

# İki veya Daha Fazla Gruplu Denemelerde Gruplar Arasındaki Başlangıç Değerleri Farklılığının Etkisinin Düzeltilmesi

## Adjusting the Effect of Baseline Differences Between Groups in Trials with Which Have Two or More Groups

Dr. Handan ANKARALI,<sup>a</sup>  
Dr. Seyit ANKARALI<sup>b</sup>

<sup>a</sup>Biyostatistik AD,  
<sup>b</sup>Fizyoloji AD,  
Zonguldak Karaelmas Üniversitesi  
Tıp Fakültesi, Zonguldak

Geliş Tarihi/Received: 05.05.2008  
Kabul Tarihi/Accepted: 08.09.2008

Yazışma Adresi/Correspondence:  
Dr. Handan ANKARALI  
Zonguldak Karaelmas Üniversitesi  
Tıp Fakültesi, Biyoistatistik AD,  
Zonguldak,  
TÜRKİYE/TURKEY  
hankarali@yahoo.com

**ÖZET Amaç:** Birçok klinik ve deneysel çalışmalarda, araştırmacılar sürekli bir değişkenin başlangıç değerlerini ölçüp, muamele sonrası yeniden ölçüm olarak muamele etkisini değerlendirirler. Başlangıç değerleri açısından gruplar arasında anlamlı fark olduğunda, muamele etkisinin tahmininde uygulanan bazı istatistik yöntemler hatalı sonuçlara neden olabilir. Çalışmanın amacı, başlangıç değerleri açısından anlamlı fark olduğunda muamele etkisinin tahmininde hangi istatistik yöntemlerin daha uygun olduğunu açıklamaktır. **Gereç ve Yöntemler:** Muamele etkisinin tahmininde kullanılan farklı istatistiksel yöntemler kısaca açıklanarak, iki grup arasında ilgilenilen değişkenin başlangıç değerleri açısından anlamlı farklılıkların bulunduğu bir hipotetik veri setinde uygulanmıştır. Ayrıca birkaç sınırlı koşul için uygun istatistik yöntemin belirlenmesi amacıyla kısıtlı bir simülasyon denemesi yapılmıştır. **Bulgular:** Hipotetik veri setinde gruplar arasında başlangıç değerleri açısından farklılığın yanı sıra başlangıç ve muamele sonrası değerler arasındaki korelasyonun da düşük olduğu görülmüştür. Bu koşulda muamele etkisini tahminde, başlangıç ve muamele sonrası değerler arasındaki farkların karşılaştırılması en iyi yöntem olarak bulunmuştur. Simülasyon denemesi sonucunda korelasyonun düşük ve örnek genişliğinin küçük olduğu durumlar için muamele etkisini tahminde farklara ait testin gücü (%85) kovaryans analizine ait güçten (%40) daha yüksek bulunmuştur. Bu iki yöntemin gücü orta büyüklükteki örnek genişliğinde birbirine yakın ve yüksek olarak bulunmuştur. Korelasyon yüksek olduğunda ise hem küçük hem de orta büyüklükteki örnek genişliklerinde her iki yöntemin gücü oldukça yüksek düzeyde idi. **Sonuç:** İki ya da daha fazla gruplu denemelerde gruplara deneklerin dağıtımı rastgele yapılmış olsa dahi incelenen özelliğin başlangıç değerleri açısından gruplar arasında anlamlı farklılığın olup olmadığı mutlaka kontrol edilmelidir. Ayrıca başlangıç değerleri ile muamele sonrası değerler arasındaki korelasyonun derecesi de dikkate alınmalıdır. Fark var ve korelasyon düşük ise farkların analizi en uygun sonucu verecektir. Korelasyon yüksek ise muamele sonrası değerlere kovaryans analizi uygulanması daha doğru sonuçlar verecektir.

**Anahtar Kelimeler:** Epidemiyolojik araştırma denemeleri, araştırma düzenleri, rastgele atama, randomize kontrollü denemeler

**ABSTRACT Objective:** In many clinical and experimental trials, researchers assess the effect of treatment by measuring the value of a continuous variable before and after the treatment. If there is an imbalance in baseline values between groups, some statistical comparisons may result with mistakes in estimation of the treatment effect. The aim of this study was to explain which statistical methods were more suitable in the estimation of the treatment effect when there was an imbalance for the baseline values between groups. **Material and Methods:** Different statistical methods, which are used in estimation of treatment effects, were briefly explained and were applied to a hypothetical data set, which had significant differences between groups according to baseline values of the related variable. In addition, a limited simulation study for several conditions was carried out to determine suitable statistical methods. **Results:** Baseline values were different between two groups and correlation was low between baseline and follow up values of related variable in each group for hypothetical data set. In this condition, comparison of simple differences between baseline and follow up values was the best method for the estimation of treatment effect. In the simulation study, the power of the test for simple differences was higher (85%) than the value in the analysis of covariance (40%) when correlations were low and sample sizes were small in each group. Moreover, the powers of these two tests were high and similar to each other, when sample sizes were moderate. When the correlation was high, the powers of both tests were high in both small and moderate sample sizes. **Conclusion:** The presence of a significant difference should be sought between groups according to baseline values of the related variable even though groups are randomly assigned. In addition, the degree of the correlation between baseline and follow up values should be taken into consideration. When significant differences exist between baseline values and the correlation is low, we suggest that the classical methods should be used to determine the significance of the effect; however, when the correlation is high, covariance analysis is a suitable method.

**Key Words:** Epidemiologic research design, research design, random allocation, randomized controlled trial

**K**linik ve deneysel arařtırmalarda etkisi denenilen ila veya tedavi muamele olarak adlandırılır. Bu tip arařtırmalarda genellikle bir kontrol ve bir ya da daha fazla muamele grubu vardır ve muamele gruplarının kontrol grubu ile farkı muamele etkisini gösterir. Bazı klinik ve deneysel alıřmalarda ise etkisi daha nceden gsterilmiş standart muamele grubu, plasebo veya sham grubu veya bazı ilalar iin sadece zc madde-lerin verildiđi gruplar da oluřturmak gerekebilir. Btn bu kořullarda dzenlenen arařtırmaların amacı tedavinin/ilacın etkisini arařtırmak veya birbirleriyle karřılařtırmaktır.

Gruplarda farklı denekler yer alacağı iin muamele etkisine denekler arası farklılıkların etkisi de karışır. Bu nedenle muamele etkisini daha iyi tahmin etmek amacıyla muamele uygulamadan nce her bir gruptan zerinde durulan zelliđe ait lmler alınır ki bu lmlere bařlangı lmleri adı verilir. Bu durumda yukarıda tanımlanan ayrı bir kontrol grubu yerine bařlangı lmleri kontrol deđerleri olarak kullanılır. Ancak denemede hem bařlangı deđerlerinin hem de ayrı bir kontrol grubunun bulunması gerekli olabilir. nkn bařlangı deđerleri ile muamele sonrası elde edilen deđerler arasındaki farklar sadece muamele etkisinden kaynaklanmayabilir. Yani kontrol grubunda zamanla kendiliđinden meydana gelen bir deđiřim varsa, saf muamele etkisi muamele grubundaki nce-sonra deđiřimi ile kontrol grubundaki nce-sonra deđiřimi arasındaki farktır.<sup>1</sup>

Muamele etkisini tahmin etmek iin muamele grubunun plasebo ya da bilinen standart bir tedavi grubu ile karřılařtırıldıđı denemeler kontroll denemeler olarak adlandırılır. Kontroll denemelerin en stn sonu vereni ise rastgele kontroll denemelerdir. Bu deneme dzeninde kontrol ve muamele gruplarına denekler rastgele atanırlar. rastgele atamaya rađmen gruplardaki deneklerin her zaman benzer zellikte olması sađlanamaz veya bu duruma etik kurallar izin vermez veya bazı zellikler aısından benzerlik sađlanırken diđer bazı zelliklerde farklılıklar gzlenebilir. Ayrıca denek sayısının az olması veya alıřılan zellikte varyasyonun byk olması gruplara rastgele atamayı etkileyebilir yani grupların bařlan-

gı deđerlerinde farklılık oluřmasına neden olabilir.<sup>2,3</sup>

Daha nce yapılan birok arařtırmada, rastgele kontroll deneme dzenlerinde zerinde durulan zelliđin bařlangı deđerleri aısından grupların benzer olduđu varsayımı kabul edilmiř ve bu gibi durumlarda muamele etkisinin tahmininde farklı yntemler karřılařtırmalı olarak incelenmiřtir.<sup>2,4</sup> Ancak yukarıda belirtildiđi gibi her zellik aısından rastgele dađıtım her zaman sađlanamamıř olabilir. Bu gibi durumlarda grupların bařlangı deđerleri anlamlı farklılık gsterebilir. Acaba bu durumda muamele etkisini hatasız tahmin edebilmek iin ne yapmak gerekir? Bir bařka ifadeyle biyolojik dengelemenin yapılamadıđı veya homojen dađıtımın sađlanamadıđı kořullarda, muamele etkisi iine karışan bařlangı farklılıklarını nasıl gidermek gerekir? Bu soruları cevaplayan eřitli rastgeleleřtirme ve/veya istatistik yntemler geliřtirilmiř, bunlardan hangisinin bu etkileri gidermede daha bařarılı olduđu kısmen arařtırılmıřtır. Ancak hangi durumlarda hangi istatistik yntemlerle muamele etkisinin dođru bir řekilde tahmin edilebileceđi konusunda net sonulara ulařılamamıřtır.<sup>3,5,6</sup>

Sunulan alıřmanın amacı bařlangı lmleri aısından farklılık gsteren normal dađılımlı iki grubun muamele etkilerini karřılařtırmak iin planlanan denemelerde, muamele etkilerinin hangi istatistik yntemlerle daha iyi tahmin edilebileceđini aıklamaktır. Aıklamaların kolay anlaşılması iin yntemler hipotetik bir veri seti zerinde uygulanarak verdiđi bilgiler tartıřılacaktır. Ayrıca uygun yntemin seiminde en ok sıkıntıya dřlecek bir kořul iin kısıtlı bir simlasyon alıřması yapılarak bu kořula uygun yntem nerilecektir. alıřmada bu amala kullanılan istatistik yntemlerin uygunluđu eřitli kořullar iin karřılařtırmalı olarak incelenerek konu ile ilgili pratikte yapılan hatalar aıklanacaktır.

## GERE VE YNTEMLER

### rnek Deney Tasarımı

alıřmanın kolay anlaşılması aısından en basit bir deney tasarımı ele alınmıř ve hipotetik uygulama

TABLO 1: Deney planı.

Gruplar	Denekler	Muamele öncesi ölçüm	Muamele sonrası ölçüm
1	1	$Y_{11}$	$Y_{21}$
	2	$Y_{12}$	$Y_{22}$
	.	.	.
	$n_1$	$Y_{1n1}$	$Y_{2n1}$
2	1	$Y_{11}$	$Y_{21}$
	2	$Y_{12}$	$Y_{22}$
	.	.	.
	$n_2$	$Y_{1n2}$	$Y_{2n2}$

verisi de bu tasarıma göre oluşturulmuştur. Bu deney tasarımında iki bağımsız grup ve iki bağımlı ölçüm mevcuttur ve deney planı Tablo 1'deki gibidir.

Bu deney tasarımında 1 numaralı grup kontrol, sham, plasebo ya da diğer bir muamele grubu iken, 2 numaralı grup muamele grubudur. Muamele uygulamadan önce bütün deneklerden başlangıç ölçümleri alınır ve daha sonra muamele uygulanarak aynı özelliğe ait muamele sonrası ölçümler elde edilir. Tanımı yapılan bu denemelerde muamele etkisinin anlamlı olup olmadığını araştırmak amacıyla kullanılan birçok istatistik yöntem vardır ve bunlardan muamele etkisini en iyi tahmin eden yöntem seçilmelidir.

#### Kullanılan İstatistiksel Yöntemler

Yukarıda tanımlanan deney düzeninde, başlangıç ölçümleri açısından iki grup arasında anlamlı farklılık varsa bu iki grubu, sadece muamele sonrası ölçümleri dikkate alarak karşılaştırmak yanıltıcı sonuçlar verir. Bu gibi durumlarda muamele etkisini tahmin etmek için başlangıç ölçüm farklarını göz önünde bulunduran çeşitli istatistiksel yöntemler kullanılmalıdır. Ancak hangi yöntemi kullanmanın daha uygun olacağı konusunda sıklıkla hatalar yapılmaktadır. Şimdi bu amaç doğrultusunda kullanılan istatistiksel testlere sırasıyla göz atalım.

Birinci yaklaşımda her bir grupta başlangıç değerleriyle muamele sonrası değerler arasında bağımlı iki örneklem ortalaması arasındaki farka ait t-testi yardımıyla ayrı ayrı her bir grup için muamele etkisi incelenir. Daha sonra gruplara ait muamele etkileri birbiriyle karşılaştırılır. Ancak bu

yaklaşımda bütün karşılaştırmalar birbirinden bağımsızmış gibi yapılmaktadır. Bu durum tahminlerdeki hata payını artırır.

İkinci yaklaşımda, iki bağımsız grup ve bu gruplara ait başlangıç ve muamele sonrası ölçüm değerleri tek bir model içinde incelenir. Bu amaçla en sık kullanılan yöntemler tekrarlanan ölçümlü varyans analizi, kovaryans analizi veya önce-sonra farklarına ait t-testidir.

Başlangıç değerlerinin gruplar arasında anlamlı farklılık gösterdiği durumlarda tekrarlanan ölçümlü varyans analizi modeli muamele etkisini tahmin için tek başına yeterli bilgi vermeyebilir ve bu analizin sonuçları hatalı yorumlamalara neden olabilir.

Bir diğer yaklaşımda her iki grup için başlangıç ve muamele sonrası ölçümler arasındaki farklar alınır ve bu farklar bakımından muamele ve kontrol grubu karşılaştırılır. Bu karşılaştırmada farklar muamele etkisinin büyüklüğünü ve başlangıç ölçümlerine göre düzeltilmiş değerini gösterir. Bu amaçla iki bağımsız örneklem ortalaması arasındaki farka ait t-testi kullanılır ve genellikle doğru sonuçlar verir. Ancak farklar ile başlangıç ölçümleri arasında yüksek bir korelasyon söz konusu ise farkları analiz etmek avantajlı değildir. Mesela başlangıç değeri büyük olan bireyin farkı küçük veya başlangıç değeri küçük olan bireyin farkı büyük ise muamelenin gerçek etkisi olduğundan az tahmin edilir. Bunun tersi olan pozitif korelasyon durumunda ise muamele etkisi gerçek etkisinden fazlaymış gibi tahmin edilir.<sup>7</sup> Bazen basit farklar yerine yüzde değişim değerleri de kullanılmaktadır ancak bu ölçü, çarpık bir dağılım gösterdiği için t-testi uygulanması sakıncalı olan bir ölçüdür. Ayrıca muamele sonrası birden fazla ölçüm alınıyorsa ve her bir ölçüm için farklar hesaplanıp birbirinden bağımsızmış gibi analiz ediliyorsa tahminlere ait hata ihtimali artar. Böyle bir durumda da tekrarlanan ölçümlü kovaryans analizinin kullanılması gerekir.

Bu çalışmada tanımlanan amaç için literatürde en sık önerilen bir diğer test kovaryans analizi (ANCOVA)'dir. Bu analizde kovaryat olarak başlangıç değerleri alınır ve muamele sonrası alınan

ölçümler başlangıç değerlerine göre düzeltilir. Böylece muamele etkisi açısından iki grup karşılaştırılacağı zaman başlangıç değerlerindeki farklılıklara göre düzeltilmiş ölçümler kullanılır. Ancak bu düzeltmenin yapılabilmesi için başlangıç ve muamele sonrası ölçümler arasında mutlaka korelasyon olmalıdır.

Daha yeni ve daha detaylı bir diğer yaklaşım ise eğilim skorlarına (propensity scores) dayalı bir yöntemdir. Bu yöntemde başlangıç değerlerine göre gruplara dengeli dağıtımın yapılmasını sağlamak için bireylere ait eğilim skorları hesaplanır ve dağıtım bu skorlara göre yapılır. Eğilim skorları birbirine yakın olan iki denegin biri bir gruba, diğeri öbür gruba atanır ve daha sonra muamele uygulanarak elde edilen ölçümler bakımından iki grup t-testi ile karşılaştırılır. Böylece gruplara homojen dağıtım istatistiksel yolla sağlanmış olur. Anlaşılacağı üzere bu yöntem önce-sonra denemelerinin başlangıç ölçümleri alındıktan sonra uygulanabilirken önce ve sonra ölçümlerinin her ikisi de alınmış yani tamamlanmış veri setlerine geriye dönük uygulanması mümkün değildir.

#### Hipotetik Veri Seti

Bu veri seti, A maddesinin sıçanlarda saldırganlık davranışına etkisinin araştırıldığı bir deneye ait olsun. Saldırganlık davranışının çalışıldığı deneysel hayvan modellerinde en çok kullanılan ölçülerden birisi ilk saldırıya kadar geçen zaman olarak tarif edilen ilk saldırı latensidir. Bu değişkene ait değerler normal dağılım gösteren bir popülasyondan alınmıştır ve dolayısıyla bu örnekteki veriler normal dağılım göstermektedir.

Şimdi A maddesinin serum fizyolojik içerisinde uygulandığını varsayalım. Saldırganlıkta zamanla kendiliğinden meydana gelebilecek bir değişiklik, alışma etkisi gibi nedenlerle A maddesinin uygulandığı grubun yanı sıra sadece serum fizyolojik verilen ayrı bir grubun da oluşturulması uygun olacaktır. Yaş, ağırlık ve cinsiyet açısından benzer özelliklere sahip 12 sıçan 6'şarlı olarak rastgele bu iki gruba atanır. Deneklerin başlangıç saldırı latensi değerleri ölçülerek 1. gruptakilere serum fizyolojik, 2. gruptakilere A maddesi verilerek ölçümler alınır.

Deneysel davranış çalışmalarında bireyler arasında farklılıklar fazla olabilir ve bu deneyler oldukça dikkat isteyen deney tasarımları gerektirir. Bireysel farklılıkların fazla olduğu bu gibi durumlarda rastgele dağıtıma rağmen grupların başlangıç saldırganlık düzeyleri farklılık gösterebilir.

#### Simülasyon Çalışması

“Başlangıç değerleri anlamlı düzeyde farklı olan iki grubun etkisi karşılaştırılırken muamele etkisinin en iyi tahmini hangi yöntemle yapılmalıdır?” sorusuna cevap verebilmek için kısıtlı bir simülasyon denemesi planlanmış ve bu denemede pratikte en çok hata yapılan koşullar ele alınmıştır.

Simülasyon çalışmasında, ortalamaları arasında 1 standart sapmalık (SD) fark bulunan ve standart normal dağılım gösteren iki popülasyondan iadeli olarak örnekler alınmıştır. Bu popülasyonların birincisinin ortalaması “0” standart sapması “1” iken, ikinci popülasyonun ortalaması “1” ve standart sapması da “1” dir. Alınan örneklerdeki değerler iki grubun başlangıç değerleri olarak kabul edilmiştir. Bu durumda gruplar arasında başlangıç değerleri açısından anlamlı farkın olduğu bir koşul oluşturulmuştur. Daha sonra yine normal dağılım gösteren ve yukarıda tanımlanan birinci popülasyondan iadeli iki örnek daha alınmıştır ve bu örnek değerleri sırasıyla birinci ve ikinci grubun muamele sonrası ölçümleri olarak kabul edilmiştir. Uygulanacak testlerin farklı örnek genişliklerindeki gücünü belirlemek için küçük ( $n=10$ ) ve orta ( $n=50$ ) büyüklükte örnekler olmak üzere iki farklı örnek büyüklüğü ile çalışılmıştır. Ayrıca başlangıç ölçümleri ile muamele sonrası ölçümler arasındaki korelasyonun etkisini görmek için düşük ( $r=0.15$ ) ve yüksek ( $r=0.90$ ) olmak üzere iki farklı korelasyon düzeyi oluşturulmuştur. Tanımlanan bu koşullarda muamele etkisinin tahmininde, farklar kullanılarak yapılan t-testinin mi yoksa ANCOVA modelinin mi daha güçlü olduğu incelenmiştir. Bu nedenle gerçekte aralarında anlamlı fark olan iki popülasyondan alınan örneklere t-testi ve ANCOVA ayrı ayrı uygulandığında anlamlı düzeyde fark bulma yüzdeleri testlerin gücü olarak hesaplanmıştır. Simülasyon çalışmasında her bir koşul 1000 defa tekrarlanmıştır. Koşulları tanımlanan bu simü-

lasyon algoritması bir MINITAB (sür. 14.0) makrosunda yazılarak sonuçlar elde edilmiştir.

## BULGULAR

### Hipotetik Veri Setine Ait Analiz Sonuçları

Gruplarda serum fizyolojik ve A maddesi verilmeden önceki ve sonraki ilk saldırı latenslerine ait tanımlayıcı istatistikler Tablo 2'de verilmiştir.

Şimdi çeşitli açılardan veri setimizi inceleyelim. Serum fizyolojik grubunda başlangıç değerlerle muamele sonrası değerler arasında anlamlı bir değişikliğin olmadığı görülmüştür ( $p= 0.832$ ). A maddesi verilen grupta da başlangıç değerleri ile muamele sonrası değerler karşılaştırıldığında anlamlı bir farklılık bulunmamıştır ( $p= 0.644$ ). Serum fizyolojik verilen grupta böyle bir farkın olmaması klinik açıdan saldırganlıkta zamanla veya tekrar karşılaşma nedeniyle alışma sonucu saldırganlıkta değişim olmadığını gösterebilir.

İki grubun başlangıç ölçümlerinin dikkate alınmadığını kabul ederek sadece muamele sonrası ölçümler bakımından iki grubu karşılaştırdığımızda gruplar arasında anlamlı farkın olmadığı görülmüştür ( $p= 0.139$ ). Bu sonuca bakarak klinik yorum yapmaya kalkarsak A maddesinin saldırganlık davranışını etkilemediğini söyleriz ancak bu yaklaşım hatalı sonuç doğurabilir. Çünkü başlangıç değerlerindeki farklılık göz ardı edilmiştir. Gerçekte iki grubun başlangıç ölçümleri karşılaştırıldığında anlamlı fark olduğu belirlenmiştir ( $p= 0.047$ ). Bu durumda gruplara atamada bu özellik açısından rastgeleliğin sağlanmadığını ve bu sonucun denek sayısının azlığından veya çalışılan özellikteki varyasyonun büyüklüğünden kaynaklanabileceğini söyleyebiliriz. Görüldüğü gibi başlangıç değerlerini karşılaştırmaksızın sadece muamele sonrası ölçümler bakımından yapılan karşılaştırmalar güvenilir sonuç vermeyecektir.

Başlangıç değerleri ile muamele sonrası değerler arasındaki farklar açısından iki grup karşılaştırılmış ve muamele etkisi benzer bulunmuştur ( $p= 0.604$ ). Yani serum fizyolojisinin saldırganlıkta meydana getirdiği değişiklikle A maddesinin meydana getirdiği değişiklik arasında anlamlı fark yoktur. Farklarla başlangıç ölçümleri arasında her iki grupta da yüksek korelasyon hesaplanmış ( $r= 0.80$ ) ancak denek sayısının az olması nedeniyle tam sınırla anlamlı bulunmuştur. Bu sonuç farklarla başlangıç değerlerinin ilişkili olduğunu ve tek başına farklar kullanılarak muamele etkisini tahmin etmenin yanıltıcı olabileceğini gösterir.

Tekrarlanan ölçümlü varyans analizi uygulandığında ise başlangıç ve muamele sonrası ölçümler arasında anlamlı fark bulunmamış ( $p= 0.700$ ) ve bu farkın gruptan gruba değişmediği belirlenmiştir ( $p= 0.604$ ). Yine aynı analiz sonucunda gruplar arasında anlamlı fark tespit edilmiş ( $p= 0.013$ ) ve bu farkın da başlangıç ve muamele sonrası ölçümlerde aynı kaldığı görülmüştür ( $p= 0.604$ ). Bu sonuçlara göre hem serum fizyolojik ve hem de A maddesinin saldırganlık davranışını değiştirmediklerini söyleyebiliriz. Ancak serum fizyolojik verdiğimiz grup ile A maddesi verdiğimiz grup arasında hem başlangıç hem de muamele sonrası saldırganlık açısından anlamlı fark çıkmıştır ve bu muamele etkisini yansıtmaktan ziyade iki grubun saldırganlık özelliği açısından farklı olduğunu göstermektedir. Bu yöntemle muamele etkisini değerlendirirken sonuçların çok dikkatli yorumlanması gerekmektedir.

Uygun istatistiksel yönteminin seçiminde dikkat edilmesi gereken bir diğer husus başlangıç değerleri ile muamele sonrası değerler arasındaki korelasyon derecesidir. Tablo 2'nin son sütunu incelendiğinde her iki grupta da başlangıç ölçümleri ile muamele sonrası ölçümler arasında anlamlı bir korelasyon bulunmamıştır.

**TABLO 2:** Tanımlayıcı istatistikler.

Muamele Grupları	n	Başlangıç Değerleri	Muamele Sonrası	Farklar	p	Korelasyon (Başlangıç- muamele sonrası)
		Ortalama $\pm$ SD (saniye)	Ortalama $\pm$ SD (saniye)	Ortalama $\pm$ SD (saniye)		
Serum fizyolojik	6	262.5 $\pm$ 76.4	280.8 $\pm$ 176.6	-18.3 $\pm$ 201.0	0.832	-0.124 ( $p= 0.814$ )
A maddesi	6	670.3 $\pm$ 435.2	547.5 $\pm$ 365.8	122.8 $\pm$ 613.2	0.644	-0.166 ( $p= 0.753$ )

ANCOVA modeli kullanıldığında ise başlangıç ölçümleri ile muamele sonrası ölçümler arasında zayıf korelasyon olması nedeniyle yeterli ölçüde düzeltme yapılamamış ve başlangıç ölçümlerini dikkate almadan sadece muamele sonrası iki grubun karşılaştırılmasıyla elde edilen sonuca çok yakın bir sonuç elde edilmiştir ( $p= 0.160$ ).

Ayrıca araştırılan özelliklere göre rastgele dağıtımın yapılamadığı durumlarda kullanılan eğilim skorları bu veri seti için de hesaplanmış ve bu skorlara göre gruplar yeniden oluşturulmuştur. Veri setindeki başlangıç değerlerine göre hesaplanan eğilim skorları yardımıyla gruplara dağıtım yaptığımız zaman gruplar arasında başlangıç değerleri açısından anlamlı farkın ortadan kalktığı görülmüştür ( $p= 0.879$ ). Bu aşamadan sonra yeni oluşan gruplar için muamele etkisi yeniden denenmelidir. Bu ise denemenin yeniden tasarlanması ve muamelelerin yeniden uygulanması demektir. Bu çalışmadaki veri seti tamamlanmış olduğu varsayılan bir deneye ait veri seti olduğu için eğilim skorlarına göre yapılan yeniden gruplama sonrasında muamele etkilerinin ne olacağının bilinmesi mümkün değildir. Bu nedenle eğilim skorları kullanılarak başlangıç değerlerinin gruplara dengeli dağıtılabilirdiği gösterilmiştir.

### Simülasyon Sonuçları

Yapılan simülasyon denemesinde başlangıç ölçümleri bakımından gruplar arasında sadece 1 standart sapmalılık fark (etki büyüklüğü= 1 SD) var iken güç değerleri incelenmiştir. Örnek genişliği küçük (her bir grupta  $n= 10$ ) ve başlangıç ölçümleri ile muamele sonrası ölçümler arasında düşük korelasyon ( $r= 0.15$ ) varken, farklar için t-testinin gücü %85 ve ANCOVA modeline ait testin gücü %40 civarında bulunmuştur. Bu sonuca göre örnek genişliği küçük ise ve başlangıç ölçümleriyle muamele sonrası ölçümler arasında düşük korelasyon var ise muamele etkisini belirlemede farkların, ANCOVA modeline göre daha başarılı olduğunu söyleyebiliriz. Ancak daha önce bahsedildiği gibi başlangıç ölçümleri ile muamele sonrası ölçümler arasındaki korelasyon düşük olduğu zaman başlangıç ölçümleri ile farklar arasındaki korelasyon yüksek çıkma eğilimindedir. Bundan dolayı muamele etkisi olarak açıklanan so-

nucun başlangıç ölçümleri ile farkların ilişkisi mi yoksa gerçek muamele etkisi mi olduğu göz önünde bulundurulmalıdır. Böyle durumlarda başlangıç değerleri ile farklar arasındaki korelasyonun anlamlı olup olmadığı incelenmeli ve eğer anlamlı ise başlangıç değerleri kovaryat olarak alınıp farklar için kovaryans analizi yapılmalıdır.

Örnek genişliği orta büyüklükte seçildiği zaman (her bir grupta  $n= 50$ ) farklara ait t-testinin gücü %100, ANCOVA modelinin gücü ise %90 olarak bulunmuştur. Örnek genişliği orta büyüklükte iken muamele etkisini doğru tahmin etmede ANCOVA modeli ve farklara ait t-testinin benzer sonuçlar vereceği söylenebilir.

Ayrıca başlangıç ve muamele sonrası ölçümler arasında yüksek korelasyon durumu da incelenmiş ( $r= 0.90$ ) ve hem farklara ait t-testinin hem de ANCOVA modelinin %100'e yakın güç değerine sahip olduğu görülmüştür. Bu koşullarda muamele etkisini tahmin etmek amacıyla her iki yöntemde kullanılabilceğini söyleyebiliriz.

## TARTIŞMA

rastgele dağıtımın her zaman sağlanmadığı/sağlanmadığı durumlarda başlangıç değerleri açısından gruplar arasında farklılık görülebilir. Çünkü üzerinde durulan konuyla ilgili veya konuyu etkileyen çok faktörlü denemelerde rastgele dağıtım bazı faktörlere göre sağlanırken diğer bazı faktörlere göre sağlanamayabilir. Çeşitli nedenlerle üzerinde çalışılan denek sayısı az ve özelliğin varyasyonu geniş ise rastgele dağıtım daha da zor uygulanabilir veya daha az verimli olabilir. Bu nedenle hiçbir zaman tek başına muamele sonrası ölçümler kullanılarak muamele etkisi tahmin edilmemelidir. İşte bu gibi durumlarda başlangıç ölçümlerindeki farklılıklarının yanı sıra başlangıç ölçümleri ile muamele sonrası ölçümler arasındaki korelasyonun derecesini de dikkate alarak uygun test yöntemi ile muamele etkisi tahmin edilmelidir.<sup>1,2</sup>

Şu ana kadar yapılmış çalışmaların bir bölümünde deneklerin diğer başlangıç karakteristikleri (yaş, cinsiyet, ağırlık gibi) açısından gruplara rastgele dağıtımın yapılamadığı koşullarda iki grubun muamele etkilerini karşılaştırmada alternatif

yöntemler incelenmiştir. Diğer bir bölümünde ise üzerinde durulan özelliğin başlangıç ölçümlerinin her iki grupta da benzer olduğu durumlar ele alınmıştır. <sup>8-11</sup> Bu çalışmalarda başlangıç ölçümleri ile muamele sonrası ölçümler arasındaki korelasyon çok düşük veya anlamlı değil ise sadece muamele sonrası ölçümler kullanılarak veya muamele sonrası ölçümlerle başlangıç ölçümleri arasındaki farklar test edilerek muamele etkisinin doğru bir şekilde tahmin edilebileceği gösterilmiştir. Korelasyon derecesi arttıkça ANCOVA modeli daha uygun sonuçlar vermektedir, ancak ANCOVA modelinin her muamele grubunda korelasyonun benzer büyüklükte ve yönde olması koşulunda kullanılabilirliği akıldan çıkarılmamalıdır.<sup>3</sup>

Başlangıç ölçümleri bakımından gruplar arasında anlamlı farkın olduğu durumlarda başlangıç ve muamele sonrası ölçümler arasındaki korelasyon derecesi çok düşük veya anlamlı değil ise ANCOVA modeli yeterli düzeltme yapamadığı için muamele etkisini özellikle küçük örneklerde genellikle hatalı tahmin edecektir. Bu durumda farkların analizi en iyi alternatif yöntem olacaktır.<sup>1,3,4</sup> Ancak başlangıç değeri küçük olan bireylerin daha az, başlangıç değeri büyük olan bireylerin daha fazla farka sahip olması bekleniyor ise farklar ile başlangıç değerleri korelasyonlu olacağı için muamele etkisi yine bir miktar yanlı tahmin edilecektir.

Korelasyon katsayısı orta büyüklükte ise muamele etkisini tahmin etmek için ANCOVA iyi bir yaklaşım olacak ve korelasyon derecesi arttıkça farkların analizi ile muamele etkisi daha iyi tahmin edilecektir. Farklar yerine bazı araştırmalarda yüzde değişim veya orantı gibi ölçüler de kullanılmaktadır ancak bu ve benzeri ölçülerin dağılımı normal dağılım olmadığı için bu ölçülere burada açıklanan testlerin uygulanmaması önerilmektedir.<sup>4</sup> Bu gibi durumlarda “parametrik olmayan testler tercih edilmelidir” şeklinde görüşler olsa da istatistiksel analizde önerilen bu testlerin uygunluğu da tartışmalıdır.

Ayrıca gruplara rastgele dağıtım yapılmadığı durumlarda, deneklere ait eğilim skorları hesaplanarak bu skorlara göre denekler rastgele dağıtılarak gruplar yeniden oluşturulur. Daha sonra muamele etkisi bu yeni gruplandırma sonucunda incelenir. Bu durumda iki gruptaki deneklerin başlangıç değerleri arasında anlamlı fark olmayacağı için yukarıda açıklandığı gibi sadece başlangıç değerleri ile muamele sonrası ölçümler arasındaki korelasyon derecesine göre uygun yöntem kullanılarak muamele etkisi yansız tahmin edilmiş olur. Bu yöntemin avantajı, sadece başlangıç ölçümlerine göre değil aynı zamanda deneklere ait diğer bazı karakteristikler açısından da gruplara dengeli dağıtım sağlayabilmesidir. Ancak birçok özelliğin açısından dengeyi sağlayabilmek için buna paralel olarak denek sayısını da artırmak gerekir ki bu durum yöntemin en önemli dezavantajıdır. Ayrıca anlaşılacağı gibi bu yöntemin tamamlanmış deneyler için bir düzeltme yöntemi gibi uygulanabilirliği yoktur.<sup>12</sup>

## SONUÇ

Sonuç olarak, muamele etkilerinin karşılaştırılmak istendiği araştırmalarda başlangıç ölçümleri kendi kontrolü olarak seçildiyse veri analizinde birçok yöntem önerilmekte veya kullanılmaktadır. Bu yöntemlerin hepsi değişkenlerin sürekli yapıda olmasını ve hatta normal dağılım gösteren değişkenler olmasını gerekli kılmaktadır. Amacın belirlenmesinde, deney tasarımının planlanmasında ve veri analizinde uygun yöntem seçiminde değişik koşullar göz önüne alınmalı ve yorumlamalarda da dikkatli olunmalıdır. Özet olarak, burada tanımlanan deney tasarımı için, başlangıç ve muamele sonrası ölçümler arasındaki korelasyonun yüksek olması durumunda kovaryans analizi, düşük olması durumunda ise farkların analizi daha uygun sonuçlar verecektir.

## KAYNAKLAR

1. Senn S. Change from baseline and analysis of covariance revisited. *Stat Med* 2006;25(24): 4334-44.
2. Roberts C, Torgerson DJ. Understanding controlled trials: baseline imbalance in randomised controlled trials. *BMJ* 1999;319 (7203):185.
3. Boshuizen HC. Re: Twisk and Proper: evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up. *J Clin Epidemiol* 2005;58(2): 209-10.
4. Kaiser L. Adjusting for baseline: change or percentage change? *Stat Med* 1989;8(10): 1183-90.
5. Vickers AJ. The use of percentage change from baseline as an outcome in a controlled trial is statistically inefficient: a simulation study. *BMC Med Res Methodol* 2001;1:6.
6. Twisk J, Proper K. Evaluation of the results of a randomized controlled trial: how to define changes between baseline and follow-up. *J Clin Epidemiol* 2004;57(3):223-8.
7. Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. *Stat Med* 1992;11(13):1685-704.
8. Samuels ML. Use of analysis of covariance in clinical trials: a clarification. *Control Clin Trials* 1986;7(4):325-9.
9. Liang KY, Zeger SL. Longitudinal data analysis of continuous and discrete responses for pre-post designs. *Sankhya-Indian Statist Inst Series B (Biostatistics)* 2000;62(1):134-48.
10. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ* 2006;332(7547):969-71.
11. Wei L, Zhang J. Analysis of data with imbalance in the baseline outcome variable for randomized clinical trials. *Drug Inf J* 2001;35(4): 1201-14.
12. Becker SO, Ichino A. Estimation of average treatment effects based on propensity scores. *The Stata Journal* 2002;2(4):358-77.