

Randomize Kontrollü Deneyler

Randomized Controlled Trials

(Davetli Derleme)

Hacettepe Üniversitesi Hemşirelik Fakültesi Dergisi 2017, 4(1), 73-92

Belgin AKIN¹, Deniz KOÇOĞLU²

¹Selçuk Üniversitesi, Sağlık Bilimleri Fakültesi, Hemşirelik Bölümü, Halk Sağlığı Hemşireliği Anabilim Dalı, Konya

ÖZ

Bu derleme kanıta dayalı uygulama bakış açısının gelişmesiyle birlikte önemli bir araştırma türü haline gelen Randomize Kontrollü Deneylerle ilgili temel kavram ve yaklaşımların açıklanması amacıyla yapılmıştır. Randomize Kontrollü Deneylerin başarısı ve kalitesi ölçülen sonucun ortaya çıkmasında uygulanan müdahalenin gerçek etkisinin gösterilmesine bağlıdır. Bu neden-sonuç bağlantısının kurulması Randomize Kontrollü Deneylerin bir protokol çerçevesinde yürütülmesi, random atamanın yapılması, yanlılıkların önlenmesi, iç ve dış geçerliliğin sağlanması, körleme tekniklerinin kullanılması ve uygun istatistiksel değerlendirmenin yapılmasıyla olasıdır. Bu derleme bu aşamalara ilişkin yapılan hataları ve alınabilecek önlemlere ilişkin yaklaşımları içermektedir.

Anahtar Kelimeler: *Dış geçerlik, iç geçerlik, nedensellik, randomize kontrollü deneyler, yanlılık*

ABSTRACT

This review aims to explain the basic concepts and approaches related to randomized controlled trials, which is an important type of research with the development of viewpoint of evidence-based care. The success and quality of randomized controlled trials hinge upon observing the actual effect of the applied intervention on the emergence of the measured result. Establishing this cause-effect relation can be possible if randomized controlled trials are conducted within a protocol, use random assignment, avoid bias, ensure internal and external validity, use blind techniques and appropriate statistical evaluations. This review includes the mistakes made in these stages and the precaution to be taken.

Key Words: *Bias, causality, external validity, internal validity, randomized controlled trials*

GİRİŞ

Kanıtı dayalı hemşirelik, hemşirelerin en iyi bakımı sunmak için bilimsel araştırmalardan elde edilen kanıtları hasta ve sağlıklı bireylerin durumu, tercihleri ve kaynaklarının uygunluğuna göre değerlendirmesi olarak tanımlanır¹. Belirli bir girişimin uygulanabilirliğinde kanıt niteliğinde bilginin kaynağı olarak gösterilen Randomize Kontrollü Deneyler (RKD) uygulamaya yönelik kararlar verilmesinde ikinci düzeyde kanıt olarak kabul edilir ve birinci düzeyde kanıt olarak kabul edilen sistematik derleme ve meta-analizler için de temel kaynaktır²⁻³. İlk RKD Bradford A. Hill (1897-1991) tarafından gerçekleştirilmiş ve sağlık girişimlerinin etkinliğini değerlendirmede altın standart olarak kabul edilmiştir. Archie Cochrane RKD'nin sağlam bir destekçisi olmuş ve kariyerinin önemli kısmını bunun araştırmalarda kullanımının geliştirilmesine adanmıştır³⁻⁴.

Kanıtın düzeyi çalışma tasarımının türüne göre değerlendirilmekle birlikte her bir çalışmanın zayıflık ve güçlülük yönünden değerlendirilmesi gerekir. Örneğin çalışma katılımcılarının ilkelere uygun olarak seçilmediği bir RKD'nin sonuçları yanlılık içereceğinden katılımcıları random atama ilkelerine tam uyularak seçilen bir çalışmanın bulguları kadar güçlü kabul edilmez⁶. RKD'lerden elde edilen kanıtın kalitesi çalışmanın her aşamasında uygulanan metodolojik kaliteyle ve yan tutma (bias) riskinin kontrol edilme derecesi ile sıkı sıkıya bağlıdır ve yetersizliği durumunda yanlı sonuçlar elde edilmesine yol açacaktır^{7,8,9}. Deneysel çalışmada elde edilecek etkinin doğru yorumlanabilmesi çalışmanın geçerliğine ve güvenilirliğine bağlıdır ve elde edilen bulgunun kanıt sayılabilmesinde vazgeçilmezdir⁸. Hemşirelik araştırmalarının, hemşirelik girişimlerinin etkinliğini değerlendirmek üzere tanımlayıcı olmaktan ziyade açıklayıcı gücü yüksek gerçek yaşama genellenebilecek bilgiler üreten deneyleri içermesi gerektiği üzerinde durulmaktadır¹⁰⁻¹³. Mantzoukas (2009)¹⁴ hemşirelikte az sayıda çalışmada (%6) randomizasyonu içeren deneysel metodolojinin kullanıldığını belirtmekte, uygulamaya yönelik kanıtların geliştirilmesinde hemşire araştırmacıların deneysel çalışmalar konusunda kendini geliştirmesinin teşvik edilmesi ve özellikle RKD'lerin sayısının arttırılmasının gereği üzerinde durmakta ve hemşirelikte sınırlı sayıda RKD'nin nedeninin bilgi eksikliği ve deney karşıtı görüşle bağlantılı olabileceğini vurgulamaktadır. Baldi ve arkadaşları (2014)¹⁵ hemşirelik alanında yayınlanan deneysel çalışmaların son 20 yılda artış gösterdiğini ve bunların yarısından fazlasını (%67) RKD'lerin oluşturduğunu bildirmişlerdir. RKD'lerin hemşirelik yönünden önemi bu çalışmaların hemşireler tarafından kullanılan girişimlerin etkinliğini ortaya koymaya yarayacak nedensel sonuçlar çıkarma olanağı sunmasıdır¹⁶. Ancak RKD'lerin hemşirelik araştırmalarında kullanımı araştırmacılar arasında hala tartışma konusudur¹⁵. Hemşirelik biliminde RKD'lere karşı çıkışın nedeni hemşireliğin karmaşık doğası nedeniyle ana unsurlarını tanımlama ve anlamlı bir karşılaştırma yapmaya yetecek kontrol sağlamanın güçlük içermesidir. Karmaşık girişimler RKD'yi zorlaştıran pek çok faktörden birisi olmakla birlikte bu güçlük tıp ve psikoloji alanındaki pek çok çalışma için de geçerlidir ve bu soruna yönelik stratejiler geliştirilmektedir¹⁷. Hemşirelikte yer alan karmaşık girişimler randomize yöntemlerden kaçınmak yerine bunların etkisini değerlendirmeye yönelik sağlam yöntemler kullanmayı daha da önemli hale getirmektedir^{11,13}. Hemşire araştırmacıların RKD'lerde sosyal çevreyi ele almaları¹⁸ teoriye dayalı yaklaşımı kullanmaları¹³ ve karmaşık hemşirelik girişimlerini

değerlendirmeye yönelik olarak geliştirilmiş modellerden¹⁶ yararlanmaları önerilmektedir. Fridlund ve arkadaşları (2014)¹³ teoriye dayalı (%51) RKD'lerin etkililiği (effectiveness) ortaya koymada teoriye dayalı olmayanlara (%38) göre daha başarılı olduğunu bildirmektedir. Rolfe (2009)¹⁹ hemşirelikte RKD'lerin kullanılmasının sınırlılıklarını tartıştığı yazısında sorunun karmaşık girişimler yanında bireyselliğin yeterince ele alınamaması olduğunu belirtmekte ve araştırmanın elde etmek istediği bilginin toplumsal olması durumunda RKD'nin kullanılabileceğinden söz etmektedir. Tümevarımsal yöntemi savunanlarla hipotez kanıtlamaya dayalı tümdengelim yöntemini savunanlar arasında süren tartışmanın gereksiz olduğunu belirten Persson ve Sahlin (2009)² hemşirelik biliminin bu iki yaklaşımın harmanlanmasından oluştuğunu vurgulamaktadır. Hemşirelikte RKD'lere karşı olanlardan bazılarının hastaların ya da hemşirelik girişimlerinin ilaç araştırmalarındaki benzer şekilde standardize edilemeyeceğini iddiasına²⁰ karşı yanıtın pragmatik yaklaşım olduğu belirtilmektedir¹². Richards ve Hamers (2009)¹¹ eylemsel (pragmatic) RKD'nin bu karmaşık girişimlerin etkinliğini değerlendirmede en iyi seçenek olduğunu belirtmekte ve RKD'lere karşı çıkışın altında yatan nedenin RKD'de yer alan etik ve teorik süreçlerin iyi anlaşılabilmesi olduğunu belirtmektedir.

Sağlık alanındaki araştırmacılarda kanıt düzeyleri ve klinik alanda uygulanabilirliği ile ilgili farkındalık düzeyinin artmasına bağlı olarak kanıt oluşturabilecek çalışmaların gerçekleştirilmesine yönelik ilgi ve farkındalık da artmaktadır. Hemşireler de uygulamalarına rehberlik etmesi için eskiye göre literatüre daha fazla başvurmaktadır, bu nedenle araştırmacılar da uygulamayı geliştirmeye dayalı iyi tasarlanmış ve kanıt oluşturabilecek özellikte çalışmalar yapmak konusunda daha fazla sorumluluk taşımaktadırlar.

Bu derlemenin amacı RKD'lerin temel amacı doğrultusunda planlanması, yürütülmesi, verilerin analizi ve yorumlanmasında dikkate alınması gereken noktalar ve sık yapılan hataların gözden geçirilmesidir.

RKD'nin Planlanması ve Çalışma Protokolü

Başlangıçta çalışmanın nasıl yürütüleceğine ilişkin bir protokol hazırlanması en temel aşamadır. Protokolün gerekliliklerini yerine getirmede yeterli uzmanlık ve finansal destek ana unsurlardır. Deneysel çalışmanın tanımlanmasında önemli üç temel noktanın deney/girişim grubu ya da birimi, tedavi/girişim ve tedavi/girişimin değerlendirilmesi olduğu belirtilmektedir^{21,22}. Bu nedenle öncelikle çalışma katılımcılarının kimlerden oluşacağı, hangi girişim/tedavilerin hedeflendiği ve her bir katılımcının girişime/ tedaviye yanıtının nasıl ölçüleceğine karar verilmelidir.

Deneysel bir çalışmanın planlamasında aşağıdaki kavramlar göz önünde bulundurulmalıdır²³:

- *Deneysel çalışmanın etkinliğinin mi (efficacy) yoksa etkililiğinin mi (effectiveness) ölçüleceğine karar verme; Açıklayıcı (explanatory) deneyler genel olarak ideal deney koşullarında belirli bir girişimin etkinliğini (efficacy) değerlendirmeye yönelik iken eylemsel/pratik (pragmatik) deneyler (practical clinical trials, practice-based research, clinical effectiveness research) bir girişimin gerçek dünyada etkililiğini (effectiveness) değerlendirme amacını taşır ve dış geçerlik önemli bir amaçtır^{24,25}.*

- Deneysel çalışmanın kavramsal yapısına karar verme; RKD'nin amacı bir girişimin diğerinden daha üstün olduğunu göstermekse üstünlük deneyi (superiority trial), iki girişimin eşitliğini göstermekse eşitlik deneyi (equivalence trial) ve bir girişimin diğerinden daha kötü olmadığını göstermekse daha kötü olmama deneyi (non-inferiority trial) olarak adlandırılır²⁶. Bu deneylerin her birinin amacı farklı olduğundan metodolojik yönden de farklılıkları olduğu ve bu konuda farkındalığın düşük olduğu bildirilmektedir²⁷. Çalışmanın kavramsal yapısı istatistik sonuçların yorumlanması, tedavi amacına yönelik (intention-to-treat/ITT) analizlerinin yönetilmesi vb. konularda farklılıklar getirecektir²⁸.
- Çalışmanın tasarımında en uygun yaklaşımı seçme; paralel, faktoriyel (factorial), küme (cluster), çapraz (cross-over) tasarımlar vs.
- *Yanlılık riskini azaltmaya yönelik neler yapılabileceğini/yapılamayacağını düşünme*; (Randomizasyon ve körleme ve ITT)
- *Sonuç çıktılarının belirlenmesi (birincil ve ikincil çıktılar/bağımlı değişkenler)*; ölçme, geçerlik/güvenilirlik ve pratik sonuçlarla uygunluk yönünden değerlendirme
- Çalışma grubunun büyüklüğü ve istatistiksel güç için gerekli çalışmaları yerine getirme ve
- *Kullanılacak istatistik çözümlenmeleri planlama.*

Çalışma Protokolü

Klinik bir deneyin protokolü; çalışmanın planlanması, yürütülmesi, raporlama ve uygulamaya aktarılması gibi konularda temel rol oynar. İyi yazılmış bir protokol deney başlamadan önce deneyin yürütülmesinin sağlamlığı, uygunluğu, deney tamamlandıktan sonra sonuçların uygulanması gibi bilimsel, etik ve güvenlik konuları konusunda değerlendirme yapmayı sağlar²⁹. Deneysel bir çalışma protokolünün ana unsurları; özet, giriş, araştırmanın amaçları, tasarım ve girişimin detayları, kimlerin çalışmaya alınacağı, dahil etme ölçütleri, randomizasyon, ölçme/değerlendirme ve veri toplama, istatistikler, etik konular, örgütsel yapı, yayın hedefleri ve ekler başlıklarını içerecek şekilde oluşturulmalıdır²⁹. Klinik çalışma protokolünde yer alması gereken standart gereklilikleri açıklayan uluslararası bir ekip tarafından oluşturulmuş SPIRIT 2013 (Standart Protokol Maddeleri: Girişimsel Deneyler için Öneriler) protokol hazırlayacak araştırmacılar için bir rehber olarak önerilmektedir³⁰⁻³¹. Chan ve arkadaşları (2013)³⁰ klinik deney protokollerinde önemli eksiklikler olduğu ve içeriklerinin yer alması gereken bilgiler konusunda yeterince açık olmadığını bildirmektedir. Gray ve arkadaşları (2016)³² bir hemşirelik dergisinde yayınlanan çalışmaların yalnızca dördünün (%9) tam protokollerinin yayınlandığı ve diğerlerinin ise birisi hariç çalışma protokollerinin bulunmadığını bildirmektedir.

RKD İçin Temel Kavramlar

RKD' lerde Nedensellik

Araştırmalar öncelikle iki değişken (bir girişim/tedavi/maruziyet ve bir sonuç/çıktı) arasındaki ilişkiyi ölçmeyi amaçlar. Bu birisi girişimi alan (çalışma/deney grubu)

ve diğeri girişimi almayan (kontrol grubu) iki grubun karşılaştırıldığı bir tasarım ile başarılabilir. Karşılaştırmanın ardından sonuç (çıkıtı) karşılaştırılır. Ancak bir girişimin etkisinin değerlendirilmesinde en önemli gerekliliklerden birisi girişim ve kontrol grubunun uygulanan girişim dışında kalan diğer tüm özellikler (karıştırıcılar) yönünden benzer olmasıdır. Karıştırıcı (confounder) üzerinde çalışılan girişim dışında kalan ancak ölçülmeye çalışılan sonucu (çıkıtı) etkileyebilen değişken olarak tanımlanabilir. Araştırmadaki her çıkıtı için deneyin planlama aşamasında göz önünde bulundurulması gereken çok sayıda karıştırıcı vardır. Karıştırıcı değişkenlerin her iki grupta da (girişim ve kontrol) eşit/dengeli şekilde dağılması gerekir³³. İyi bir randomizasyonun bunu sağladığı düşünülse de grupların atanma aşamasında basit rasgele yapılan randomizasyona ilişkin eleştiriler de bulunmaktadır³⁴. En temel eleştiri ise deney ve kontrol grubunda sayı olarak ortaya çıkabilecek dengesizliktir. Bunun kontrol altına alınmasında blok rastgele yöntemin tercih edilebileceği belirtilmektedir. Ayrıca özellikle bireylerin bazı sosyodemografik ya da hastalık özellikleri deneyin sonucu üzerinde etki edebilecek etkiye sahipse tabakalı rastgeleleştirme yöntemi önerilebilir. Bu aşamada tabaka sayısı artıçça random atamanın bozulabileceği dikkate alınarak tabaka sayısı minimum düzeyde tutulmalıdır³⁵. Deney ve kontrol grupları oluşturulduktan sonra randomizasyonun sağlanıp sağlanmadığının istatistiksel yöntemlerle gösterilmesi de önemlidir ve istatistik paket programlarının birçoğu bu tür analizler için kolaylaştırıcıdır³⁶. Nedenselliğin ikinci şartı deney ve kontrol grubunda zamana karşı meydana gelen değişimin müdahaleden kaynaklandığının açıkça ortaya konabilmesidir. Bunun sağlanabilmesi, çalışmalarda körleme tekniklerinin kullanılması, araştırma konusunun uygun olması koşuluyla faktoriyel (factorial), küme (cluster), çapraz (cross-over) tasarımlarından birisinin tercih edilmesiyle olası hale gelebilir³⁷. Nedenselliğin gösterilmesinde önemli olan diğer bir unsur ise istatistik değerlendirmelerin kapsamıdır. Etki büyüklüğü sonuçlarının verilmesi, değerlerin %95 güven aralığı içinde sunulması, p değeri 0.05 büyük ya da küçük ifadeleri yerine değerlerin açık yazımı, verilerin özelliklerinin uygun olması durumunda cox regresyon, Manova, multivariate regresyon, tekrarlayan ölçümlerde çift yönlü varyans analizi gibi değerlendirmelerin yapılması önemlidir^{36,38,39}. Nedensellik kararını etkileyebilecek durumların kontrol altına alınmasında araştırma sürecinde ortaya çıkabilecek yanlılıkların önlenmesi de önemlidir.

Yanlılık (Bias)

Klinik deneylerde hata tümüyle rastgele bir hata (değişkenlik-variability) ve sistematik bir hata (yanlılık-bias) olmak üzere iki şekilde olabilir. Örnek büyüklüğünü arttırmak ya da deneyi tekrarlamak ölçüm ortalamalarındaki değişkenliği azaltarak tahminlerdeki doğruluğu arttırabilir ve böylece sistematik hatalar önlenir. Yan tutma diğer tüm rastgele değişkenlik nedenlerine bağlı olarak gerçek deney etkilerinden sapmaları tanımlar. Yan tutma bir klinik deney sırasında deneyin tasarımı ve uygulanmasından veri analizine ve sonuçların yorumlanmasına kadar pek çok noktada ortaya çıkabilir. Yanlılık nedenlerini iyi bir şekilde anlamak yanlılığı azaltmak ve akla uygun bir etki tahmini yapmakta oldukça önemlidir^{9,40,41}. Yanlılık küçük ya da büyük çapta olmak üzere çeşitli düzeylerde olabilir ve farklı yan tutmalar çoğunlukla sonucun gerçekte olduğundan daha olumlu ya da bazen olumsuz yönde ortaya çıkmasına yol açabilir.

Bilimsel gerçeklik asla tam olarak değerlendirilemeyeceğinden olası yanlışlık riski ifadesini kullanmak daha doğru kabul edilmektedir^{8,42}. Yanlılık oldukça geniş kapsamlı ele alınabilir ve 35 den fazla yanlışlık türü olabileceği bildirilmektedir. Ancak bir çalışmanın sonucunu önemli oranda etkileyebilecek temel yanlışlıklar vardır ve çalışmalarda özellikle bunların kontrol edilmesi beklenir³³.

Temel yanlışlık kaynaklarından söz etmeden önce yanlışlık tartışmalarında sık yer almayan ancak özellikle deneysel çalışmalar ve bu çalışmalardan üretilen sistematik derleme ve meta analizler için önem taşıyan bir yanlışlık kaynağı olarak yayın yanlılığından (publication bias) söz etmek gerekir. Özellikle sonucu hipotezi destekleyecek şekilde olumsuz çıkan çalışmaların araştırmacılar tarafından daha az yayınlama eğiliminde olduğu ve dergi editörlerinin de böyle çalışmaları genelde yayınlamaya değer görmediği bazen de çalışmanın tamamlanmadığı ve bu nedenlerle yayınlanan çalışmaların gerçeği yansıtmada konusunda yanlışlık riski taşıdığından söz edilmektedir⁴³. Bunun yanında küçük örneklerle elde edilen yüksek etki büyüklüğü de yayın yanlılığının diğer önemli bir boyutu olarak belirtilmektedir^{44,45}. Bu nedenlerle yayınlanmış çalışmalardan sentezlenerek sonuçlar çıkarılmasının da yanlışlık içerdiği vurgulanmaktadır⁴⁴. Tüm bunlar göz önünde bulundurulduğunda deneysel çalışmalarda araştırma probleminin belirlenmesi ve sonuçların yorumlanması da yanlışlık riski taşıyacaktır.

Son yıllarda deneysel çalışmaların tercihen planlama aşamasında (ya da bazen araştırma raporu tamamlandıktan sonra) açıkça ulaşılabilecek şekilde bu amaçla oluşturulmuş web sitelerine kaydedilerek kayıt numarası alınması beklenen bir durum olarak karşımıza çıkmaktadır^{29,46}. Klinik araştırmaların kaydedilmesinin yayın yanlılığının azaltılması yanında sistematik derleme ve diğer literatür incelemelerine kolaylık sağlama, olası katılımcılar ve klinik çalışanlarına bilgi sağlama ve çalışma sonuçlarının standardize bir formatta herkese açık sunulmasına katkı sağlaması gibi yararlarından söz edilmektedir⁴⁷. Dünya Sağlık Örgütü Uluslararası Klinik Deneyler Kayıt Platformu (International Clinical Trials Registry Platform-ICTRP) tüm klinik deneylerin kaydedilmesinin bilimsel, etik ve moral bir sorumluluk olduğunu⁴⁹ ve Uluslararası Tıbbi Dergi Editörleri ise klinik deneylerden elde edilen sonuçların yayınlanabilmesi için bir gereklilik olduğunu bildirmektedir⁴⁶. Klinik deneylerin kaydedilmesi ilaç araştırmaları için ise bir zorunluluk olarak tanımlanmaktadır⁴⁹. Kayıt numarası RKD raporlarının yazım denetimi için geniş kabul gören (CONSolidated Standards of Reporting Trials) CONSORT⁵⁰ gibi kontrol listelerinde de yer almakta ve pek çok dergi çalışmaların dergiye kaydedilmesinde aynı şeyi beklemektedir⁵¹. Yayınlanmış hemşirelik araştırmalarının değerlendirildiği çalışmalarda incelenen araştırmaların çok azında kayıt numarasının yer aldığı bildirilmiştir^{51,52}. Bir hemşirelik dergisinde 2011-2016 yılları arasında yayınlanan çalışmaların üçte ikisinden fazlasının kayıtlı olmadığı, kayıtlı olanların %54' ünün çalışma başlatıldıktan sonra kaydedildiği ve yalnızca %9' unun tam protokolünün yayınlandığı belirtilmektedir³². Bir çalışmada⁵¹ incelenen 15 hemşirelik dergisinden yedisinin CONSORT değerlendirmesini beklediği bunlardan üçünün deneylerin kaydedilmesini talep ettiği belirtilmektedir. Editörler, hakemler ve yazarların CONSORT kriterlerinin farkında olması ve kayıtlı olmayan çalışmaların yayınlanmamasının önemi üzerinde durulmaktadır^{32,51}. Bu bilgiler hemşire araştırmacıların bu sorunun farkındalığı konusunda yetersizlikleri olduğunu düşündürmektedir.

İç Geçerlik (Internal Validity)

Bir RKD'nin sonuçlarının gerçeği yansıtma düzeyi olası yan tutma kaynaklarından ne kadar uzak kalabildiğine bağlıdır. RKD'nin yanlılıklardan uzak olması iç geçerliliğinin (internal validity) yüksek olduğu yönünde bir güvencedir. Bir sağlık girişiminin gerçek etkisi bilinmeyen bir durumdur ve belirli bir örnek katılımcıdan daha geniş bir hedef nüfusa genellenebilir sonuçlar elde etmek için yanlılığın tahmin edilmesi, saptanması ve kontrol edilmesine çalışılır^{8,53}. Araştırma probleminin oluşturulmasından yayınlanma sürecine kadar pek çok yanlılık kaynağı olmasına rağmen temel olarak yanlılıklar katılımcıların çalışma grubuna atanmasından, girişimlerin gerçekleştirilmesine ve sonuçların/çıktıların ölçülmesine kadar bir deneyin gerçekleştirilmesi sürecinde ortaya çıkar^{8,53,54}. RKD'lerin kalitesini değerlendiren çok sayıda araç bulunmaktadır. Kalitenin değerlendirilmesinde göz önünde bulundurulmuş alanlar/özellikler; giriş, hedefler, tasarım, çalışma grubunun seçimi, atanması, randomizasyon ve gizlenmesi, körleme, girişimler, veri kaybı, izlemler/protokole uyumsuzluk, çıktılar ve istatistikler olduğu belirtilmektedir⁵⁵. Kalite değerlendirme araçlarının ağırlıklı olarak iç geçerliliğe odaklandığı ancak etik boyut vb. farklı alanları da değerlendirebildiği bu nedenle yanlılık riski değerlendirmesinin çalışmanın kalitesinin değerlendirilmesinden farklı olduğu bildirilmektedir. Kalite değerlendirmesinin çalışmanın yürütülmesi ile ilgili olduğu ve puanlamayı içerebileceği geçerliliğin ise çalışmanın araştırma sorularını yanıtlama yeterliliği ile ilgilendiği vurgulanmaktadır^{8,56}. Araştırmanın yanlılık riskinin iç ve dış geçerlik boyutuyla değerlendirildiği Cochrane Collaboration yanlılık riski değerlendirme aracı (Cochrane Collaboration's tool for assessing risk of bias)⁸ yanlılığın değerlendirilmesinde geniş kabul gören ve önerilen bir araçtır⁵⁷. Hemşirelikte yapılan sistematik derlemelerde çalışmaların değerlendirilmesinde yan tutmadan ziyade çalışmaların kalitesinin değerlendirildiği bildirilmekte ve Cochrane Collaboration yanlılık riski değerlendirmesinin kullanımı önerilmektedir⁵⁸. Bu yazıda da yanlılık riski konusunun özetlenmesinde Cochrane Collaboration yanlılık riski değerlendirmesi esas alınmıştır.

Tüm olası yanlılık kaynakları için yanlılığın olası büyüklüğü ve yönü göz önünde bulundurulmalıdır. Bazı durumlarda olası yanlılıklara rağmen elde edilecek kanıtlar girişimin etkili olduğunu gösterebilir. Eğer çalışmanın tüm metodolojik boyutlarında yanlılık söz konusu ise sonuçta ortaya çıkacak etkinin gerçeği yansıtma olasılığı da ortadan kalkabilecektir. Bu nedenlerle çalışmanın sonuçlarının gerçeği yansıtmasını güvenceye almak ve kanıtlamak için olası yanlılık kaynakları ve bunların nasıl kontrol edileceği titizlikle planlanmalı, yürütülmeli ve araştırma raporunda da açıkça ifade edilmelidir⁴⁰. RKD'lerle ilgili çeşitli yanlılıklar özellikle çalışma grubunun seçimi, ölçümler ve analizler aşamasında ortaya çıkabilir^{40,54}. Cochrane yanlılık riski değerlendirmesinde yanlılıklar çalışma grubunun seçimi sırasında seçim yanlılığı (selection bias), çalışmanın yürütülmesinde performans yanlılığı (performance bias), ölçümler sırasında saptama yanlılığı (detection bias), analiz aşamasında eksilme yanlılığı (attrition bias) ve araştırma raporunun yazılması sürecinde ortaya çıkabilecek raporlama yanlılığı (reporting bias) olmak üzere altı başlıkta toplanır ve her bir yanlılıktan kaçınmak için geliştirilmiş çeşitli yöntemler bulunmaktadır⁸.

Seçim yanlılığı ve Randomizasyon (randomization):

Çalışma gruplarının belirlenmesi aşamasında ortaya çıkabilecek seçim yanlılığı (selection bias) karşılaştırılan çalışma gruplarının temel özellikler yönünden farklılığı anlamına gelir. Kimin deney kimin kontrol grubunda yer alacağına karar verilmesi sürecinde örneğin girişime daha etkin yanıt vereceklerin ya da uyumsuz olacağı düşünülenlerin bilerek ya da bilmeyerek ayrılması söz konusu olabilir. Seçim yanlılığını kontrol etmenin kabul edilen etkili yöntemi randomizasyondur^{8,53}. Randomizasyonun temel üstünlüğü, başarılı bir şekilde gerçekleştirildiğinde katılımcıların gruplara atanmasında seçim yanlılığını önlemesidir^{22,43}. Gerçek bir randomizasyonla tüm katılımcılara çalışma gruplarına atanmada eşit şans verilir. Randomizasyonun başarısı bazı ilişkili adımların başarıyla gerçekleştirilmesine bağlıdır. Ancak en iyi şekilde yapılmış bir randomizasyonda bile seçim yanlılığı önlenmiş sayılmaz. Çünkü katılımcıların çalışma sürecinde çalışmaya katılma onamından vazgeçmesi ya da başka şekillerde çalışmadan ayrılması nedeniyle çalışma guruplarının benzerliği bozulabilir. Bu nedenle çalışmada analiz aşaması da dahil randomizasyonun etkisinin sürdürülmesi yönünde önlemler almak gerekecektir⁴³.

RKD'lerde randomizasyondan söz etmeden önce random seçme (random selection) ve random atamanın (random allocation) aynı şeyler olmadığından söz etmek gerekir. Random seçmede belirli bir evrenden seçilecek kişilere eşit şans verme söz konusudur ve çalışmanın evrene genellenebilirliği ile ilgilidir ve dış geçerliliği (external validity) arttırma yönünde bir etkisi vardır. Random atamada ise belirlenen çalışma grubunun girişim ve kontrol gruplarına atanmasında eşit şans verilmesinden söz edilir ve çalışmada ortaya çıkacak seçim yanlılığını önleyerek çalışmanın iç geçerliliğini (internal validity) arttırır. Randomizasyon RKD'lerde en temel gereklerden birisi olarak kabul edilir ve eksikliğinde çalışmanın kalitesi ve içsel geçerliliği büyük oranda etkilenir. Randomize olmayan karşılaştırmalı çalışmalardan elde edilen bilgi randomize çalışmalardakine göre daha zayıf bir kanıt olarak değerlendirilmektedir^{9,43}. Randomizasyonun doğru bir şekilde uygulanmadığı çalışmalarda girişim etkisine ilişkin daha yüksek/abartılı (%30-40) tahminler elde edildiği bildirilmiştir⁵⁹.

Randomizasyonun etkili olabilmesi için iki önemli adımın gerçekleştirilmesi beklenir. Birincisi rastgele atama (random allocation) ve ikincisi ise atama sürecinin çalışmada yer alan araştırmacılardan gizlenmesidir (allocation concealment). Random atama oluşturmada önerilen yöntem rastgele sayılar tablosu ya da bilgisayar ortamında oluşturulmuş rastgele sayılar vb. bir sürecin kullanılmasıdır. Katılımcılardan onam alındıktan ve dahil etme kriterleri değerlendirildikten sonra randomizasyon için geliştirilmiş yöntemlerden (basit randomizasyon, blok randomizasyon vb.) çalışma için en uygun olanın kullanılması ve bu sürecin tamamen araştırmacılar dışındaki kişiler tarafından yürütülmesi önemlidir. Doğum tarihi, başvuru sırası, kayıt numarası gibi yöntemler seçim için uygun olmayan yöntemler olarak tanımlanır. Atamanın gizlenmesi sürecinde merkezi bir atama ofisi ya da kapalı opak zarflar vb. yöntemlerin kullanılması önerilir. Rastgele sayıların açık bir tablosu, vaka kayıt sırası, doğum tarihi vb. açık listelerin yer aldığı yöntemler yanlılık riskini arttıran ve uygun olmayan teknikler olarak kabul edilir^{8,22,33,42,60,61}. Çalışmayı yürütenlerin ve katılımcıların çalışma başlayana kadar atama sürecinde kimin girişim kimin kontrol grubunda olduğunu bilmemesi esastır^{8,9,22,33,61}. Bu gizleme süreci kör teknik olarak adlandırılabilir,

ancak körleme (blinding) ile karıştırılmamalıdır^{60,62}. Körleme çalışmanın yürütülmesi sırasında ve raporlama sürecinde çalışma içinde yer alanların körlendiği bir kontrol yöntemidir.

Performans ve saptama yanlılığı ve Körleme (blinding/masking)

Çalışmanın yürütülmesi ve ölçümler aşamasında ortaya çıkabilecek yanlılıklardan birisi performans yanlılığı (performance bias) olarak adlandırılır ve çalışma girişimlerini gerçekleştirenlerin ve çalışma katılımcılarının bilerek ya da bilmeyerek çalışma sonucunu etkilemesi olarak tanımlanabilir. Örneğin çalışma girişimlerini yürüten kişi girişim grubu lehine davranabilir (rosenthal effect) ya da hangi girişim grubunda olduğunu bilen kişi beklenen yönde davranışlar (hawthorne effect) sergileyebilir. Saptama yanlılığı (detection bias) performans yanlılığına benzer bir etki ile çıktı ölçümlerini yapan kişinin ölçüm yaptığı gruptan beklenen etki yönünde değerlendirme yapması olarak tanımlanabilir^{8,42}. Performans ve saptama yanlılığı ortak bir ifade ile tespit yanlılığı (ascertainment bias) olarak da adlandırılabilir⁵³. Randomizasyonla gruplar arasındaki farklılıklar azaltılmış olsa bile sonrasında bu grupların yanlı tahminlere yol açacak şekilde farklı girişimler almaları ya da bu girişimlerin etkisinin ölçülmesinde farklılıklar olmasının garantisi değildir. Çalışma içinde yer alan kişilerin çalışma sonucunu etkileyecek şekilde yanlı davranması riskinin en aza indirilmesi RKD'lerde aranan durumlardan birisidir. Bu tür yanlılıkların önlenmesinde etkili olduğu kabul edilen teknik körleme (blinding/masking) tekniğidir. Körleme RKD'lerde araştırma çalışmasının içinde yer alan bir ya da daha fazla kişinin çalışma gruplarından hangisinin deney hangisinin kontrol grubu olduğunu bilmemesini içerir^{8,9,22,54,60,62}. Körlemenin randomize olmayan yarı deneysel çalışmalarda da kullanılmasının yanlılıkları azaltmada önemli olduğu belirtilmektedir¹².

Körlemenin çalışmanın özelliğine göre olabildiğince çok yönlü olarak uygulanması RKD' nin kalitesi ve içsel geçerliliği yönünden önemli bir kanıt olarak kabul edilir. Bu nedenle çalışmalarda mümkün olduğunca fazla kişinin körlenmesi şiddetle önerilen bir yöntemdir⁶². Araştırmacılar raporlarında çalışmada kimlerin, nasıl körlendiğini ve bunun nasıl denetlendiğini açıkça ifade etmeli ve kimlerin körlenemediğini ve nedenlerini açıklamalıdır^{54,60,63}. Körleme yöntemi girişimi ve ölçümleri uygulayanlar, katılımcılar ve istatistikçinin körlendiği üçlü/dörtlü körleme, girişimi/ ölçümleri uygulayanlar ve katılımcıların körlendiği ikili körleme ve yalnızca girişimi/ ölçümleri uygulayanların ya da katılımcıların körlendiği tekli körleme yöntemleri ile yapılabilir^{8,42,60,62}. İkili ve üçlü körlemeler çalışmada tespit yanlılığının en düşük düzeyde olduğu güçlü bir yöntem olarak kabul edilir⁶². Çift körleme yapılmayan çalışmalarda girişim etkisine ilişkin daha yüksek/abartılı (%17) tahminler elde edildiği bildirilmiştir⁵⁹. Yanlılık istatistik analizler aşamasında da ortaya çıkabileceğinden tüm analizler tamamlanana kadar istatistiksel analizleri yapan kişilerin de körlenmesi önerilen bir durumdur^{8,22,42,60,62}. Açıklayıcı deneylerde (explanatory trial) çiftli körleme ve plasebo kontrollü yaklaşımın kullanılması gerektiği belirtilmekle birlikte eylemsel/pratik deneylerde (pragmatic trial) körlemenin ve plasebo uygulamasının etik, pratik vb. yönden mümkün olmayabileceğinden söz edilmektedir²². Hemşirelik mesleğinin amacı laboratuvar sınırlarının ötesinde gerçek dünyada uygulamayla ilişkili bilgi üretmek olduğundan açıklayıcı (explanatory) olmaktan ziyade eylemsel (pragmatic)

yaklaşımın hemşirelik uygulamasını daha doğru şekilde yansıtaacağı belirtilmektedir⁶⁴. Bazı yazarlar da hemşirelikte açıklayıcı (explanatory) çalışmaların eksikliğinden ve öneminden söz etmekte¹² ve hemşirelik araştırmalarında laboratuvar çalışmalarının yeterince kullanılmayan ve önemsenmesi gereken bir alan olduğundan söz etmektedir¹⁹. Plasebo'nun açıklayıcı deneylerde kullanılması ile ilgili bile literatürde etik temelli çok sayıda tartışma yapılmakla birlikte⁶⁵ son zamanlarda eylemsel deneylerde de plasebo kullanımının başladığını görmekteyiz. Bu çalışmalarda genellikle deney, kontrol ve plasebo gruplarında ön - son test değerlendirmesine dayalı bir desen oluşturulduğu ve deney grubuna müdahale uygulanırken plasebo grubuna da başka bir programın uygulandığı yaklaşımlar olduğu görülmektedir. Özellikle; araştırmada, araştırmacının varlığının veya deneğin gözlemlendiğinin farkında olması nedeniyle, deneğin araştırma sonucuna uygun davranışlarını değiştirmesi olarak ortaya çıkan Hawthorne etkisi'nin⁶⁶ kontrol altına alınmasında plasebo grubunun olması önemli olabilir. Böylece sonucun uygulamadan mı yoksa deneğin gözlemlendiğinin farkından olmasından mı kaynaklandığı ayırt edilebilir. Hemşirelik araştırmalarına plasebo uygulamalarının nasıl entegre edileceği düşünülmelidir ve böylece hemşirelik çalışmalarında körlemenin yeterli düzeyde uygulanamaması nedeniyle ortaya çıkan dezavantajlar kontrol altına alınabilir. Eylemsel (pragmatic) deneylerde körleme genelde zor (etik, pratik vb. nedenler) hatta bazen olanaksız olabileceğinden çıktılar değerlendirilen²⁴ ve veri analizini gerçekleştirenlerin⁶⁷ körlenmesi önerilebilecek bir yöntemdir. Hrobjartsson ve arkadaşları (2013)⁶³ özellikle ilaç dışı araştırmalarda yaygın olarak kullanılan çıktıyı değerlendirenlerin körlenmesinin etki büyüklüğü tahminlerindeki yanlılığı azaltmada yararlı olduğunu bildirmektedir. Çıktı ölçümlerinin subjektif olması durumunda çıktıyı değerlendirenlerin körlenmesinin en uygun yöntem olduğu belirtmektedir^{12,22,60}. Hemşirelik çalışmalarının çıktı ölçümlerinin çoğunlukla subjektif olduğu göz önünde bulundurulduğunda çıktı ölçümlerini yapanların körlenmesinin uygulanabilir bir körleme yöntemi olması yanında tercih edilmesi gereken bir yöntem olduğunu söyleyebiliriz. Eylemsel (pragmatic) çalışmalarda dış geçerliliği arttırmak önemli bir hedef olmakla birlikte iç geçerliliğin düşük olmasının dış geçerliliği etkileyeceği unutulmamalıdır. Treweek and Zwarenstein (2009)²⁵ iç geçerliliği olan ancak uygulanabilirliği düşük ya da bunu değerlendirmeye yetecek nitelikte raporlanmamış deneylerin klinik uygulamanın etkililiği ile ilgili kaçırılmış bir fırsat olduğunu ve pragmatik bakış açısının tüm denemelerde olması gerektiğini vurgulamaktadır. Bunun yanında hemşire araştırmacıların klinik alana uygulanabilir sonuçlar üretmesinin iç ve dış geçerlik arasında bir denge sağlamakla başarılabilirliği belirtilmektedir⁶⁸.

Eksilme yanlılığı ve tedavi amacına yönelik analiz (Intention-to-treat analizi-ITT)

Çalışma sırasında bazı katılımcılar çeşitli nedenlerle (çalışmaya katılmaktan vazgeçmesi, çalışmaya uyum göstermemesi, sağlık sorunları vb.) çalışmadan ayrılabilir ve çalışma sonlandığında bu kişilere ait verilerde eksiklik olur. Randomizasyon sonrasında çalışma sürecinde sonuç/çıktı ölçümlerinde ortaya çıkan veri eksikliğine bağlı yanlılık eksilme yanlılığı (attrition bias) olarak adlandırılır^{8,42}. Daha önce söz edildiği gibi seçim yanlılığını önlemeye yönelik olarak gerçekleştirilen randomizasyon sonucu elde edilmiş olan denge (grupların önemli özellikleri yönünden benzerliği) gruplardan eksilen kişiler

olması nedeniyle bozulmuş olacaktır⁶⁹. Randomizasyonun etkisinin sürdürülmesinde ve eksilme yanlılığının önlenmesinde önerilen temel yöntem tedavi amacına yönelik analizdir (intention-to-treat analiz) (ITT)^{8,69}. ITT analizi randomizasyon sonrasında ortaya çıkan ayrılma, uyumsuzluk, tedavi/girişimi alıp almama vb. hiçbir gerekçeyi dikkate almadan her katılımcının atandığı grupta analizlere dahil edilmesi olarak tanımlanır. Temel ilkesi “nasıl randomize ettiysen öyle analiz yap” (as randomized so analysed) şeklinde özetlenebilir. ITT’ nin temel avantajı randomizasyon dengesini korumak, yan tutmalardan kaçınmak ve gücü arttırmaktır. Dezavantajları girişimin etkisini azaltabilmesi ve eksik verilerin tamamlanmasını gerektirmesidir^{41,70}.

Girişimi almayan ya da tamamlamayan kişinin girişimin sonuçları yönünden değerlendirmeye alınmasının önemli gerekçeleri arasında toplumda/gerçek yaşamda her zaman uyumsuz bireyler olacağı, çalışmadan ayrılmanın bizzat girişimin kendisinden kaynaklanan olumsuz etkiler ya da yararını görmeme gibi nedenlerden kaynaklanabileceği ve çalışmadaki kayıpların aynı zamanda çalışmanın etkili bir uygulama olup olmadığı ile ilgili bir durum olabileceği yer almaktadır. Bu nedenlerle çalışmanın planlanmasında çalışma grubundan kayıpları önleyecek/kontrol edecek yöntemler geliştirilmesi birinci önlem olmalıdır ancak özellikle uzun süreli çalışmalarda kayıpların olması neredeyse kaçınılmazdır^{41,71}. Deneysel bir çalışmada kayıpları %10⁷² ya da %15’ in altında tutmak⁴¹ yanlılığı azaltmada önemli kabul edilmekte ve kayıplara bağlı yanlılığı kontrol etmede örnek büyüklüğünü arttırmanın yararlı olduğu vurgulanmaktadır⁷³. Eksik verilerin şekli ve dengesi de önemlidir. Örneğin deney grubundan daha fazla kayıp varsa (dengesiz eksilme) ve çalışma körlenmemişse deney grubundakilerin çalışmayı bırakma nedenleri girişimden kaynaklanan memnuniyetsizlik ya da işe yaramadığı inancı olabilir. Bu ihmal edilmemesi gereken bir eksik olarak adlandırılır. Belirli bir duruma bağlı ya da belirli bir temel özelliğe (örn.yaş) sahip katılımcılar ayrıldıysa bu özellikle değerlendirilmesi gereken bir kayıp şeklidir^{41,55}. Çalışmadaki tüm katılımcıların analize dahil edilebilmesi için verilerinin eksiksiz olması gerekir. Çalışma grubundan kayıplar olması durumunda ortaya çıkan en önemli sorunlardan birisi eksik verilerin nasıl yönetileceğidir. Eksik verilerin tamamlanması hipotez testinin istatistiksel gücünün arttırılmasında yararlı bir yöntem olarak kabul edilmektedir^{55,74}. Araştırmacıların eksik verileri tamamlamak üzere çok sayıdaki istatistik stratejiden en az birisini kullanması gerekir. Longitudinal mix model, cox regresyon gibi istatistik modeller, son gözlemin kaydedilmesi (last observation carried forward), en iyi ya da en kötü değerlerin kullanılması (extreme case analysis) ve ayrıca eksik değerleri tamamlamada farklı senaryolarda çok sayıda analiz setlerinin gerçekleştirilmesini içeren duyarlılık analizleri (sensitivity analysis) gibi yöntemlerden en uygun olanını kullanmak gerekir. En iyi iki yöntemin çoklu tamamlama (multiple imputation) ve maksimum olası tamamlama (maximum likelihood imputation) (Expectation-maximization imputation) yöntemleri olduğunu bildirmektedir^{55,74}. Duyarlılık analizinin ise çalışma sonuçlarının sağlamlığını saptamada özellikle dengesiz kayıp durumunda oldukça yararlı olduğu belirtilmektedir^{39,41}.

Eksik veri tamamlamada kullanılacak yöntemin seçiminde kayıp verinin miktarı, denge durumu, elde bazı verilerin olup olmaması ve çalışmanın modeli (superiority, equality, non-inferiorty) gibi özellikleri göz önünde bulundurulmalıdır. Örneğin kayıp veri miktarının büyük olması durumunda ve çalışma superiority modelde ise en iyi-en kötü değer yaklaşımının sorun oluşturabileceği, H_0 hipotezini reddetme olasılığı ve

Tip 1 hata eğilimini artıracak şekilde belirtilmektedir⁷⁵. Farklı istatistiksel analitik stratejilerin kullanılmasının etki tahminlerinde önemli bir farklılık oluşturmadığından söz edilmekle birlikte⁷⁶ eksik verileri yönetmede var olan çok sayıda yöntemin etki tahminlerinde oldukça farklı sonuçlar verdiği özellikle vurgulanmaktadır⁷⁴. En uygun yöntem karar verilmesinde en belirleyici nokta olarak gruplardan kayıpların rastgele ya da rastgele olmayan şekilde olması belirtilmektedir⁴¹

ITT analizinin fazla tedbirli ve Tip 2 hata eğilimini arttırıcı olduğu ve özellikle açıklayıcı (explanatory-efficacy) çalışmalarda randomize edilen kişilerin girişimi/tedaviyi almadığı durumda girişimin gerçek pozitif etkisini azalttığı ve bu durumun eylemsel (pragmatic-effectiveness) çalışmalardan daha önemli olduğu bildirilmektedir⁷⁰. ITT analizine alternatif yöntemler olarak uyarlanmış/modified ITT (MoITT) ve per-protokol (PP) analizlerinden söz edilebilir. MoITT çalışmanın başında hiçbir girişim almadan çalışmadan ayrılanların analizden çıkarılmasına dayanan bir yöntemdir. Per-protokol (PP) analizi yalnızca gerçekten girişimi alanların analize dahil edildiği yöntemdir ve temel avantajı gerçek deney performansını yansıtmasıdır ve özellikle kayıpların tümüyle rastgele olmadığı durumlarda önemli düzeyde yanlışlık riski taşımaktadır^{41,75}. Bu nedenle PP analizi ve ITT nin birlikte kullanılması sonuçların sağlamlığını kontrol etmede önerilen bir yöntemdir⁵⁵. Superiority çalışmalarda ve pragmatik yaklaşıma dayalı klinik deneylerin analizinde intention to treat analizi büyük önem taşır⁷⁷. ITT'ye alternatif uygulamalar olmakla birlikte önemli oranda yanlışlık içerdiğini unutmamak gerekir⁷⁸. Eksik verilerle baş etmenin en iyi yolunun eksik veri olmamasına çalışmak olduğu ve deneyin tasarım, uygulama, analiz ve raporlama aşamalarında bu durumun göz önünde bulundurulması gerektiği vurgulanmaktadır^{39,79}. ITT analizi RKD'lerde sıklıkla kullanılan bir terim olmakla birlikte araştırma raporlarında gerçek anlamda kullanımında sorunlar olduğu, yapılan analizler ve kişileri analize katmama nedenleri, oranları ve eksik verinin nasıl tamamlandığı gibi konularda önemli eksiklikler olduğu bildirilmektedir^{79,80,81}. Araştırmacıların kayıplar, nedenleri ve eksik verilerin ele alınmasında kullanılan istatistik yöntemleri açıkça ifade etmeleri gereklidir⁵⁵. Çalışma katılımcılarından kayıplar olması durumunda katılımcıların temel özelliklerinin randomizasyon öncesi ve sonrası tablo olarak verilmesi gerektiği ancak çoğu çalışmada buna yer verilmediği belirtilmektedir. Duyarlılık analizlerinin çok az ve yetersiz biçimde uygulandığı bu durumun yanlış sonuçlara neden olduğu vurgulanmaktadır⁸². Ağrı ile ilgili RKD'lerin değerlendirildiği bir sistematik derlemede⁸¹ çalışmaların % 47'si ITT ve % 5'i MoITT kullandığını bildirmiş ve % 67'si analiz dışı bırakma nedenlerinden söz etmiştir. Hemşirelik araştırmaları ile ilgili bir çalışmada incelenen çalışmaların % 14,5'inde ITT ve bir çalışmada da MoITT kullanılmış olduğu bildirilmiştir⁸⁰. Gravel ve ark. (2007)⁸³ ITT kullandığını belirten makalelerin yalnızca %39'unda tüm katılımcıların randomize edildiği şekilde ve sayıda analize katıldığı ve tüm katılımcı ifadesinin ulaşılabilen tüm katılımcılar olarak algılandığını belirtmektedir. LANCET, JAMA, BMJ gibi önemli dergilerde yayınlanan çalışmaların incelendiği derlemede⁷⁹ çalışmaların %95'inde eksik veriler olduğu, % 47' sinde kayıp oranının %10' dan fazla olduğu ve eksik verilerin ele alınmasında en yaygın kullanılan yöntemin (%45) yalnızca veri eksikliği olmayanların analize dahil edilmesi olduğu bulunmuştur. Aynı çalışmada 2001 yılındaki çalışmalarla karşılaştırıldığında ITT kelimesinin (duyarlılık analiziyle birlikte) daha fazla kullanıldığı ancak eksik veri oranları ve yaklaşımların benzer olduğu bildirilmektedir.

Raporlama yanlılığı ve Raporu yazan kişinin körlenmesi

Araştırma raporunun yazılması aşamasında ortaya çıkan raporlama yanlılığı (reporting bias) araştırma raporunu yazan kişinin araştırma bulgularını ve yorumlamasını araştırma hipotezi yönünde yazma eğilimidir. Önerilen yöntem araştırma raporu yazılıp bitene kadar raporu yazan kişinin hangi grubun girişim hangi grubun kontrol olduğu bilgisi yönünden körlenmesidir^{8,42}.

Dış Geçerlik (External validity)

Kanıta dayalı uygulamaların temel sorununun iç geçerliliği yüksek çalışmaları dikkate alan sınırlı bilimsel bakış açısı olduğu belirtilmektedir. Bu yöntemin keşif aşamasında yararlı olabileceği ancak uygulamaya aktarma konusunda yetersiz kalacağı vurgulanmakta uygulamaya aktarmada dış geçerliliğin önemine dikkat çekilmektedir⁸⁴. Shuttleworth' in ⁸⁵ bildirdiğine göre Campbell ve Stanley (1966) dışsal geçerliliği genelde belirli bir ortamda ve az sayıda kişiyle elde edilen bulguların gerçek yaşam ortamındaki daha fazla kişiye genellenebilmesi olarak tanımlamakta ve toplumsal geçerlik (population validity) ve ekolojik geçerlik (ecologic validity) olmak üzere iki tip dış geçerlikten söz etmektedir. Toplumsal geçerlik seçilen çalışma grubunun evreni temsil yeteneği ve örnekleme özelliklerine işaret eder. Ekolojik geçerlik çalışmanın yürütüldüğü çevre/ortam ve bunun bireylerin davranışlarını ve sonuçları nasıl etkileyebileceğine odaklanır. Deney ortamı gerçek dünyadakin benzemiyorsa genelleme güçleşecektir. Son yıllarda RKD'lerin elde ettiği bilgilerin uygulamaya aktarılmasındaki sorunlara yönelik eleştiri ve değerlendirmeler olduğu bildirilmektedir⁸⁶. Khorsan ve Crawford (2014)⁸⁷ iç geçerliliğin dış geçerlik için gerekli ancak yeterli olmayan bir koşul olduğunu RKD'lerin dış geçerliliğinin ayrıca değerlendirilmesi gerektiği savunmaktadır. RKD'lerle yapılan bir çalışmada ⁸⁸ dahil edilen tüm çalışmaların dış geçerliliğe yönelik yetersizlikleri olduğu ve katılımcıların daha geniş grupları temsiliyeti, katılımcıların kaydedilmesinin ayrıntıları, maliyeti ve toplumsal uygulamada girişimin nasıl uygulamaya aktarılacağı gibi önemli bilgilerin eksik olduğu bildirilmektedir.

Sistematik derlemelerde iç geçerlik önemsense de sonuçları kullanacak kişiler bulguların nasıl genellenebileceği ile daha fazla ilgilenecektir⁸⁶. RKD'ler iç geçerliliği yüksek ve aynı zamanda sonuçları belirli bir alandaki tanımlanmış bir grupla bağlantılı şekilde klinik olarak yararlı olmalıdır. Dış geçerliliği değerlendirmeye yönelik bazı konular eylemsel (pragmatic) ve açıklayıcı (explanatory) deneyler arasındaki ayrımla ilişkilidir. Ancak eylemsel deneylerin açıklayıcı deneylere göre daha yüksek bir dışsal geçerliliğe sahip olduğunu söylemek doğru olmaz. Örneğin çalışma grubunun seçim kriterlerinin geniş olması, temel özelliklere ilişkin verilerin sınırlı olması, farklı bir bakım ve uzmanlık sunan merkezlerin dahil edilmesinin çalışmayı yürütmede pek çok avantajı olabilecek ancak girişimin etkisine ilişkin elde edilen bulgunun genellenebilirliğini oldukça sınırlandıracaktır⁸⁹. RKD'lerin dış geçerliliğini arttırmada dışlama kriterlerinin azaltılması özellikle önerilen bir yöntemdir⁹⁰. Araştırma sonuçlarının genellenebilmesinde araştırma örneğinin evrenden rastgele seçilmesinin de (random selection) yeterli olmayacağı çünkü yalnızca seçildiği evreni temsil edebileceği ve tüm evrene genellenemeyeceği belirtilmektedir⁸⁶.

RKD'lerin kalitesini ya da yanlışlık riskini değerlendiren araçlar ve ölçümlerde genel eğilim girişimin ideal koşullarda etkin olup olmadığını (etkinlik-efficacy) etkileyecek faktörlere yönelik olarak iç geçerliliğe odaklı değerlendirme şeklindedir. İç geçerlik esas olarak araştırma bulguları ve sonuçlarının araştırmanın çalışma grubu için geçerli olduğuna ve nedensel ilişkinin uygun şekilde başarıldığına ilişkin bilgi sunar⁸⁷. Cochrane Collaboration'da ⁸ kalite değerlendirme araçlarının araştırmanın metodolojik olarak yürütülme kalitesinden ziyade raporlama yöntemine odaklandığını belirtmektedir. Bu nedenlerle RKD'lerin kalitesini değerlendirmeye yönelik araçlar klinik karar verme ve gerçek dünyada kullanıma ilişkin sınırlılıklar içermektedir⁸⁷.

Dergi editörlerinin ortak katılımını içeren bir toplantıda dış geçerlik için belirlenen önemli koşulların, çalışma grubu, yeri ve girişimleri yürütenlerin toplumu temsil etme özellikleri yanında uygulanan programın tutarlılığı ve yoğunluğu, çıktıların kalite, maliyet ve olumsuz sonuçlar gibi etkileri, izlem kayıplarını da içeren izlem raporları, programın sürdürülebilirliği ve girişimin uzun süreli etkileri olduğu vurgulanmıştır⁹¹. RKD'lerin tek merkezli çalışmalar olması, küçük örneklerle ve sınırlı özelliklere sahip bireylerle çalışılması, kısa izlem süreleri gibi bazı özelliklerinin çalışmanın dış geçerliliği yönünden sınırlılıklar yarattığı vurgulanmaktadır⁹². RKD raporlarında çalışma yeri ve özelliklerinden söz etmenin dış geçerliliği değerlendirmede ve böylece bulguların klinik uygulamaya aktarılabilmesinde önemi üzerinde durulmaktadır⁹³. Avelar ve arkadaşları (2016)⁸⁶ dış geçerliliğin geniş bir anlam taşıdığını ve genellenebilirlik (generalizability), uygulanabilirlik (applicability) ve uygunluk (feasibility) olmak üzere üç boyutta ele alınması gerektiğini vurgulamaktadır.

RKD'lerde İstatistik Kullanımı

RKD'lerde istatistik kullanımı araştırma sürecinin farklı aşamalarında karşımıza çıkmaktadır. Deneysel araştırmalarda istatistik kullanımı örneklem sayısının belirlenmesi ile başlamaktadır. Örneklem sayısının belirlenmesi araştırma sırasında ortaya çıkabilecek Tip 1 ve Tip 2 hatanın denetlendiği ve örneklemde elde edilen sonuçların yokluk hipotezinde tanımlanan beklentilerden sapma düzeyini gösteren etki büyüklüğünün göz önüne alınmasına dayalı olmalıdır ve p değerinin tek yönlü ya da çift yönlü mü kabul edileceği bilgisine göre bir örneklem sayısı belirlenmelidir. Çok sayıda istatistik paket programında önemlilik değeri, güç değeri ve etki büyüklüğü değerlerini göz önünde bulundurularak deney ve kontrol grupları için gerekli örneklem büyüklüğünü belirlenebilmektedir. Bu aşama sadece örneklem sayısının yeterliliği hakkında bilgi vermektedir. RKD'lerin temel koşulu olan random atama sonucunda oluşturulan grupların temel özellikler yönünden karşılaştırılması ve araştırma sonucunu etkileyeceği düşünülen karıştırıcı faktörlerin her iki grupta eşit dağılımının sağlanıp sağlanmadığının istatistik testleri ile değerlendirilmesi istatistik kullanımının ikinci aşamasıdır. Eğer bu değerlendirmede herhangi bir karıştırıcı faktörlerin dağılımıyla ilgili sorunlar belirlenirse yapılacak ileri analizlerle bu karıştırıcı faktörün kontrol altına alınması gerekmektedir³⁶. Verilerin analizi aşamasında ise yapılan karşılaştırmaların sadece p değeri ile değerlendirilmemesi gerekmektedir. P değeri gruplar arasında elde edilen farklılığın şansa bağlı ortaya çıkmadığı ile ilgili bilgi vermektedir. p değeri sonucu ile ilgili yapılan eleştiriler iki önemli noktaya işaret etmektedir. Birincisi p değeri anlamlı bulunsa bile fark klinik açıdan önemsiz olabilir, ikincisi ise p değerinin

anlamsız çıkması durumunda örneklem grubunun sayısal yetersizliği nedeniyle var olan bir farklılık belirlenememiş olabilir⁹⁴. Bu nedenle özellikle deneysel çalışmaların klinik anlamlılık yönünden de değerlendirilmesi gerekmektedir. Bu ulaştığımız tartışma da araştırma sonuçlarının etki büyüklüğü değerlerinin hesaplanması sonucunu ortaya çıkarmaktadır. Etki büyüklüğü değişkenler arasındaki ilişkinin ya da uygulama etkisinin büyüklüğünü yansıtan bir değerdir. Analizlerde kullanılan değerler ne olursa olsun ortalama, standart sapma, odds değerleri, korelasyon katsayısı, varyans analiz sonuçları gibi her değerlendirmenin etki büyüklükleri hesaplanabilmektedir. Bu nedenle deneysel çalışmalarda sadece p değerleri değil, etki büyüklük sonuçlarının da rapor edilmesi gerekmektedir^{94,95,96}. Araştırmada p değeri bize tip 1 hatayı ne kadar kontrol altına aldığımız, etki büyüklüğü ise sonuçların uygulamaya aktarıldığındaki etkisinin ne olacağı hakkında bilgi verirken, Tip 2 hatayı değerlendirmek için analiz değerlerine dayalı güç düzeyinin (power değeri) de verilmesi gerekmektedir. Ayrıca deney grubunda meydana gelen değişimin bütün denekleri kapsayıp kapsamadığının da değerlendirilmesi önemlidir. Bunun yanı sıra nedensellik yönünden daha iyi yorum getiren karıştırıcı faktörlerin denetimini sağlayan, değişimleri zaman*grup etkileşimi içerisinde değerlendiren cox regresyon, Manova, multivariate regresyon, tekrarlayan ölçümlerde çift yönlü varyans analizi gibi analiz türlerinin tercih edilmesi ve alt grup istatistiklerinin yapılması önemlidir. Yapılan bütün analizlerin ITT ve Per-protokol (PP) analizi şeklinde ayrı ayrı yapılması ve sonuçlarının belirtilmesi önemli bir etik sorumluluktur. Deneysel çalışmalarda dikkat edilmesi gereken diğer bir nokta ise hipotezin yönüne göre p değerinin yorumlanma zorunluluğudur. Çalışmadaki amaç daha iyi ya da daha kötü olmama durumlarından birini göstermekse p değeri tek yönlü, her iki durumda da araştırma hipotezi için olası sonuç olarak tanımlanmışsa p değeri çift yönlü kabul edilmelidir^{36,38,97,98}. RKD'lerde istatistik kullanımı için bir özetleme yapmak gerekirse, analizlerin çalışmadan ayrılan kişileri kapsayıcı ve kapsamayan şekilde ayrı ayrı yapılması, analiz sonuçlarının p değeri, güven aralığı, güç düzeyi, etki büyüklüğünü içerecek şekilde yazılması ve istatistik analizlerinin seçiminde verilerin uygunluğu durumunda karıştırıcı faktörleri kontrol altına alan ve grup içi ve gruplar arası etkileşimleri değerlendiren analizlerin tercih edilmesi gerektiğini söyleyebiliriz.

SONUÇ

Sonuç olarak randomize kontrollü çalışmaların planlanması, yürütülmesi, verilerin değerlendirilmesi ve raporlanması aşamalarında uyulacak ilkeler ve alınacak önlemlerin çalışmanın kanıt değerini ortaya koyduğunu söyleyebiliriz. Temel ilke iç ve dış geçerliliği etkileyebilecek yanlılıkların kontrol edilmesine yönelik gerekli önlemleri almak, uygun istatistiksel teknikleri kullanmak ve yapılan/yapılmayan şeyleri araştırma raporunda açıkça belirtmektir. Bunun yanında araştırma protokolü oluşturulması ve çalışmanın uygun şekilde kaydedilmesi önemle üzerinde durulan konulardır.

Araştırmacıların randomize kontrollü deneysel çalışmalarında bir çalışma protokolünü oluşturmaları ve bu protokolü tasarım türünün, random atamanın uygun yapılması, yanlılıkların kontrol edilmesi, iç ve dış geçerliliğin sağlanması, uygun istatistiklerinin seçimiyle eleştirilere karşı güçlü hale getirmeleri gerekmektedir. Araştırmacıların deney öncesi ve raporlama öncesi çalışma protokollerini uluslararası geçerliliği olan kontrol listelerine göre değerlendirmeleri ve çalışma protokollerini yine uluslararası kayıt sistemlerine kayıt etmeleri önerilmektedir.

KAYNAKLAR

1. Ciliska D, Cand N, Brian H. Evidence-based nursing, USA: Wiley; 2013. p.15-48
2. Persson J, Sahlin NE. A philosophical account of interventions and causal representation in nursing research. *International Journal Of Nursing Studies* 2009; 46(4): 547-56.
3. Cochrane Consumer Network. Levels of evidence. 2016. URL:<http://consumers.cochrane.org/levels-evidence>: 10.12.2016
4. Shah HM, Chung KC. Archie Cochrane and his vision for evidence-based medicine. *Plast Reconstr Surg* 2009; 124(3): 982-8.
5. Johnson MH. The early history of evidence-based reproductive medicine. *Reprod Biomed Online* 2013;26(3):201-9.
6. Peterson MH, Barnason S, Donnelly B, Hill K, Miley H, Riggs L, Whiteman K. Choosing the best evidence to guide clinical practice: Application of AACN levels of evidence. *Critical Care Nurse* 2014; 34(2):58-68.
7. Nichol AD, Bailey M, Cooper DJ. Challenging issues in randomised controlled trials. *Injury* 2010;41: 20-3
8. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions* version 5.1.0. England: The Cochrane Collaboration; 2011.
9. Flecha OD, Douglas de Oliveira DW, Marques LS, Gonçalves PF. A commentary on randomized clinical trials: How to produce them with a good level of evidence. *Perspect Clin Res* 2016;7:75-80.
10. Forbes A. Clinical intervention research in nursing. *International Journal of Nursing Studies* 2009; 46: 557-68
11. Richards DA, Hamers JP. RCTs in complex nursing interventions and laboratory experimental studies. *International Journal of Nursing Studies* 2009;(46): 588-92
12. Borglin G, Richards DA. Bias in experimental nursing research: Strategies to improve the quality and explanatory power of nursing science. *International Journal of Nursing Studies* 2010;47:123-8
13. Fridlund B, Jönsson AC, Andersson EK, Bala SV, Dahlman GB, Forsberg, A, et al. Essentials of nursing care in randomized controlled trials of nurse-led interventions in somatic care: A Systematic Review. *Open Journal of Nursing* 2014; 4: 181-97.
14. Mantzoukas S. The research evidence published in high impact nursing journals between 2000 and 2006: A quantitative content analysis. *International Journal of Nursing Studies*. 2009;46: 479-89
15. Baldi I, Del Lago E, De Bardi S, Sartor G, Soriani, N, Zanotti, R. Trends in RCT nursing research over 20 years: mind the gap *British Journal of Nursing* 2014;23(16), 895-99.
16. Blackwood B. Methodological issues in evaluating complex healthcare interventions *Journal of Advanced Nursing* 2006; 54(5): 612-2.
17. Thompson C. Fortuito us phenomena: on complexity, pragmatic randomised controlled trials, and knowledge for evidence-based practice. *World Views on Evidence-Based Nursing* 2004;1: 9-17.
18. Glasdam S, Sivberg B, Wihlborg, M. Nurse-led interventions in the concept of randomized controlled trials – critical perspectives on how to handle social contexts. *International Journal of Multiple Research Approaches* 2015; 9(1); 1-23.
19. Rolfe G. Complexity and uniqueness in nursing practice: Commentary on Richards and Hamers (2009). *International Journal of Nursing Studies* 2009;46:1156-8
20. Lindsay B. Randomized controlled trials of socially complex nursing interventions: creating bias and unreliability. *Journal of Advanced Nursing* 2004;45 (1): 84-94.
21. Chow, SC. *Design and analysis of clinical trials*, USA:Wiley; 2013, p.124-98
22. Pocock SJ. *Clinical trials*, USA:Wiley; 2013, p.16-89, p.321-80

23. Institute of Medicine Staff. Small clinical trials, National Academies Press; 2001. p.3-89
24. Alford L. On differences between explanatory and pragmatic clinical trials. *New Zealand Journal of Physiotherapy* 2006;35(1):12-6.
25. Treweek S, Zwarenstein M. Making trials matter: pragmatic and explanatory trials and the problem of applicability. *Trials* 2009;10:37
26. Lesaffre E. Superiority, equivalence, and non-inferiority trials. *Bulletin of the NYU Hospital for Joint Diseases* 2008;66:150-4.
27. Christensen E. Methodology of superiority vs. equivalence trials and non-inferiority trials. *Journal of Hepatology* 2007; 46:947-54.
28. Gonzalez CD, Bolaños R, de Sereday M. Editorial on hypothesis and objectives in clinical trials: superiority, equivalence and non-inferiority. *Thrombosis Journal* 2009; 7(1):3.
29. Fayers PM, Machin D. *Randomized clinical trials. Incorporated: Wiley&Sons, 2010.*
30. Chan AW, Tetzlaff JM, Gøtzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013;346:75-86.
31. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić, K. et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Annals of Internal Medicine* 2013;158(3): 200-7.
32. Gray R, Badnapurkar A, Thomas D. Reporting of clinical trials in nursing journals: how are we doing? *JAN* 2016; Editorial:1
33. Attia A. Bias in RCTs: Confounders, selection bias and allocation concealment. *Middle East Fertility Society Journal* 2005; 3(10): 258-61.
34. *Research and Innovation . Randomization in clinical trials, England: University of the West of England ; 2011.*
35. Xiao L, Lavori PW, Wilson SR, Ma J. Comparison of dynamic block randomization and minimization in randomized trials: A simulation study. *Clin Trials* 2011;8:59-69.
36. Field A. *Discovering statistic using IBM SPSS statistics, 4th Edition. Los Angels: SAGE Publication; 2009*
37. Shively WP. *Causal thinking and design of research the craft of political research. 8th edition. New Jersey: Copyright © by Pearson Education Upper Saddle River; 2009.*
38. Çağatay P, Şenocak M, Baykal İE. A proposal for scoring of the evaluation of randomized clinical trials. *Cerrahpaşa J Med* 2000; 31 (1): 49-55.
39. Dziura JD, Post LA, Zhao Q, Fu Z, Peduzzi P. Strategies for dealing with missing data in clinical trials: From design to Analysis. *Yale Journal of Biology and Medicine* 2013; 86:343-58.
40. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg* 2010;126:619-25.
41. Ron K, Faltin F, Ruggeri F. *Statistical methods in healthcare, edited by Ron Kenett. USA:Wiley; 2012.*
42. European Network for Health Technology Assessment (EUnetHTA). *Guideline - levels of evidence - internal validity of randomized controlled trials, Final version; 2013 URL:https://eunethta.fedimbo.belgium.be/sites/5026.fedimbo.belgium.be/files/Internal_Validity.pdf ,29.11.2016*
43. Machin D, Fayers PM. *Randomized clinical trials. USA:John Wiley&Sons; 2010*
44. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to meta-analysis, publication bias USA: John Wiley & Sons, Ltd; 2009*
45. Song F, Hooper L, Loke YK. Publication bias: what is it? How do we measure it? How do we avoid it? *Open Access Journal of Clinical Trials* 2013;5: 71-81
46. International Committee Medical Journal Editors-ICMJE. *Clinical Trials Registration 2008 URL:: http://www.icmje.org/about-icmje/faqs/clinical-trials-registration/ 04.12.2016*

47. Tse T, Williams RJ, Zarin DA. Reporting “basic results” in Clinicaltrials. CHEST Journal 2009; 136(1):295-303.
48. WHO. International standards for clinical trial registries 2012. URL: http://apps.who.int/iris/bitstream/10665/76705/1/9789241504294_eng.pdf?ua=1&ua=1. 04.12.216
49. Food and Drug Administration. Amendments act (FDAAA) 2007 URL: <https://www.gpo.gov/fdsys/pkg/PLAW-110publ85/pdf/PLAW-110publ85.pdf#page=82> Erişim tarihi: 04.12.216
50. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;63:2237.
51. Jull A, Aye PS. Endorsement of the CONSORT guidelines, trial registration, and the quality of reporting randomised controlled trials in leading nursing journals: A cross-sectional analysis. International Journal of Nursing Studies 2015; 52(6): 1071–1079
52. Guo JW, Sward KA, Beck SL, Stagers N. Quality of reporting randomized controlled trials in cancer nursing research. Nursing Research 2014; 63(1): 26–35.
53. Jadad AR, Enkin MW. Bias in randomized controlled trials in Randomized controlled trials: Questions, answers and musings. 2nd Edition. USA: Blackwell Publishing 2007.
54. Turlik M. Evaluating the internal validity of a randomized controlled trial. The Foot and Ankle Online Journal 2009; 2 (3): 5-9
55. Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: a descriptive analysis. BMC Medical Research Methodology 2013; 13(1): 116.
56. Smith AS, Oldman A, McQuay H, Moore R. Teasing apart quality and validity in systematic reviews: an example from acupuncture trials in chronic neck and back pain. Pain 2000;86:119-32
57. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. Journal of Evidence-Based Medicine 2015; 8(1): 2-10.
58. Zhang J, Wang J, Han L, Cao X, Shields L. Tools to assess risk of bias in systematic reviews of nursing intervention in China: Global implications of the findings. Nursing Outlook; 2016
59. Schulz KF. Randomised trials, human nature, and reporting guidelines. Lancet 1996;348:5968.
60. Karanicolas PJ, Farrokhyar F, Bhandari M, Blinding: Who, what, when, why, how? Can J Surg 2010;53(5):345-8.
61. Bagiella E. Randomization in handbbook for clinical research. Medical Publishing ; 2014.
62. Schulz F, Grimes DA. Blinding in randomised trials: hiding who got what. The Lancet 2002;359(9307):696-700.
63. Hróbjartsson A, Thomsen ASS, Emanuelsson F, Tendal B, Hilden J, Boutron I, et al. Observer bias in randomized clinical trials with measurement scale outcomes: a systematic review of trials with both blinded and nonblinded assessors. Canadian Medical Association Journal 2013;1: 185:279
64. King, KM, Thompson DR. Guest editorial: Pragmatic trials: is this a useful method in nursing research?. Journal of Clinical Nursing 2008; 17(11): 1401-2
65. Hróbjartsson A. What are the main methodological problems in the estimation of placebo effects. Journal of Clinical Epidemiology 2002;55: 430-5
66. Macefield R. Usability studies and the Hawthorne effect. Journal of Usability Studies 2007; 2(3): 145-54.
67. Polit DF. Blinding during the analysis of research data. International Journal of Nursing Studies 2011; 48(5): 636-41.

68. Hallberg IR. Moving nursing research forward towards a stronger impact on health care practice? *International Journal of Nursing Studies* 2009; 46(4): 407-12
69. Armijo-Olivo S, Warren S, Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. *Physical Therapy Reviews* 2009;14 (1):36-49.
70. Fergusson D, Aaron SD, Guyatt G, Hebert P. Postrandomisation exclusions: the intention to treat principle and excluding patients from analysis. *BMJ* 2002;325:652-4.
71. Gupta SK. Intention-to-treat concept: A review. *Perspect Clin Res* 2011; 2(3): 109-12.
72. Silva Filho CR, Saconato H, Conterno LO, Marques I, Atallah AN. Assessment of clinical trial quality and its impact on meta-analyses. *Rev Saude Publica* 2005;39:865-73.
73. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ* 2001;165:1339-41.
74. Cheema JR. Some general guidelines for choosing missing data handling methods in educational research. *Journal of Modern Applied Statistical Methods*. 2014;13(2): 3.
75. Akl EA, Shawwa K, Kahale LA, Agoritsas T, Brignardello-Petersen R, Busse JW. et al. Reporting missing participant data in randomised trials: systematic survey of the methodological literature and a proposed guide. *BMJ* 2015;5(12): 8431.
76. Weinstein JN, Lurie JD, Tosteson TD, Skinner JS, Hanscom B, Tosteson AN et al. Surgical vs nonoperative treatment for lumbar disk herniation: The Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA* 2006;296:2441- 50.
77. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. *BMJ* 2010;3: 40
78. Sainani KL. Making sense of Intention-to-Treat, *American Academy of Physical Medicine and Rehabilitation* 2010; 2:209-213.
79. Bell ML, Fiero M, Horton NJ, Hsu CH. Handling missing data in RCTs; a review of the top medical journals. *BMC medical research methodology*, 2014;14(1): 118.
80. Polit DF, Gillespie BM. The use of the intention-to-treat principle in nursing clinical trials. *Nursing Research* 2009; 58(6): 391-9.
81. Gewandter JS, Mc Dermott MP, McKeown A, Smith SM, Pawlowski JR, Poli JJ et al. Reporting of intention-to-treat analyses in recent analgesic clinical trials: ACTION systematic review and recommendations, *Pain*, 2014;155(12):2714-19.
82. Joseph R, Sim J, Ogollah R, Lewis, M. A systematic review finds variable use of the intention-to-treat principle in musculoskeletal randomized controlled trials with missing data. *Journal of Clinical Epidemiology* 2015; 68(1): 15-24.
83. Gravel J, Opatrny L, Shapiro S. The intention-to-treat approach in randomized controlled trials: are authors saying what they do and doing what they say?. *Clinical Trials* 2007; 4(4): 350-6.
84. Fernandez A, Sturmberg J, Lukersmith S, Madden R, Torkfar G, Colagiuri R et al. Evidence-based medicine: is it a bridge too far? *Health Research Policy and Systems* 2015; 13(1):66.
85. Shuttleworth M. External Validity. (Aug 7, 2009) URL: <https://explorable.com/external-validity>. 14.01.2017
86. Avelar SA, Thomas J, Kleinman R, Sama-Miller E, Woodruff SE, Coughlin R et al. External validity. The next step for systematic reviews? *Evaluation Review* 2016; 1:43.
87. Khorsan R, Crawford C. External validity and model validity: A conceptual approach for systematic review methodology. *Evidence Based Complementary and Alternative Medicine* 2014; 1: 12 -15
88. Partridge SR, Juan S J-H, McGeechan K, Bauman A, Allman-Farinelli M. Poor quality of external validity reporting limits generalizability of overweight and/or obesity lifestyle prevention interventions in young adults: a systematic review. *Obesity Reviews* 2015; 1(6):13-31.

89. Rothwell PM. Factors that can affect the external validity of randomized controlled trials. PLoS Clin Trials 2006; 1(1):e9
90. Buntrock C, Ebert D, Lehr D, Riper H, Smit F, Cuijpers P et al. Effectiveness of a web-based cognitive behavioural intervention for subthreshold depression: Pragmatic Randomised Controlled Trial. Psychother Psychosom 2015; 84:348–58.
91. Steckler A, McLeroy KR. (Editorial) The Importance of external validity. American Journal of Public Health. 2008; 98(1): 9-10.
92. Donald F, Kilpatrick K, Reid K, Carter N, Martin-Misener R, Bryant-Lukosius D et al. A Systematic review of the cost-effectiveness of nurse practitioners and clinical nurse specialists: What is the quality of the evidence? Nursing Research and Practice 2014;1:28
93. Sigurdardottir KR, Oldervoll L, Hjermsstad MJ, Kaasa S, Knudsen AK, Lohre ET et al. How are palliative care cancer populations characterized in Randomized Controlled Trials? A Literature Review. Journal of Pain and Symptom Management 2014; (47) 5: 906-14.
94. Ferreira JC, Patino CM,3 What does the p value really mean? J Bras Pneumol 2015;41(5):485
95. Alpar R. Spor, Sağlık ve eğitim bilimlerinden örneklerle uygulamalı istatistik ve geçerlik-güvenirlilik. Ankara: Detay Yayıncılık. 7. Baskı. 2016. S.227-233, s.232-233, s.314
96. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Meta analizi giriş. Ankara: Anı Yayıncılık.2013. s.17-60
97. West A, Spring B. Randomized controlled trials. URL: http://www.ebbp.org/course_outlines/randomized_controlled_trials/ 14.01.2017
98. Peduzzi P, Henderson W, Hartigan P, Lavori P. Analysis of randomized controlled trials. Epidemiologic Reviews 2002; 24(1):12-5