

Çoklu Sonlanım Noktalarına Sahip Klinik Denemelerde Denetleyici Test Stratejileri

Gatekeeping Test Strategies in the Clinical Trials with Multiple Endpoints: Review

Bahar TAŞDELEN^a

^aBiyostatistik AD,
Mersin Üniversitesi Tıp Fakültesi,
Mersin

Geliş Tarihi/Received: 13.12.2012
Kabul Tarihi/Accepted: 22.02.2013

Bu derleme, XIII. Ulusal Biyoistatistik Kongresi (12-14 Eylül 2011, Ankara)'nde sunulmuştur.

Yazışma Adresi/Correspondence:
Bahar TAŞDELEN
Mersin Üniversitesi Tıp Fakültesi,
Biyostatistik AD, Mersin,
TÜRKİYE/TURKEY
bahartasdelen@gmail.com

ÖZET Bu derlemede, çoklu sonlanım noktasına sahip klinik denemelerde test sonuçlarının geçerliliği için denetleyici test stratejilerinin teorisi ve kullanım amaçları tartışılmaktadır. Çoklu sonlanım noktalarına sahip klinik denemelerde, Tip I hatanın kontrol altında tutulabilmesi deneme sonuçlarının istatistiksel analizinde en temel sorundur. Klinik denemelerde, çalışmanın asıl amacı ile ilişkili olan birincil sonlanım noktalarına ait hipotezler, ilacın veya muamelenin etkinliği ile ilgili olup önce test edilir. Çoğu kez birincil sonlanım noktalarının yanı sıra, ikincil sonlanım noktaları da kullanılır ve hipotezler test edilen sonlanım noktasının önemine göre aileler içinde gruplandırılır. Bu tip hipotez ailelerini analiz etmek için denetleyici test stratejileri kullanılır. Eğer ikincil hipotezleri test etmek için birincil hipotezlerin tümünün red edilmesi gerekiyorsa, aile seri denetleyicidir. Eğer birincil ailedeki hipotezlerden en az birinin red edilmesi ikincil hipotezleri test etmek için yeterli ise aile paralel denetleyici olarak adlandırılır. Denetleyici test stratejileri genellikle Bonferroni tabanlı prosedürleri içerse de sonlanım noktaları arasındaki korelasyon yapısına karşı daha sağlam ve özellikle çoklu doz ve çoklu sonlanım noktalarının olduğu durumlarda tercih edilen yöntemler de mevcuttur. Çoklu sonlanım noktaları olan doz-cevap çalışmalarında çoklu dozların kontrolle kıyaslanması için Dunnett tabanlı paralel denetleyici test stratejisi geliştirilmiştir. Denetleyici test stratejileri kullanılırken, hipotez ailelerinin sayısı, sonlanım noktalarının sayısı ve aralarındaki korelasyon yapısı, test prosedürlerinin seçimi önemli konulardır.

Anahtar Kelimeler: Çoklu sonlanım noktaları; aile başına hata; denetleyici test stratejileri; CONSORT

ABSTRACT In this review, the theory and purpose of gatekeeping test strategies for the validity of test results in the clinical trials with multiple endpoints are discussed. In the clinical trials with multiple endpoints, control of Type I error is the main problem in statistical analysis of experiment results. In the clinical trials, the hypothesis for primary endpoints associated with main purpose of the study are related with the effectiveness of the drug or treatment is tested before. In addition to primary endpoints, secondary endpoints are frequently used and the hypothesis are grouped into families according to the importance of endpoint tested. The gatekeeping test strategies are used to analyse this type of hypothesis families. A family is termed a serial gatekeeper if one must reject all primary hypothesis to test secondary hypothesis. If rejecting at least one primary hypothesis is sufficient to test secondary hypothesis, the family is called as paralel gatekeeper. Although gatekeeping strategies generally includes Bonferroni-based procedures, there are more robust methods to correlation structure between endpoints. They are preferred especially in multiple dose and multiple endpoints. In dose-response studies with multiple endpoints, a Dunnett-based paralel gatekeeping test strategy was developed to compare multiple dose-to-control. The number of families, the number of endpoints and the correlation structure between them and selection of test procedures are significant subjects while using gatekeeping test strategies.

Key Words: Multiple endpoints; family wise error; gatekeeping test strategies; CONSORT

Klinik denemelerin amacı, alternatif tedavi yöntemlerinin birbirlerine üstünlüğünü, eş-değerliğini veya daha kötü olmadığını göstermektir ve bu denemelerde klinik olarak anlamlı olduğu düşünülen sonlanım değişkenlerinin istatistiksel olarak da anlamlı olup olmadığı araştırılır. Sonlanım değişkenleri sıklıkla birincil (primary) ve ikincil (secondary) olarak adlandırılmaktadır. Randomize kontrollü klinik denemelerin yürütülmesi sırasında mutlaka dikkate alınması gereken CONSORT (Consolidated Standards of Reporting Trials) kriterlerine göre, birincil ve ikincil sonlanım değişkenlerinin ne zaman ve nasıl belirlendiği açıklanmalıdır. Analiz ve yorumlama sıkıntısı yaratması nedeniyle, birden fazla birincil sonlanım değişkeninin kullanılması (multiplicity) önerilmektedir.¹

Çalışmanın asıl amacı ile ilişkili olan birincil sonlanım noktalarına ait hipotezler, ilacın veya muamelenin etkinliği ile ilgili olup diğer hipotezlerden önce test edilir. İkincil hipotezler ise, test edilen ilaç veya muamele hakkında destekleyici bilgiye sahiptir veya test edilen ilacın etkinlik profilinin belirlenmesinde önemli rol oynarlar. İkincil hipotezlerden elde edilen bulgular ürün prospektüsünde kullanılır.² Bununla birlikte, ikincil hipotezlerden elde edilen bulgu, birincil hipotezle birlikte bir anlam taşırken, tek başına her hangi bir iddiada bulunmak için yeterli değildir.

D'Agostino'ya göre iki tip ikincil sonlanım noktası vardır: 1) Birincil amacın unsurları olan sonlanım noktaları. Örneğin, belirli bir hastalık nedeniyle biyokimya parametrelerindeki değişim birincil sonlanım noktası iken, hastalığa bağlı olarak ortaya çıkan semptom gelişimi ikincil sonlanım noktası olabilir veya 2) Birincil sonucu yorumlamaya yardımcı yaşam kalitesi gibi sonlanım noktaları olabilir.^{3,4}

Bir klinik deneme dizayn edilirken, hiyerarşik olarak sıralanmış hipotezleri test etmek, birincil bulguları destekleyici nitelikteki ikincil sonlanım noktalarını analiz etmek için geliştirilmiş, tek başına hipotezlerin değil, hipotez ailelerinin test edildiği test stratejileri mevcuttur. Bu stratejiler genellikle denetleyici (gatekeeping) veya hiyerarşik test stratejileri olarak adlandırılır.^{3,5,6}

Klinik denemelerde çoklu sonlanım noktalarına ait hipotezler test edilirken, Tip I hatanın kontrol altında tutulabilmesi sonuçların istatistiksel analizinde en temel sorundur. Bu tip denemelerde, doğru olan en az bir yokluk hipotezinin yanlışlıkla red edilmesi olasılığı aile başına hata oranı [family-wise error (FWE) rate,] olarak adlandırılır.⁷ Bu çalışmada, çoklu sonlanım noktasına sahip klinik denemelerde test sonuçlarının güvenilirliği için denetleyici test stratejilerinin teorisi ve kullanım amaçları, uygun test stratejisi seçerken dikkat edilmesi gereken unsurlar tartışılmaktadır.

■ DENETLEYİCİ (GATEKEEPING) TEST STRATEJİLERİ

Klinik denemelerde, test edilen hipotezin önemine göre hiyerarşik yapı dikkate alınarak, hipotezler aileler içinde gruplandırılır (F_1, \dots, F_m). Birincil ve ikincil hipotez aileleri $F_1 = \{H_1, \dots, H_k\}$ ve $F_2 = \{H_{k+1}, \dots, H_m\}$ olmak üzere F_1 hipotez ailesi denetleyicidir ve F_2 ailesi için hiçbir düzeltme yapılmaksızın test edilir. Yani bu ailedeki hipotezler test edilirken diğer ailelerdeki hipotezler dikkate alınmaz.^{2,6} F_2 ailesindeki hipotezler ise ancak F_1 ailesi geçilmiş ise test edilir.⁸ Amaç, her iki hipotez ailesi test edilirken FWE'yi güçlü bir şekilde kontrol altında tutmaktır. Alt ailelerdeki testlerde daha önce değerlendirilen aile sayısını dikkate alan düzeltmeler yapılır. Buradaki ardışık test stratejisinde, zincirin ilk sıralarındaki daha önemli hipotezlerin gücü yüksektir.

Denetleyici test stratejileri aşağıdaki koşulları sağlayan test yöntemleridir:

1. Tip I hata oranı kontrol altında tutulur.

2. F_i ($i=1, \dots, m-1$) hipotez ailesi eğer seri denetleyici ise F_{i+1} 'deki hipotezler ancak F_i deki hipotezlerin tümü red edilirse test edilebilir. F_i ailesi paralel denetleyici ise F_i 'deki hipotezlerden en az birinin red edilmiş olması F_{i+1} 'deki hipotezlerin test edilebilmesi için yeterli koşuldur. Test edilmeyen hipotezler kabul edilmiş varsayılır.

3. F_i 'deki çıkarımlar ($i=1, \dots, m-1$), F_{i+1} 'deki çıkarımlardan bağımsızdır (Bağımsızlık koşulu). Yani birincil hipotezlerin (daha önemli) analiz sonuçları

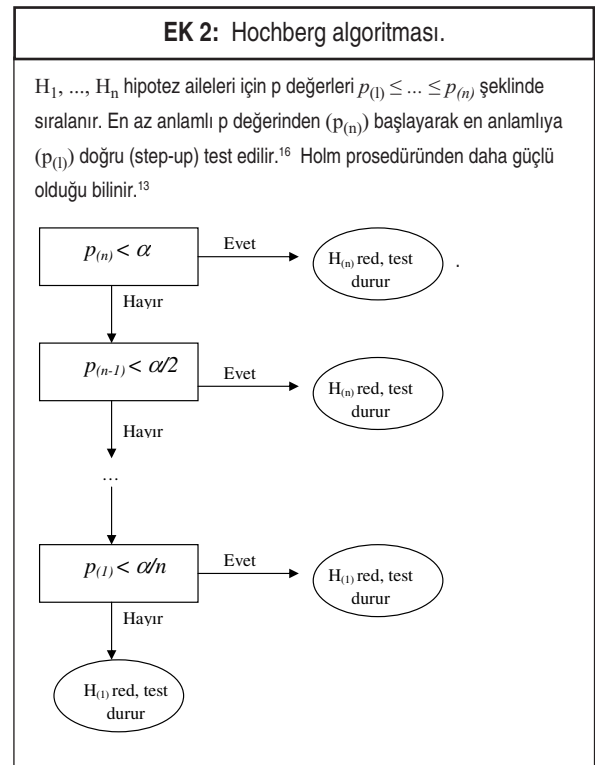
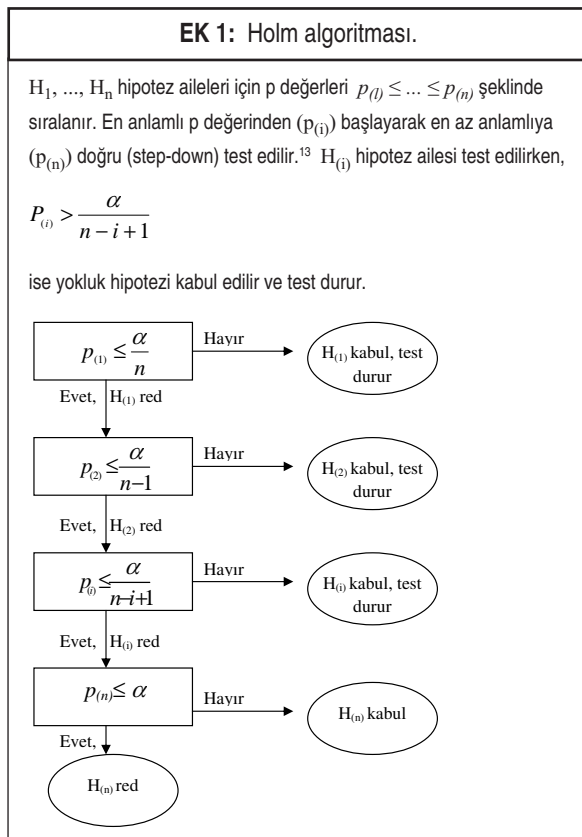
ikincil hipotezlerin (az önemli) sonuçlarına bağlı değildir.⁶

SERİ DENETLEYİCİ TEST STRATEJİSİ

Seri denetleyici test stratejilerinde, FWE oranının α düzeyinde kontrol edildiği varsayılır.⁹ F_1 ve F_2 hipotez ailelerinden F_1 seri denetleyici olmak üzere, F_1 'deki her bir hipotez için α düzeyinde ayrı ayrı test yapılır ve F_2 'deki testlerin yapılabilmesi için F_1 'deki tüm testlerin red edilmesi beklenir.^{6,7} F_2 'deki testler için Holm veya Hochberg algoritmaları önerilir (Ek 1-2).^{6,10,11} Seri denetleyiciler kısıtlayıcıdır ve ancak ikincil sonlanım noktalarının birincil amacın unsurları olduğu durumlarda uygulanması uygundur.

PARALEL DENETLEYİCİ TEST STRATEJİSİ

Paralel denetleyici test stratejilerinin temeli bilinen çoklu karşılaştırma yöntemlerinden Bonferroni testine dayanır. Aile başına hatayı güçlü bir şekilde kontrol altına aldığı belirtilen Holm algoritması



bunlardan birisidir.³ Klasik Bonferroni testinde p değerleri α/n ile karşılaştırılırken Holm algoritmasında (ardışık red eden Bonferroni testi) sıralı p değerleri $\alpha/n, \alpha/n-1, \dots, \alpha/1$ ile karşılaştırılır.¹⁰

Paralel denetleyici test stratejilerinin muamele gruplarının yanı sıra, muamele ve kontrol gruplarının karşılaştırılmasında da kullanılabileceği gösterilmiştir.² Bu amaçla öncelikle Dunnett testi denetleyici test stratejilerine modifiye edilmiş (Ek-4) daha sonra ise Bonferroni tabanlı Dunnett algoritması geliştirilmiştir.¹²

ADIMSAL PARALEL DENETLEYİCİ TEST STRATEJİSİ

Bağımsızlık koşulunu ve paralel denetleyici koşulunu sağlaması için ağırlıklandırma yöntemi ile adimsal bir algoritma (Fallback algoritması) kullanılır. F_i ($i=1, \dots, m-1$) ailesindeki hipotezler $\rho_k \alpha$ düzeyinde Bonferroni testi kullanılarak test edilir. F_m ailesindeki yokluk hipotezleri ise $\rho_m \alpha$ düzeyinde Holm algoritması ile test edilir. Burada ρ_k "H₀ red kazanç faktörü" olarak adlandırılır ve önceki aşamalarda red edilen hipotezlerin anlamlılığına ve sayısına bağlı olarak hesaplanır.

$$\rho_1 = 1 \quad \rho_k = \prod_{i=1}^{k-1} (\sum_{j=1}^{n_{ij}} r_{ij} w_{ij}) \quad (k=2, \dots, m) \text{ ve}$$

$$r_{ij} = \begin{cases} 1, & H_0 \text{ red} \\ 0, & H_0 \text{ kabul} \end{cases}$$

Burada w_{ij} test edilen hipotezin önemine göre verilen ağırlık değeridir. Eşit ağırlıklı hipotezler için $w_{ij}=1/n_{ij}$ olarak belirlenir.

Böylece kazanç faktörü kullanılarak i . ailedeki ($i=2, \dots, m$) j . hipotez (H_{ij}) için düzeltilmiş p değeri $\widetilde{p}_{ij} = p_{ij}/\rho_i$ olarak yeniden hesaplanır. Eğer ρ_k pozitif ise yani önceki ailelerdeki hipotezlerden en az biri red edilmiş ise sonraki aileler test edilir. $\rho_k=1$ ise önceki ailelerdeki tüm hipotezler red edilmiş demektir.

Yukarıda bahsedilen Holm algoritması kullanılırken Tip I hata oranı her aşamada harcanır. Bu dezavantajından dolayı Bonferroni ve Holm testlerinin konveks kombinasyonu olan budanmış (truncated) Holm testi geliştirilmiştir.

“Kullan veya at” (use it or lose it) prensibinin geçerli olduğu bu yöntemde, red edilen hipotezlerin α değeri daha sonraki hipotezler için de kullanılır.¹³

AĞAÇ TİPİ DENETLEYİCİ TEST STRATEJİSİ

Paralel ve seri denetleyici prosedürleri tek boyutludur. Ancak çoğu kez, klinik denemeler daha karmaşık çok boyutlu stratejilerin kullanılmasını gerektirir. Özellikle birden fazla dozun bir arada denendiği, birden fazla sonuçla aynı anda ilgilenildiği ve buna ilaveten üstünlük (superiority) veya eşdeğerlik (non-inferiority) hipotezlerinin test edildiği klinik denemelerde bu durumla karşılaşılır. Bu tip karmaşık durumlarda tek boyutlu test stratejisi kul-

lanmak yanlıtıcı olabileceğinden Dimitrienko tarafından geliştirilmiş karar ağaçları benzeri denetleyici prosedür kullanmak daha uygun olabilir.^{6,14}

TARTIŞMA

Denetleyici test stratejilerinde, metodolojik olarak göze çarpan en temel nokta Tip I hata oranının kontrol altında tutulmasıdır. Bunun için aile başına hata oranı belirlenir. Eğer denetleyici, alttaki aileye (birincil) ait tüm hipotezler red edildiği takdirde üstteki ailenin (ikincil) hipotezlerinin test edilmesine izin veriyorsa seri denetleyicidir. Eğer birincil ailedeki hipotezlerden en az birinin red edilmesi ikincil hipotezleri test etmek için yeterli ise paralel denetleyicidir.

Birincil ve ikincil sonlanım noktalarının farklı iddiaları içerdiği durumda seri denetleyicilerin uygulanması doğru olmayabilir. Çünkü ikincil hipotezlerin test edilebilmesi için birincil hipotezlerin tümünün red edilmiş olması gerekir. Ayrıca doz belirleme çalışmalarında test ilacın olası tüm dozları deneniyorsa, yüksek dozdan başlayarak paralel denetleyici test stratejisi kullanılması önerilir. Yüksek dozlardan en az biri kontrolden farklıysa bir sonraki düşük doz denenebilir.³

Ancak bununla birlikte denetleyici test stratejilerinin kullanılabilmesi için ikincil hipotezlerin, farklı iddiaları içerse de destekleyici rolü olmalıdır. Denetleyici test stratejileri genellikle Bonferroni tabanlı prosedürleri (Holm prosedürü) içerse de daha güçlü prosedürler de mevcuttur. Örneğin, Simes tabanlı prosedürler (Hochberg prosedürü) hem güç kazancı olması hem de sonlanım noktaları arasındaki korelasyon yapısına karşı daha sağlam olması

EK 3: Fallback algoritması.

Bu yöntemde hipotezler, önceden belirlenen (a priori) önemine göre sıralanır (H_1, \dots, H_n).^{17,18} En önemli hipotezden en az önemli olan hipoteze doğru (step-down) ardışık olarak test edilir. Toplam yanılma payı, $\sum_{i=1}^n \alpha_i = \alpha$ olacak şekilde hipotezlere paylaştırılır ($\alpha_1, \dots, \alpha_n$).

Adım 1. H_1 hipotezi $\alpha_1=\alpha/n$ seviyesinde test edilir. Eğer $p_1 \leq \alpha_1$ ise $H_{0(1)}$ yokluk hipotezi red edilir.

Adım 2. H_2 hipotezi için $p_2 \leq (2-t)\alpha/n$ ise $H_{0(2)}$ yokluk hipotezi red edilir. Burada, eğer $H_{0(1)}$ hipotezi red edilmiş ise $t=1$, değilse $t=0$ 'dir.

....

Adım i. H_i hipotezi $(i-t)\alpha/n$ seviyesinde test edilir. Burada t , en son kabul edilen hipotezin indeksidir. Önceki hipotezlerin hiçbiri kabul edilmemiş ise $t=0$ 'dir.

EK 4: Dunnett algoritması.

En basit haliyle Dunnett paralel denetleyici test algoritmasını açıklamak için, iki doz bir kontrol ve iki sonlanım noktası durumunu ele alalım.

Birincil sonlanım noktası için $F_1 = \{H_{11}, H_{12}\}$ hipotez ailesinde, H_{11} düşük doz ile kontrol kıyaslamasına ait hipotezi, H_{12} yüksek doz ile kontrol kıyaslamasına ait hipotezi göstermektedir. İkincil sonlanım noktası için $F_2 = \{H_{21}, H_{22}\}$ hipotez ailesinde, H_{21} düşük doz ile kontrol kıyaslamasına ait hipotezi, H_{22} yüksek doz ile kontrol kıyaslamasına ait hipotezi göstermektedir. Her bir hipotez için test istatistikleri $t_{11}, t_{12}, t_{21}, t_{22}$ olmak üzere, mümkün hipotezler kesişim hipotezlerinden oluşmaktadır. Bunlar; $H_{1111}, H_{1110}, H_{1101}, H_{1100}, H_{1011}, H_{1010}, H_{1001}, H_{1000}, H_{0111}, H_{0110}, H_{0101}, H_{0100}, H_{0011}, H_{0010}, H_{0001}$ 'dir. Eğer test edilen hipotezler kümesi, birincil ve ikincil hipotezlerin kesişimini içeriyorsa, birincil hipotez red edilmedikçe ikincil hipotez red edilemez.²

Dimitrienko tarafından çoklu sonlanım noktaları için Dunnett testi yaparken kullanılan karar matrisi aşağıdaki gibidir. Bu tabloya göre, örneğin H_{12} yokluk hipotezinin red edilebilmesi için, 1-4, 9-12 nolu kesişim hipotezlerinin tamamının red edilmiş olması gerekir.² Birincil hipotezler anlamlı değilse H_{21} ve H_{22} ikincil hipotezleri red edilemezler.

No	Kesişim Hipotezi	Karar Kuralı
1	1111	$t_{11} > c_1$ veya $t_{12} > c_1$
2	1110	$t_{11} > c_1$ veya $t_{12} > c_1$
3	1101	$t_{11} > c_1$ veya $t_{12} > c_1$
4	1100	$t_{11} > c_1$ veya $t_{12} > c_1$
5	1011	$t_{11} > c_1$ veya $t_{22} > c_2$
6	1010	$t_{11} > c_1$
7	1001	$t_{11} > c_1$ veya $t_{22} > c_2$
8	1000	$t_{11} > c_1$
9	0111	$t_{12} > c_1$ veya $t_{21} > c_2$
10	0110	$t_{12} > c_1$ veya $t_{21} > c_2$
11	0101	$t_{12} > c_1$
12	0100	$t_{12} > c_1$
13	0011	$t_{21} > c_1$ veya $t_{22} > c_1$
14	0010	$t_{21} > c_3$
15	0001	$t_{22} > c_3$

Kritik değerler c_1, c_2 ve c_3 hesaplanırken aşağıdaki adımlar izlenir.²

Öncelikle, $m =$ doz sayısı ve $n =$ her gruptaki denek sayısı olmak üzere c_1 kritik değeri, m ve $(m+1)(n-1)$ serbestlik dereceli Dunnett tablo değerlerinden elde edilir. Sonra c_2 kritik değeri c_1 'e bağlı olarak, c_3 kritik değeri ise c_1 ve c_2 'ye bağlı olarak elde edilir.

Adım 1. $r_1(H)$, F_1 ailesindeki yokluk hipotezi sayısı olmak üzere, $r_1(H) = \sum_{j=1}^m \delta_j(H) > 0$

ise kritik değer hesaplanır. Değilse sonraki adıma geçilir.

$$P(R_1 > c_1(H)) = \alpha$$

Burada $R_1 = \max(t_{11}, \dots, t_{1m})$ 'dir.

Adım 2. $r_2(H)$, F_2 ailesindeki test edilebilir yokluk hipotezi sayısı olmak üzere, $r_2(H) = \sum_{j=1}^m \delta_{2j}(H)(1 - \delta_j(H)) > 0$ ise kritik değer hesaplanır.

Aksi takdirde test durur.

$$P(R_1^* > c_1(H) \text{ veya } R_2^* > c_2(H)) = \alpha$$

Burada,

$$R_1^* = \max(t_{11}\delta_{11}(H), \dots, t_{1m}\delta_{1m}(H)),$$

$$R_2^* = \max(t_{21}(1 - \delta_{11}(H))\delta_{21}(H), \dots, t_{2m}(1 - \delta_{1m}(H))\delta_{2m}(H))$$

Yani, R_2^* aynı dozun birincil ve ikincil hipotezleri içermektedir. H_{1j} red edilmeden H_{2j} 'nin red edilmemesi için ağırlığı sıfırdır.

Kritik değer bulunurken kullanılacak dağılım tipi için 3 durum söz konusudur.

Durum 1.

c_1 kritik değeri m ve $(m+1)(n-1)$ serbestlik dereceli Dunnett dağılımı yardımıyla bulunur. Örneğin, 2 dozun 1 plaseboyla karşılaştırıldığı denemede 2 ve $3(n-1)$ serbestlik dereceli Dunnett tablosu kullanılarak kritik değer elde edilir. Bu durumda, 1-4, 6, 8, 11-13 numaralı kesişim hipotezleri için kritik değer bu şekilde elde edilir.

Durum 2.

Sadece ikincil sonlanım noktalarına ait 14 ve 15 numaralı kesişim hipotezleri için c_3 kritik değeri $m(n-1)$ serbestlik dereceli birikimli t dağılımından elde edilir.

Durum 3.

İki farklı kritik değere bağlı olarak karşılaştırma yapılan 5, 7, 9 ve 10 numaralı hipotezler için hem t dağılımı hem de Dunnett dağılımından faydalanılır. Dunnett dağılımından yararlanarak c_1 , iki değişkenli t dağılımından yararlanarak c_2 kritik değerleri elde edilir.¹⁹ Bu durumda iki sonlanım noktası arasındaki korelasyon da dikkate alınmış olur.

nedeniyle özellikle çoklu doz ve çoklu sonlanım noktalarının olduğu durumlarda tercih edilebilir.³

Holm prosedürünün ayırabilme (separable) özelliği zayıftır, yani doğru olan yokluk hipotezini red etme sayısı fazladır. Bonferroni çoklu test prosedüründe her bir hipotez ailesi için α/k ile karşılaştırma yapılmaktadır. Bonferroni ile red edilen hipotez Holm prosedürü ile de red edilmektedir ancak, $\alpha/(k-i+1)$ ile α/k arasında kalan $p_{(i)}$ değerleri için Bonferroni ile red edilmeyen hipotez Holm ile red edilebilmektedir.¹⁵ Bu nedenle budanmış (truncated) versiyonu kullanılır ve böylece Bonferroni'den hem daha güçlü hem de ayırabilir test yapılır. Benzer durum, Hochberg, Fallback ve Dunnett prosedürleri için de geçerlidir.¹⁵

Doz-cevap çalışmalarında çoklu sonlanım sorununa ek olarak, çoklu dozların kontrolle kıyaslanması söz konusudur. Bu sorunun çözümü olarak, paralel denetleyici test stratejileri Dunnett testi için genelleştirilmiştir. Birincil hipotezde anlamlı bulunan dozlar için, ikincil değişkenlere ait hipotezler test edilir.²

SONUÇ VE ÖNERİLER

Çoklu karşılaştırmalarda Bonferroni düzeltmesi yapmak bilinen bir yöntemdir. Paralel denetleyici

cilar de aile başına hatayı kontrol altında tutan Bonferroni testinden türetilmiş prosedürleri içerir. Bunlar; Holm prosedürü, Hochberg prosedürü gibi yöntemlerdir. Seri denetleyici stratejilerde ise birincil hipotezler birleşim-kesişim prosedürü ile test edildikten sonra ikincil hipotezlerin Holm veya Hochberg prosedürü ile test edilmesi önerilir. Ağaç tipi denetleyici stratejilerde ise ilaçlar ve kombinasyonları için yapılan karşılaştırmalar daha kötü olmama (non-inferiority) testleri ile başlayıp üstünlük (superiority) testleri ile devam etmektedir.

Klinik denemelerde çoklu sonlanım noktalarına ait hipotezler test edilirken, Tip I hatanın kontrol altında tutulabilmesi için, hipotez ailelerinin sayısı, test edilecek sonlanım noktalarının sayısı, sonlanım noktaları arası korelasyon yapısı, test stratejisi ve buna bağlı olarak kullanılacak test tipi dikkatli seçilmelidir. Aksi takdirde, yanıltıcı test sonuçları ile karşılaşılabilir.

Ayrıca, deneme sonuçlarının raporlanması aşamasında da CONSORT kriterleri arasında önemli bir yeri olan birincil ve ikincil sonlanım noktalarının analizi ve kullanılan düzeltme yöntemleri mutlaka belirtilmelidir.

KAYNAKLAR

1. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ* 2010;340:c869.
2. Dmitrienko A, Offen W, Wang O, Xiao D. Gatekeeping procedures in dose-response clinical trials based on the Dunnett test. *Pharm Stat* 2006;5(1):19-28.
3. Dmitrienko A, Offen WW, Westfall PH. Gatekeeping strategies for clinical trials that do not require all primary effects to be significant. *Stat Med* 2003;22(15):2387-400.
4. D'Agostino RB Sr. Controlling alpha in a clinical trial: the case for secondary endpoints. *Stat Med* 2000;19(6):763-6.
5. Li JD, Mehrotra DV. Gatekeeping testing via adaptive alpha allocation. *Biom J* 2008;50(5):704-15.
6. Dmitrienko A, Tamhane AC. Gatekeeping procedures with clinical trial applications. *Pharm Stat* 2007;6(3):171-80.
7. Bretz F, Maurer W, Brannath W, Posch M. A graphical approach to sequentially rejective multiple test procedures. *Stat Med* 2009;28(4):586-604.
8. O'Neill RT. Secondary endpoints cannot be validly analyzed if the primary endpoint does not demonstrate clear statistical significance. *Control Clin Trials* 1997;18(6):550-6; discussion 561-7.
9. Westfall PH, Krishen A. Optimally weighted, fixed sequence, and gatekeeping multiple testing procedures. *Journal of Statistical Planning and Inference* 2001;99(1): 25-40.
10. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 1979;6(2):65-70.
11. Hochberg Y. A sharper Bonferroni procedure for multiple tests of significance. *Biometrika* 1988;75(4):800-2.
12. Xu H, Nuamah I, Liu J, Lim P, Sampson A. A Dunnett-Bonferroni-based parallel gatekeeping procedure for dose-response clinical trials with multiple endpoints. *Pharm Stat* 2009; 8(4):301-16.
13. Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. *Biom J* 2008;50(5):667-77.
14. Dmitrienko A, Wiens BL, Tamhane AC, Wang X. Tree-structured gatekeeping tests in clinical trials with hierarchically ordered multiple objectives. *Stat Med* 2007;26(12):2465-78.
15. Holland BS, Copenhaver MD. Improved Bonferroni-Type Multiple testing procedures. *Psychological Bulletin* 1988;104(1):145-9.
16. Huang Y, Hsu JC. Hochberg's step-up method: Cutting corners off Holm's step-down method. *Biometrika* 2007;94(4):965-75.
17. Wiens B. A fixed-sequence Bonferroni procedure for testing multiple endpoints. *Pharmaceutical Statistics* 2003;2(3):211-5.
18. Wiens BL, Dmitrienko A. The fallback procedure for evaluating a single family of hypotheses. *J Biopharm Stat* 2005;15(6):929-42.
19. Genz A, Bretz F. Methods for the Computation of Multivariate t-Probabilities. *J Comp Graph Stat* 2002;11(4):950-71.