

Antrenman Bilimi Araştırmalarında En Zayıf Halka: Rastgeleleştirme Yöntemi

The Weakest Link in the Training Science Researches: Randomization Method: Review

Ekim PEKÜNLÜ^a

^aAntrenörlük Eğitimi Bölümü,
Ege Üniversitesi Beden Eğitimi ve
Spor Yüksekokulu, İzmir

Geliş Tarihi/Received: 12.01.2012
Kabul Tarihi/Accepted: 17.02.2012

Yazışma Adresi/Correspondence:
Ekim PEKÜNLÜ
Ege Üniversitesi Beden Eğitimi ve
Spor Yüksekokulu,
Antrenörlük Eğitimi Bölümü, İzmir,
TÜRKİYE/TURKEY
ekimpekunlu@gmail.com

ÖZET Rastgeleleştirme yöntemi (RY), bilimsel araştırmalarda bir altın standart olarak kabul edilmektedir. Ancak RY'nin altın standart olarak tanımlanabilmesi için ön koşul olarak RY'nin kullanıldığı araştırmaların büyük bir örnekleme sahip olması gerekmektedir. Antrenman bilimi araştırmalarındaki örneklem genellikle küçük olmasına rağmen, RY'nin bu araştırmalarda hiçbir sorgulama yapılmadan kullanıldığı gözlemlenmektedir. Dolayısıyla RY'nin bu koşullar altında kullanılması araştırma sonuçlarının geçerliliği üzerinde şüphe oluşturmaktadır. Küçük örnekleme sahip bağımsız grup tasarımı araştırmalarda, araştırma sonuçlarının gerçeği yansıttığını garanti altına alabilmek için araştırma yapısına en uygun grup oluşturma yönteminin seçilmesi gerekir. Araştırma sonuçları üzerinde etkisi olacağı düşünülen temel değişkenlere göre eşleme yönteminin kullanılması bir seçenek olarak göz önünde bulundurulmalıdır. Bu yöntem kullanılarak araştırmadaki yanlışlıklar büyük oranda önlenebilir. Küçük örneklemlerle araştırma sonuçlarının genelleştirilmesi oldukça zordur. Dolayısıyla bu araştırmalardan elde edilen sonuçların "doğru" olup olmadığını belirlemenin tek geçerli yolu, araştırmaların aynı yöntem ve aynı değişkenler kullanılarak farklı örneklemde birbir tekrar edilmesi ve elde edilen sonuçların karşılaştırmalı olarak incelenmesidir.

Anahtar Kelimeler: Rastgeleleştirme; eşleme; örneklem büyüklüğü; yanlışlık

ABSTRACT Randomization method (RM) is regarded as a gold standart in scientific researches. However, in order to define RM as a gold standart, the researches in which RM is used should have a large sample size as a prerequisite. Although the sample sizes in the training science researches are generally small, it is observed that RM is used in these researches without any questioning. Therefore, the use of RM under these circumstances raises doubts about the validity of the research results. In the researches with small sample size and independent group design, the grouping method that fits the research design best should be selected to ensure that the research results are "true". The use of matching method based on important characteristics of participants that might affect the outcomes of the research should be considered as an alternative. Research bias could be prevented to a large extent by the use of this method. Generalization of the results of researches with small sample sizes are quite difficult. Therefore, the only way to determine whether or not the results of a research using a small sample size are true is to repeat that research just as it is with different samples by using the same method and same variables, and to analyze the obtained results of these repeated researches in a comparative manner.

Key Words: Randomization; matching; sample size; bias

Türkiye Klinikleri J Sports Sci 2012;4(2):92-104

Bilim, doğa olayları arasındaki ilişkileri inceleyerek bilgi edinmeye ya da bu olaylar arasında var olan ilişkileri kanıtlamaya yönelik gerçekleştirilen çalışmalar bütünüdür. Bu çalışmalar kendine özgü sistematik bir mantığa sahiptir.¹⁻³ Bilim insanları, olayları nedenselliğe yönelik

güvenilir ve geçerli kanıtlar doğrultusunda tekrarlanabilir deney ve gözlemlerle sadece bilimsel yöntemi kullanarak değerlendirirler.²⁻⁵ Bilimde mutlak kesinlik yoktur. Bilim, elimizde bulunan kanıtlar doğrultusunda ve mevcut zaman içinde neyin en geçerli olduğunu belirler. Olgular için “Kesinlikle doğrudur” şeklinde bir tanımlama yapılamaz. Olgular sadece desteklenebilir, güçlendirilebilir veya onaylanabilir.³ Dolayısıyla bilimsel bir bakış açısıyla değerlendirildiğinde, bir araştırmadan elde edilen sonuçların gerçeği yansıtıp yansıtmadığının bilinmesi çok büyük önem taşır.⁶

Gerçek bir bilim insanı, herhangi bir araştırmadan elde edilen sonuçları değerlendirirken şüpheli bir yaklaşım sergilememelidir. Bu sonuçların geçerliliği çok önemlidir, çünkü bu geçerlilik test edilen hipotezlerin doğruluğu ve yanlışlığı konusunda bir temel oluşturmaktadır.⁷ Bilim insanı, araştırmacının herhangi bir aşamasında herhangi bir nedenle gözden kaçırılmış bir değişkenin, araştırmacının sonucuna etkide bulunup bulunmadığını değerlendirmelidir.⁸ Araştırma yöntemleriyle ilgili varsayımları sorgulamak ve bu varsayımları deneysel olarak test etmek, bilim insanının temel önceliklerinden biri olmalıdır.⁹

Bilimsel araştırmalarda katılımcıların uygulama ve kontrol gruplarına ayrılma aşamasında genellikle “rastgeleleştirme yöntemi (RY)” kullanılmaktadır. Ancak bu yöntemin kullanılması konusunda çoğunlukla hiçbir şekilde sorgulamada bulunulmamaktadır.⁶ RY'nin kullanımı, gerek uygulama boyutunda gerekse kuramsal boyutta sınırlılıklar gösterir. Buna ek olarak RY'nin yöntemsel değeri konusunda da çeşitli şüpheler bulunmaktadır.⁹

RY'nin başarılı sonuçlar verebilmesi için araştırmadaki örneklem grubunun yeterince büyük olması gerekir.¹⁰⁻¹³ Bağımsız grupların kullanıldığı antrenman bilimi araştırmalarında genellikle örneklem gruplarının büyük olmamasına rağmen grupların RY ile oluşturulması ve “olasılıksal olarak denk” kabul edilmesinin sağlam bir temele dayandığı söylenemez. Gruplar oluşturulurken gruplar arası denklik (benzerlik); değişkenlerin araştırma sonucu üzerindeki olası etki düzeyi ve değişkenlerdeki dengesizliklerin ölçüsü göz önünde

bulundurularak sağlanmalıdır.¹⁴ Bu nedenle, araştırmalarda gruplar arası dengenin en yüksek düzeyde oluşmasına olanak sağlayacak temel değişken(ler) hassasiyetle belirlenmelidir.

DENEYSEL ARAŞTIRMALAR

Deneysel araştırmalar, belirli bir olayın ortaya çıktığı koşulları belirlemek ve bu olayla, var olan koşullar arasındaki neden-sonuç ilişkisini incelemek amacıyla düzenlenen kontrollü araştırmalardır.^{15,16} Antrenman bilimi araştırmaları; temel araştırmalar ve deneysel araştırmalar arasında yer alan bir konumdadır. Temel araştırmalar, belirli bir olayın mekanizmasıyla bağlantılı olan kuramları kanıtlamak ve bu alanda bilgi edinmekle ilgilenir. Deneysel araştırmalar ise temel araştırmalardan elde edilen bilgiler doğrultusunda gerçek dünya koşullarındaki değişkenler üzerinde etkili olan etkenleri araştırır.¹⁷

Deneysel araştırmalarda araştırmacı, belirli bir anda sadece tek bir değişken üzerinde farklılıklar yaratarak bu değişkenin bağımlı değişkenler üzerindeki etkisini belirlemeye çalışır. Bu şekilde, istenmeyen değişkenleri denk tutarak veya en aza indirgeyerek hassasiyeti artırmayı hedefler.^{15-16,18} İç geçerliliği olumsuz yönde etkileyebilecek tüm etkenleri ortadan kaldırabilen araştırma, gerçek deneysel araştırma olarak tanımlanmaktadır.¹⁹ Dolayısıyla deneysel araştırmaların tasarımında aşağıda belirtilen konulara dikkat edilmelidir:⁵

- Gruplar arasındaki tek fark, gruplar üzerinde gerçekleştirilen uygulama olmalıdır.
- Karşılaştırılabilir gruplar yeterince büyük olmalıdır.
- Uygulama ve araştırma sonuçları arasındaki ilişkinin doğru tanımlanması için uygun olasılıksal yöntemler kullanılmalıdır.

Araştırmaların en önemli hedefi, uygulamaları en yüksek düzeyde bir tamlik, doğruluk, hassaslık ve geçerlilikle karşılaştırmaktır.¹³ Araştırmacının ortaya çıkardığı sonuçların uygulamadan kaynaklandığının söylenebilmesi; sadece araştırma öncesi grupların tamamen birbirine denk olması ve sistematik açıdan farklılık göstermemesi koşulunda geçerlidir.^{13,20} Bunun tersi bir durumda araştırma

sonuçları yanlılık gösterir.¹³ Dolayısıyla iyi bir araştırmanın en önemli yönü, karşılaştırılan grupların “karşılaştırılabilir” özellikte olmasıdır.

Araştırma gruplarında farklı bireyler bulunduğu için bilin(e)meyen veya ölçül(e)meyen değişkenler temelinde bireysel farklılıkların araştırma sonuçları üzerinde etkisi olacağı açıktır.^{10,21-23} Dolayısıyla gruplar arasındaki farkın en düşük düzeyde tutulması, gerçek sonuçlara ulaşmak için bir gerekliliktir. Gruplar arasındaki farkı ortadan kaldırmak için araştırmalarda çeşitli kontrol yöntemleri kullanılmaktadır. Bu yöntemler arasında en sık kullanılanı rastgeleleştirme yöntemidir.

■ RASTGELELEŞTİRME YÖNTEMİ

RY, katılımcıları uygulama veya kontrol gruplarına tamamen rastgele bir şekilde dağıtma yöntemidir.^{13,24-27} Bu yöntemde her bir katılımcının herhangi bir gruba yerleştirilme olasılığı aynıdır.^{13,26} RY'nin temelinde, katılımcıların araştırma gruplarına rastgele olarak dağıtılması ve bu şekilde gruplar arası denge sağlanarak dağıtım sırasındaki olası yanlılığın önlenmesi yatar.^{2,13-14,20-21,23,25,28-30} Bu yöntem kullanılarak, sonuca etki edebilecek bütün değişkenlerin “ortalamada” her gruba benzer şekilde dağılması sağlanmış olur.^{20,25,31-33} Böylece uygulamaların aynı türden katılımcıların oluşturduğu karşılaştırılabilir gruplar üzerinde gerçekleştirildiği söylenebilir. Bundan yola çıkarak, araştırma sonucunda elde edilen farkın, uygulamadan kaynaklandığı da garantilenmiş olur.^{20,29}

RY, çoğu bilimsel araştırma için bir temel oluşturmaktadır. RY, uygulama sonuçlarının değerlendirilmesinde en geçerli, en güvenilir ve optimal bir yöntem olarak kabul edilmektedir.^{10,13,30,34,35} Ayrıca bu yöntem, araştırmalardaki “altın standart” olarak tanımlanmakta ve bu tanımlama RY'nin sahip olduğu aşağıdaki özelliklerden kaynaklanmaktadır:^{10,13-14,19,21-23,27-28,36,37}

- Karşılaştırılabilir gruplar oluşturması.

- Bilin(e)meyen ve ölçül(e)meyen değişkenlerin dengesizliğinden kaynaklanan sistematik hataları büyük oranda önlemesi (Ayrıca RY bilinen ve ölçülebilen değişkenleri de gruplar arasında kontrollü bir şekilde dağıtabilen en temel yöntem olarak kabul edilmektedir).^{9,31,32,38}

- İstatistiksel testlerdeki geçerliliği arttırması.
- Sonuçları ve veri yorumlamasını kuvvetlendirmesi.

- Uygulama sonrası sonuçların değerlendirilmesindeki etkinlik ve etkililiği arttırması.

Büyük örneklem grubuna sahip araştırmalarda katılımcıların uygulama ve kontrol gruplarına rastgele dağıtılması -gruplarda bulunan çok çeşitli katılımcı özelliklerinden dolayı- RY'ye güçlü bir dış geçerlilik sağlar. Bu durum da, uygulama sonuçlarının genelleştirilmesine olanak verir.

■ RASTGELELEŞTİRME YÖNTEMİNİN KULLANIMINDAKİ SINIRLILIKLAR VE BU SINIRLILIKLARIN KAYNAKLARI

Rastgeleleştirme süreci, araştırma grupları arasında karşılaştırılabilirlik sağlama ve denklik oluşturma amacıyla uygulanıyor olsa da bu amaca ulaştığı söylenemez.^{30,31} Var olan kuramsal ve uygulamalı veriler, RY'nin kullanımıyla ilgili çeşitli sınırlılıkların bulunduğunu göstermektedir.⁹ Ayrıca RY'nin yönetsel önemi de çeşitli şüpheler taşımaktadır. RY, araştırmalarda en çok kullanılan yöntem olsa da, araştırmalardan elde edilen sonuçların gerçeği yansıttığını garanti altına alamaz.¹² RY'nin gerçeği yansıttığını söyleyebilmek için çok sayıda katılımcıya ve de sistematik ve sistematik olmayan sapmaların en düşük düzeye indirgenmesine ihtiyaç vardır.^{12,19} RY'de gerçek sonuçtan sistematik olmayan sapmalar, rastgele hatalardan kaynaklanır (örneğin: gruplardaki bireylerin verilerindeki şans etkeni).¹² Ancak, iyi düzenlenmiş bir rastgeleleştirmede bile sistematik dengesizliklerle karşılaşılacağı ve RY'de önemli değişkenlerdeki farklılıkların şans eseri de ortaya çıkabileceği göz ardı edilmemelidir.^{29,39}

Bir araştırmada “kontrol”, “sistematik hataların önlenmesi” ve “rastgele hataların en aza indirgenmesi” geçerli bir değerlendirme için gerekli olan üç temel özelliktir. Araştırmalarda genellikle RY kullanıldığı için sistematik hataların önlenmediği düşünülür. Ancak bir uygulamada RY'nin kullanılmış olması, sistematik hataların büyük ölçüde önlenmiş olduğu anlamına gelmez. Araştırma sonucu üzerinde etkili olacağı düşünülen değişkenlerin gruplar arasında denk dağılmış olması durumunda, sistematik hataların önlenmediği düşünülebilir.¹⁰

RY'nin kullanıldığı ve çok sayıda katılımcı içeren araştırmalarda "Belirlenebilir değişkenler gruplar arasında denk dağılıyorsa bilin(e)meyen ya da ölçül(e)meyen değişkenler de gruplar arasında denk dağılmıştır" "varsayımı", "umut"u ve "beklenti"si bulunur.^{10,11,25} Ayrıca bu tür araştırmalarda, karşılaştırılan gruplar -sadece rastgele farklılık gösterdikleri için- yeterli düzeyde karşılaştırılabilir özelliktedir; "sav"ı, "tahmin"i, "beklenti"si ve "umut"u vardır.^{8,11,22,25,30,40-42} Bu sav, tahmin, beklenti ya da umudun doğru olduğu söylenemez; doğru olsa bile başlangıç değişkenlerindeki rastgele dengesizlikler, sonuçların yorumlanmasındaki ve uygulanabilirliğindeki geçerlilik üzerinde şüpheler oluşturur.^{22,33}

RY'nin diğer yöntemlerden daha güçlü oluşu yukarıda belirtilen varsayıma dayanmaktadır.^{6,10,23} Ancak sonuç üzerinde etkili olduğu bilinen değişkenleri gruplar arasında denk dağıtamayan bir RY'nin, bilin(e)meyen/ölçül(e)meyen değişkenleri gruplar arasında denk dağıtmada başarılı olacağı söylenemez.^{10,42} Ayrıca belirlenebilir değişkenleri dengelemek için kullanılan diğer yöntemlerin, belirlenemeyen değişkenlerle ilgili olarak RY ile karşılaştırıldığında daha kötü bir durum ortaya çıkarmayacağı belirtilmektedir.²⁵ Buna ek olarak RY'nin, sonuca etki edebilecek tüm değişkenleri gruplar arasında denk dağıtıp dağıtmadığı konusunun da tam bir kesinlikle bilinmeyeceği ifade edilmektedir.^{8,32,43} Bütün bunlara rağmen, temel değişkenler araştırma grupları arasında denk dağılmamış olsa bile değişkenlerin denk dağılmış olduğu varsayılır.⁸

RY ile oluşturulan gruplar arasında fark olmadığı görüşü, ortalamada geçerli olan bir ifadedir. Büyük bir örnekleme RY ile gruplar oluşturma işlemi çok sayıda tekrarlandığında, işlemlerin çoğunda gruplar arası farkın olmadığı görülür.⁴⁴ Ancak bu durum, işlemlerin hepsinde gruplar arasında fark olmayacağı anlamına gelmez.^{43,44} Dolayısıyla bu farklılık araştırma sonuçları üzerinde etkide bulunabilir.⁴³

RY'nin kullanıldığı araştırmalarda başlangıç değerlerindeki farklılıkların kesinlikle rastgele olduğu varsayımı, "geniş çaplı bir inanış" olarak tanımlanmaktadır.^{40,45,46} RY'yi kullanan iyi düzenlenmiş bir araştırmanın, geçerli bir karşılaştırma özelliğini sağlayamaması ve başlangıç değerleri açı-

sından önemli sistematik dengesizlikler oluşturması da olasıdır.^{10,14,21,22,31,42,43,45-47} RY'deki başlangıç değerlerindeki farklılıkların oluşumu sadece şansla açıklandığında, uygulama sonrası karşılaştırma sonuçları da sadece şans eseri ortaya çıkmış sonuçlar olarak tanımlanabilir.^{12,45,48} Araştırmalarda bazı farklılıklar şans eseri oluşuyor olsa da, gruplar arasında ortaya çıkan orantısız düzeydeki bazı dengesizlikler de araştırmalarda "rastgele ve önemsiz" olarak tanımlanmaktadır.^{12,45} Araştırmalarda grupların başlangıç değerlerinde önemli bir fark varsa, bu farkın ne büyüklükte olduğu ve sonucu ne ölçüde etkileyeceği göz önünde bulundurulması gereken bir konudur.²³

Yanlılık, gözlemlenen araştırma sonucuna uygulamanın dışında başka bir bilinen ya da bilinmeyen değişkenin etki etmesi veya etki etmiş olabileceği durumu ifade eden istatistiksel bir terimdir.⁴⁹ "Katılımcıların seçilimi", "uygulama", "sonuçların elde edilmesi", "değerlendirme", "araştırmanın bitirilmesi" ve "yayın" aşamalarında yanlılıklarla karşılaşılabilir.^{11,20} RY sırasında ortaya çıkan yanlılık, karşılaştırılan grupların başlangıç özellikleri arasındaki sistematik farklılıklardan oluşabilir.^{20,33} Böyle bir durum rastgeleleştirmede kusurların varlığına işaret eder ve bu durumda rastgeleleştirmenin sağlanmadığı söylenir.^{14,29,33,42} RY'nin katılımcıların seçilmesi ve dağıtımında bütün yanlılığı ortadan kaldırdığı ifadesi bir gerçekten çok bir "hayal" olarak tanımlanmaktadır.⁴⁵ Katılımcıların gruplara dağıtılmasında bir yanlılık varsa araştırma sonuçları, uygulanan yöntemlerin etkileri konusunda abartılı sonuçlar üretebilir ve aslında herhangi bir etkisi olmayan bir uygulamanın çok etkili bir yöntem olarak algılanmasına neden olabilir.^{33,45} Bu yanlılık, araştırmanın sonuçlarının uygulama etkisinden mi, yoksa temel değişkenlerdeki farklılıklardan mı kaynaklandığı sorusunu ortaya çıkarır.^{33,42} "Gözlemlenen uygulama etkisi, katılımcıların dağılımından kaynaklanan yanlılık nedeni ile oluşmuştur." önermesi, yanlışlanana kadar bu etkilerin uygulamadan kaynaklandığı söylenemez.⁴⁵

RY'nin kullanıldığı araştırmalarda -özellikle sonuca etki edecek değişkenler açısından heterojen bir örneklem evreni varsa- rastgele hataların ortaya çıkma probleminin önüne geçmek için çok fazla sayıda katılımcıya ihtiyaç duyulur.^{10,12,21} Büyük ör-

neklem gruplarında ($n > 400$), ($n > 200$) gruplar arasındaki temel değişkenlerin denk olması beklenir.^{13,33,44} Dolayısıyla bu durumlarda RY'nin çok başarılı sonuçlar vereceği söylenebilir.^{10,13} Uygulama sonrasında gruplar arasında ortaya çıkacak olası bir farkın uygulamadan kaynaklandığı sonucuna ulaşılabılır.²⁷

Küçük örneklem gruplarında örneklemin heterojenlik düzeyi RY'nin uygun olup olmadığını belirleyen özelliklerden biridir.¹⁰ RY'nin kullanıldığı araştırmalarda örneklem grubu küçük ($n < 100$) ve heterojen bir yapıdaysa, sonuç üzerinde etkili olan önemli değişkenlerin gruplar arasında denk dağılmış olması düşük bir olasılığa sahiptir.^{10,12-14,23,28,32,33,49,50} Bu durum da yanlılığı ortaya çıkarır ve RY'nin etkisiz bir yöntem olarak tanımlanmasına neden olur.^{13,38,51}

Örneklem belirleme ve örneklemin gruplara ayrılma sürecinin doğası nedeni ile, araştırmacılar araştırmalarında elde ettikleri sonuçların doğru olduğu ya da ölçtükleri etkinin gerçeği yansıttığı konusunda emin olamazlar.⁵² Özellikle bağımsız grup tasarımı antrenman bilimi araştırmalarında küçük örneklem-

ler kullanılmasına rağmen grupların RY ile oluşturulması, araştırma sonuçlarını şüphe altında bırakabilecek bir etken olarak tanımlanabilir. Bu derleme çalışması kapsamında "Journal of Strength and Conditioning Research (JSCR)" dergisinde 2005-2011 yılları arasında yayımlanmış ve bağımsız grupların RY kullanılarak oluşturulduğu araştırmalar incelenmiştir (Erişim Adresi: <http://search.proquest.com/publication/30912>). Bu çalışmalarda örneklem büyüklüklerine yönelik bazı tanımlayıcı istatistikler aşağıda gösterilmiştir Tablo 1'de görülmektedir. Tablodaki verilerden de anlaşılacağı gibi bu çalışmalarda kullanılan örneklem, "büyük örneklem" ($n > 400$), ($n > 200$) olmaktan çok uzaktır.^{13,33,44}

DENGELİ DAĞITIM

"Araştırmada yer alan katılımcılar", "araştırma evreninin özellikleri", "araştırmanın süresi ve maliyeti"; araştırma gruplarının hangi teknikle oluşturulacağı konusunda göz önünde bulundurulması gereken bazı konulardır.⁶ Uygun olmayan grup oluşturma yöntemlerinin kullanılmasının

TABLO 1: JSCR dergisinde 2005-2011 yılları arasında yayımlanan RY yöntemi kullanılmış bağımsız grup tasarımı araştırmalarla ilgili tanımlayıcı istatistikler.

		*2005	2006	2007	2008	2009	2010	2011	**Hepsi
Bağımsız 2 grup	^a Arş. sayısı	19	11	19	22	32	30	17	150
	^b n (°X)	18,8	18,4	13,4	13,6	13,3	15,1	10,3	14,4
	^d s (±)	17,8	16,2	8,6	12,2	10,4	13,1	4,5	12,1
	En az	4	6	6	4	5	6	4	4
	En fazla	83	60	51	70	63	62	26	83
Bağımsız 3 grup	Arş. sayısı	8	6	9	12	17	14	6	72
	n (X)	11,8	19,5	11,6	11,8	11,5	9,7	11,4	11,9
	s (±)	2,6	18,7	7,5	6,4	4,3	3,3	5,0	7,5
	En az	8	8	5	6	5	6	7	5
	En fazla	18	60	33	27	20	19	21	60
Bağımsız 4 grup	Arş. sayısı	2	3	2	4	4	5	3	22
	n (X)	7,8	11,8	8,9	12,5	12,3	15,3	14,8	12,4
	s (±)	3,4	1,1	2,0	3,3	4,7	11,8	5,7	6,8
	En az	4	10	7	7	6	8	8	4
	En fazla	13	14	11	16	19	47	20	47
Bağımsız 5 grup	Arş. sayısı	-	-	-	-	1	1	-	2
	n (X)	-	-	-	-	17,4	20,8	-	19,1
	s (±)	-	-	-	-	1,3	2,7	-	2,7
	En az	-	-	-	-	16	18	-	16
	En fazla	-	-	-	-	19	24	-	24

*2005 yılının 2. sayısında 1967 katılımcıyla gerçekleştirilen araştırma⁵³, ortalamalar üzerinde olumsuz etki oluşturmaması amacıyla değerlendirilmeye alınmadı.

**2005-2011 yıllarına ait bağımsız grup tasarımı araştırmaların toplam istatistiksel verileri.

^aArş.: Araştırma; ^bn: Bağımsız grupların her birinde yer alan ortalama katılımcı sayısı; ^c: Aritmetik ortalama; ^ds: Standart sapma.

temel nedenlerinden biri, daha önce yapılan benzer çalışmalarındaki yöntemin sorgulanmadan aynen kullanılmasıdır.⁶

Özellikle küçük örneklem grubuna sahip araştırmalarda, RY'den daha uygun ve daha iyi yöntemlerin kullanılması gerekmektedir.¹³ Eşit sayıda ve benzer değişken özelliklerine sahip katılımcının gruplara sistematik bir şekilde denk dağıtılması RY'nin olumsuz yönlerini önler.⁵¹ Yapılan dengeli dağıtım, araştırma başlangıcında gruplar arası karşılaştırılabilirliğin arttırılmasını sağlamaya yöneliktir. Bu yöntemle araştırma sonucuna etki edebilecek güçlü değişkenler gruplar arasında denk bir şekilde dağıtılır. Dört yüzden az katılımcının kullanıldığı araştırmalarda dengeli dağıtım yönteminin kullanılması önerilmektedir. Bu yöntemle -belirlenebilir değişkenlerle bağlantılı olması koşuluyla- belirlenemeyen değişkenlerin dengeli dağıtılması ve daha dengeli grupların oluşturulması amaçlanır. Ancak değişkenler arasında bir bağlantı yoksa başlangıç aşamasındaki karşılaştırılabilirlik eşit düzeyde oluşmayabilir.⁵⁴

EŞLEME YÖNTEMİ

Gruplar arasında dengeli dağıtımın sağlanması ve RY'nin kullanımındaki sınırlılıkların önlenmesi için çeşitli yöntemler kullanılmaktadır. Bunlar arasında yer alanlardan biri de eşleme yöntemidir. Genel anlamda, uygulama grupları veya uygulama ve kontrol grubu arasındaki değişken dağılımını eşitlemeyi (dengelemeyi) amaçlayan herhangi bir yöntem, eşleme yöntemi olarak tanımlanmaktadır. Eşleme yöntemi genellikle nedensellik ilişkisi belirlemek ve uygulama etkisini belirlemedeki yanlılıkları önlemek amacıyla kullanılır.⁴⁰ Küçük örneklem gruplarında gerçekleştirilecek çalışmalarda en iyi yöntem başlangıç değerleri arasındaki farkı en aza indirgeyen eşleme (matching) yönteminin kullanılmasıdır.^{10,11,25,47} Ancak eşleme için kullanılan değişkenler, araştırma sonucuna etki edebilecek kadar güçlü ise bu geçerlidir.²⁵ Eşleme yöntemindeki amaç, katılımcıları -belirlenebilir bütün değişkenleri göz önünde bulundurarak- aralarında en az fark bulunan ikililere ayırıp bu ikilileri rastgele olarak gruplara dağıtmaktır.^{8,10,34}

Bu yöntemin olumsuz taraflarından biri -özellikle birden fazla değişkene bağlı olarak eşleme ger-

çekleştiriliyorsa- ikililer için yeterli düzeyde bir eşlemeyi gerçekleştirilememesi olasılığı ve belirlenemeyen yanlılıkların oldukça fazla oluşudur.^{10,34,40} Eşleme yapılırken hangi değişkenlerin göz önüne alınması gerektiği konusu da önemlidir. Fikir birliğine varılmış ve gerçekleştirilmekte olan araştırmanın sonucuna kesin etkide bulunacağı düşünülen değişkenlere öncelik verilmelidir.^{6,55} Araştırmalarda bazı değişkenler arasında neden-sonuç ilişkisi bulunabilir, eğer araştırmacı bunu bilmiyorsa, incelenmesi gereken değişkenleri incelemediği için hatalı yorumlarda bulunabilir.⁶ Araştırmadaki katılımcı sayısı azsa, eşleme yönteminin kullanılması çok yararlıdır. Gruplar arası fark çok az olduğunda uygulama etkileri görece olarak az sayıda bir katılımcıyla da gözlemlenebilir.³⁴ Ayrıca RY'nin içine eşleme yönteminin yerleştirilmesi, araştırma sonuçlarının yorumlanmasında ortaya çıkabilecek büyük olumsuzlukları önleyebilir.⁵⁵

Bu derleme çalışması kapsamında JSCR dergisinde 2005-2011 yılları arasında yayımlanmış ve bağımsız grupların, eşleme yöntemi kullanılarak oluşturulduğu araştırmalar incelenmiştir. Bu araştırmalardaki örneklem büyüklüklerine yönelik bazı tanımlayıcı istatistikler Tablo 2'de görülmektedir. Elde edilen veriler göz önünde bulundurulduğunda, belirtilen yıllar arasında bağımsız gruplarda gerçekleştirilen araştırmaların sadece %25'inde dengeli gruplar oluşturmak için eşleme yöntemi kullanılmıştır.

RASTGELELEŞTİRME YÖNTEMİNE YÖNELİK ELEŞTİRİLER

Literatürde RY'nin bağımsız grup tasarımı araştırmalarda, araştırma sonucuna etki edebilecek temel bağımsız değişkenleri araştırma grupları arasında denk dağıtıp dağıtmadığına yönelik çeşitli çalışmalar bulunmaktadır.^{34,54,55} Bu çalışmalar, rastgeleleştirmenin kullanıldığı ve kullanılmadığı yöntemleri karşılaştırmak için çeşitli bilgisayar simülasyonları, meta-analizler ve yeniden örnekleme gibi uygulamaları kapsamaktadır.⁵⁶ Bu çalışmalar, bağımsız değişkenler temelinde hangi yöntemin denk gruplar oluşturma konusunda daha verimli olduğunu belirlemeye yönelik gerçekleştirilmektedir. Bu denklik -doğal olarak- sadece incelenen değişkenler açısından gerçekleştirilebilmektedir.

TABLO 2: JSCR dergisinde 2005-2011 yılları arasında yayımlanan eşleme yöntemi kullanılmış bağımsız grup tasarımı araştırmalarla ilgili tanımlayıcı istatistikler.

		Bağımsız 2 grup	Bağımsız 3 grup	Bağımsız 4 grup	Bağımsız 5 grup
Araştırma sayısı		51	24	4	2
*n ($\bar{X}\pm s$)		15,7±7,5	17,3±22,1	24,3±21,4	8,3±1,4
Eşleme için kullanılan değişken sayısı	$\bar{X}\pm s$	2,2±1,6	2,2±1,1	3,5±1,0	3,0±0,0
	En az	1	1	2	3
	En fazla	9	5	4	3

* Bağımsız grupların her birinde yer alan ortalama katılımcı sayısı.

RY'nin sonuca etki edebilecek tüm değişkenleri gruplar arasında denk dağıtıp dağıtmadığının tam bir kesinlikle bilinmeyeceği belirtilmektedir.^{8,32,43} RY ile temel değişkenler gruplar arasında denk dağılırsa belirlenemeyen/ölçülemeyen değişkenlerin de gruplar arasında denk dağılmış olacağı görüşünün sadece bir "varsayım", "umut" ve "beklenti" olduğu ifade edilmektedir.^{10,11,25} Bu konuyla ilgili olarak Berger ve Weinstein RY'nin, başlangıç değerlerindeki yanlışlıkların tümünü ortadan kaldırdığı inanın sorgulanmadan kabul edilen bir inanış olduğunu vurgulamışlardır. Bununla bağlantılı olarak da araştırmalarda elde edilen ve gruplar arası incelenen değişkenlerdeki istatistiksel açıdan anlamlı farkı ifade eden yeterli küçüklükteki "p" değerlerinin ($p<0,05$) (diğer bir deyişle gerçekte gruplar arasında bir fark olmamasına rağmen, sadece şans eseri gözlemlenen farkın ya da daha büyük bir farkın ortaya çıkma olasılığının) tek kaynağı olarak araştırmalarda etkisi incelenmekte olan değişkenin gösterildiğini ifade etmişlerdir. Ancak bu süreçte, elde edilen yeterli küçüklükteki "p" değerlerinin gruplar arası farktan ya da küçük başka bir yöntemsel hatadan kaynaklanmış olabileceğine yönelik olasılığın tamamen göz ardı edildiği noktasına da vurgu yapmışlardır.⁴⁵

Gruplar arasında temel değişkenler denk olsa bile, belirlenemeyen/ölçülemeyen değişkenler araştırma sonucu üzerinde dolaylı veya dolaysız yoldan etkide bulunabilecek özellikteyse bu denkliğin işlevsel olduğu söylenemez. "Denk olduğu düşünülen" araştırma gruplarının hepsinde aynı uygulamanın aynı yöntemle gerçekleştirilmesi sonucunda incelenen bağımlı değişkende ortaya çıkacak olası değişimin gruplar arasında benzer olduğu kanıtlanmadığı sürece, grupların denk olduğu (temel bağımsız değişkenler denk olsa bile) görüşü sadece bir varsayım olarak kalacaktır.

Bağımsız gruplarda gerçekleştirilen araştırmalarda genel olarak farklı uygulamaların kısa, orta veya uzun vadeli kronik etkileri incelenmektedir. Araştırma sonunda farklı uygulamalar, denk olduğu varsayılan grupların özelliklerinde değişiklikler yaratacağından, bir grupta belirli bir etkiyi yaratan uygulamanın diğer grupta da aynı etkiyi yaratıp yaratamayacağı test edilemez. İnsan organizması gibi karmaşık bir sistemde yer alan çok sayıda değişken, zamana bağlı olarak küçük değişimler gösterir. Çok sayıda "küçük değişim" in büyük farklılıklara neden olabileceği göz önünde bulundurulduğunda, insan merkezli çalışmalarda denk gruplar oluşturmanın zorluğu çok daha iyi anlaşılabilir.⁴² Dolayısıyla çaprazlama yönteminin kullanılamayacağı küçük örneklerde ($n<400$), RY yerine temel değişkenler arasındaki farkı en aza indirmeye yönelik yöntemlerin kullanılmasının daha uygun olacağı belirtilmektedir.^{10,11,21,33}

Bilimsel araştırmalarda yanlılık sadece uygulama aşamasında değil, yayın aşamasında da ortaya çıkmaktadır.^{11,20} RY'nin kullanıldığı araştırmalarda bu yanlılıkların önlenmesi amacıyla "Consolidated Standards of Reporting Trials (CONSORT)" adı altında bir rapor yayımlanmıştır (Erişim Adresi: <http://www.consort-statement.org/consort-statement/overview/>). Bu rapor, RY'nin raporlaştırılma düzeyini geliştirmeyi ve RY kullanımında bir standart oluşturmayı amaçlamaktadır.^{11,25} RY ile yapılan çalışmaların yayın haline getirilmesi aşamasında, bu raporun içerdiği maddelere ek olarak, farklı maddeler hakkında da bilgi verilmesi yönünde görüş ve öneriler bulunmaktadır.⁵⁷⁻⁵⁹ Bu şekilde, araştırmada kullanılan RY'ye ait tasarım, uygulama, değerlendirme ve yorumlama aşamalarının daha anlaşılabilir hale gelmesi ve okuyucuların, araştırma sonuçlarının geçerliliğini daha rahat

yorumlamaları planlanmıştır. Araştırmacıların yayınlarında kullandıkları RY'yi ayrıntılı ve tamamen şeffaf bir şekilde açıklamaları, kullanılan yöntemin değerlendirilebilmesi için bir gerekliliktir.^{11,25}

JSCR dergisinde 2005-2011 yılları arasında yayımlanmış ve bağımsız grupların RY kullanılarak oluşturulduğu küçük örneklemlerli araştırmalarda, gruplar arası katılımcı sayısı dengesizliklerinin önlenmesi için araştırma gruplarının eşit sayıda veya birbirine yakın sayıda katılımcıdan oluşturulduğu görülmektedir. Ancak bu araştırmaların hemen hemen hiçbirinde RY'nin nasıl uygulandığı açıklanmamıştır.

RY ile ilgili en büyük yanlış anlaşılmalardan biri de, bir evrenden rastgele örneklem seçme (random sampling) ve araştırma örneklemini rastgele gruplara ayırma (random allocation/assignment) işlemlerinin aynı süreçler olarak algılanmasıdır.³² RY aşağıdaki süreçleri içerir:^{20,23,60}

- Kuramsal evrenin belirlenmesi (Araştırma sonuçlarının genelleştirileceği nüfus),
- Araştırma evreninin belirlenmesi (Araştırma için ulaşabileceğiniz nüfus),
- Örneklem belirleme [Araştırma evreni içinden rastgele seçtiğiniz ve araştırmaya katılma özelliklerine uyan örneklem (random sampling)],
- Örneklem grubu (Belirlenen örneklem içinden araştırmaya katılanlar),
- Araştırma grupları [Örneklem grubunun rastgele araştırma gruplarına ayrılması (random allocation/assignment)],

Antrenman bilimi literatüründe yer alan çalışmalarda genellikle “rastgele örneklem seçme” işleminin nasıl gerçekleştirildiği (araştırmada eğer böyle bir işlem gerçekleştirilmişse) açıklanmamaktadır. Buna ek olarak, araştırma örnekleminin gruplara nasıl ayrıldığı konusunda açıklama sağlayan çalışma sayısı da çok azdır. Araştırmacılar çalışmalarında kullandıkları RY'yi açıklamak için sadece “Katılımcılar rastgele gruplara ayrılmıştır” veya bu anlama gelen farklı ifadeler kullanmakla yetinmektedirler. Dolayısıyla antrenman bilimi literatüründe RY'nin kullanıldığı çalışmaların çoğu, aslında RY'nin “sözde kullanıldığı” çalışmalar olarak tanımlanabilir. JSCR dergisinin 2005-2011 yılları arasındaki incelenen sayıları içinde sadece

Koenig ve ark.nın çalışmasında, katılımcıların, “bilgisayar tarafından oluşturulan rastgele sayı tablosu” kullanılarak gruplara dağıtıldığı rapor edilmiştir.⁶¹ Paradisis ve Cooke'nin çalışmasında ise RY'nin nasıl uygulandığı rapor edilmemekle birlikte, kendi araştırmalarında kullandıkları örneklem grubunun küçük olmasına atıfta bulunarak, RY'nin kullanıldığı küçük örneklemlerde bağımsız grupların temel değişkenler açısından farklılık gösterebileceği konusuna değinilmiştir.⁶²

RY'nin kullanımında üzerinde durulması gereken diğer bir konu da, rastgele dağıtım sonrası gerçekleştirilen istatistiksel kontrol aşamasıdır. Altman, katılımcıların araştırma gruplarına çok iyi düzenlenmiş bir rastgeleleştirmeye dağıtılması sonucunda, gruplar arasındaki herhangi bir farkın tamamen şans eseri ortaya çıkacağını ifade etmiştir.⁶³ Bununla birlikte, gerçekleştirilen rastgeleleştirme sonrası başlangıç değerleri açısından gruplar arasında bir fark olup olmadığını belirlemek için istatistiksel yöntemlerin kullanılmasının, zaten şans eseri ortaya çıkmış olabilecek bir durumun şans eseri ortaya çıkma olasılığını belirlemek anlamına geldiğini ve böyle bir prosedürün de tamamen saçmalık olduğunu belirtmiştir.⁶³ Benzer şekilde, Berger'in kendi kitabında Rothman'dan yaptığı bir alıntıda da [Rothman KJ (1977) *Epidemiologic methods in clinical trials*. Cancer, 39, 1771-1775] Altman'ın yorumuna benzer bir yorumla karşılaşmaktadır.⁶⁴

Altman ayrıca, incelenen temel değişkenler arasında istatistiksel olarak fark bulunmamasının, bu değişkenlerin, araştırma sonuçları üzerinde etkide bulunmadığı anlamına gelmediğini belirtmektedir. Açıklama olarak da, araştırma sonucu üzerinde büyük etkiye sahip bir değişkendeki gruplar arası küçük bir farkın, araştırma sonucu üzerinde büyük etkisi bulunmayan bir değişkendeki çok büyük ve anlamlı bir farkla karşılaştırıldığında çok daha önemli olduğu görüşünü öne sürmektedir.⁶³ Özellikle insan organizması gibi karmaşık sistemlerde bu etkiyle daha sık karşılaşılacağı göz önünde bulundurulmalıdır.⁴² Altman, değişkenlerdeki dengesizliklerin, istatistiksel anlamlılık boyutunda değil, uygulamayla ilişkisi boyutunda incelenmesi gerektiğini de belirtmiştir.⁶³

Her katılımcı bir grup değişkeni kapsadığı ve gruplara dağılım kişi temelinde yapıldığından,

gruplar arası denklik işlemleri de kişilere ait değişkenler temelinde gerçekleştirilmelidir.¹⁰ RY bu işlemi grup ortalamaları temelinde gerçekleştirmeyi hedeflediğinden, kişi temelinde denk olmayan bir dağılım, ortalama açısından denk bir dağılım olarak ortaya çıkabilir.

Rastgeleleştirme sonrası ortaya çıkabilecek olumsuz durumları önlemek için kullanılan yöntemlerden biri, ortak değişken düzeltme yöntemidir. Bu yöntem çeşitli değişkenleri ölçer ve matematiksel düzenlemelerle bu değişkenlerin etkisini bağımlı ve bağımsız değişken arasındaki tahmini gerçek ilişkiyi belirlemek üzere ortadan kaldırır. RY ile birlikte kullanılan bu ortak değişken düzeltme yönteminin temel amacı, karşılaştırılabilir alt gruplar bulmak ve uygulama karşılaştırmalarını bu alt gruplarla sınırlandırmaktır. Dolayısıyla ortak değişken düzeltme yöntemi karşılaştırılabilir gruplar yaratmaz.⁴⁷ Ayrıca, ortak değişken düzeltme yöntemiyle sadece ölçülebilir değişkenler düzeltilir.^{32,47,65} Bu yöntem ölçül(e)meyen bir değişken üzerinde düzeltme yapamayacağı için çeşitli sınırlılıklar içerir.^{22,32} RY’de gruplar arasında oluşan dengesizlikleri düzeltme çabalarının araştırma sonucunun geçerliliğiyle ilgili belirsizliklere neden olacağı da ifade edilmektedir.²¹ Ayrıca, ortak değişkenlerin dengesizliği öncellenemeyen etkileşimlere sıklıkla yol açtığından, düzeltme sonrasında ortaya çıkan durumun yorumlanmasının oldukça zor olduğu, vurgulanan diğer bir konudur.¹³ Elde edilen sonuç ve ortak değişkenler arasında gerçekte orta düzeyde doğrusal olmayan bir ilişkinin bulunması durumunda, doğrusal regresyonla gerçekleştirilen düzeltmelerin tahmin edilen uygulama etkisinde yanlılıkları arttırabileceği de belirtilmektedir. Bu durumun özellikle, araştırma grupları arasında ortak değişken ortalamaları ve varyansları açısından büyük farklar bulunduğu söz konusu olduğu öne sürülmektedir.⁴⁰ Ayrıca, kontrol edilen değişken sayısı arttıkça ölçümlerin ve istatistiksel değerlendirmelerin karmaşıklığı da artmaktadır. Bununla bağlantılı olarak da araştırmanın dış geçerliliğinin azaldığı ifade edilmektedir.⁴⁷

Feinstein, tıp alanında uygulanan çeşitli tedavilerin etkisini belirleme ve sonuçları değerlendirme konusunda yanlılık içermeyen bir karşılaştırmanın bir istatistiksel yöntemde mutlaka aranılan şart ol-

duğunu ve bu nedenle RY’nin tıp alanında tek bilimsel strateji olarak kullanıldığını belirtmiştir.⁶⁶ Bu durumun, araştırmalardaki bilimsel klinik hassaslık (tamlık) için bir tehdit oluşturduğunu vurgulamıştır. Bazı araştırmalar çelişkili sonuçlar ortaya çıkardığında, sonuçlar ortak değişkenlerden etkilenmiş olabileceği için elde edilen verilerin ek istatistiksel analizlerle değerlendirilmesi gerekebileceğine dikkat çekmiştir. Ancak RY’nin bütün yanlılıkları önleyebileceği kabul edildiğinden, genellikle araştırmalarda önemli klinik farklılıkların kaydedilmediğini, sınıflandırılmadığını ya da analiz için mevcut olan verilere eklenmediğini belirtmiştir. Bu nedenle de araştırmadaki çelişkilerin nadiren çözülebildiğini savunmuştur. Feinstein’in klinik araştırmalarla ilgili görüşlerinin antrenman bilimi araştırmaları ve diğer araştırmalar için de geçerli olduğunu söylemek hatalı olmayacaktır. Feinstein, RY’de kullanılan istatistiksel stratejilerin araştırma sonuçları üzerindeki şüpheyi ortadan kaldırmada yeterince başarılı olmayacağını belirtmiş ve RY’yi bilimsel bir yöntem olarak değil; “belirsizliğin matematiksel sömürüsü için değersiz bir istatistiksel strateji” olarak tanımlamıştır.⁶⁶

Dış geçerlilik, araştırma sonuçlarının genelle uygulanabilirliğini ifade eden bir terimdir.¹⁹ Araştırmalarda RY kullanımının temel hedeflerinden biri de, araştırma sonuçlarına güçlü bir dış geçerlilik sağlamaktır. Bununla birlikte insan araştırmalarında RY’nin dış geçerlilikten çok, iç geçerliliği sağladığı ifade edilmektedir.^{19,32} Büyük bir araştırma örnekleme yoksa ve örneklem, araştırma sonuçlarının genelleştirilmek istendiği bilinen bir evrenden seçilmemişse elde edilen sonuçlar çok büyük bir önem taşımaz.¹⁹ Benzer şekilde, araştırma evreni sadece araştırmaya katılanlardan oluşuyorsa (çoğu antrenman bilimi araştırmasında olduğu gibi) RY’nin bir “altın standart” oluşturduğu söylenemez. Todman ve Dugard, elde edilen araştırma sonuçlarının dış geçerliliğinin ancak araştırmanın farklı örneklemelerde tekrar edilmesiyle ve istatistiksel olmayan mantıksal düşünce sistemiyle sağlanabileceğini ifade etmişlerdir.¹⁹

Literatürde, benzer yöntemlerle gerçekleştirilmiş olmalarına rağmen çelişkili sonuçlar ortaya koyan çok sayıda antrenman bilimi araştırması bu-

lunmaktadır. Ancak bu araştırmalarda -RY kullanılmış olduğu için- oluşturulan grupların denk olmama olasılığı hiçbir zaman göz önünde bulundurulmamaktadır. Hiçbir mantıklı açıklamaya dayanmayan “denk gruplar varsayımı”, çelişkili sonuçların belki de en büyük kaynağı olabilecek “denk olmayan gruplar” seçeneğinin sınanmasına olanak tanımamaktadır. Dolayısıyla araştırma sonuçlarındaki çelişkilerin, sürekli başka nedenlere dayandırılarak açıklanması tercih edilmektedir. Bunun sonucunda, genelleştirilmesi oldukça zor bulgularla dolup taşan ve uygulama alanına katkısı tartışılabilir birçok yayın, hakemli dergilerin sayfaları içerisinde literatürde yerini almaktadır.

RASTGELELEŞTİRME YÖNTEMİ YERİNE EŞLEME YÖNTEMİNİN KULLANILMASI

Bağımsız grup tasarımı araştırmalarda eşleme yapılırken hangi değişkenlerin göz önüne alınması gerektiği konusunun önemli olduğu ve fikir birliğine varılmış, gerçekleştirilmekte olan araştırmanın sonucuna kesin etkide bulunacağı düşünülen değişkenlere öncelik verilmesi gerektiği belirtilmektedir.^{6,55} Araştırmalarda, denk gruplar oluşturma konusunda en uygun yöntemi belirlemek amacıyla çeşitli RY ve dengeli dağıtım yöntemleri; simülasyon çalışmaları yoluyla karşılaştırılmıştır.^{67,68} Daha çok tıp alanında gerçekleştirilen bu çalışmalarda, araştırma gruplarına düşen hasta sayıları ve grupların bazı temel özelliklerinin denkliği incelenmiştir.⁶⁷

Denk gruplar oluşturmak için kullanılan eşleme yönteminin temel bağımsız değişkenleri gruplar arasında denk dağıtımdaki verimliliği de çeşitli simülasyon araştırmalarında ve benzeri araştırmalarda incelenmiştir.⁵⁶ Bu tür araştırmalarda genellikle grupların denkliği, incelenen bağımsız değişkenlerin gruplar arasında ne düzeyde denk olduğuyla belirlenmiştir.

Yapılan bazı simülasyon çalışmalarının sonuçları, küçük örneklem (n=20, n=50) gruplarıyla ve hatta 100-300 arası katılımcıdan oluşan örneklem gruplarıyla gerçekleştirilen araştırmalarda bile dengeli dağıtımın RY'ye göre daha geçerli ve hassas sonuçlar verdiğini göstermiştir.^{21,25,54} Zwinderman ve ark. tıp alanında artık RY'nin tercih edilmediğini,

tıbbi uygulamalar arasındaki karşılaştırılma hassasiyetini arttırmak için araştırmalarda artık laboratuvar verileri açısından birbirine benzer özellikte olan hastaların kullanıldığını belirtmişlerdir. Bununla beraber, bu yöntemin normal olmayan veri riskini arttırabileceğine de vurgu yapmışlardır.⁶⁹

RY'nin yanlılıkları önleyen bir yöntem olduğu savı hiçbir sorgulama yapılmadan kabul edilmektedir.⁴⁵ Ayrıca, RY'nin küçük örneklem gruplarında kullanılmasının, yanlılıkları ortaya çıkardığı da ifade edilmektedir.^{13,51} RY kullanılmayan araştırmalarda, bir uygulamanın araştırma sonuçları üzerinde yaratabileceği yanlılıkları önlemek için genellikle bu araştırma tasarımlarının belirli bölümlerinde RY kullanılmaktadır.^{9,55} Kunz ve Oxman, yanlılıkların boyutundaki tahmin edilemezliği önlemek için uygulama tasarımının içine rastgeleleştirme yoluyla tahmin edilemezlik katılmasını bir paradoks olarak tanımlamışlardır.⁹

EŞLEME YÖNTEMİNDE KULLANILAN DEĞİŞKENLER

Denk gruplar oluşturma amacıyla kullanılan bir yöntem, araştırma sonucuna etki edebilecek ne kadar fazla sayıda bağımsız değişkeni gruplar arasında denk dağıtırsa yöntemin o kadar verimli olduğu kabul edilmektedir. Ancak bu değişkenlerin kendi aralarındaki doğrusal olmayan karmaşık ve nedensel etkileşimlerinin araştırma sonucuna ne düzeyde etki ettiği belirlenemez.⁸ Quigley'in yayımladığı raporda, eşleme değişkeni olarak birden fazla değişkenin kullanılmasının, belirlenemeyen düzeyde yanlılık oluşturabileceği belirtilmiştir.³⁴ Bu nedenle denk gruplar oluşturmada birden fazla değişken kullanılmıyorsa, araştırma sonucuna etki edebilecek tüm değişkenlerin (ölçülebilen ve ölçülemeyen) toplam etkisini yansıtabileceği düşünülen tek bir değişkene bağlı olarak katılımcıların gruplar arasında denk dağıtılması, araştırmalarda çok daha verimli sonuçlar ortaya koyabilir.

Glaister ve ark.nın çalışmasında, antrenman bilimi araştırmalarının çoğunun bağımsız grup tasarımıyla gerçekleştirilmesine rağmen, araştırma sonucuna etki edebilecek temel değişkenler açısından herhangi bir eşleme yapılmadığı konusuna vurgu yapılmaktadır.⁷⁰ Bununla birlikte, çoğu araştırmada

-araştırma sonucuna çok büyük bir etkisi bulunmayacak olsa da- boy, kütle, yaş gibi ölçümünde hiçbir zorluk yaşanmayan değişkenlerin sıklıkla eşleme değişkeni olarak kullanıldığı görülmektedir.

Antrenman bilimi, “hareket”le ilişkili “nasıl” sorusuna yanıt arayan bir bilim dalıdır. Hareket sistemimiz, kasların tendonlar aracılığıyla kemikleri çektiği ve eklemlerle desteklenen bütünleşik bir kaldıraç sisteminden oluşmaktadır. Dolayısıyla hareketin -temel olarak- kaslar ve kemikler arasındaki etkileşimden kaynaklandığı söylenebilir. Kaslar tarafından üretilen kuvvetin çevreye hareket olarak yansması, -temel hareket kanunlarına dayalı olarak- hareketin gerçekleştiği vücut parçalarının (kemiklerin) antropometrik ölçüleriyle yakından ilişkilidir. Dolayısıyla araştırmalarda beden kitlesi değişkeni yerine, hareketin gerçekleştiği vücut bölümlerine ait yağsız beden kitlesi değişkeninin ve de boy uzunluğu değişkeni yerine, harekete katılan vücut bölümleri arasındaki oransal antropometrik ölçüm değerlerinin, eşleme değişkeni olarak kullanılması çok daha mantıklı bir yaklaşım olacaktır [2005-2011 yılları arasında JSCR’de yer alan bağımsız grup tasarımı araştırmaların 19 (%23)’unda beden kütlesi değişkeni, 21 (%26)’inde de boy değişkeni eşleme amacıyla kullanılmıştır].

Zaman kavramı, gündelik yaşamı kolaylaştırmak için insanlar tarafından oluşturulmuştur. Zaman kavramı temelinde tanımlanan kronolojik yaş, yıl cinsinden ifade edilir ve yıl, dünyanın güneş çevresinde tam bir turunu tamamlaması için geçen süre olarak tanımlanmıştır. Dolayısıyla kronolojik yaş, bir insanın doğum anından günümüze kadar dünyanın güneş çevresinde kaç defa döndüğünü ifade eden bir sayıdan ibarettir. Antrenman bilimiyle ilgili bilimsel araştırmaların merkezinde insan ve insan organizmasına ait biyolojik sistemler bulunmaktadır. Bu nedenle araştırmalarda gruplar arası denklik değişkeni olarak “dünyanın güneş çevresinde dönme sayısından” çok, bu “sayı”nın organizmadaki biyolojik sistemler üzerinde yaratmış olduğu toplam etkiyi ifade edebilecek değişkenler kullanılmalıdır. Aynı kronolojik yaşa sahip, ancak yaşam tarzları ve koşulları açısından büyük farklılık gösteren (dolayısıyla farklı biyolojik sistem verimliliğine ve farklı adaptasyon kapasitelerine sahip) katılımcıların “yaş” değişkeni açısından denk kabul

edilmesi mantıksal açıdan sorunlu bir yaklaşımdır [2005-2011 yılları arasında JSCR’de yer alan bağımsız grup tasarımı araştırmaların 36 (%44)’sında yaş değişkeni eşleme amacıyla kullanılmıştır].

Özellikle bazı araştırmalarda eşleme yönteminin sadece yukarıda belirtilen demografik özelliklere dayalı olarak gerçekleştirilmiş olması -Shadish ve ark.nın da belirttiği gibi- bu araştırmaların denk gruplar üzerinde gerçekleştirilmiş olduğu anlamına gelmemektedir [2005-2011 yılları arasında JSCR’de yer alan bağımsız grup tasarımı araştırmaların 9 (%11)’unda eşleme değişkeni olarak “sadece” yukarıda belirtilen üç değişkenden en az biri kullanılmıştır].⁵⁶

SONUÇ

RY, çoğu bilimsel araştırma için bir “altın standart” olarak kullanılmaktadır. Ancak RY’nin bir altın standart oluşturması için araştırmalarda gerekli olan en önemli ön koşulun (büyük bir araştırma örneklemini) sağlanıp sağlanmadığı konusu göz ardı edilmekte ve antrenman bilimi araştırmalarında RY, hiçbir sorgulama yapılmadan kullanılmaktadır. Dolayısıyla bu durum, araştırma sonuçlarının geçerliliği üzerinde şüphe oluşturmaktadır.

Antrenman bilimi alanında gerçekleştirilen bağımsız grup tasarımı araştırmalar genellikle bir uygulamanın kısa, orta ve uzun vadeli etkilerini ortaya çıkarmak için düzenlenir. Dolayısıyla bu tür araştırmaların büyük örneklemelerde gerçekleştirilmesi ve araştırma sonuçları üzerinde yanlışlıklar oluşturmaya- cık şekilde idare edilmesi oldukça zordur. Bu nokta, antrenman bilimi araştırmalarındaki en büyük sınırlılığı oluşturmaktadır. Özellikle elit düzeye ulaşmış, milli takımlarda yer alan veya araştırmanın yapısı gereği belirli performans özelliklerine sahip olması beklenen katılımcılarla gerçekleştirilen araştırmalarda, büyük örneklemeler (n>200, n>400 vb.) üzerinde çalışmak hemen hemen imkânsızdır.

Aynı antrenman uygulamasına yönelik gerçekleştirilen antrenman bilimi araştırmalarında benzer yöntemler kullanılmış olmasına rağmen araştırmalarda çelişkili sonuçların elde edilmesi sıklıkla karşılaşılan bir durumdur. RY’nin kullanıldığı antrenman bilimi araştırmalarından elde edilen sonuçların, gerçek antrenman etkisini mi

yanıttığı, yoksa RY ile oluşturulan gruplar arasındaki olası farklılıklardan kaynaklanan büyük yanlışlıklar mı taşıdığı konusu aydınlatılması gereken bir noktadır. Bu çelişkiler ve araştırma gruplarındaki olası farklılıklar arasında bir nedensellik ilişkisi olup olmadığının belirlenmesi önemli bir araştırma konusudur. Bu konunun açıklığa kavuşturulabilmesi için literatürde var olan antrenman bilimi araştırmalarının, aynı araştırma yöntemi ve aynı değişkenler kullanılarak farklı örneklemeler üzerinde birebir tekrarlanması gerekir. Bu yaklaşım, yaratıcılıktan uzak gibi görünse de, elde edilen araştırma sonuçlarının “doğru” sonuçlar olduğunu test etmenin tek geçerli yoludur. Dolayısıyla bu tür çalışmaların sistemli bir şekilde planlanması ve gerçekleştirilmesi,

literatürdeki büyük bir eksikliği gidermekle kalmayacak, aynı zamanda araştırma sonuçlarının genelleştirilebilmesine de olanak sağlayacaktır.

Diğer taraftan, antrenman bilimi alanındaki bağımsız grup tasarımı araştırmalarda RY’yi vazgeçilmez bir yöntem olarak kabul etmektense, araştırma yapısı için daha uygun bir yöntem kullanmak daha sağlıklı sonuçlar sağlayabilir. Özellikle araştırma sonuçlarına etki edebilecek olası tüm değişkenlerin toplam etkisini yansıtabileceği düşünülen temel bir değişken belirlenerek, araştırma gruplarının bu değişkene göre denk olarak oluşturulması, araştırmalarda incelenen uygulama etkisini daha açık bir şekilde ortaya koyabilir.

KAYNAKLAR

- Srinagesh K. Experimental research in science: its name and nature. *The Principles of Experimental Research*. 1st ed. Amsterdam; Boston: Elsevier/Butterworth-Heinemann; 2006. p.1-14.
- Cipani E. Science and the scientific method. *Practical Research Methods for Educators: Becoming An Evidence-Based Practitioner*. 1st ed. New York: Springer; 2009. p.1-34.
- Gould JE. Philosophy of science-the scientific approach to knowledge. *Concise Handbook of Experimental Methods for the Behavioral and Biological Sciences*. 1st ed. Boca Raton: CRC Press; 2002. p.3-29.
- Trochim W, Donnelly JP. *Foundations. The Research Methods Knowledge Base*. 3rd ed. Ohio, United States: Atomic Dog Publishing; 2006. p.1-37.
- WHO. *Experimental studies and clinical trials. Health Research Methodology: A Guide for Training in Research Methods*. 2nd ed. Manila: World Health Organization; 2001. p.55-70.
- Ercan İ, Yazıcı B, Yang Y, Özkaya G, Cangür Ş, Ediz B, et al. Misusage of statistics in medical research. *Eur J Gen Med* 2007;4(3):128-34.
- Strickland OL. Impact of unreliability of measurements on statistical conclusion validity. *J Nurs Meas* 2005;13(2):83-5.
- Watt JH, van den Berg SA. Testing hypotheses. Confounds and controls. *Research Methods for Communication Science*. 1st ed. Boston: Allyn and Bacon; 1995. p.39-49.
- Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *Br Med J* 1998;317(7167):1185-90.
- Blair E. Gold is not always good enough: the shortcomings of randomization when evaluating interventions in small heterogeneous samples. *J Clin Epidemiol* 2004;57(12):1219-22.
- Tai SS, Iliffe S. Considerations for the design and analysis of experimental studies in physical activity and exercise promotion: advantages of the randomised controlled trial. *Br J Sports Med* 2000;34(3):220-4.
- Devereaux PJ, Yusuf S. The evolution of the randomized controlled trial and its role in evidence-based decision making. *J Intern Med* 2003;254(2):105-13.
- Kang M, Ragan BG, Park JH. Issues in outcomes research: an overview of randomization techniques for clinical trials. *J Athl Training* 2008;43(2):215-21.
- Ankaralı H, Ankaralı S. [Adjusting the effect of baseline differences between groups in trials with which have two or more groups]. *Türkiye Klinikleri J Med Sci* 2009;29(1):91-8.
- Singh YK. *Experimental method. Fundamental of Research Methodology and Statistics*. 1st ed. New Delhi: New Age International Publishers; 2006. p.134-46.
- Yeadon MR. What are the limitations of experimental and theoretical approaches in sports biomechanics? In: McNamee MJ, ed. *Philosophy and the Sciences of Exercise Health and Sport: Critical Perspectives on Research Methods*. 1st ed. London; New York: Routledge; 2005. p.126-36.
- Cooper SM, Nevill AM. Do statistical methods replace reasoning in exercise science research? In: McNamee MJ, ed. *Philosophy and the Sciences of Exercise Health and Sport: Critical Perspectives on Research Methods*. 1st ed. London; New York: Routledge; 2005. p.111-25.
- Gould JE. Forms of scientific observation and research. *Concise Handbook of Experimental Methods for the Behavioral and Biological Sciences*. 1st ed. Boca Raton: CRC Press; 2002. p.31-50.
- Todman JB, Dugard P. Data analysis in single-case and small-n experiments. *Single-Case and Small-N Experimental Designs: A Practical Guide to Randomization Tests*. 1st ed. Mahwah NJ: Lawrence Erlbaum Associates; 2001. p.1-9.
- Matthews JNS. What is a randomized controlled trial? *Introduction to Randomized Controlled Clinical Trials*. 2nd ed. Boca Raton: Chapman & Hall/CRC; 2006. p.1-13.
- Scott NW, McPherson GC, Ramsay CR, Campbell MK. The method of minimization for allocation to clinical trials: a review. *Controlled Clin Trials* 2002;23(6):662-74.
- Berger V. A review of methods for ensuring the comparability of comparison groups in randomized clinical trials. *Rev Recent Clin Trials* 2006;1(1):81-6.
- Viera AJ, Bangdiwala SI. Eliminating bias in randomized controlled trials: Importance of allocation concealment and masking. *Fam Med* 2007;39(2):132-7.
- Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance not choice. *Lancet* 2002;359(9305):515-9.
- Bruhn M, McKenzie D. In pursuit of balance: randomization in practice in development field experiments. *Am Econ J Appl Econ* 2009;1(4):200-32.
- Gould JE. Control in experiments-Additional techniques and principles. *Concise Handbook of Experimental Methods for the Behavioral and Biological Sciences*. 1st ed. Boca Raton: CRC Press; 2002. p.208.

27. Torgerson DJ, Torgerson C. Background to controlled trials. *Designing Randomised Trials in Health, Education and the Social Sciences: an Introduction*. 1st ed. Basingstoke England; New York: Palgrave Macmillan; 2008. p.1-8.
28. Chia KS. Randomisation: magical cure for bias? *Ann Acad Med Singapore* 2000;29(5): 563-4.
29. Treasure T, MacRae KD. Minimisation: the platinum standard for trials? *Randomisation doesn't guarantee similarity of groups; minimisation does*. *BMJ* 1998;317(7155):362-3.
30. Schmelzer M. Understanding the research methodology: should we trust the researchers' conclusions? *Gastroenterol Nurs* 2000;23(6):269-74.
31. Berger V. Part I: Is there a problem with reliability in medical studies? An evolution of comparative methodology. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. 1st ed. Baltimore County: John Wiley & Sons Ltd; 2005. p.1-15.
32. Torgerson DJ, Torgerson C. What is special about randomisation? *Designing Randomised Trials in Health, Education and the Social Sciences: an Introduction*. 1st ed. Basingstoke England; New York: Palgrave Macmillan; 2008. p.22-43.
33. Gurusamy KS, Gluud C, Nikolova D, Davidson BR. Assessment of risk of bias in randomized clinical trials in surgery. *Br J Surg* 2009; 96(4):342-9.
34. Quigley DD. Using Multivariate Matched Sampling That Incorporates the Propensity Score to Establish a Comparison Group CSE Technical Report No. 596. Center for the Study of Evaluation University of California, National Center for Research on Evaluation, Standards, and Student Testing Graduate School of Education & Information Studies University of California, Los Angeles; 2003. p.17.
35. Raaijmakers M, Koffijberg H, Posthurnus J, van HB, van EH, Matthys W. Assessing performance of a randomized versus a non-randomized study design. *Contemp Clin Trials* 2008;29(2): 293-303.
36. 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.
37. Gracely E. So, why do I have to correct for multiple comparisons? Concepts and commentary on Turk et al. *Pain* 2008;139(3):481-2.
38. Ameringer S, Serlin RC, Ward S. Simpson's paradox and experimental research. *Nurs Res* 2009;58(2):123-7.
39. Berger V. Part II: Actions to be taken to improve the reliability of medical studies-preventing selection bias in randomized trials. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. 1st ed. Baltimore County: John Wiley & Sons Ltd; 2005. p.105-22.
40. Stuart EA. Matching methods for causal inference: a review and a look forward. *Stat Sci* 2010;25(1):1-21.
41. Blessing LTM, Chakrabarti A. *Descriptive study methods. DRM, a Design Research Methodology*. 1st ed. Dordrecht; London: Springer; 2009. p.239-73.
42. Rickles D. Causality in complex interventions. *Med Health Care Phil* 2009;12(1):77-90.
43. Torgerson DJ, Torgerson C. Sources of bias within randomised trials. *Designing Randomised Trials in Health, Education and the Social Sciences: an Introduction*. 1st ed. Basingstoke England; New York: Palgrave Macmillan; 2008. p.44-70.
44. Imai K, King G. Misunderstandings between experimentalists and observationalists about causal inference. *JR Statist Soc A* 2008; 171(2): 481-502.
45. Berger V, Weinstein S. Ensuring the comparability of comparison groups: is randomization enough? *Controlled Clin Trials* 2004;25(5): 515-24.
46. Berger V. Part I: Is there a problem with reliability in medical studies?-Susceptibility of randomized trials to subversion and selection bias. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. 1st ed. Baltimore County: John Wiley & Sons Ltd.; 2005. p.17-35.
47. Simon SD. Is the randomized clinical trial the gold standard of research? *J Androl* 2001; 22(6):938-43.
48. Kalof L, Dan A, Dietz T. *Collecting the data. Essentials of Social Research*. 1st ed. New York: Open University Press; 2008. p.103-46.
49. Torgerson DJ, Torgerson C. *Glossary of terms. Designing Randomised Trials in Health, Education and the Social Sciences: an Introduction*. 1st ed. Basingstoke England; New York: Palgrave Macmillan; 2008. p.XI-XV.
50. Blessing LTM, Chakrabarti A. *Writing up: publishing results. DRM, a Design Research Methodology*. 1st ed. Dordrecht; London: Springer; 2009. p.215-29.
51. Sweeney C, Egger MJ. The bias coin design. *J Nurse Midwifery* 1985;30(3):175-7.
52. Nevill AM, Holder RL, Cooper SM. Statistics, truth, and error reduction in sport and exercise sciences. *Eur J Sport Sci* 2007;7(1):9-14.
53. Knapik J, Darakjy S, Scott SJ, Hauret KG, Canada S, Marin R, et al. Evaluation of a standardized physical training program for basic combat training. *J Strength Cond Res* 2005;19(2):246-53.
54. Rovers MM, Straatman H, Zielhuis GA. Comparison of balanced and random allocation in clinical trials: a simulation study. *Eur J Epidemiol* 2000;16(12):1123-9.
55. Greevy R, Lu B, Silber JH, Rosenbaum P. Optimal multivariate matching before randomization. *Biostatistics* 2004;5(2):263-75.
56. Shadish WR, Clark MH, Steiner PM. Can non-randomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *J Am Stat Assoc* 2008;103(484):1334-43.
57. Ioannidis JPA, Evans SJW, Gotzsche PC, O'Neill RT, Altman DG, Schulz K, et al. Better reporting of harms in randomized trials: an extension of the CONSORT statement. *Ann Intern Med* 2004;141(10):781-8.
58. Campbell MK, Elbourne DR, Altman DG, Grp C. CONSORT statement: extension to cluster randomised trials. *Br Med J* 2004;328(7441): 702-8.
59. Zwarenstein M, Treweek S, Gagnier JJ, Altman DG, Tunis S, Haynes B, et al; CONSORT group; Pragmatic Trials in Healthcare (Practihc) group. Improving the reporting of pragmatic trials: an extension of the CONSORT statement. *BMJ* 2008;337:a2390.
60. Trochim W, Donnelly JP. *Sampling. The Research Methods Knowledge Base*. 3rd ed. Ohio, United States: Atomic Dog Publishing; 2006. p.41-59.
61. Koenig CA, Benardot D, Cody M, Thompson WR. Comparison of creatine monohydrate and carbohydrate supplementation on repeated jump height performance. *J Strength Cond Res* 2008;22(4):1081-6.
62. Paradisi GP, Cooke CB. The effects of sprint running training on sloping surfaces. *J Strength Cond Res* 2006;20(4):767-77.
63. Altman DG. Comparability of randomised groups. *Statistician* 1985;34(1):125-36.
64. Berger V. Part II: Actions to be taken to improve the reliability of medical studies-Detecting selection bias in randomized trials. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. 1st ed. Baltimore County: John Wiley & Sons Ltd; 2005. p.123-55.
65. Berger V. Part II: Actions to be taken to improve the reliability of medical studies-Adjusting for selection bias in randomized trials. *Selection Bias and Covariate Imbalances in Randomized Clinical Trials*. 1st ed. Baltimore County: John Wiley & Sons Ltd; 2005. p.157-69.
66. Feinstein AR. Clinical judgment revisited-The distraction of quantitative models. *Ann Int Med* 1994;120(9):799-805.
67. Schrimpf D, Plotnicki L, Pilz LR. Choice and simulation of the randomization procedure for clinical trials. *Int J Clin Pharm Th* 2011;49(1): 91-2.
68. Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: a simulation study. *Clin Trials* 2011;8(1):59-69.
69. Zwiderman AH. Clinical trials do not use random samples anymore. *Clin Res Regul Aff* 2006;23(2):85-95.
70. Glaister M, Lockey RA, Abraham CS, Staerck A, Goodwin JE, McInnes G. Creatine supplementation and multiple sprint running performance. *J Strength Cond Res* 2006;20(2): 273-7.