.

Nadir görülen ve uzun seyirli hastalıklarla yapılan çalışmalarda yeterli sayıda hasta bulmak güç olacağı için, çapraz çalışmalar tercih edilmektedir.

Ancak güçlü çalışmalar, potansiyel tehlikeleri de beraberinde barındırır. Paralel - çapraz çalışma düzeni arasında seçim yaparken ve bir çapraz çalışmaya başlamadan önce şu beş özellik dikkate alınmalıdır:

1. Tedavilerin taşınma (carry-over) ve periyod etkisi var mıdır?
2. Tedavi sırası ve hastaların gruplara seçilmesi nasıl belirlenecektir?
3. Çaprazlama kuralları nasıl olacaktır ve ölçümler ne zaman yapılacaktır?
4. Eksik veri, kontrole gelmeyen hasta gibi veri ile ilgili sorunların ne boyutta olduğu tahmin edilmektedir?
5. Örneklem genişliği ve uygulanacak istatistik analiz yöntemi ne olacaktır?

### 1. Taşınma (carry-over) ve Periyod Etkisi

İlk verilen tedavinin etkisi, ikinci tedavinin verildiği dönemde de devam edebilir. Buna ilacın etkisinin taşınması (carry-over), ya da taşınma etkisi (carry-over effect) adı ve-

rilir. Bunu önlemek için ikinci tedaviden önce bir süre ara verilmelidir. Tedaviler arasındaki ilaçsız geçen süreye arınma dönemi (washout period) adı verilir.

Arınma döneminin süresi, her klinik durum ve ilaç için farklı olup, araştırmacı tarafından belirlenmelidir. Bu konudaki kararı verirken ilacın farmakokinetik özellikleri, hedef hücre, doku ya da organdaki etki süresi, ilaç kesildikten sonra hastalık bulgularının ortaya çıkma süresi, hastalarda ilacın metabolizmasını etkileme olasılığı bulunan özellikler vb. dikkate alınmalıdır. Örneğin 300 mg aspirinin yarı ömrü yaklaşık 2.5 saat olduğu halde, plateletler üzerine olan antiagregan etkileri irreversibl olduğu için aspirin sonrası en az iki hafta ara verilmelidir. Öte yandan uzun etkili bile olsa, bir antihipertansif ilacın son dozundan 36-48 saat sonra kan basıncı yükselmeye başlayacağı için 4-5 günlük bir arınma dönemi yeterli olacaktır.

Bazı hastalıklarda, hastalığın şiddeti zamanla artabilir, azalabilir ya da dalgalanmalar gösterebilir. Bu tip seyir gösteren hastalıklarda tedaviler birer kez değil, bir çok kez verilirse, bu etki kısmen arttırılabilir. Örneğin Behçet hastalığının seyri önceden kestirilemeyecek dalgalanmalar gösterebilir. Kolşisin ile steroidin karşılaştırıldığı bir çapraz çalışmada steroidin verildiği dönemlerde hastalığın şiddetinde kendiliğinden azalmalar olması durumunda bunun nedeninin ilacın etkisi olduğu sonucuna varılabilir. Kolşisin ve steroid birer dönem yerine, örneğin üçer dönem verilirse, hata payı kısmen azalacaktır.

Çapraz çalışmalarla ilgili bir diğer sorun da tedavilerden biri ya da birkaçının hastalığın doğasında kalıcı ya da çok uzun süreli değişiklikler meydana getirmesidir.

İdrar yolu infeksiyonunda sulbaktam-

ampisilin ile trimetoprim-sulfametoksazolün çapraz düzenle karşılaştırılmak isteniyor. Hastaların yarısına önce sulbaktam-ampisilin ve arınma döneminden sonra trimetoprim-sulfametoksazol, diğer yarısına ise önce trimetoprim-sulfametoksazol, sonra sulbaktam-ampisilin verilmesi planlanıyor.

Önce sulbaktam-ampisilin, sonra trimetoprim-sulfametoksazol verilecek bir hastada büyük bir olasılıkla tam iyileşme sağlanacağı için (yani hastalığın doğasında kalıcı değişiklik sağlanacağı için) trimetoprim-sulfametoksazol verilemeyecektir. Aynı durum önce trimetoprim-sulfametoksazol, sonra sulbaktam-ampisilin verilecek hastalarda da gözleneceği için, çalışma planı aksayacaktır.

Bu nedenle çapraz çalışma düzeni akut hastalıklarda kesinlikle uygulanamaz. Çapraz çalışma düzenine en uygun hastalık grubu, diyabet, hipertansiyon gibi kronik seyirli ve hangi tedavi olursa olsun, kesildiğinde makul bir süre içinde hastalık bulguları (kan şekerinde, kan basıncında yükselme gibi) ortaya çıkan hastalıklardır.

Taşınma etkisi, hastalığın seyrinin zamanla dalgalanmalar göstermesi, ilaçların hastalığın doğasında kalıcı ya da uzun etkili değişiklikler meydana getirmesi gibi durumlara genel olarak sıra etkisi (order effect) adı verilir. Yeterli uzunlukta arınma dönemi, gerekirse bir kaç dönemli çalışma planlanması, çalışma gruplarındaki hastalıkların kronik seyirli olması gibi konulara özen gösterilerek planlanan bir çalışmanın sonuçları üzerindeki sıra etkisi uygun istatistik modellerle, tedavi etkisinden ayrılabilir. Eğer sıra etkisi, tedavi etkisine yakın ya da daha önemli ise, her tedavinin ayrı ayrı etkilerinden söz edilemez.

Örneğin A tedavisi ile B tedavisinin karşılaştırıldığı bir çapraz çalışmada sıra et-

kisi, tedavi etkisinden daha önemli ise, A'nın B'den, ya da B'nin A'dan daha iyi olduğu öne sürülemez. Bu durumda A tedavisinin ve B tedavisinin etkileri yerine, A-B sırasıyla verilen kombine tedavi ile B-A sırasıyla verilen kombine tedavinin etkisi karşılaştırılabilir. Yani çapraz çalışma bu iki tedavi protokolünün karşılaştırıldığı bir paralel çalışmaya dönüşür. Bu durumda çalışmanın gücü en fazla paralel bir çalışmanın gücü kadar olur, hatta daha az güçlü bir çalışma olması olasılığı da vardır.

Kronik refrakter immün trombositopenik purpura (ITP) steroid ve splenektomi çapraz çalışma ile karşılaştırılıyor. Hastaların yarısına (I. grup) 3 ay süreyle steroid veriliyor ve daha sonra steroid kesildikten 3 ay sonra splenektomi uygulanıyor. Diğer yarısı ise (II. grup) doğrudan splenektomiye veriliyor ve splenektomiden 3 ay sonra steroid başlanıyor. Çalışma düzeni Tablo XIII'te, sonuçlar Tablo XIV'te gösterilmiştir.

Tablo XIV'te görüldüğü gibi steroidden önce ya da sonra olması farketmeksizin, trombosit sayısı splenektomi ile ortalama 60.000/mm<sup>3</sup> artmıştır. Öte yandan ortalama trombosit sayısı artışı, steroid splenektomiden önce verildiğinde 20.000/mm<sup>3</sup> iken, splenektomiden sonra verildiğinde 100.000/mm<sup>3</sup> olarak gözlenmiştir. Tedavilerin uygulanma sırası dikkate alınmadığında her iki tedavi ile de ortalama trombosit sayısı artışının 60.000/mm<sup>3</sup> olduğu görülmektedir. Hangi te-

**Tablo XIII.** Kronik ITP Çalışmasındaki Çalışma Düzeni.

	I. dönem	II. dönem
I. grup	Steroid	Splenektomi
II. grup	Splenektomi	Steroid

**Tablo XIV.** Kronik ITP Çalışmasında Trombosit Sayısındaki Ortalama Artma Miktarları (/mm<sup>3</sup> Olarak).

	<b>I. dönem</b>	<b>II. dönem</b>
I. grup	20.000	60.000
II. grup	60.000	100.000
	<b>Steroid</b>	<b>Splenektomi</b>
I. grup	20.000	60.000
II. grup	100.000	60.000
Ortalama	60.000	60.000

davi yöntemi daha iyidir, ya da aralarında fark var mıdır? Bu soruların basit bir yanıtı yok. Çünkü sıra etkisi, tedavi etkisinden daha önemli gibidir ya da en azından tedavi etkisi kadardır. Bu nedenle tek tek tedavilerden çok, "önce steroid, sonra splenektomi" ile "önce splenektomi, sonra steroid" şeklindeki tedavi protokollerinde "önce splenektomi, sonra steroid" tedavisinin daha etkili olduğu, ayrıca splenektomi sonrasında steroide yanıtta artma olduğu gibi sonuçlar çıkarılabilir. Her durumda da artık çapraz çalışma düzeni geçerliliğini yitirmiştir.

## 2. Tedavi Sırası ve Hastaların Gruplara Seçimi

Çapraz çalışmaya alınan hastaların her biri önce tedavi gruplarından birine alındıktan sonra, eğer ikiden çok tedavi verilecekse, ikinci ve daha sonraki tedavileri verilme sırası belirlenmelidir. Hastaların hepsinin tedavileri aynı sırada alması sonuçların karşılaştırılabilmesini geçersiz kılabilir.

Hastaların tedavi gruplarına seçimi şu dört yöntemle yapılabilir:

a. Tedavileri tüm hastalara aynı sırada vermek.

b. Tedavileri "random" sıra ile vermek.

c. Dengeli tedavi seçimi. Örneğin birinci

hastaya AB, ikinci hastaya BA, üçüncü hastaya yine AB, .... gibi belirli bir sıra ile vermek.

d. Kontrolsüz, tamamen rastgele vermek.

Bu yöntemlerden en geçerlisi, duruma göre "bloklama" da ekleyerek, tedavileri "random" vermektir.

## 3. Çaprazlama Kuralları ve Ölçüm Zamanları

Genellikle, bir tedaviden diğerine geçmek (çaprazlama) için iki kuraldan birine uyulur:

a. Belirli bir süre sonunda (zaman-bağımlı = *time dependent*)

b. Hastanın klinik özelliklerine göre (hastalığa-bağımlı = *disease-state dependent*)

Bu kurallar çalışmanın genel bilimsel kalitesi ve sıra etkisinin miktarı ve yorumlanması açısından farklıdır. Zaman-bağımlı çaprazlama, bilimsel olarak kabul görmüş bir yöntemdir. Hastalığa-bağımlı çaprazlama sakıncalı olabilir. Örneğin semptomların başlaması ile tedaviye başlanması ya da semptomlar ortadan kalkınca tedavinin kesilmesi ya da değiştirilmesi, tedavi etkisinin değerlendirilmesinin imkansızlaştırabilir.

Çapraz çalışmalarda mümkünse, hem hastalar, hem de araştırmacı "kör" olmalıdır. Aksi takdirde verilecek tedavinin bilinmesi tedavinin etkisini ve tedaviye yanıtın değerlendirilmesini etkileyerek, sıra etkisini artırabilir.

## 4. Eksik Veri, Kontrole Gelmeyen Hasta Gibi Verilerle İlgili Sorunlar

Her çalışmada verilerle ilgili sorunlar önemli olsa da, çapraz çalışmada her hastaya ait veriler, çalışmanın toplam verilerine oranı diğer çalışma düzenlerine göre daha fazla olduğu için, çapraz çalışmalar, başlangıçtaki plandan sapmalardan çok etkilenir. Konunun başındaki örneğe dönersek, dört hastadan birinin A tedavisi sırasında kan basıncındaki düşme miktarının ölçülememiş olması, B tedavisi sırasındaki değeri de kullanılamaz hale

getirir, çünkü iki değer arasındaki fark hesaplanamaz. Bu nedenle bu hasta olduğu gibi çalışmadan çıkarılmak zorunda kalınır. Bu durumda hesaplanan p değeri önemli ölçüde değişir. Oysa bu çalışmaya eşdeğer olan 56 hastalı bir paralel çalışmada bir hastanın çalışmadan çıkması ile p değeri çok az değişir.

Çapraz çalışmalarda hastaların izlem süreleri, paralel çalışmalara göre daha uzun olduğu için, çalışma dışı kalma oranı yüksek olabilir. Çalışma dışı kalmanın bir nedeni de verilen tedavilerden biri ya da daha fazlasının önemli yan etkisinin olmasıdır. Her tedavi sırasında ya da sonrasında çalışma dışı kalma oranı farklı değilse, verilerin tamam olduğu tedaviler için bu hastalar da analize dahil edilebilir. Öte yandan belirli tedaviler sırasında ya da sonrasında çalışma dışı kalma, diğer tedavilerden fazla ise, sonuçların yorumlanması oldukça güçleşir.

Bu nedenle çapraz çalışmalarda olası eksik veriler nedeniyle başlangıçtaki denek sayısı yeterince fazla tutulmalıdır. İstatistik analize eksik verilerin de tanımlanması dahil edilmelidir.

### 5. İstatistik Analiz ve Örneklem

#### Genişliği

Aynı hastada farklı zamanlarda ölçülen değerler arasındaki değişkenlik, farklı hastalarda ölçülen değerlerin değişkenliğinden daha az olduğu için, bunu dikkate alan çapraz çalışma analizleri, daha komplike olmakla birlikte, paralel çalışma analizlerinden daha güçlüdür. En önemlisi, çapraz çalışma analizlerinin birim verisi, paralel çalışmalarda olduğu gibi bir hastaya ait tek bir ölçüm değil, hastanın kendisidir.

İki kez ölçüm yapılan çapraz çalışmalarda, bağımlı gruplar için Student t testi kullanılabilirse de, genel olarak çok değişkenli regresyon ve varyans analizlerinin kullanılması gerekir. Çok değişkenli analiz yöntem-

leri, çok sayıda t testinin sonuçlarını birleştirmek prensibine dayalıdır.

### Kendi Kendine Kontrollü Çalışmalar

Kendi kendine kontrollü çalışmalarda tedavi verilen deneklerin tedavi öncesi ve sonrası ölçümleri karşılaştırılarak, tedavinin etkileri analiz edilir. Kendi kendine kontrollü çalışmalarda tedavi öncesi ölçümler kontrol ölçümleri olarak ele alınır. Eğer tedavinin verilmediği dönem kesin kriterlerle belirlenmişse tedavi dönemi ve kontrol dönemi olarak iki dönemi olan bir çapraz çalışma olarak adlandırmak daha doğru olur. Çapraz çalışmalar için söz edilen beş sorun, kendi kendine kontrollü çalışmalarda da vardır. Ancak başka sorunları da dikkate almak gerekir.

### Asimetri

Çapraz çalışmalarda hastalar hem tedavi döneminde, hem de kontrol döneminde eşit koşullarda izlenir, yani hasta izlemi simetriktir. Kendi kendine kontrollü çalışmalarda ise izlem süresi, ölçülen parametreler, tedaviye devam etme ya da kesme kararı gibi açılardan iki dönem farklıdır, yani hasta izlemi asimetriktir. Bu ise kimi zaman tedavi lehine, kimi zaman aleyhine sonuçlanabilecek önyargılara (bias) yol açabilir.

### Çalışılan Sorunun Doğası

Genellikle kendi kendine kontrollü çalışmalar, yeni tedavi yöntemleri ya da ilaçların çok sayıda parametre üzerine etkilerinin ilk kez araştırıldığı çalışmalardır. Öte yandan çapraz çalışma düzeninde ise ölçülen ya da izlenen parametre sayısı bir-iki tanedir. Çok sayıda parametre izlemenin sakıncaları, özellikle istatistik analizde çoklu karşılaştırma ve çok değişkenli analiz teknikleri ile çözülebilir.



### Refrakter Hastalar

Çapraz çalışmaların aksine, kendi kendine kontrollü çalışmaların önemli bir kısmında refrakter hastalığı olan denekler çalışılır. Bunun nedenlerinden en başta geleni, kendi kendine kontrollü çalışmaların, yeni ve dolayısıyla etkinliği kesin olmayan tedavileri kullanılması ile ilgili etik sorunlardır.

Refrakter hastalarla çalışmanın neden olduğu çeşitli sorunlar vardır. Bunlardan başta geleni hastanın çalışmanın konusu olan tedavi verilmeden önce almakta olduğu çok sayıda ilacın etkilerinin artırılması için geçmesi gereken arınma dönemine karşın, hastaların buna çoğu kez katlanamamalarıdır. Bu nedenle yetersiz uzunlukta arınma dönemini takiben, hatta hiç arınma dönemi olmadan çalışılacak tedaviye başlanmak zorunda kalınabilir.

Bir diğer sorun ise, "ortalamaya doğru eğilim", (regression toward the mean) dir. Ortalamaya doğru eğilim, hiç değilse bazı deneklerde başlangıçta ölçülen aşırı değerlerin tekrar ölçülmesi ile ortalamaya, yani normale daha yakın olarak bulunmasıdır. Örneğin 100 kişinin kan basınçlarının ölçüldüğünü ve ortalama sistolik kan basıncının 140 mmHg olduğunu düşünelim. Bir hafta sonra aynı 100 kişinin kan basınçları tekrar ölçüldüğünde yine ortalama 140 mmHg civarında bulunacaktır. Ancak kan basıncı 140 mmHg'nın biraz üzerinde olan bir kaç kişinin kan basıncının daha düşük olduğu ve 140 mmHg'nın biraz altında olan bir kaç kişinin kan basıncının ise daha yüksek olduğu görülecektir. Çünkü başlangıçtaki değerlerin aşırılığında hastalığın şiddetinin yanında "random" varyasyonun da payı olabilir. Ortalamaya doğru eğilim nedeniyle, bir kez ölçüm sonucunda çalışmaya alınma kriterlerini tutturduğu düşünülen deneklerin bir kısmına hiç bir tedavi verilme de,

ölçülen değer ortalamaya yaklaşacaktır. Bu nedenle hastaların çalışmaya alınıp alınmayacaklarına karar vermeden önce, hiç değilse ikinci bir kez daha ölçüm yapılmalıdır.

Refrakter hastalarla ilgili bir sorun da bu hastalardaki klinik durumun, belki de hastalığın refrakter bir varyantına sahip olmaları ya da genel durumlarının hiç bir tedaviye yanıt vermeyecek kadar bozulmuş olmalarıdır. Her iki durumda da belki de etkili olabilecek yeni bir tedavi yönteminin etkinliği adaletli bir şekilde test edilmemiş olur. Yine de etik açıdan almakta olduğu tedavi iyi yanıt vermiş ve hastalığı kontrolde olan denekler yerine refrakter hastaların kullanılması daha doğrudur. Hiç değilse, yanıt veren ve vermeyen hastalar arasındaki farklara dikkat çekilmesi ile ileride planlanacak olan randomize kontrollü çalışmalara ışık tutulmuş olur.

### Diğer Tedavilerle Doğrudan Karşılaştırma Yapılamaması

Kendi kendine kontrollü çalışmalarda tek tedavi yöntemi denendiği için diğer tedavi yöntemleriyle karşılaştırma ancak diğer çalışmaların sonuçları ile yapılabilir. Farklı çalışma gruplarındaki deneklerin demografik özellikleri, hastalıklarının şiddeti vb. özellikler açısından farklı olmaları nedeniyle bu karşılaştırmaların ne kadar güvenilir olduğu oldukça tartışmalıdır. Bu nedenle tedavi etkinliğinin en objektif olarak değerlendirileceği randomize çalışma düzenlerinden biri seçilmelidir.

Geliş tarihi : 20.07.1998

Yayına kabul tarihi : 24.07.1998

Yazışma adresi:

Dr. Oktay ÖZDEMİR

Omega Araştırma Organizasyon Eğitim Danışmanlık Ltd. Şti.  
ANKARA

