

Meta Analizde Cochran Q Heterojenlik Testi Sonucuna Göre Heterojenlik Ölçümleri İçin Kesim Noktalarının Belirlenmesi: Bir Simülasyon Çalışması

Determination of Cut-Off Values for Heterogeneity Measurements According to Cochran Q Heterogeneity Test Result in Meta-Analyses: A Simulation Study

Semra ERDOĞAN,^a
E. Arzu KANIK^a

^aBiyostatistik ve Tıbbi Bilişim AD,
Mersin Üniversitesi Tıp Fakültesi,
Mersin

Geliş Tarihi/Received: 22.02.2011
Kabul Tarihi/Accepted: 06.04.2011

Yazışma Adresi/Correspondence:
E. Arzu KANIK
Mersin Üniversitesi, Tıp Fakültesi,
Biyostatistik ve Tıbbi Bilişim AD, Mersin,
TÜRKİYE/TURKEY
arzukanik@gmail.com

ÖZET Amaç: Bu çalışma ile meta analizi çalışmalarında kullanılacak olan heterojenlik testinin ve heterojenlik ölçümlerinin I. ve II. Tip hatadan etkilenme durumlarının ortaya konması, heterojenlik testine ait anlamlılık seviyeleri ve bu anlamlılık seviyelerinde heterojenlik ölçümlerine ait en uygun kesim noktalarının belirlenmesi hedeflenmiştir. **Gereç ve Yöntemler:** Bu çalışmada, etki büyüklüğü olarak OR değerinin beklenen değerinin 1'e eşit ve meta analizine dahil edilen çalışmaların homojen ve heterojen yapıda olduğu durumlar göz önünde bulundurularak, farklı çalışma sayıları ve örneklem büyüklükleri için binom dağılımı gösteren rasgele çalışmalar üretilmiştir. Üretilen bu çalışmalar için heterojenlik testine ait anlamlılık değeri ve heterojenlik ölçümleri hesaplanmıştır. Bu işlemler 1000 kez tekrar edilmiştir. **Bulgular:** Meta analizi çalışmalarında Cochran Q testine ait 0.05 veya 0.10 anlamlılık seviyesi için, gerçek homojen ve heterojenleri en iyi ayırt edebilen heterojenlik ölçümü olarak τ^2 test istatistiğinin kullanılması önerilmektedir. Ayrıca, meta analizine dahil edilen çalışma sayısının 20 olması durumunda, Cochran Q testine ait 0.05, 0.10 ya da 0.30 anlamlılık seviyesi için testin gücünün %100'lere ulaştığını söyleyebiliriz. Örneklem büyüklüğünün 1000'in altında olduğu durumlarda tüm anlamlılık seviyesinde testin gücü %90'ın altına düşerken, 2000'in üzerindeki örneklem büyüklükleri için testin gücü %100'e ulaşmaktadır. **Sonuç:** Bu çalışmada, tedavi ve kontrol gruplarında, örneklem büyüklükleri eşit olmak üzere, birleştirilmiş çalışmalarda örneklem büyüklükleri dikkate alınmıştır. Gruplar arasında örneklem büyüklüklerinin dengeli olmadığı durumlar da incelenebilir.

Anahtar Kelimeler: Heterojenlik; meta analizi; Cochran Q, heterojenlik ölçümleri, çalışmalar arası varyans

ABSTRACT Objective: In this study, it is aimed to put forward heterogeneity test that will be used in meta analysis study and how measures of heterogeneity are affected by tip 1 and 2 error, and to determine the significant levels of heterogeneity test and the most appropriate cutting points of heterogeneity measurements in these significant levels. **Material and methods:** In this study, randomized studies showing binomial distribution for different number of studies and sample sizes were produced by the fact that the conditions in which the studies included in meta-analysis are homogeneous and heterogeneous and as effect size, OR, the value of the expected value is equal to 1 are taken into account. For these produced studies, significant value for heterogeneity test and measures of heterogeneity were calculated. This process was repeated 1000 times. Results: In Meta-analysis studies, for significance level 0.05 or 0.10 of Cochran's Q test, τ^2 test statistic is recommended to use as a measure of heterogeneity which is the best to distinguish between real homogeneous and heterogeneous. Also, if the number of studies included in meta-analysis is 20, for significance level 0.05, 0.10 or 0.30, it can be said that the power of the test has reached to 100%. While the power of the test is below 90% at all the significant value in the sample size is under 1000 cases, the power of the test is reached 100% in the sample size is upper 2000 cases. Conclusion: In this study, as sample sizes are equal, in treatment and control groups, sample sizes in the combined studies were taken into consideration. Between the groups in which the sample sizes are not balanced can also be investigated.

Key Words: Heterogeneity; meta analysis; Cochran Q, measures of heterogeneity, between study variance

Çalışmalar arasındaki varyans veya heterojenliğin değerlendirilmesi meta analizinde önemli bir konudur. Çalışmalar arasındaki gerçek heterojenliğin olup olmadığını değerlendirmede kullanılan en basit ve çok yaygın olan yaklaşım Cochran tarafından önerilen (k-1) serbestlik dereceli Ki-Kare heterojenlik testidir ve bu test Q istatistiği olarak bilinmektedir ve aşağıdaki gibi hesaplanır. Literatürde standart χ^2 testi olarak da geçmektedir.¹

$$Q = \sum_{i=1}^k w_i Y_i^2 - \frac{\left(\sum_{i=1}^k w_i Y_i \right)^2}{\sum_{i=1}^k w_i} \quad (1.1)$$

Eşitlik 1.1'de yer alan Y_i , etki büyüklüğü değerini göstermektedir. w_i , her bir çalışmanın ağırlığı olup, o çalışmaya ait varyansın tersi olarak hesaplanır.

Heterojenliğin test edilmesi, Q istatistiğinin düşük güce sahip olmasından ya da heterojenlik ölçümü miktarının önemsiz bulunmasından dolayı yanlış anlaşılmalıdır. Q istatistiğinin gücü meta analizine dahil edilen çalışma sayısından etkilenmekte ve çalışma sayısı arttıkça Q değeri de artmaktadır. Meta analizine dahil edilen çalışma sayısı az ise testin gücü düşük, fazla ise testin gücü yüksek olmaktadır. Baujat, anlamlılık testi sonucunun heterojenlik için kuvvetli bir kanıt olarak değerlendirilmesi gerektiği üzerinde durmaktadır.² Ancak bu durum meta analizine dahil edilecek olan çalışmaların sayısı az olduğunda doğrudur. Bununla birlikte eğer meta analizine pek çok çalışma dahil edilmişse (çalışma sayısı çok fazla ise) meta analizi için önemsiz olan küçük bir miktardaki heterojenlik istatistiksel olarak anlamlı bulunabilir. Yani pek çok çalışma meta analizine dahil edildiğinde heterojenlik miktarı çok küçük de olsa bu değer test edildiğinde istatistiksel olarak anlamlı kabul edilebilir. Böyle bir durumda heterojenliğin olup olmamasına da klinik olarak önemli olup olmamasına da karar vermek zorlaşacaktır.³⁻⁵

Meta analizi çalışmalarında en uygun heterojenlik ölçümü, çalışma sayısından ve örneklem büyüklüğünden etkilenmemelidir. Simülasyon bulgularına göre, çalışma sayısından ve örneklem

büyükliğinden etkilenmeyen heterojenlik ölçümleri I^2 ve τ^2 istatistikleridir. Ancak, bu heterojenlik ölçümlerinin güven aralıkları ile ilgili problemler olduğu gözlenmiştir. Buradan yola çıkarak, meta analizi çalışmalarında kullanılacak olan heterojenlik testinin ve heterojenlik ölçümlerinin I. ve II. Tip hatadan etkilenme durumlarının ortaya konması ve heterojenlik ölçümlerine ait kesim noktalarının belirlenmesinin gerekliliği sonucuna varılmıştır.

Literatürde Cochran Q istatistiğinin anlamlılık seviyesi üzerine pek çok tartışma yer almakta ve anlamlılık seviyesi için 0.05 yerine 0.10 alınması önerilmekte ve ayrıca heterojenliği sadece istatistiksel test ile değerlendirmek değil, aynı zamanda meta analizi üzerinden heterojenliğin etkisinin de tanımlanması tavsiye edilmektedir.⁶

Bu amaçla, homojen ve heterojen yapıda çalışmaların meta analizine alındığı durumlara ait veriler üretilerek, Cochran Q heterojenlik testine ait anlamlılık seviyesi (p) ve heterojenlik ölçümleri için ROC analizleri yapılmış ve uygun kesim noktalarının belirlenmesi amaçlanmıştır.

GEREÇ VE YÖNTEMLER

Heterojenliğin varlığında heterojenliğin miktarının belirlenmesi için pek çok heterojenlik ölçümü ortaya atılmıştır. Literatürde 4 farklı ölçüm yer almaktadır. H^2 , R^2 , τ^2 ve I^2 istatistikleridir ki bunlardan ilk ikisinin karekökü alınarak verilmektedir.³

H İSTATİSTİĞİ

Higgins ve Thompson tarafından elde edilen bir indekstir ve bu indeks Birge'nin oranı olarak da adlandırılmaktadır. Cochran'ın Q istatistiğinden yararlanılarak aşağıdaki gibi tanımlanmakta olup 1 ile ∞ arasında bir değer almaktadır.⁷

$$H^2 = \begin{cases} \frac{Q}{(k-1)}, & Q > df \\ 1, & Q \leq df \end{cases} \quad (2.1)$$

Formülde k, meta analizine dahil edilen çalışma sayısını göstermektedir. H'nin 1 olması homojenliğin mükemmel olduğunu göstermektedir. H istatistiğine ait %95 güven aralığı,

$$\exp[\ln(H) \pm 1.96 \times SE(\ln(H))] \quad (2.2)$$

şeklinde tanımlanmaktadır. $SE(\ln(H))$ için, Cochran Q istatistiğinin çalışma sayısından büyük veya küçük oluşuna göre farklı formüllerle hesaplanmaktadır. Buna göre, $Q > k$ olduğu durumda,

$$SE(\ln(H)) = \frac{1}{2} \left(\frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{(2k-3)}} \right) \quad (2.3)$$

şeklinde, $Q \leq k$ olduğu durumda ise,

$$SE(\ln(H)) = \sqrt{\frac{1}{2(k-2)} \left(1 - \frac{1}{3(k-2)^2} \right)} \quad (2.4)$$

şeklinde hesaplanmaktadır.^{3,7} H'nin değeri çalışma sayısı küçük olduğunda ($k < 8$) küçük bir yanlılık gösterebilir. H'nin değişkenliği çalışma sayısı az olduğunda büyüktür, böylece pratikte orta derecede olan bir heterojenliği bile ayırt etmek zorlaşabilmektedir.⁷

H istatistiği heterojenlikten dolayı varyans şişirme faktörü olarak yorumlanmaktadır. H^2 'nin düzenlenmiş hali ise aşağıdaki gibi verilmekte olup 0 ile ∞ arasında bir değer alır.³

$$H_M^2 = \begin{cases} \frac{Q - (k-1)}{(k-1)}, & Q > df \\ 0, & Q \leq df \end{cases} \quad (2.5)$$

τ^2 İSTATİSTİĞİ

Çalışmalar arası varyans olarak ifade edilir ve aşağıdaki şekilde hesaplanır.

$$\hat{\tau}^2 = \begin{cases} \frac{Q - (k-1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} & Q > (k-1) \\ 0 & Q \leq (k-1) \end{cases} \quad (2.6)$$

τ^2 istatistiğine ait %95 güven aralığının alt sınır ve üst sınır değerleri Eşitlik 2.7 ve 2.8'deki gibi hesaplanmaktadır.

$$LL_{\tau^2} = \frac{(k-1) * (L^2 - 1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \quad (2.7)$$

$$UL_{\tau^2} = \frac{(k-1) * (U^2 - 1)}{\sum w_i - \frac{\sum w_i^2}{\sum w_i}} \quad (2.8)$$

Formüllerde yer alan L ve U, Eşitlik 2.9 ve 2.10'daki gibi hesaplanır.

$$L = \exp(0.5 * \ln(\frac{Q}{k}) - 1.96 * B) \quad (2.9)$$

$$U = \exp(0.5 * \ln(\frac{Q}{k}) + 1.96 * B) \quad (2.10)$$

Eşitlik 2.9 ve 2.10'a ait formüllerde bulunan B, Cochran Q istatistiğinin çalışma sayısından büyük veya küçük oluşuna göre farklı şekillerde hesaplanmaktadır. Buna göre, eğer $Q > k$ ise,

$$B = 0.5 * \frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{(2k-3)}} \quad (2.11)$$

şeklinde $Q \leq k$ ise,

$$B = \sqrt{\frac{1}{2(k-2) \left[1 - \left(\frac{1}{3(k-2)^2} \right) \right]}} \quad (2.12)$$

şeklinde hesaplanır. τ^2 , alt sınır veya üst sınır değerlerinden herhangi biri "0"dan küçük ise gerçek değeri yerine 0 değeri atanır. Eğer alt sınır değeri 0'dan büyük ise τ^2 istatistik olarak anlamlı kabul edilir.⁸

R İSTATİSTİĞİ

H istatistiğine benzerdir ve τ^2 'den hesaplanır ve aşağıdaki şekilde formüle edilir.

$$R^2 = \frac{\tau^2 + \sigma^2}{\sigma^2} \quad (2.13)$$

R=1 olması homojenliğin mükemmel olduğunu göstermektedir. R'nin değeri ve %95 güven aralıkları hesaplamaları ve yorumlamaları H istatistiği ile çok benzerdir. Buna göre R'ye ait %95 güven aralığı aşağıdaki şekilde hesaplanmaktadır.⁷

$$\exp[\ln(R) \pm 1.96 * SE(\ln(R))] \quad (2.14)$$

$SE(\ln(R))$ için, Cochran Q istatistiğinin çalışma sayısından büyük veya küçük oluşuna göre farklı formüllerle hesaplanmaktadır. Buna göre, $Q > k$ olduğu durumda,

$$SE(\ln(R)) = \frac{1}{2} \left(\frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{(2k-3)}} \right) \quad (2.15)$$

şekilde, $Q \leq k$ için;

$$SE(\ln(R)) = \sqrt{\frac{1}{2(k-2)} \left(1 - \frac{1}{3(k-2)^2} \right)} \quad (2.16)$$

şeklinde hesaplanmaktadır.⁷

I² İNDEKSİ

Son yıllarda, Higgins ve Thompson tarafından önerilen ve Cochran'ın Q indeksinden ve H istatistiğinden yararlanılarak oluşturulmuş bir indeks olup aşağıdaki şekillerde formüle edilmektedir.^{4,5,7}

$$I^2 = \begin{cases} \frac{Q - (k-1)}{Q} * 100\%, & Q > (k-1), \\ 0, & Q \leq (k-1) \end{cases} \quad (2.17)$$

$$I^2 = \begin{cases} \frac{H^2 - 1}{H^2} * 100\%, & Q > (k-1), \\ 0, & Q \leq (k-1) \end{cases} \quad (2.18)$$

$$I^2 = \begin{cases} \left(\frac{\sum w_i - \frac{\sum w_i^2}{\sum w_i}}{Q} \right) \tau^2 * 100\% & Q > (k-1) \\ 0 & Q \leq (k-1) \end{cases} \quad (2.19)$$

I² için %95 güven aralığına ait alt sınır değeri Eşitlik 2.20'de, üst sınır değeri Eşitlik 2.21'de yer alan formüllere göre hesaplanmaktadır.⁸

$$LL_{I^2} = \frac{(L^2 - 1)}{L^2} * 100\% \quad (2.20)$$

$$UL_{I^2} = \frac{(U^2 - 1)}{U^2} * 100\% \quad (2.21)$$

Formüllerde yer alan L ve U, Eşitlik 2.22 ve 2.23'deki gibi hesaplanmaktadır.⁸

$$L = \exp(0.5 * \ln(\frac{Q}{k-1}) - 1.96 * B) \quad (2.22)$$

$$U = \exp(0.5 * \ln(\frac{Q}{k-1}) + 1.96 * B) \quad (2.23)$$

Eşitlik 2.22 ve 2.23'e ait formüllerde bulunan B, Cochran Q istatistiğinin çalışma sayısından büyük veya küçük oluşuna göre farklı formüllerle hesaplanmaktadır. Buna göre, eğer Q > k ise,

$$B = 0.5 * \frac{\ln(Q) - \ln(k-1)}{\sqrt{2Q} - \sqrt{(2k-3)}} \quad (2.24)$$

şeklinde Q ≤ k ise,

$$B = \sqrt{\frac{1}{2(k-2) \left[1 - \left(\frac{1}{3(k-2)^2} \right) \right]}} \quad (2.25)$$

şeklinde hesaplanmaktadır. I², alt sınır veya üst sınır değerlerinden herhangi biri "0"dan küçük ise bu değere 0 değeri atanır. Eğer alt sınır 0'dan büyük ise I² istatistik olarak anlamlıdır.⁸

I² indeksi yüzde olarak ifade edilmekte olup, bir ölçek üzerinde doğal olarak 0 ile %100 arasında görülür. %100'e yakın değerler için heterojenlik çok büyük, sıfıra yakın değerler için ise küçük heterojenlik şeklinde değerlendirilir. I² değeri %100'e asla ulaşmaz ve çok nadiren %90'ın üzerinde bir değer alır.⁵ Fakat bu iki uç değer arasındaki değerler nasıl yorumlanmalıdır? Higgins ve Thompson, I² değeri için üç kategori belirlemişlerdir ve %25 düşük, %50 orta ve %75 ise yüksek heterojenliği ifade etmektedir. %50'nin üzerindeki değerler için heterojenliğin önemli olduğu ifade edilmektedir.⁷ Patsopoulos, %25'in altında ise düşük heterojenlik, %25-50 arası orta heterojenlik, %50'nin üzerinde olduğunda ise heterojenliğin yüksek olduğu şeklinde yorumlanmaktadır.⁹

Simülasyon çalışması: Bu çalışmada, meta analizine dahil edilen her bir çalışmanın sonuç değişkeninin sadece 0 ve 1'lerden oluşan ikili yapıda olduğu, ikili yapıda kullanılan etki büyüklüğü OR değerinin beklenen değerinin 1'e eşit olduğu durumlar göz önünde bulundurularak simülasyonlar iki bölümde ele alınmıştır. İlk bölümde meta analizine dahil edilen çalışmaların homojen olduğu, ikinci bölümde de çalışmaların heterojen olduğu durumlara ait farklı örneklem büyüklükleri ve tedavi etkinlik oranlarında veriler üretilmiş ve bu verilere ait Cochran Q heterojenlik testine ait anlamlılık seviyesi (p) ve heterojenlik ölçümleri için kesim noktalarının belirlenmesine çalışılmıştır.

Bu amaçla, tedavi ve kontrol grubu için tedavi etkinliği oranının her iki grupta eşit olduğu (homojen) bazı kombinasyonlar (0.05; 0.10; 0.20; 0.50; 0.80; 0.90 ve 0.95) göz önünde bulundurularak, farklı çalışma sayıları (3, 5, 10 ve 20) için MINITAB 15.0 paket programının demo versiyonunda her iki grupta örneklem büyüklükleri eşit olmak üzere (50 ve 500) binom dağılım gösteren 56 (7x4x2) farklı kombinasyonda rasgele çalışmalar üretilmiştir. Ayrıca, tedavi ve kontrol gruplarında tedavi etkinliği oranları eşit olmadığı (heterojen), 7 farklı tedavi etkinliği oranları için, 4 farklı çalışma sayısında ve 2 farklı örneklem büyüklüğünde 56 farklı kombinasyona sahip çalışmalar üretilmiştir. Excel programında oluşturulan bir makro yardımıyla üretilen

bu verilerden, Cochran Q heterojenlik testine ait anlamlılık seviyesi (p) ve heterojenlik ölçümleri hesaplanmıştır. Tüm işlemler 1000 kez tekrar edilmiştir.

BULGULAR

56 homojen yapıda 56 heterojen yapıda olmak üzere toplam 112 kombinasyon sonucuna göre, literatürde Cochran Q heterojenlik testine ait 0.05 ve 0.10 anlamlılık seviyesi için heterojenlik ölçümlerine ait kesim noktaları kaydedilmiştir. Ayrıca, bu kombinasyonlara göre anlamlılık seviyesi için yeni bir kesim noktası belirlenmiş ve bu değere göre heterojenlik ölçümlerinin ROC analizleri yapılmıştır.

Cochran Q testine göre 0.05 anlamlılık seviyesi için heterojenlik ölçümlerine ait kesim noktaları Tablo 1'de, 0.10 anlamlılık seviyesi için heterojen-

lik ölçümlerine ait kesim noktaları Tablo 2'de, homojen ve heterojen yapıda oluşturulan kombinasyonlardan elde edilen durumda, heterojenlik testine ait anlamlılık seviyesi ve heterojenlik ölçümlerinin kesim noktaları ise Tablo 3'de verilmektedir.

Cochran Q testine göre 0.05 anlamlılık seviyesinde, heterojenlik ölçümü olan I^2 test istatistiği için kesim noktası %71.43, H için 1.41, H_M için 1.1, R için 1.4 ve τ^2 için ise 0.24 olarak elde edilmekte ve bu değerlerin üzerindeki değerler için meta analizine dahil edilen çalışmaların heterojen olduğu sonucuna varılmaktadır (Tablo 1). Meta analizine dahil edilen çalışmaların gerçekte heterojen olması gerekirken, 0.05 anlamlılık seviyesinde I^2 , H, H_M ve R test istatistiklerinin bazı değerleri homojen olarak yorumlanmaktadır (Şekil 1). Yalnızca çalışmalar arası varyans olarak da bilinen τ^2 heterojen-

TABLO 1: Cochran Q'ya göre 0.05 anlamlılık seviyesinde heterojenlik ölçümlerine ait kesim noktaları.

P<0.05	Duyarlılık	Seçicilik	LR(+)	LR(-)	AUC (p)	Kesim Noktası
I^2	100.00 (93.0-100.0)	100.00 (94.1-100.0)	0.00	0.00	1.000 (< 0.001)	> 71.43
H	100.00 (93.0-100.0)	98.36 (91.2-100.0)	61.00 (59.1-63.0)	0.00	0.999 (< 0.001)	> 1.41
H_M	100.00 (93.0-100.0)	98.36 (91.2-100.0)	61.00 (59.1-63.0)	0.00	0.999 (< 0.001)	> 1.1
τ^2	92.16 (81.1-97.8)	98.36 (91.2-100.0)	56.22 (51.6-61.3)	0.080 (0.009-0.7)	0.989 (< 0.001)	> 0.24
R	100.00 (93.0-100.0)	98.36 (91.2-100.0)	61.00 (59.1-63.0)	0.00	0.999 (< 0.001)	> 1.4

TABLO 2: Cochran Q'ya göre 0.10 anlamlılık seviyesinde heterojenlik ölçümlerine ait kesim noktaları.

P<0.10	Duyarlılık	Seçicilik	LR(-)	AUC (p)	Kesim Noktası	Literatür Kesim Noktası ^{3-5,7}
I^2	100.00 (93.3-100.0)	100.00 (93.9-100.0)	0.00	1.000 (< 0.001)	> 39.85	> 50.0
H	98.11 (89.9-100.0)	100.00 (93.9-100.0)	0.019 (0.001-0.3)	0.999 (< 0.001)	> 1.41	> 1.00
H_M	98.11 (89.9-100.0)	100.00 (93.9-100.0)	0.019 (0.001-0.3)	0.999 (< 0.001)	> 1.1	> 0.0
τ^2	90.57 (79.3-96.9)	100.00 (93.9-100.0)	0.094 (0.02-0.5)	0.991 (< 0.001)	> 0.24	> 0.0
R	98.11 (89.9-100.0)	100.00 (93.9-100.0)	0.019 (0.001-0.3)	0.999 (< 0.001)	> 1.4	> 1.00

TABLO 3: Gerçek durumda heterojenlik testine ait anlamlılık seviyesi ve heterojenlik ölçümleri için kesim noktaları.

Gold standart	Duyarlılık	Seçicilik	LR (+)	LR (-)	AUC (p)	Kesim Noktası
P değeri	100.00 (93.6-100.0)	100.00 (93.6-100.0)		0.00	1.000 (< 0.001)	≤ 0.29
I ²	100.00 (93.6-100.0)	100.0 (93.6-100.0)		0.00	1.000 (< 0.001)	> 15.66
H	100.00 (93.6-100.0)	100.0 (93.6-100.0)		0.00	1.000 (< 0.001)	> 1.15
H _M	100.00 (93.6-100.0)	100.0 (93.6-100.0)		0.00	1.000 (< 0.001)	> 0.62
τ^2	100.00 (93.6-100.0)	92.86 (82.7-98.0)	14.00 (13.0-15.1)	0.00	0.992 (< 0.001)	> 0.13
R	100.00 (93.6-100.0)	100.0 (93.6-100.0)		0.00	1.000 (< 0.001)	> 1.15

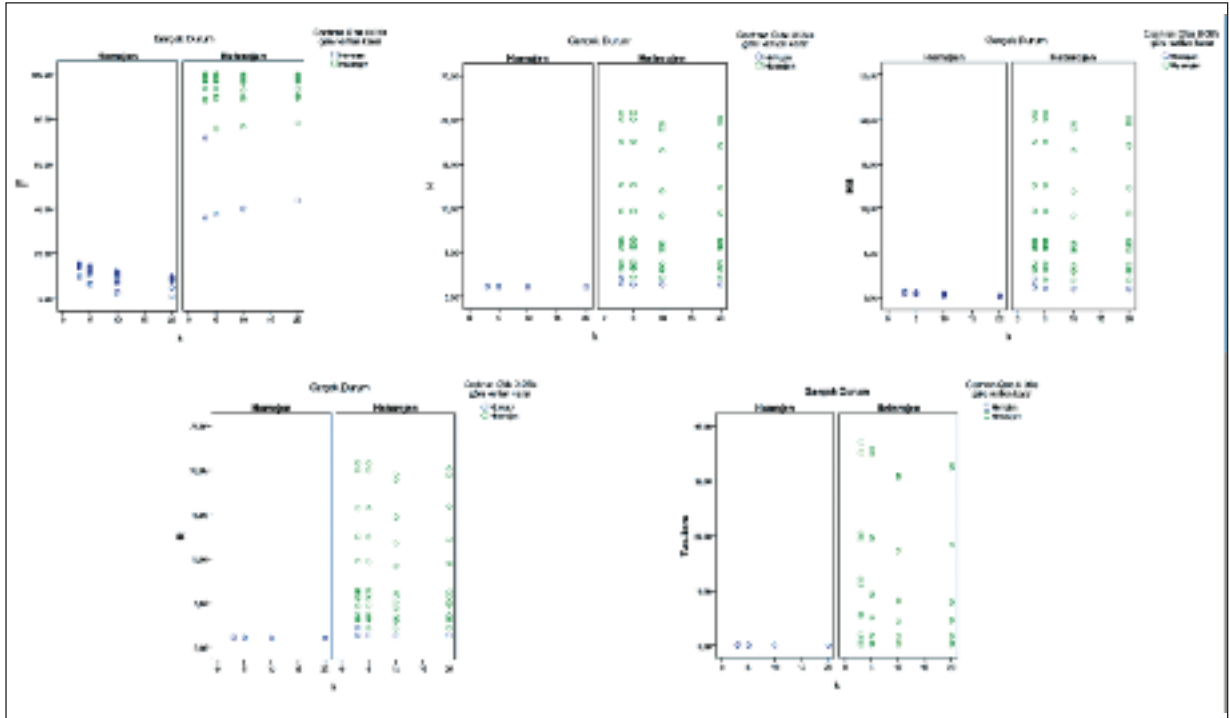
lik ölçüm değerleri, Cochran Q testinde 0.05 anlamlılık seviyesine göre karar verilirken gerçek homojen ve heterojenleri çok iyi ayırt edebilmektedir. Bu sonuçlara göre denilebilir ki, Cochran Q test istatistiğinde 0.05 anlamlılık seviyesine göre karar verilmek isteniyorsa heterojenlik ölçümü olarak τ^2 heterojenlik ölçümü kullanılması önerilebilir.

Cochran Q testine göre 0.10 anlamlılık seviyesinde heterojenlik ölçümlerine ait kesim noktaları ve literatürde yer alan kesim noktaları Tablo 2'de verilmektedir. Heterojenlik ölçümlerinden I² değeri için literatürde %50'nin üzerinde olduğunda çalışmaların heterojen olduğu sonucuna varılırken yapılan simülasyon sonucunda homojenlik sınırı %39.85 olarak belirlenmiş ve bu değer üzerindeki değerler heterojen olarak kabul edilmiştir. Diğer heterojenlik ölçümü olan H ve R test istatistikleri için, 1 olduğunda mükemmel bir homojenlik sonucuna varılırken simülasyon sonucunda 1.4 olarak belirlenmiştir. Bu değer üzerindeki tüm değerler için çalışmalar heterojen olarak kabul edilirken, yanlılığı göz önünde bulunduran H_M heterojenlik ölçümüne göre 1.1'in üzerindeki değerler heterojen olarak kabul edilmektedir. Çalışmalar arası varyans olarak bilinen τ^2 için ise 0.24'ün üzerinde olduğu durumlarda meta analizine dahil edilen çalışmaların heterojen olduğu sonucuna varılmaktadır (Tablo 2).

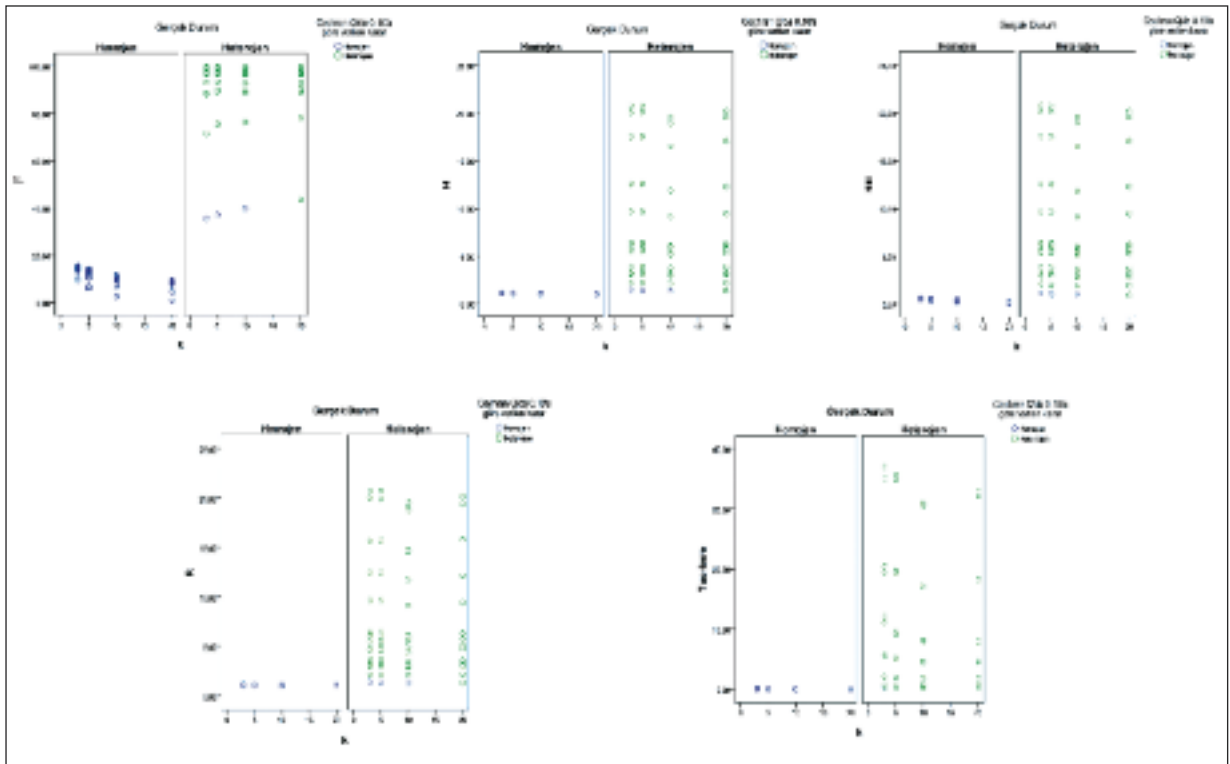
Meta analizine dahil edilen çalışmaların gerçekte heterojen olması gerekirken, 0.10 anlamlılık

seviyesine göre karar verildiğinde I², H, H_M ve R test istatistiklerinin bazı değerleri homojen olarak yorumlanmaktadır. τ^2 heterojenlik ölçüm değerleri incelendiğinde, Cochran Q testinde 0.10 anlamlılık seviyesine göre karar verilirken gerçek homojen ve heterojenleri çok iyi ayırt edebilmektedir. Bu sonuçlara göre denilebilir ki, Cochran Q test istatistiğinde 0.10 anlamlılık seviyesine göre karar verilmek isteniyorsa heterojenlik ölçümü olarak τ^2 heterojenlik ölçümü kullanılması önerilebilir (Şekil 2). Ayrıca, I² heterojenlik ölçümü hariç diğer heterojenlik ölçümlerinin kesim noktalarının 0.05 ve 0.10 anlamlılık seviyesi için değişmediği görülmektedir. Yani, Cochran Q heterojenlik testi ne göre karar verilirken 0.05 veya 0.10 anlamlılık seviyesi için de kesim noktalarının aynı olduğu gözlenmektedir.

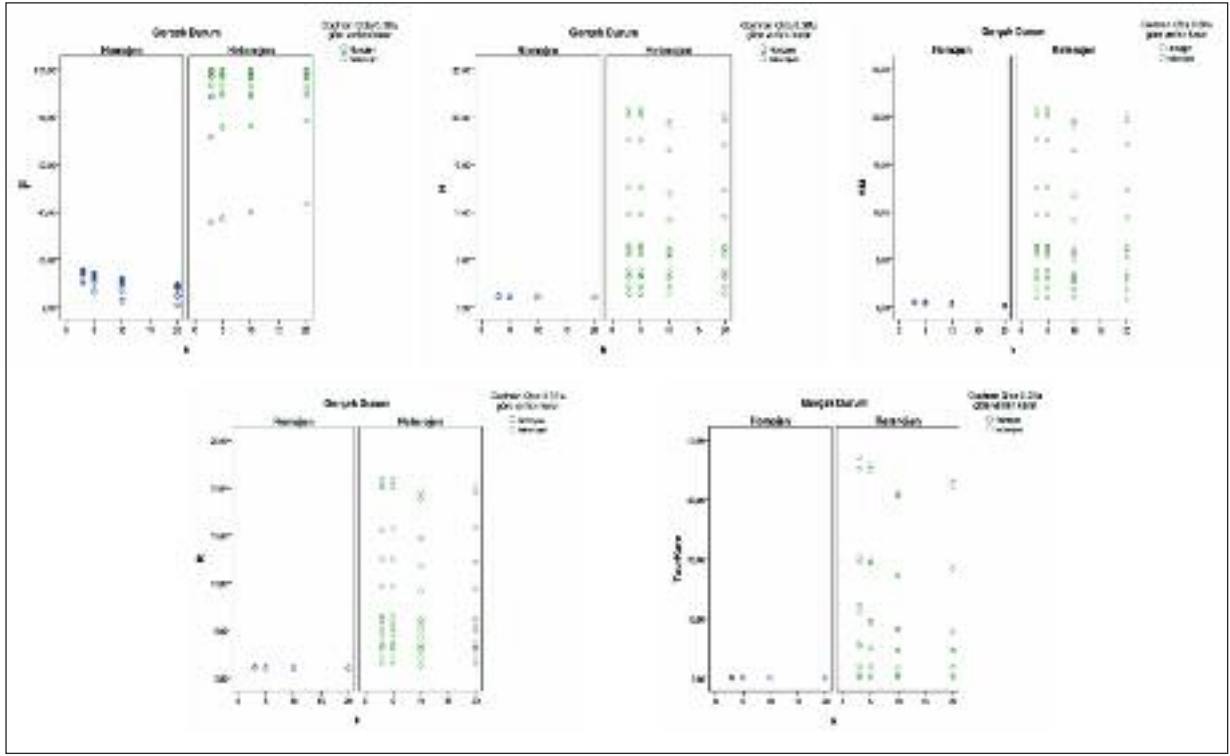
Tablo 3'e göre, meta analizine dahil edilen çalışmaların homojen olduğunu söyleyebilmek için heterojenlik testi sonucunda elde edilen anlamlılık seviyesi 0.30'un üzerinde olması gerekmektedir. Heterojenlik ölçümlerinden I² değerinin homojenlik sınırı %15.66 olarak belirlenmiş ve bu değer üzerindeki değerler heterojen olarak kabul edilmiştir. Diğer heterojenlik ölçümü olan H ve R test istatistikleri için, 1.15'in üzerindeki tüm değerler için çalışmalar heterojen olarak kabul edilirken, yanlılığı göz önünde bulunduran H_M heterojenlik ölçümüne göre 0.62'nin üzerindeki değerler heterojen olarak kabul edilmektedir. Çalışmalar arası



ŞEKİL 1: Heterojenlik ölçümlerinde Cochran Q'da 0.05 anlamlılık seviyesinde verilen kararların gerçek durumla karşılaştırılması.



ŞEKİL 2: Heterojenlik ölçümlerinde Cochran Q'da 0.10 anlamlılık seviyesinde verilen kararların gerçek durumla karşılaştırılması.

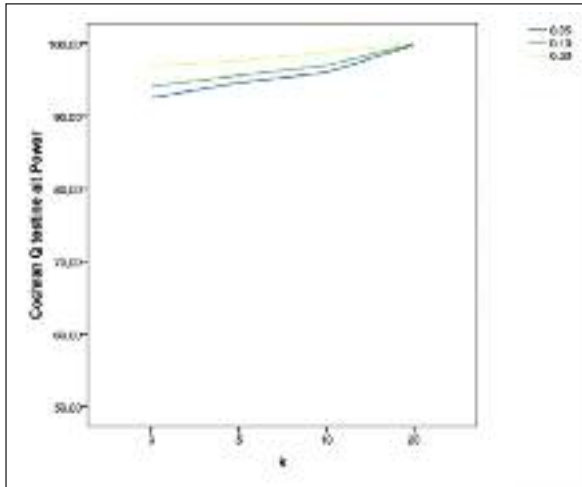


ŞEKİL 3: Heterojenlik ölçümlerinde Cochran Q'da 0.30 anlamlılık seviyesinde verilen kararların gerçek duruma karşılaştırılması.

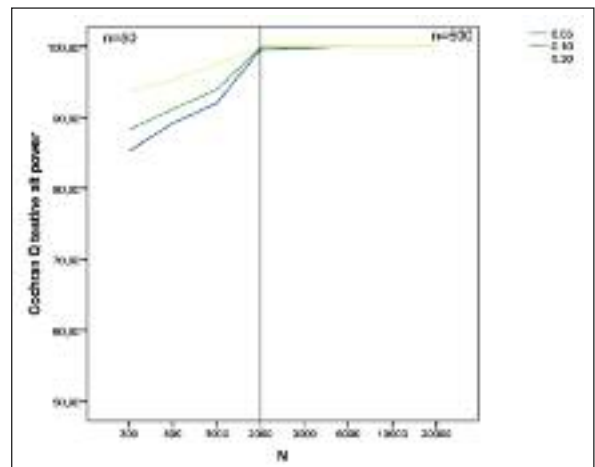
varyans olarak bilinen τ^2 için ise 0.13'ün üzerinde olduğu durumlarda meta analizine dahil edilen çalışmaların heterojen olduğu sonucuna varılmaktadır (Tablo 3).

Cochran Q testinde 0.30 anlamlılık seviyesine göre karar verilirken, tüm heterojenlik ölçüm-

leri gerçek homojen ve heterojenleri çok iyi ayırt edebilmektedir (Şekil 3). Yani, tüm heterojenlik ölçümleri için, çalışmaların gerçek homojen ve heterojen olduğu ancak Cochran Q testinde 0.30 anlamlılık seviyesinde sağlanmaktadır. Bu sonuçlardan yola çıkarak, meta analizi çalışmalarında,



ŞEKİL 4: Farklı anlamlılık seviyelerinde ve çalışma sayılarında heterojenlik testinin gücü.



ŞEKİL 5: Farklı anlamlılık seviyelerinde ve örneklem büyüklüklerinde heterojenlik testinin gücü.

çalışmaların homojen ya da heterojen olduğuna karar verebilmek için Cochran Q testinde 0.30 anlamlılık seviyesine göre karar verilmesi önerilmektedir.

Meta analizi çalışmalarında kullanılacak olan heterojenlik testinin ve heterojenlik ölçümlerinin I. ve II. Tip hatadan etkilenme durumlarının ortaya konması amacıyla, homojen oluşturulan durumlar için gerçekte homojen olması beklenirken heterojen olarak yorumlananlar (I. tip hata, α) ve heterojen oluşturulan durumlar için gerçekte reddedilmesi gerektiği halde homojen olarak yorumlananlar (2. tip hata, β) tespit edilmiştir. Farklı tedavi oranları, çalışma sayıları, örneklem büyüklükleri ve anlamlılık seviyeleri için, 1000 denemede yapılan I. tip hata yüzdeleri, yapılan 2. tip hatalardan yararlanarak testin gücü ($1-\beta$) hesaplanarak ve bunlara ait yüzdeler elde edilmiştir.

Heterojenlik testi için 0.05, 0.10 ve 0.30 anlamlılık seviyesinde çalışma sayısına bağlı olarak heterojenlik testinin gücündeki değişim Şekil 4'de, örneklem büyüklüğüne bağlı olarak heterojenlik testinin gücündeki değişim ise Şekil 5'de verilmektedir. Cochran Q heterojenlik testinde 0.05, 0.10 ve hatta 0.30 anlamlılık seviyesinde bile testin gücü %90'ın üzerindedir. Ancak, 0.05 ve 0.10 anlamlılık seviyesi için meta analizine dahil edilen çalışma sayısı 10'dan büyük olduğunda testin gücü biraz daha güçlüdür. 0.30 anlamlılık seviyesi için meta analizine dahil edilen çalışma sayısı ne olursa olsun testin gücü değişmemektedir (Şekil 4).

Birleştirilmiş çalışmalar örneklem büyüklükleri bakımından incelendiğinde ise, meta analizine dahil edilen çalışma sayısı ne olursa olsun örneklem büyüklüğü her iki grupta da 500 olduğunda tüm anlamlılık seviyeleri için Cochran Q testinin gücü %100 olarak tespit edilmiştir. Aynı şekilde, birleştirilmiş çalışmalarda meta analizine dahil edilen tüm çalışmalardaki toplam örneklem büyüklüğü 2000'in üzerinde olduğu durumlarda testin gücünün %100 olduğunu söyleyebiliriz. Tedavi ve kontrol gruplarında örneklem büyüklükleri 50 olarak alındığında ise Cochran Q testinin gücünün %100 olarak tespit edildiği durum ancak çalışma sayısının 20 olduğu durumdur. Özellikle meta analizine dahil edilen çalışma sayısı 10'dan az oldu-

ğunda Cochran Q testinin gücü %90'ın altına düşmektedir (Şekil 5).

TARTIŞMA VE SONUÇ

Meta analizinde heterojenlik testi olarak kullanılan Cochran Q testi, meta analizine dahil edilen çalışma sayısından, örneklem büyüklüğünden ve gruplardaki tedavi etkinlik oranları arasındaki farklılıklardan etkilenmektedir. Dolayısıyla heterojenlik kontrolü için heterojenlik testinin yanı sıra heterojenlik ölçümlerinin de verilmesi tavsiye edilmektedir.^{7,10} Bu amaçla, heterojenlik testinin ve ölçümlerinin çalışma sayısından ve örneklem büyüklüğünden etkilenme durumları araştırılmış ve simülasyon bulgularına göre, meta analizine alınacak olan çalışmaların, homojen veya heterojen olması durumlarında, heterojenlik ölçümü olarak I^2 ve τ^2 test istatistiklerinin kullanılması gerekliliği görülmüştür. Ancak bu heterojenlik ölçümlerinin de güven aralıklarına ait problemlerle karşılaşmıştır.

Buradan yola çıkarak, meta analizi çalışmalarında kullanılacak olan heterojenlik testinin ve heterojenlik ölçümlerinin I. ve II. tip hatadan etkilenme durumlarının ortaya konması ve heterojenlik ölçümlerine ait kesim noktalarının belirlenmesinin gerekliliği sonucuna varılmıştır. Cochran Q istatistiğinin dezavantajından dolayı 0.05 anlamlılık seviyesi yerine 0.10 anlamlılık seviyesine göre karar verilmesi ve Cochran Q heterojenlik testinin yanı sıra bir de heterojenlik ölçümlerinden olan I^2 ve H_M ölçüm değerlerinin de verilmesi tavsiye edilmektedir.³ Yani önce heterojenlik testi ile heterojen olup olmamasına karar verilecek, eğer heterojen ise heterojenlik ölçümlerinden I^2 değerinin de %50'nin üzerinde olması durumunda çalışmaların heterojen olduğu sonucuna varılacaktır.

Elde edilen bulguların sonucuna göre, meta analizi çalışmalarında 0.05 veya 0.10 anlamlılık seviyesine göre karar veriliyorsa, heterojenlik ölçümü olarak, gerçek homojen ve heterojenleri en iyi ayırt edebilen τ^2 test istatistiğinin kullanılması önerilmektedir. Ayrıca, τ^2 test istatistiğinin 0.24'ün üzerinde olduğu durumlarda meta analizine dahil edilen heterojen çalışmaları tam olarak ayırabilecektir. Diğer heterojenlik ölçümlerinden H ve R test istatistiği için 1.4'ün, H_M test istatistiği için

1.1'in üzerindeki değerler için meta analizine dahil edilen çalışmaların heterojen olduğu sonucuna varılmaktadır. Diğer bir heterojenlik ölçümü olan I^2 test istatistiği için ise, Cochran Q testine göre 0.05 anlamlılık seviyesinde karar veriliyorsa %71.43'ün üzerinde, 0.10 anlamlılık seviyesinde karar veriliyorsa %50'nin üzerinde değil %39.85'in üzerinde olduğunda çalışmaların heterojen olduğu sonucuna varılmaktadır. Bunun yanı sıra, Cochran Q test istatistiği için 0.30 anlamlılık seviyesinde karar veriliyorsa tüm heterojenlik ölçümlerinin gerçek homojen ve heterojenleri çok iyi ayırt ettiği ortaya konulmuştur. Heterojenliğin test edilmesinde, heterojenlik ölçümlerinden I^2 test istatistiği, %15.66'nın üzerinde bir değer aldığı anda, H ve R için 1.15'in üzerinde olması durumunda, H_M için 0.62'nin üzerinde olması, çalışmalar arası varyansın da (τ^2) 0.13'ün üzerinde olması durumlarında da meta analizine dahil edilen çalışmaların heterojen olduğu sonucuna varılabilmektedir.

Ayrıca, elde edilen bulgular sonucunda, meta analizine dahil edilen çalışma sayısının 20 olması durumunda, 0.05, 0.10 ya da 0.30 anlamlılık seviyesinde testin gücünün %100'lere ulaştığını söyleyebiliriz. Bunun yanı sıra birleştirilmiş çalışmaların toplam örneklem büyüklüğünün 1000'in altında olduğu durumlarda tüm anlamlılık seviyesinde testin gücü %90'ın altına düşerken, 2000'in üzerindeki örneklem büyüklükleri için testin gücü %100'e ulaşmaktadır.

Bu çalışmada, tedavi ve kontrol gruplarında, örneklem büyüklükleri eşit olmak üzere, birleştirilmiş çalışmalardaki örneklem büyüklükleri dikkate alınmıştır. Gruplar arasında örneklem büyüklüklerinin dengeli olmadığı durumlar da incelenebilir. Ayrıca, sonuç değişkeninin sürekli yapıda olduğu durumlar için heterojenlik testi ve heterojenlik ölçümlerinin çalışma sayısından ve örneklem büyüklüğünden etkilenme durumları incelenebilir.

KAYNAKLAR

1. Cochran WG. The combination of estimates from different experiments. *Biometrics* 1954; 10(1):101-29.
2. Baujat B, Mahé C, Pignon JP, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat Med* 2002;21(18):2641-52.
3. Mittlböck M, Heinzl H. A simulation study comparing properties of heterogeneity measures in meta-analyses. *Stat Med* 2006;25(24): 4321-33.
4. Huedo-Medina TB, Sánchez-Meca J, Marín-Martínez F, Botella J. Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Psychol Methods* 2006;11(2):193-206.
5. Rücker G, Schwarzer G, Carpenter JR, Schumacher M. Undue reliance on I(2) in assessing heterogeneity may mislead. *BMC Med Res Methodol* 2008;8:79.
6. Petitti DB. Approaches to heterogeneity in meta-analysis. *Stat Med* 2001;20(23):3625-33.
7. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21(11):1539-58.
8. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Heterogeneity, identifying and quantifying heterogeneity. *Introduction to Meta Analysis. Part 4. 1st ed.* West Sussex: John Wiley&Sons Inc; 2009. p.107-125.
9. Patsopoulos NA, Evangelou E, Ioannidis JP. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. *Int J Epidemiol* 2008;37(5): 1148-57.
10. Virgili G, Conti AA, Moja L, Gensini GF, Gusinu R. Heterogeneity and meta-analyses: do study results truly differ? *Intern Emerg Med* 2009;4(5):423-7.