

Tanı Doğruluğu Çalışmalarının Stard Kriterlerine Göre Türk Tıp Dizini Özelinde Değerlendirilmesi

Use of the Stard Statement in Reporting Diagnostic Test Research in Turkish Medical Directory

Mustafa Agah Tekindal¹, Can Ateş², Yasemin Yavuz², Rabia Albayrak³, Pınar Selvi³

¹ Department of Biostatistics and Medical Informatics, Faculty of Medicine, Izmir University

² Department of Biostatistics, Faculty of Medicine, Ankara University

³ Department of Animal Science Biometry and Genetics, Faculty of Agriculture, Ankara University

Amaç: Tanı testi çalışmaları incelendiğinde, testlerin ayrıcalık gücünü değerlendiren çalışmaların planlanması, yürütülmesi ve sonuçlarının raporlanmasına gereken önemin verilmediği görülmektedir. Son yıllarda meta analizi çalışmalarının artması, çalışma sonuçlarının raporlanmasında standartların oluşturulması gerekliliğini ortaya koymuştur. Bu gerekliliklerden yola çıkarak 2003 yılında bu standartları oluşturan QUADAS ve STARD isimleri ile iki kontrol listesi geliştirilmiş ve yayınlanmıştır. Genç ve arkadaşları 2012 yılında STARD kriterlerinin Türkçe uyarlamasını yayınlamışlardır. Çalışmamızda Türkçe STARD Kontrol listesi kullanılarak Türkçe yayınlanan tanı doğruluğu çalışmalarının kaliteleri değerlendirilecek ve her bir kriterin sağlanma oranları yıllara göre belirlenecektir.

Gereç ve Yöntem: Çalışmamızda “duyarlılık, özgüllük, seçicilik, işlem karakteristiği eğrisi, İKE, ROC ve tanı doğruluğu” anahtar kelimeleri kullanılarak 2007-2010 yılları arasında ULAKBİM Tıp Veri Tabanında [Türk Tıp Dizini] yayınlanmış tanı doğruluğu çalışmaları taranmış ve 459 makale belirlenmiştir. Daha sonra bu çalışmalar belirlenen kriterlere göre tekrar değerlendirilerek 75 çalışma seçilmiştir. Bu çalışmalar Türkçe STARD kontrol listesine göre değerlendirilerek her bir kriter için sağlanma yüzdeleri bulunmuştur.

Bulgular: STARD kriterlerini en çok sağlayan makaleler 25 maddenin 21’ini karşılarken kriterlere en az uyum makalelerde bu sayı 6’ya düşmüştür. Bunun yanı sıra, 75 makalenin %9’u STARD beyanına %50’den daha az uyum gösterirken, makalelerin %81’i, %50 - %75 arasında bir uyum sergilemektedir. Çalışmaların büyük bölümünde tanı doğruluğuna ilişkin ölçüler verilirken (%91), sadece %32’sinde bunlara ilişkin Güven Aralıkları verilmiştir. Değerlendirilen makalelerde, çalışmaya alınan deneklere ilişkin kriterler genel olarak yüksek oranda karşılanırken (ortalama %84) bu oranlar yıllar içinde de benzerdir. Değerlendirilen çalışmaların %96’sında sonuçların klinik uygulanabilirliğini tartışılmıştır. Bu kriter en çok sağlanan kriterdir.

Sonuç: 2007-2010 yılları arasında ULAKBİM Tıp Veri Tabanında [Türk Tıp Dizini] yayınlanmış tanı doğruluğu çalışmalarının STARD kriterlerini sağlama oranlarının çok yüksek düzeyde olamasa da yeterli olduğu söylenebilir. STARD kontrol listesi Türkçe versiyonunun kullanıma sunulması Türkçe yayınlanan tanı doğruluğu çalışmalarının standarda kavuşturulması için çok önemli bir araç olarak kullanılabilir. Böylece, Türkçe literatürdeki tanı doğruluğu çalışmalarının niteliğinde daha fazla iyileşme sağlanabilir.

Anahtar Sözcükler: **STARD Kriterleri, Tanı Doğruluğu Çalışmaları, Sistematik Derleme**

Aim: When diagnostic testing studies are examined, it is seen that due importance is not attached to planning and conducting studies that evaluate the discriminating power of tests and reporting the results of such studies. Recent increase in the number of meta-analysis studies has indicated the need to create standards for the reporting of results. Based on this need, two statements making up such standards were developed and published in 2003 under the names of QUADAS and STARD. STARD criteria were adapted to Turkish by Genç et al. in 2012. The aim of our study is to identify the quality of the diagnostic accuracy studies published in Turkish Medical Directory by using the Turkish version of the STARD statement and to determine the rate of satisfaction of each criterion by year.

Materials and Methods: To find appropriate studies, the articles that were published on ULAKBİM Medical Database [Turkish Medical Directory] between 2007 and 2010 and included the words, “sensitivity, specificity, discrimination receiver operating characteristic, ROC, diagnostic accuracy” were searched, and 459 articles were determined. As a result of the electronic search, 459 articles were determined. Then those studies were re-evaluated based on the specified criteria. Finally, 75 appropriate articles were chosen. Those studies were evaluated according to the Turkish STARD statement, and percentage of satisfaction was found for each criterion.

Results: The most compliant articles with the STARD statement satisfied 21 items out of 25 whereas the least compliant articles satisfied only 6 items. Moreover, 9% of 75 articles have less than 50% relevance to the STARD statement while 81% of 75 articles are in 50-75% relevance range. A great majority of the studies (91%) provide measures related to diagnostic accuracy, but only 32% give confidence intervals regarding them. In the evaluated articles, the criteria concerning the subjects included in the studies were mostly satisfied at a high rate (84% in average). These rates of satisfaction are similar over the years. In 96% of the evaluated studies, the clinical applicability of the results was discussed. This is the criterion satisfied most.

Conclusions: To conclude, the rate of satisfaction of the STARD criteria by the diagnostic accuracy studies published on the ULAKBİM Medical Database [Turkish Medical Directory] between 2007 and 2010 is adequate, though not very high. The introduction of the Turkish version of the STARD statement may serve as a very important tool for the standardization of the diagnostic accuracy studies published in Turkish. This may lead to an improvement in the quality of the diagnostic accuracy studies in the Turkish literature.

Key Words: **STARD Statements, Diagnostic Accuracy Studies, Systematic Review**

Geliş tarihi : 16.06.2015 • Kabul tarihi: 12.09.2015

İletişim

Arş. Gör. Mustafa Agah TEKİNDAL

Tel: +90 (232) 246 49 49

GSM:+90 555 399 05 99

E-posta: matekindal@gmail.com

Izmir Üniversitesi Tıp Fakültesi Biyoistatistik Anabilim Dalı
35350 İzmir Türkiye

İndeks testler (tanı testleri), hasta ve sağlıklı bireylerin oluşturduğu heterojen bir kitlede bireylerin gerçek durumunu (gerçekten hasta olup olmadıklarını) ortaya çıkarmak amacıyla kullanılır. Doğruluğu kesin olarak ka-

nitlanmış referans standart (altın standart test) testler ile bireylere “kesin hasta” ya da “kesin sağlıklı” tanısı konulabilir. Fakat bu testlerin uygulanmalarının zor, maliyetlerinin yüksek ve bazı hastalıkların oluşmasında

etkili olmaları nedeniyle her şüpheli durumda kullanılmaları mümkün olmayabilir. Bu sebeple, birçok bilim dalında referans standartlara alternatif olacak indeks testler (tanı testleri) geliştirilmeye çalışılır. Belirli kriterlere göre seçilmiş bir grup kişiye, referans standart ve indeks test uygulanarak; indeks testin ayırıcılık gücünü gösteren “doğruluk ölçütleri” elde edilir. Yeni geliştirilen indeks testlerin ayırıcılık gücünü belirlemek ya da testlerin ayırıcılık gücünü karşılaştırmak amacıyla yapılan çalışmalara “tanı doğruluğu” çalışmaları adı verilir (1). Tanı testi çalışmaları incelendiğinde, testlerin ayırıcılık gücünü değerlendiren çalışmaların planlanması, yürütülmesi ve sonuçlarının raporlanmasına gereken önemin verilmemesi görülmektedir. Son yıllarda meta analizi çalışmalarının artması, çalışma sonuçlarının raporlanmasında standartların oluşturulması gerekliliğini ortaya koymuştur. Bu gerekliliklerden yola çıkarak 2003 yılında bu standartları oluşturan QUADAS (Quality Assessment of Diagnostic Accuracy Studies) ve STARD (The STAndards for Reporting of Diagnostic Accuracy) isimleri ile iki kontrol listesi geliştirilmiş ve yayınlanmıştır (2-7).

QUADAS, geçerliliği kanıtlanmış bir kalite kontrol listesi olup 14 maddeden oluşmaktadır. Bu maddelerin dikkate alınmasıyla; en önemli yanlışlık kaynakları ve tanı doğruluğu çalışmalarında gözlenen varyasyonun engellenmesi öngörülmüştür. Kontrol listesinin oluşturulması sürecinde, epidemiyolog ve biyoistatistikçilerin de aralarında bulunduğu çalışma grubu üyeleri tarafından 4 kez biraraya gelinerek (Delphi prosedürü); 28 maddeden oluşan liste, verilen ortak kararlar 14 maddeye indirilmiştir. QUADAS kontrol listesindeki her bir madde için “EVET”, “HAYIR” ve “BELİRSİZ” olmak üzere üç ayrı yanıt seçeneği mevcuttur. Hazırlanan bu kontrol listesinin amacı, tanı doğruluğu çalışmalarının kalitelerinin artırılması olarak belirlenmiş ve bu alanda çalışma yapan araştırmacıların kullanımına sunulmuştur (6, 7).

“The STAndards for Reporting of Diagnostic Accuracy-STARD” adıyla bilinen kontrol listesi ise, tanı doğruluğu çalışmalarının raporlanmalarına yönelik kalitenin artırılması amacıyla geliştirilmiştir. Çoğunluğunu epidemiyolog ve biyoistatistikçilerin oluşturduğu bir grup araştırmacı tarafından kontrol listesinin geliştirilmesi sürecinde, tanı doğruluğu çalışmaları için yayınlanmış rehberler incelenmiş ve 33 ayrı liste bulunmuştur. STARD grup üyeleri tarafından bu listelerin irdelenmesi sonucunda, 75 maddeden oluşan yeni bir liste hazırlanmıştır.

Daha sonra geniş katımlı bir toplantının ardından liste 25 maddeye indirilmiştir. Oluşturulan yeni STARD bildirim 2003 yılının Ocak ayında aynı anda 8 tıbbi dergide (Radiology, American Journal of Clinical Pathology, Annals of Internal Medicine, British Medical Journal, Clinical Biochemistry, Clinical Chemistry, Clinical Chemistry of Laboratory Medicine, and Lancet) yayınlanarak ilan edilmiştir. STARD bildirim 2003’te yayınlanmasından bu yana 200’den fazla süreli yayında “Yazarlara Bilgi/Yazım Kuralları” kısmında bu kriterlerin kullanılması önerilmiştir. Çünkü tanı doğruluğu çalışmalarının tam ve doğru olarak raporlanması, okuyucuya sonuçlarda var olabilecek yanlışlığı anlama fırsatı sağlayabileceği gibi elde edilen sonuçların uygulanabilirliği ve genellebilirliği hakkında da bilgi vermektedir. Her iki kontrol listesi de literatürde yavaş yavaş kabul görmeye başlamış olup Nisan 2008’de, yazarlar için STARD ifadesinin kullanımını teşvik eden tahmini olarak 200’den fazla biyomedikal dergi olduğu belirlenmiştir(8).

2012 yılında Genç ve arkadaşları STARD kontrol listesinin geçerlilik-güvenirlilik çalışmalarını yaparak Türkçe uyarlamasını gerçekleştirmişlerdir(9).

Çalışmamızda Türkçe STARD Kontrol listesi kullanılarak Türkçe yayınlanan tanı doğruluğu çalışmalarının kaliteleri değerlendirilecek ve her bir kriterin sağlanma oranları yıllara göre belirlenecektir

METOT VE YÖNTEM

Çalışmada 2007-2010 yılları arasında ULAKBİM Tıp Veri Tabanında [Türk Tıp Dizini] yayınlanmış tanı doğruluğu çalışmaları taranmıştır. Taramada “duyarlılık, özgüllük, seçicilik, işlem karakteristiği eğrisi, İKE, ROC ve tanı doğruluğu” kelimeleri kullanılarak, makalelerin başlık, özet ve anahtar kelimeleri taranmıştır. Sadece Türkçe dilde yazılmış insanlar üzerinde yapılan çalışmalar seçilmiştir.

Elektronik tarama sonucu 459 makale belirlenmiştir. Bu makaleler basılmış ve belirlenen beş kriterlere göre tekrar bir eleme süreci başlatılmıştır. Bu kriterler (1) Vaka raporu (olgu sunumu) olan çalışmalar, (2) duyarlılık ve seçicilik ölçüleri raporlanmamış çalışmalar, (3) Duyarlılık ve seçicilik ölçülmemiş veya yorumlanmamış çalışmalar, (4) Kurum içi yada klinik şüphe nedeniyle test protokolü değiştirilmiş çalışmalar, (5) Tekrarlanan (orjinal olmayan) test çalışması olarak belirlenmiş çalışmalar. Bu eleme sonucunda 75 çalışma değerlendirme için uygun bulunmuştur.

BULGULAR

ULAKBİM Tıp Veri Tabanında 2007-2010 yılları arasında yayınlanmış toplam 75 tanı doğruluğu makalesi Genç ve ark. (9) tarafından STARD Kriterlerinin Türkçe uyarlamasına göre değerlendirilmiş ve her bir kriterin yıllara göre ve toplam karşılanma sayı ve yüzdeleri Tablo 1’de verilmiştir.

“Başlık, Özet ve Anahtar Kelimeler” bölümünde STARD kriterinin sağlanması için makalenin bir tanı doğruluğu çalışması olarak tanımlanması gerekmektedir. Değerlendirilen 75 makalenin 70’i (%93) bu kriteri sağlamaktadır. Yıllara göre ilgili kriterin sağlanma yüzdeleri incelendiğinde ise 2007, 2008, 2009 ve 2010 yılları için değerler sırasıyla %94, %94, %87 ve %100 olarak bulunmuştur. Yıllara göre kriterlerin sağlanma yüzdelerinde değişim olmadığı gözlenmektedir.

Tablo 1: Değerlendirilen makalelerde STARD kriterlerinin karşılanma sayısı ve yüzdeleri

Bölüm ve Konu	No		Karşılanan Kriter sayısı ve yüzdeleri				
			2007 (n=36) n(%)	2008 (n=18) n(%)	2009 (n=15) n(%)	2010 (n=6) n(%)	TOPLAM (n=75) n(%)
BAŞLIK/ ÖZET/ ANAHTAR KELİ- MELER	1	Makaleyi bir tanı doğruluğu çalışması olarak tanımlayın (MeSH başlığının duyarlılık ve seçicilik olması önerilir).	34(94)	17(94)	13(87)	6(100)	70(93)
GİRİŞ	2	Tanı doğruluğunun tahmin edilmesi, tanı testlerinin karşılaştırılması ya da tanı doğruluğunun alt gruplarda karşılaştırılması gibi ifadeler kullanarak çalışmanın amacını ve/veya hipotezlerini belirtin.	32(89)	10(56)	12(80)	5(83)	59(79)
METOD							
Denekler	3	Çalışma popülasyonunu tanımlayın: Dahil etme ve çıkarma kriterleri, verilerin toplandığı ortam ve mekanlar.	32(89)	13(72)	14(93)	6(100)	65(87)
	4	Katılımcıların çalışmaya alınma prosedürünü tanımlayın: Katılımcı alımı, var olan semptomlara mı, önceki testin sonuçlarına mı, yoksa referans standart (altın standart test) ve/veya indeks test (tanı testi) sonuçlarına mı dayanmaktadır?	35(97)	15(83)	14(93)	6(100)	70(93)
	5	Katılımcı örneklemini tanımlayın: Madde 3 ve 4 deki seçilme kriterlerini sağlayan katılımcılar, çalışmaya ardışık olarak mı alındı? Eğer değilse katılımcıların nasıl seçildiğini belirtin.	33(92)	10(56)	11(73)	3(50)	57(76)
	6	Verilerin toplama biçimini tanımlayın: Veri toplama biçimi indeks test (tanı testi) ve referans standart (altın standart test) uygulanmadan önce mi (ileriye dönük planlanmış çalışma) uygulandıktan sonra mı (geriye dönük planlanmış çalışma) belirlendi?	33(92)	16(89)	12(80)	5(83)	66(88)
Test Yöntemleri	7	Referans standardı (altın standart testi) ve bilimsel dayanağını tanımlayın.	29(81)	10(56)	9(60)	5(83)	53(71)
	8	Materyal ve metodun teknik özelliklerini, ölçümlerin ne zaman ve nasıl yapıldığını da belirtilerek detaylı olarak açıklayın ve/veya indeks test (tanı testi) ve referans standart (altın standart test) için referanslar gösterin.	31(86)	10(56)	9(60)	4(67)	54(72)
	9	İndeks test (tanı testi) ve referans standart (altın standart test) için kullanılan birimleri, kesim noktalarını ve belirlenen kategorilerin geçerliliğini gerekçeli olarak tanımlayın.	32(89)	12(67)	11(73)	6(100)	61(81)
	10	İndeks testi (tanı testini) ve referans standardı (altın standart test) uygulayan kişilerin sayısını, eğitimini ve uzmanlık düzeyini belirtin.	14(39)	7(39)	4(27)	1(17)	26(35)
	11	İndeks testi (tanı testini) ve referans standardı (altın standart test) uygulayan gözlemcilerin, katılımcıların klinik bilgilerine ve diğer test sonuçlarına kör olup olmadıklarını belirtin.	10(28)	11(61)	3(20)	2(33)	26(35)
İstatistiksel Yöntemler	12	Tanı doğruluğu ölçütlerinin hesaplanmasında veya karşılaştırılmasında kullanılan yöntemleri ve belirsizlik tahmini (örneğin, %95 güven aralığı) için kullanılan istatistiksel yöntemleri belirtin.	31(86)	12(67)	10(67)	5(83)	58(77)
	13	Eğer yapılmış ise testin tekrar edilebilirliğini hesaplamak için kullanılan yöntemleri belirtin.	1(3)	0(0)	1(7)	0(0)	2(3)
SONUÇLAR							
Katılımcılar	14	Çalışmanın yapıldığı zamanı, katılımcı alımına başlama ve katılımcı alımını sonlandırma tarihleri ile birlikte belirtin.	21(58)	10(56)	7(47)	5(83)	43(57)
	15	Çalışma popülasyonunun klinik ve demografik özelliklerini (örneğin, yaş, cinsiyet, var olan semptomların spektrumu, eşlik eden hastalıklar, uygulanan tedaviler, tedavi merkezleri) raporlayın.	28(78)	12(67)	11(73)	6(100)	57(76)
	16	İndeks test (tanı testi) ve/veya referans standart (altın standart test) uygulanıp/uygulanmayıp, dahil etme kriterlerini sağlayan katılımcıların sayısını bildirin. Katılımcıların neden her iki testi de almadıklarını tanımlayın (akış diyagramı önerilir).	32(89)	9(50)	9(60)	6(100)	56(75)
Test sonuçları	17	İndeks testler (tanı testleri) ile referans standart (altın standart test) arasında geçen zamanı ve bu süreçte herhangi bir tedavi uygulanıp uygulanmadığını belirtin.	15(42)	4(22)	6(40)	1(17)	26(35)
	18	Hedef koşullara sahip olanlarda hastalık şiddetinin (kriter tanımla) dağılımı ve hedef şartlara uymayan katılımcılara ait diğer teşhisleri raporlayın.	15(42)	6(33)	10(67)	5(83)	36(48)
	19	İndeks testlerin (tanı testlerinin) sonuçlarını (belirlenemeyen ve kayıp sonuçlar da dahil olmak üzere) referans standardın (altın standart testin) sonuçları ile karşılaştırarak tablo halinde verin; sürekli sonuçlar için ise test sonuçlarının referans standart (altın standart test) sonuçlarına göre dağılımını verin.	32(89)	13(72)	14(93)	5(83)	64(85)
	20	İndeks test (tanı testi) veya referans standart (altın standart test) uygulanırken karşılaşılan olumsuz durumları raporlayın.	5(14)	13(72)	0(0)	3(50)	21(28)
Tahminler	21.A	Tanı doğruluğu tahminlerini	33(92)	17(94)	13(87)	5(83)	68(91)
	21.B	İstatistiksel belirsizlik ölçütlerini (örneğin, %95 güven aralığı) raporlayın.	12(33)	8(44)	2(13)	2(33)	24(32)
	22	İndeks testlerde (tanı testlerinde) şüpheli sonuçların, eksik (kayıp) yanıtların ve sapan değerlerin nasıl ele alındığını raporlayın.	5(14)	3(17)	4(27)	3(50)	15(20)
	23	Eğer yapıldıysa, katılımcıların alt grupları, gözlemciler veya merkezler arasında tanı doğruluğunun değişkenlik tahminlerini raporlayın.	4(9)	2(11)	1(7)	0(0)	7(9)
	24	Eğer yapıldıysa, testlerin tekrarlanabilirlik tahminlerini raporlayın.	1(2)	0(0)	1(7)	0(0)	2(3)
TARTIŞMA	25	Çalışma bulgularının klinik uygulanabilirliğini tartışın.	35(74)	16(89)	15(100)	6(100)	72(96)

Tanı doğruluğu makalelerinde iki nolu STARD kriterinin sağlanması için Giriş bölümünde, çalışmanın amacının ve/veya hipotezlerinin “tanı doğruluğunun tahmin edilmesi”, “tanı testlerinin karşılaştırılması” ya da “tanı doğruluğunun alt gruplarda karşılaştırılması” gibi ifadeler kullanılarak belirtilmesi gerekmektedir. Değerlendirilen 75 makalenin 59(%79)’unda ilgili kriter sağlanmıştır.

Metod bölümünde, deneklere ait özelliklerin tanımlanmasına ilişkin dört kriter (No:3-6) bulunmaktadır. STARD kriterlerine göre tanı doğruluğu çalışmalarında çalışma popülasyonu, katılımcıların çalışmaya alınma prosedürü, katılımcı örnekleme, verilerin toplanma biçimi raporlanmalıdır. Değerlendirilen 75 makalede bu kriterlerin sağlanma oranları sırasıyla %87, %93, %76 ve %88’dir. Yine STARD kriterlerine göre metod bölümünde Referans standart ve İndeks test’in uygulanmasına ilişkin standartları içeren 5 kriter (No:7-11) bulunmaktadır. Değerlendirilen çalışmalarda bu beş kriterlerin karşılanma yüzdeleri %35 ile %81 arasında değişmektedir. Tanı doğruluğu çalışmalarında tanı doğruluğu ölçülerinin hesaplanmasında veya karşılaştırılmasında kullanılan yöntemler ve Güven Aralıkları raporlanmalıdır. Değerlendirilen çalışmaların 58(%77)’inde bu istatistiksel yöntemler belirtilmiştir. Ayrıca eğer yapılmış ise testin tekrar edilebilirliğini hesaplamak için kullanılan yöntemleri de belirtilmelidir. Bu kriter makalelerin sadece %3’ünde sağlanmıştır.

STARD kontrol listesinde tanı doğruluğu çalışmalarının SONUÇ bölümünde katılımcılara ilişkin sağlanması gereken üç (No:14-16) kriter belirlenmiştir. Buna göre makalede çalışmanın yapıldığı zaman, katılımcı alınma başlama ve sonlandırma tarihleri belirtilmelidir. Bunun yanında makalede çalışma popülasyonunun klinik ve demografik özellikleri raporlanmalı ve İndeks test ve Referans test uygulanma sürecine ilişkin akış diyagramı çizilmesi önerilmektedir. Değerlendirilen çalışmalarda bu kriterlerin sağlanma oranları sırasıyla %57, %76 ve %75’tir.

STARD kontrol listesinde test sonuçlarının nasıl verilmesi gerektiği konusunda dört kriter (No:17-20) önerilmiştir. Değerlendirilen makalelerde bu kriterlerin sağlanma oranları %28 ile %85 arasında değişmektedir.

STARD kontrol listesinde indeks testin sonuçlarına ilişkin elde edilen tahminlerin nasıl raporlanması gerektiğine ilişkin dört (No:21-24) kriter yer almaktadır. Buna göre tanı doğruluğu çalışmalarında duyarlılık, seçicilik, pozitif tahmini değer, negatif tahmini değer v.b. ölçüler ve bunlara ilişkin Güven Aralıkları raporlanmalıdır. Değerlendirilen çalışmaların %91’inde tanı doğruluğu tahminleri yer alırken sadece %32’sinde bunlara ilişkin güven aralıkları raporlanmıştır. Makalelerin %20’sinde indeks test için (tanı testlerinde) şüpheli sonuçlar, eksik (kayıp) yanıtlar ve sapan değerler raporlanmıştır. Makalelerin %9’unda katılımcıların alt grupları, gözlemciler veya merkezler arasında tanı doğruluğunun değişkenlik tahminleri 2 (%3) makalede ise tekrarlanabilirlik tahminleri raporlanmıştır.

STARD kontrol listesinin son maddesi Tartışma bölümünde yer alması gerekenler hakkındadır. Değerlendirilen çalışmaların %96’sında sonuçların klinik uygulanabilirliğini tartışılmıştır.

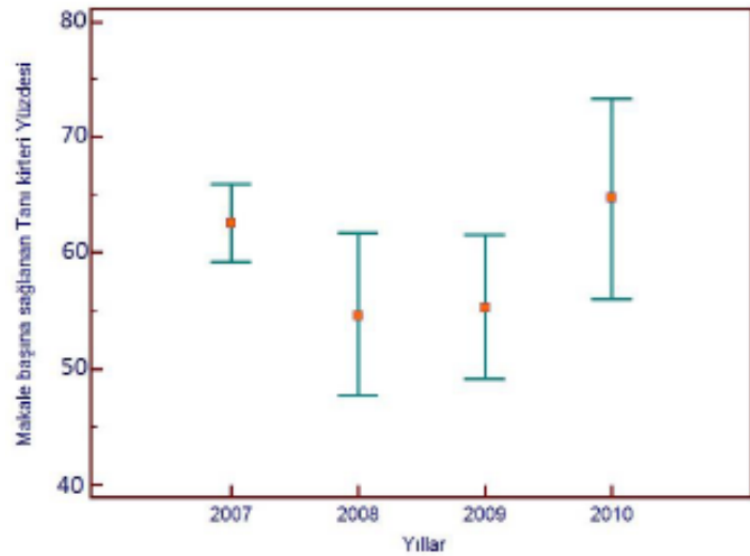
Şekil 1’ye bakıldığında makale başına sağlanan STARD Kriterlerini sağlama yüzdelerine bakıldığında yıllara göre en düşük sağlama oranı %55.4 ile 2009 yılında en yüksek ise %64.7 ile 2010 yılında olduğu belirlenmiştir.

SONUÇ ve ÖNERİLER

Çalışmamızda 2007 yılında yayınlanmış 36, 2008 yılında 18, 2009 yılında 15 ve 2010 yılında yayınlanmış 6 olmak üzere toplam 75 makale Türkçe uyarlaması yapılmış STARD kontrol listesine göre değerlendirilmiştir. En yüksek oranda sağlanan kriter(No:25) %96 ile sonuçların klinik uygulanabilirliğinin tartışılmış olmasıdır.

Çalışmaların büyük bölümünde tanı doğruluğuna ilişkin ölçüler verilirken (%91), sadece %32’sinde bunlara ilişkin Güven Aralıkları verilmiştir. Fakat şüpheli sonuçların, eksik yanıtların ve sapan değerlerin nasıl ele alındığı düşük oranda (%20) raporlanmıştır. Testlerin tekrarlanabilirlik tahminleri sadece 2 çalışmada raporlanmış ve hangi istatistiksel yöntemle elde edildiği de belirtilmiştir. Fakat tekraredilebilirlik tahminlerinin toplam kaç çalışmada yapıldığı bilinmemektedir.

Değerlendirilen makalelerde, çalışmaya alınan deneklere ilişkin kriterler genel



Şekil 1: Yıllara göre makale başına sağlanan STARD kriteri yüzdeleri

olarak yüksek oranda karşılanırken (ortalama %84) bu oranlar yıllar içinde de benzerdir.

Makalelerde testleri uygulayan kişilerin sayısının, eğitiminin ve uzmanlık düzeylerinin belirtilme oranlarının düşük olduğu (%35), benzer şekilde testleri uygulayan gözlemcilerin, katılımcıların klinik bilgilerine ve diğer test sonuçlarına kör olup olmadıklarının belirtilme oranlarının(%35) da düşük olduğu saptanmıştır.

Yıllar içinde daha fazla sayıda STARD kriterinin raporlandığını görmek sevindirici olsa da bu düzeyin yeterli olmadığı düşünülmektedir. Değerlen-

dirilen 75 çalışmanın hiçbirinde akış diyagramına yer verilmemiştir. Detaylı bir biçimde hazırlanan akış diyagramı, özellikle metot bölümünde yer alan kriterlerden bir çoğunu içerecektir.

Çalışmamızda elde edilen sonuçlara göre, 2007-2010 yılları arasında ULAKBİM Tıp Veri Tabanında [Türk Tıp Dizini] Türkçe dilde yayınlanmış tanı doğruluğu çalışmalarının STARD kriterlerini sağlama oranlarının çok yüksek düzeyde olmasa da yeterli olduğu gözlenmiştir. Benzer olarak Smidt ve arkadaşları (2005), MEDLINE veri tabanında etki faktörü 4 ve üzerindeki dergilerde İngilizce dilde yayınlanmış

124 tanı doğruluğu çalışmasını değerlendirdiklerinde, çalışmaların %41'nde STARD kriterlerinin %50'sinden fazlasının sağlandığını bulmuşlardır (10).

Türkçe uyarlaması yapılmış STARD kriterlerinin, ülkemizde tanı doğruluğu çalışması yapacak olan araştırmacıların yanı sıra dergi hakem ve editörleri için de bir kılavuz olması umut edilmektedir. Çalışmalarda metodolojik standartlara uygun ortak bir ölçüt kullanılması, bireysel çalışmaların yanında meta analizi çalışmalarının da kalitelerinin artmasını sağlayacaktır

KAYNAKLAR

1. Genç Y. Tanı Testi Çalışmalarında Metodolojik Standartların Kullanılması. Ankara Üniversitesi Tıp Fakültesi Mecmuası. 2003;56:4: 259-264.
2. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research: Getting better but still not good. JAMA 1995;274:645-651.
3. Lijmer JG, Mol BW, Heisterkamp S.et al. Empirical evidence of deign-related bias in studies of diagnostic tests. JAMA 1999;282:1061-1066.
4. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: A new meta-analytic method. Med Decis Making 1993; 13:313-321.
5. Bossuyt PM, Reitsma JB, Bruns DE. et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. Clinical Chemistry 2003; 49:1 1-6.
6. Whiting P, Rutjes AW, Reitsma JB.et al. The development of QUADAS: A tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003; 3:25.
7. Whiting PF, Weswood ME, Rutjes AW.et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol 2006; 6:9.
8. Bossuyt PM, Reitsma JB, Bruns DE. et al. The STARD statement for reporting studies of diagnostic accuracy: Explanation and elaboration. Clinical Chemistry 2003;49:1 7-18.
9. Genç Y., Albayrak R., Ateş C. et al. Tanı Doğruluğu Çalışmalarının Kalitelerinin Değerlendirilmesi: STARD Kriterlerinin Türkçe Uyarlaması, Ankara Üniversitesi Tıp Fakültesi Mecmuası 2012; 65:137-145
10. Smidt N, Rutjes AWS, Van der Windt AWM.et al. Quality of reporting of diagnostic accuracy studies. Radiology 2005;235:347-353..

