

# TIPTA BAĞIMSIZ KONTROLLÜ (PARALEL) İLAÇ ARAŞTIRMALARI NASIL YAPILIR?

İki ya da daha fazla hasta grubunun bir çalışmanın parçası olarak, fakat farklı gruplarda yer aldığı çalışmalar bağımsız kontrollü çalışmalar (paralel çalışmalar) 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 çalışma olarak adlandırılır.

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 B tedavisinin verildiğini düşünün. Bu durumda bu çalışma çapraz (cross-over) çalışma olarak adlandırılır.

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. Bu yazıda bağımsız kontrollü çalışma düzeninin ayrıntıları anlatılacaktır.

Paralel çalışmalarda farklı kişilere farklı ilaçlar verilir, 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 tutmaya yada yanlılığa (*bias*) neden 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 etmek doğru değildir. Örneğin iki gruba 25'erden 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 önemi.

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 kriterleri açısından değerlendirilir. Kriterlere 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 arařtırıcının inisiyatifinde olmasının önemli sakıncaları vardır. Örneğin yeni ve etkinliği pek de bilinmeyen bir antidepresif ilacın plaseboyla karşılaştırıldığı bir çalışmada arařtırmacılar bilinçli ya da bilinçsiz olarak, depresyon bulgularının ağır olduđu bir hastalarda ilacı, hafif olduđu 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 rastlantısal 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 olduđu zaman basit randomizasyonla bile gruplar arasında dengesizlik olabilir. Denek sayısı arttıkça, gruplar arasında dengesizlik olma olasılığı azalır.

Denek sayısının az olduđu bir çalışmada basit randomizasyon tek başına

dengesizliğin azaltılmasında yeterli olmayabilir. Bu durumda randomizasyona řu iki yöntemden biri ya da her ikisi birlikte eklenmelidir:

1. Tabakalandırma.
2. Bloklama.

Arařtırmacı, 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, diğeri diğeri 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 kendiliğinden dengeli olur. Yazı-tura yerine istatistik kaynak kitaplar ya da bilgisayarlardan elde edilebilecek rastgele sayılar tablosundan yararlanarak da randomizasyon yapılabilir.

Tablo 1'de cinsiyetin tabaka değıřkeni olduđu ve blokların ikişer kişiden olduđu 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 1'deki randomizasyon şemasına göre ilk erkeğe ilaç, ikinci ve üçüncü erkeğe plasebo, dördüncü erkeğe ilaç verilecektir.

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 kullanılmasının 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, operasyon 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 1'deki örnekte "cinsiyet" in iki adet kategoriden oluşacağına kuşku yoktur. Ancak yaş, vertebra sayısı gibi sayısal değişkenler için tabakaların kategorilerinin araştırmacı tarafından belirlenmesi gerekecektir.

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

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 solunum yolu

**Tablo 1.** Cinsiyet açısından tabakalandırılmış ve her blokta iki hasta olmak üzere bloklama yapılmış bir çalışmadaki randomizasyon şeması.

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 2).

Çalışma süresince bu 18 kombinasyondan bazılarının görülmemesi 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

**Tablo 2.**  
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
		Doğumsal	2	2
		Edinsel	3	3
20-49	1-5	Doğumsal	1	1
		Edinsel	2	2
	6-10	Doğumsal	3	2
		Edinsel	1	1
		Doğumsal	2	2
		Edinsel	4	4
>49	1-5	Doğumsal	-	-
		Edinsel	2	2
	6-10	Doğumsal	-	-
		Edinsel	1	1
		Doğumsal	-	-
		Edinsel	3	3
TOPLAM			32	32

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 (Tabii yeri gelmişken, araştırma planlanması aşamasında çok ciddi bir literatür taraması yapılması gerektiğini vurgulamakta yarar var).

Bu durumda prognostik etkenlerin dengesizliğinin çözümünde kullanılacak bir yöntem, analizi tabakalandırmaktır. Tabakalı analiz, çalışma grupları içindeki prognostik 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 infarktüsü (MI) gelişmesini önlemek üzerindeki etkilerinin karşılaştırılması amacıyla düzenlenen 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 3'ten Tablo 8'e kadar yer alan tablolarda bu çalışmadaki verilere ilişkin çapraz tablolar görülmektedir. Beta-bloker grubunda 21, kalsiyum antagonisti grubunda 30 hastanın MI geçirdiği gözleniyor ve ki-kare analizi ile beta-bloker grubundaki MI oranının kalsiyum antagonisti grubundan istatistiksel olarak önemli derecede az olduğu görülüyor ( $p=0.03$ ) (Tablo 3).

Miyokard infarktüsü			
	Geçiren	Geçirmeyen	Toplam
Beta-bloker	21	19	40
Kalsiyum antagonisti	30	10	40
Toplam	51	29	80

**Tablo 3.**  
İki tedavi grubunda MI geçiren hasta sayılarının gösterildiği çapraz tablo.

	Erkek	Kadın	Toplam
Beta-bloker	10	30	40
Kalsiyum antagonisti	30	10	40
Toplam	40	40	80

**Tablo 4.**  
İki tedavi grubundaki deneklerin cinsiyete göre dağılımının gösterildiği çapraz tablo.

**Tablo 5.**  
*Beta-bloker grubunda farklı cinslerde MI geiren denek sayılarının gösterildiđi apraz tablo.*

	Miyokard infarktüsü		
	Geiren	Geirmeyen	Toplam
Erkek	8	2	10
Kadın	13	17	30
Toplam	21	19	40

Ancak gruplardaki deneklerin cinsiyetlere göre dađılımlarının farklı olduđu görülüyor. Beta-bloker grubunda 10 erkek, 30 kadın, kalsiyum antagonisti grubunda 30 erkek, 10 kadın var (Tablo 4). Bu nedenle farklı cinslerde ilaçların etkilerinin farklı olabileceđi, ya da verilen ilaç ne olursa olsun farklı cinslerde hastalığın seyrinin farklı olabileceđi gibi sorular akla geliyor. Gerekten de iki farklı tedavi grubundaki denekler cinsiyet ve prognoz arasındaki iliřki aısından incelendiđi zaman her iki ilaç grubunda da erkeklerde MI oranının daha yüksek olduđu saptanıyor (Tablo 5 ve 6).

Acaba beta-bloker grubunda infarktüs oranının düşük olmasının nedeni ilacın daha etkili olması mıdır, yoksa bu grupta hastalığın daha kötü seyrettiđi anlařılan

cinsten, yani erkeklerden daha az denek bulunması mıdır? Acaba her iki tedavi grubundaki deneklerin cinsiyet aısından dađılımı benzer olsaydı, sonuç yine beta-bloker lehine mi bulunacaktı? Bu sorunun özümü için tabakalı analiz yöntemi kullanılabilir. İnfarktüs oranlarının, tedaviye yanıt ya da hastalığın seyri aısından farklı olan altgruplarda, yani erkeklerde ve kadınlarda tedavi gruplarındaki ayrı ayrı incelenmesi grupların cinsiyet aısından dengesiz olmasının neden olduđu yorumlama güçlüđünü ortadan kaldıracaktır. Gerekten de hem erkeklerde (Tablo 7), hem de kadınlarda (Tablo 8) hangi ilaç kullanılırsa kullanılsın, prognozun farklı olmadığı görülmektedir.

**Tablo 6.**  
*Kalsiyum antagonisti grubunda farklı cinslerde MI geiren denek sayılarının gösterildiđi apraz tablo.*

	Miyokard infarktüsü		
	Geiren	Geirmeyen	Toplam
Erkek	27	3	30
Kadın	3	7	10
Toplam	30	10	40

Miyokard infarktüsü			
	Geçiren	Geçirmeyen	Toplam
Beta-bloker	8	2	10
Kalsiyum antagonistisi	27	3	30
Toplam	35	5	40

**Tablo 7.**  
İki tedavi grubunda MI geçiren erkek denek sayılarının gösterildiği çapraz tablo.

Bu örnekte olduğu gibi çalışma bittikten sonra altgrupların belirlenmesine “sonradan-tabakalandırma” (*poststratification*) 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 araştırmada, iki grupta tedaviye yanıt oranı arasında fark olduğu varsayımı araştırmada 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 daha doğru bir anlatımla, bu araştırma aynı sayıda hasta ile ve aynı koşullarda 100 kez yapılsa, iki ilaç arasında 12 araştırmada fark bulunamayacak, 88’inde fark bulunacaktır.

Öte yandan gerçekte fark olduğu halde, yapılan araştırmadan 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 araştırmanın analizini 6 tane çapraz tablo ile yapabildik. MI geçirmeyi önleme etkisi açısından beta-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 çalışmanın gücü önemli derecede azaldı. Eğer araştırmanın başlangıcında cinsiyetin prognoz üzerine etkisi olabileceği öngörülse ve buna göre hastalar tabakalandırılarak

Miyokard infarktüsü			
	Geçiren	Geçirmeyen	Toplam
Beta-bloker	13	17	30
Kalsiyum antagonistisi	3	7	10
Toplam	16	24	40

**Tablo 8.**  
İki tedavi grubunda MI geçiren kadın denek sayılarının gösterildiği çapraz tablo.



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

**Tablo 9.**  
Gruplardaki hastaların cinsiyete göre dağılımı.

randomize edilseydi, 80 hastayla yapılmış olan bu araştırmadaki 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, fark olmadığı halde fark varmış gibi sonuç elde edilmesidir.

Örneğin Tablo 3'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 (Bunun teorik açıklaması bu yazının amaçları dışında olduğu için ayrıntıya girilmeyecektir). Bu durumda Tablo 3'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 9). Bu nedenle kan basıncındaki düşme miktarı olarak tanımlanan ilaca yanıt kriteri erkek ve kadınlar için ayrı ayrı analiz edilmiştir (Tablo 10). 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 edilebilir.

	A ilacı	B ilacı	Plasebo
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

**Tablo 10.**  
Kan basıncında ortalama düşme miktarı (mmHg).



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 olacaktı. 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 ortalama farklı değildir. B ilacı verilen grupta gözlenen ortalama kan basıncı düşmesi 34.0 mmHg'dır. Benzer şekilde hesaplandığında 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 artılmıştır, çünkü her üçü de % 50'si erkek, % 50'si kadın olan hipotetik popülasyonlarda gözlenmesi beklenen değerlerdir.

## **Kontrol Değişkenlerinin Etkilerinin Artırıldığı İstatistik Modeller**

Gruplar arasında dengesizlik olduğu zaman kullanılabilecek bir yöntemin tabakalı analiz olduğundan ve bu yöntemin sakıncalarından yukarıda söz edildi. Dikkat ederseniz, tabakalı analizde prognostik etkene göre her altgrupun 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 artırıldığı 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 pektoris 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 enfasyonu riskini göze almıştık. Bu çalışmada alternatif olarak log-lineer analiz yapıldığında iki tedavi grubundaki hastaların cinsiyet dağılımının benzer olmadığı ( $p < 0.0001$ ), prognozun iki cinsten 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 artırılmış olmasıdır.

Prognostik etkenlerin etkilerinin artırılmasında hangi durumlarda hangi istatistik modellerin uygulanabileceği Tablo 11'de özetlenmiştir. Sonradan-tabakalandırma ile grup karşılaştırmaları sayısındaki artmaya bağlı "istatistiksel önemlilik enfasyonu", bu istatistik modellerle önlenir. Bu yöntemlerle, hem tedavinin etkisi prognostik etkenlerden artırılmış, hem de prognostik etkenlerin önemi test edilmiş olur.

Gerek tabakalı analiz, gerekse kontrol değişkenlerinin etkilerinin artırıldığı 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 artırıldığı modeller kullanıldığında ise, (eğer doğru model kullanılırsa) hem taraf tutma azalır, hem de sonuçların doğruluk derecesi artar.

Böylece istatistiksel önemliliği sınırda olan yargaların 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 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 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.

Oktay Özdemir, Doç. Dr.  
Omega CRO

İzlenen yanıt	Kontrol değişkeni	Tabakalı analiz	Alternatif model
Sayısal	Sayısal	-	Kovaryans analizi
Sayısal	Kategorik	Student t testi, Tek-yönlü varyans analizi	İki-yönlü varyans analizi
İkili kategorik*	Sayısal	-	Lojistik regresyon analizi
İkili kategorik	Kategorik	Ki-kare analizi	Lojistik regresyon analizi Log-lineer analiz
İkiden daha çok düzeyli kategorik**	Sayısal	-	Diskriminant analiz
İkiden daha çok düzeyli kategorik	Kategorik	Ki-kare analizi	Log-lineer analiz

**Tablo 11.** Kontrol değişkenlerinin etkilerinin artıldığı istatistik modeller.

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

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