

Ölçeklerin farklı uygulama yöntemlerine göre tutarlılıklarının deęerlendirilmesi

Evaluating consistencies of scales according to different application methods

Pembe KESKİNOĐLU¹, Selahattin ÇOLAKOĐLU²

¹Dokuz Eylül Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Biliřim Anabilim Dalı, İzmir

²Dokuz Eylül Üniversitesi Tıp Fakültesi

ÖZ

Amaç: Klinik arařtırmalarda tanısıl ve tedavi ile ilgili yaklařımlarda geçerlilik ve güvenilirlięi kanıtlanmış çeřitli ölçekler kullanılmaktadır. SF-36, öz bildirime dayalı arařtırmalarda kullanılmaya uygun, kısa kapsamlı, oldukça sık kullanılan bir ölçektir. Bu çalışmanın amacı SF-36 ölçeęinin farklı yöntemlerle uygulanmasını karřılařtırmak ve tutarlılıklarını deęerlendirmektir.

Yöntem: Bu çalışma yöntemsel bir çalışma olarak gerçekteřtirilmiřtir. Dokuz Eylül Üniversitesi Tıp Fakültesi Dahiliye Poliklinięine başvuran 18 yař üzerinde 81 kronik hasta çalışmaya dâhil edilmiřtir. Ölçekle ilgili soruları yanıtlamada yüz yüze görüřme ve kendi kendine yanıt olma üzere iki yöntem uygulanmıřtır.

Bulgular: Her iki grubun cinsiyete göre daęılımları benzerdir. Her iki yöntemde arařtırmaya katılmayı red oranları sırası ile %26 ve %35'tir. Reddetme sıklıęı kendi yanıtlayanlarda daha yüksektir. Her iki yöntemin de iç tutarlılıęı yüksektir. Bununla birlikte, yüz yüze görüřme uygulanan grupta iç tutarlılıęı daha yüksektir.

Sonuç: SF-36 öz bildirim ölçeęinin tutarlılıę düzeyi her iki uygulama yöntemi ile de yüksek saptanmıřtır. Yüz yüze görüřme tutarlılıę düzeyini artırmaktadır.

Anahtar kelimeler: Ölçeklerin tutarlılıęı, ölçeklerin uygulama yöntemi

ABSTRACT

Objective: Various scales with proven validity and reliability are used in diagnostic and therapeutic approaches in clinical trials. Short form health survey (SF-36) is often used in researches based on self-report. It has a short, and easily applicable content. The aim of this study was to compare the different application methods of the SF-36 scale and to evaluate the consistencies of these methods.

Methods: This study is a methodological study and 81 chronic patients aged over 18 years who applied to Dokuz Eylül University Medical School Department of Internal Medicine were included in the study. Two different methods were used as face to face interview and self administration.

Results: Gender distribution of the groups were similar. The refusal rates of the These two methods were rejected by 26% and 35% of the study participants, respectively. Percentage of the subjects who refused to fill the questionnaire were higher than the subjects who accepted. Both methods were found to have quite high degrees of consistency. However, the face to face interview had a higher consistency than the self administration (Cronbach alfa 0.934, 0.872, respectively).

Conclusion: The consistency of the SF-36 scale is quite high. Face to face interviews increase the level of consistency.

Key words: The consistency of the scale, method of administrating scale

Alındıęı tarih: 12.01.2016

Kabul tarihi: 18.01.2016

Yazıřma adresi: Doç. Dr. Pembe Keskinođlu,
Dokuz Eylül Üniversitesi Tıp Fakültesi, İzmir
e-mail: pembe.keskinoglu@gmail.com

GİRİŞ

Araştırmalarda karşılaşılan en önemli sorunlardan biri yapılan araştırmada hangi amaçların gerçekleştirileceği, hangi sorunların çözüleceği ve çözüm için nasıl bir ölçüm yapılacağı konusudur. Bu sorunun çözümünde, özellikle nitel konuların sınıflandırılarak ölçülmesinde, ölçeklerle (scale) yanıt aranır. Ölçekler ölçme göstergeleridir ⁽¹⁻⁴⁾.

Bilimsel ölçümler, belirli bir olguyla ilgili olarak ölçüt ve norm değerleri oluşturmaya yönelik veya önceden belirlenmiş olan norm ve ölçüt değerlerle karşılaştırma yapmaya yöneliktir. Ölçümlerde eğer önceden saptanmış bir ölçüt değer veya norm değeri varsa güvenilirliği hesaplamak görece kolaydır. Bu tür ölçüt değerlerin bulunmadığı durumlarda ise ölçüm verilerinin güvenilirliğinin test edilmesi gereklidir. Güvenilirlik, herhangi bir ölçüm işleminin en öncelikli karşılanması gereken temel kriteridir ⁽⁴⁻⁶⁾. Güvenirlik ölçümler arası tutarlılığın göstergesidir. Tutarlılık sağlanamayan ölçeğin ya da ölçümün geçerliliğinin test edilmesi gereksizdir.

Ölçekler, madde sayısına ve uygulanaşına göre sürekli, tekli ve çoklu ölçekler şeklindedir ⁽⁷⁾. Ölçek türlerinin fark, sıra, uzaklık ve başlangıç noktalarına göre farklı özellikleri bulunmaktadır. Ölçülen konu veya durumun farklı özellikleri için farklı alt boyutlar içerebileceği gibi tek boyutlu ölçekler de olabilir. Ölçeklerde maddelerin yanıtları isimsel (nominal), sıralı (ordinal), aralıksal (interval) ve oransal (ratio) şeklinde alınabilir ^(4,5,7). Yanıtların sayısallaştırılmasında da sorunlar bulunmaktadır ⁽⁸⁾.

Klinik araştırmalarda çeşitli alanlarda geçerlilik ve güvenilirliği kanıtlanmış birtakım ölçekler kullanılmaktadır. Kısa form yaşam kalitesi ölçeği (Short Form Health Survey-36) sağlık bakım araştırmalarında en iyi bilinen ve en yaygın kullanılan sağlıkla ilişkili genel yaşam kalitesi ölçeğidir. SF-36, öz bildirim dayalı, araştırmalarda kullanılmaya uygun, kısa kapsamlı, psikometrik özellikleri açısından güçlü genel sağlık anketidir. Her yaştaki hastaya uygulanabilmekte, hastalığın etkilerini değerlendirmede ve farklı tedavilerin yararlarını karşılaştırmada sıklıkla

kullanılmaktadır ⁽⁹⁾. Ölçeğin Türkçe geçerlilik ve güvenilirliği yapılmıştır ⁽¹⁰⁾. Ölçeklerin uygulanmasında kişilere ölçek verilerek, kendilerinin soruları okuyup, yanıtlamaları istenmektedir. Bu en yaygın uygulama yöntemidir. Kişinin algıladığı durumunu yansıtmaması için, anketi uygulayacak başka bir kişi tarafından standartlaştırılmamış soru ve açıklamalara yer vermeden, ölçeği yanıtlayan kişinin kendi kendine yanıtlaması ile en iyi sonuçların alınabileceği düşünülmektedir. Ancak yanıtlayanların okuma-anlama ve ilgilerini ölçek için yoğunlaştırma sorunlarının göz ardı edildiği ve gerçeğe yakın yanıtların alınmasının güçleşebileceği de tartışılmaktadır.

Bu çalışmada, SF-36 ölçeğinin yüz yüze görüşme ve kendi başına soruları okuyup yanıtlaması yöntemleri ile uygulanması, uygulama yöntemine göre tutarlılıklarının değerlendirilmesi amaçlanmıştır.

GEREÇ ve YÖNTEM

Bu çalışma yöntemsel bir çalışma olarak gerçekleştirilmiştir. Dokuz Eylül Üniversitesi Tıp Fakültesi Dahiliye Polikliniğine başvuran okuma-yazması olan, 18 yaş üzerinde gönüllü, kronik hastalar çalışmaya dâhil edilmiştir. Bu araştırmada öz bildirim dayalı ölçeklerden en bilineni SF-36, dâhiliye polikliniğe başvuran benzer hasta gruplarına farklı iki yöntemle uygulanmış ve tutarlılıkları incelenmiştir. Aynı günlerde gelen hastalara kura ile hangi yöntemle anketi yanıtlayacakları belirlenmesi hedeflenmiştir. Herhangi bir girişim uygulanmayacak olmasına rağmen, farklı iki yöntemle gerçekleştirilecek ölçek uygulamasında yanıtların tutarlılıklarını etkileyebilecek sonuçların gruplara benzer dağılmasını sağlamak amacıyla randomizasyon yapılmıştır. Ancak bir gruptaki hastaların yaklaşık %35'ine uygulama düzeyinde randomize örnekleme gerçekleştirilememiştir. Yaşlı hastalar kendi başına ölçeği doldurmayı reddetmişler, görüşmecinin soruları sorarak uygulamasına izin vermişlerdir.

Çalışmanın temel hipotezleri olan ölçeklerin tutarlılıkların katsayısı ve yanıtlamayı reddetme oranları değerlendirilmiştir. Katılımcıların diğer özellikleri

(yař, cinsiyet) aısından gruplar arası farklılık olup olmadığını karřılařtırılmıřtır. Temel hipotez iin karřılařtırma yapılmayacađından rnek byklđ hesaplaması yapılmamıřtır. Hastalar leđi kendi kendine ve yz yze yanıtlanmasına gre iki gruba ayrılmıřtır. Parametrik kořulların sađlanması iin her iki gruba en az 30 kiřinin alınması planlanmıřtır. alıřmada gruplar ařađıdaki řekilde oluřturulmuřtur:

1. Kendi kendine anketi yanıtlayan hasta grubu: Belirlenen veri toplama sresinde (1 ay) bu gruba dhil olan 63 kiřiye arařtırmanın amacı anlatılarak arařtırmaya katılmaya davet edilmiř, ancak 22'si (%35) katılımı reddetmiřtir.

2. Yz yze anketi yanıtlayan hasta grubu: Bu grupta, 54 kiřiye arařtırmanın amacı anlatılmıř, 14' (%26) katılımı reddetmiřtir. Veri, aynı zaman sresinde toplanmıřtır.

Bu alıřmada her iki uygulama grubundan SF-36 yařam kalitesi leđini yanıtlanması istenmiřtir. Farklı iki uygulama ynteminde benzer sonuların alınma durumu deđerlendirilmiřtir. SF-36 fiziksel ve mental olmak zere iki ana boyut ve sekiz alt lekten oluřmakta ve 36 soru iermektedir. lekte fiziksel fonksiyonlar (10 soru), fiziksel sađlıđın kısıtlılıktaki rol (4 soru), duygusal oluř (3 soru), sosyal iliřkiler (2 soru), ađrı (2 soru), genel sađlık (5 soru) sorgulanması ile ilgili sorular bulunmaktadır. 19., 23., 24., 25., 26., 27., 28., 29., 30., 31., 32., 33., 34., 35. ve 36. sorularda yanıtlan olumludan olumsuzu řeklindeki yn terstir. Kendi kendine leđi okuyup yanıtlayanların ve yzyze grřmede kendine sorularak yanıtlan alındađı uygulamalarda, likert leđindeki ters yanıtlan ne kadar anlařılarak dođru bir řekilde yanıtlandıđı uygulama yntemlerinin bařarısını ortaya koymada kriter olarak deđerlendirilmiřtir. leklere verdikleri yanıtlan dıřında tanımlayıcı olarak hastaların ayrıca yař, cinsiyet ve geliř nedeni bilgisi de kaydedilmiřtir.

alıřma iin, DETF giriřimsel olmayan Etik Kuruldan izin alınmıřtır. Katılan gnlllere alıřma ile ilgili bilgi verilmiř, bilgilendirilmiř onayları szli olarak alınmıřtır.

BULGULAR

Anketi kendisi uygulayan grupta uygulamayı reddedenlerin yař ortalaması 47,1 yıl, kabul edenlerin yař ortalaması 41,5 yıldır. Kabul etmeyenlerin yařı anlamlı dzeyde daha yksektir ($p<0,05$). Anketin grřmeci tarafından uygulandıđı grupta katılımı kabul etmeyenlerin yař ortalaması (42,6 yař) yanıtlayanların yař ortalamasından (47,8) kk bulunsa da bu fark istatistiksel olarak anlamlı deđildir. lek uygulama gruplarının belirlenmesinde yařa bađlı bir fark ortaya ıkmıřtır, genler yz yze grřmeyi, daha yařı ileri olanlar kendi kendilerine leđi yanıtlanmayı reddetmiřtir.

alıřmaya katılan farklı iki yntemle leđi yanıtlayan grupların yař ve cinsiyet dađılımları Tablo 1'de gsterilmektedir. Cinsiyete gre dađılımları anketi farklı uygulayan grupların her ikisinde benzer olduđu bulunmuřtur. leđi kendisi uygulayan gruptakilerin yař ortalaması anlamlı olarak daha kk bulunmuřtur.

Tablo 1. Arařtırma grubu hastaların sosyo-demografik zelliklerinin dađılımı.

zellikler	Kendisi uygulayan grup	Grřmeci uygulayan grup	Toplam
Cinsiyet n (%)*			
Erkek	23 (56,1)	15 (37,5)	0,094**
Kadın	18 (43,9)	25 (62,5)	
Yař (ortalama±sapma)	40,5±14,5	47,8±18,4	0,046***

*Stn yzde, **Ki-kare test, ***t-test

Kendisi yanıtlayan grupta SF-36'da leđine verdikleri yanıtlan soru temelinde %4,2'sinin yanıtlanmadan boř bırakıldıđı saptanmıřtır. Grřmecinin uyguladıđı anketlerin toplamında boř yanıt bulunmadıđı grlmřtir. Bu fark istatistik olarak anlamlıdır ($p<0,001$, ki-kare=37,35).

Her iki uygulama yntemine gre soru temelinde ortalamalar Tablo 2'de sunulmuřtur.

Ađrı, fiziksel fonksiyon, fiziksel fonksiyona bađlı rol kısıtlılıđı ve duygusal alt boyutta birer soruda her iki uygulama yntemine gre yanıt ortalamaları arasında anlamlı fark vardır. Bu sorular, anlam ve cmle yapısı bakımından yine incelendiđinde, anlařılmasında glk yaratabilecek zellikler ve anlam aısından

hastanın yanıtlamaktan çekineceği içerikleri taşımadıkları görülmüştür. Likert ölçek yanıtları açısından iki sorunun olumlu iki sorunun olumsuz yönde olduğu görüldü. Bu sorular için iki grup arasında farkın bir grup tarafından sorunun anlaşılmasından mı, yoksa katılımcıların dikkatsizliğinden mi olduğu bu analizde saptanamamıştır.

Tablo 2. Anket uygulama gruplarına göre soruların ortalamalarının karşılaştırılması.

	Görüşmeci		Kendi yapan		p
	Ortalama	Sapma	Ortalama	Sapma	
Genel sağlık	27,50	46,53	16,54	19,04	0,000
	43,33	52,08	24,51	21,02	0,123
	52,67	59,43	26,51	27,65	0,32
	60,67	69,14	23,18	23,93	0,154
	37,33	52,00	22,73	25,30	0,017
Ağrı	56,67	65,14	28,32	31,19	0,259
	52,00	70,59	35,85	24,73	0,021
	63,33	76,39	34,57	23,10	0,084
	38,33	61,11	38,69	36,08	0,016
	80,00	81,94	31,07	29,65	0,796
Fiziksel fonksiyon	51,67	56,94	44,50	43,35	0,628
	61,67	68,06	42,92	38,08	0,524
	91,67	84,72	18,95	33,42	0,294
	68,33	72,86	35,92	39,00	0,630
	56,67	70,83	43,02	36,60	0,153
	85,00	79,17	26,75	34,59	0,453
	93,33	87,50	17,29	30,18	0,330
	98,33	88,89	9,13	29,55	0,076
	50,00	76,47	50,85	43,06	0,030
	43,33	58,82	50,40	49,96	0,222
Rol kısıtlaması	56,67	68,75	50,40	47,09	0,333
	46,67	67,65	50,74	47,49	0,093
Duygusal rol kıst.	60,00	67,65	49,83	47,49	0,532
	63,33	52,94	49,01	50,66	0,409
	63,33	69,70	49,01	46,67	0,600
	58,33	61,72	29,60	34,19	0,679
	58,67	62,29	22,85	31,72	0,596
Enerji	74,00	78,29	19,76	24,43	0,445
	47,33	60,00	25,99	28,70	0,069
	58,33	64,29	31,03	30,49	0,439
	27,50	58,09	25,72	36,78	0,000
	59,17	62,88	26,65	39,59	0,662
Duygusal	58,67	61,14	22,24	27,42	0,694
	69,33	65,71	22,12	27,26	0,563
	60,00	64,12	18,19	23,50	0,441
Sosyal	72,50	75,76	26,55	31,55	0,661

Tablo 3'te gösterilen her iki uygulama yöntemine göre tutarlılığın iyi olduğu gözlenmiştir. Cronbach's Alfa değeri her iki uygulama yönteminde de 0.85 üzerinde olmasına karşın, görüşmecinin uyguladığı yöntemin tutarlılık katsayısı daha yüksek olarak saptanmıştır. Yani görüşmecinin uyguladığı yöntemin daha tutarlı olduğu söylenebilir.

Tablo 3. Yöntemlere göre tutarlılık analizi.

	Cronbach's Alpha
Görüşmeci	0,934
Kendi yapan	0,872

Ölçeğin her maddesinin tutarlılık için katkısı, madde silinirse Cronbach's Alpha katsayısında gerçekleşecek değişimle incelenir. Madde silindiğinde tüm soruların dâhil olduğu analizdeki alfa değerinden daha büyük bir değere ulaşırsa bu maddenin güvenilirliği olumsuz etkilediği düşünülür. Uygulaması iyi gerçekleştirilen ölçeklerde maddenin olumsuz katkısı ile ilgili bu durumda, sorunun anlaşılabilirliğinin iyi olmadığı ya da bu sorunun gereksiz olduğu sonucuna ulaşılır.

Çalışmamızda madde istatistikleri incelendiğinde, kendi başına uygulayan grupta 19. 28. ve 30. sorular hesaplamadan çıkarıldığında testin tutarlılığının arttığı gözlenmiştir. Başka bir deyişle bu sorular testi yapanlar tarafından, iyi okunarak yanıtlanmamıştır. Çünkü yüz yüze görüşme ile alınan sonuçlarda böyle bir sonuç ortaya çıkmamıştır.

TARTIŞMA ve SONUÇ

Ölçeklerin ilk kez yapılandırılırken, içerik, kapsam, madde temelinde ilişkiler, maddelerin sıralaması, her maddenin yanıtlarının ölçeklendirilmesindeki sorunlar üzerinde de çabalar devam etmektedir⁽¹¹⁻¹⁵⁾. Ölçeklerin uygulanma yöntemleri, bu ölçeklerin sonuçları açısından çok önemlidir. Bu sonuçlar tıbbi verilere yansıtılacak ve tıbbi kararlara destek olacaksa doğru yöntemle, doğru şekilde yapılmasına daha çok dikkat edilmelidir. Çalışmanın başında araştırmayı reddetme durumları, kabul ve reddedenlerin yaş gruplarına göre dağılımı incelendiğinde, ölçeklerin uygulandığı çalışmalarda temsiliyet açısından önemli bir sorunla karşılaşıldığı sonucuna ulaşılabilir. Red oranlarının yüksek olması ve kendisi yanıtlayanlarda daha ileri yaş gruplarının çalışmaya katılmayı istememesi, tüm ölçek çalışmalarında bu yanlılığın olabileceği ve kontrolünün de çok iyi yapılmayacağı

düşündürmüştür. Anket uygulama gruplarına göre soruların ortalamaları karşılaştırıldığında farklılık olan sorular olduğu saptanmıştır. Bu soruların içerik değerlendirmesinde anlaşılabilirlikla ilgili değil soru ve yanıtların dikkatli şekilde algılanıp yanıtlanmamasından kaynaklanabileceği sonucuna varılmıştır. Sorunun doğru anlaşılıp yanıtlanmasında kendi yanıtlayan grupta sorun olduğu başka bir bulgu ile desteklendiği görülmüştür. Ölçekteki bazı soruların ters yanıtli ve dikkatli okunması gereken sorular olduğu dikkati çekmektedir. Bu sorular çıkarıldığında uygulamayı kendi yapan grupta testin tutarlılığının arttığı görülmektedir. Bunun nedeninin soruların anlaşılmasından değil, katılımcıların soruları dikkatle okumadıklarından kaynaklandığını düşünmekteyiz.

Sonuç olarak, uygulanan SF-36 ölçeğinin tutarlılık düzeyinin yüksek olduğu, yüz yüze görüşme ile ölçeğin çalışmaya katılanların kendisi tarafından uygulanmasına göre tutarlılık düzeyinin daha yüksek olduğu görülmektedir. Klinik ve/veya topluma dayalı çalışmalarda uygulanan ölçeklerin uygulandığı toplum için geçerlilik ve güvenilirlik değerlendirmelerinin doğru yapılabilmesi için seçilen uygulama yöntemlerinin eksiklikleri belirlenmelidir. Belirlenen eksikliklerin giderilmesi için standart uygulamalar üzerinde çalışmalı, tutarlılık kesinlikle sağlanmalıdır.

KAYNAKLAR

1. Kaptan S. Bilimsel Araştırma ve İstatistik Teknikleri, Rehber Yayınevi, Ankara; 1995.
2. Tonta Y. Bilimsel Araştırma ve İstatistik Tekniklerin Kullanımı ve Bulguların Sunumu Üzerine. *Türk Kütüphaneciliği* 1999;12(2):112-124.
3. Özdamar K. Paket Programlar ile İstatistiksel Veri Analizi, Kaan Kitabevi, Eskişehir; 2004.
4. Şencan H. Sosyal ve Davranışsal Ölçümlerde Güvenilirlik ve Geçerlilik, Seçkin Yayıncılık, 2005.
5. Alpar R. (2010), Spor, Sağlık ve Eğitim Bilimlerinde Uygulamalı İstatistik ve Geçerlilik-Güvenilirlik Birinci Baskı, Ankara: Detay Yayıncılık.
6. Embretson SE, Hershberger HL. The new rules measurements, in Wright B. "Fundamental Measurement For Psychology" Psychology Press, Taylor and Francis New York 2012.
7. Velleman P, Wilkinson, L. Nominal, Ordinal, Interval and Ratio Typologies are Misleading. *The American Statistician* 1993;1:65-72.
8. Araştırmalarda Likert Yanılgıları yunus.hacettepe.edu.tr/~alerbasi/yazilar/Likert.doc erişim tarihi:04.01.2016.
9. Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36). Conceptual framework and item selection. *Med Care* 1992;30:473-83. <http://dx.doi.org/10.1097/00005650-199206000-00002>
10. Pınar R. Sağlık araştırmalarında yeni bir kavram: Yaşam kalitesi, bir yaşam kalitesi ölçeğinin kronik hastalarda geçerlilik ve güvenilirliğinin sınanması. *Hemşirelik Bülteni* 1995;9:85-95.
11. Sünbül SD. (2006). Farklı Likert Tipi Ölçek Geliştirme Teknikleri ile Geliştirilen Tutum Ölçeklerinin Psikometrik Özelliklerinin Karşılaştırılması. Yüksek Lisans Tezi.
12. Mersin Üniversitesi Sosyal Bilimler Enstitüsü, Mersin. Tavşancıl E. (2010). Tutumların Ölçülmesi ve SPSS ile Veri Analizi. (4. Baskı). Ankara: Nobel Yayın Dağıtım.
13. Likert Tipi Ölçek Geliştirme Klavuzu. Ankara: Psikologlar Derneği Yayınları. Tezbaşaran, A. (2004).
14. Likert Tipi Ölçeklere Madde Seçmede Geleneksel Madde Analizi Tekniklerinin Karşılaştırılması. *Türk Psikoloji Dergisi*. TPD Yayınları, 19(54):77-87.
15. Walsh WB, Betz NE. (1995). *Test and Assessment*. 3th. Ed. New Jersey: Prentice Hall