

Gözleme Dayalı Çalışmalarda Eğilim Skoru (Propensity Score) ve Tıp Bilimleri'nde Bir Uygulama

Propensity Score in Observational Studies and an Application in Medical Sciences

Elif Çiğdem KASPAR,^a
Nural BEKİROĞLU,^b
Mehmet GENÇELİ^c

^aBiyostatistik BD,
Yeditepe Üniversitesi Tıp Fakültesi,
^bBiyostatistik AD,
Marmara Üniversitesi, Tıp Fakültesi,
^cİstatistik Bölümü,
Yıldız Teknik Üniversitesi
Fen Edebiyat Fakültesi, İstanbul

Geliş Tarihi/Received: 06.04.2009
Kabul Tarihi/Accepted: 19.01.2010

Bu çalışma 8. Ulusal Biyoistatistik Kongresi, 2005, Bursa'da sözlü sunum olarak sunulmuştur.

Yazışma Adresi/Correspondence:
Elif Çiğdem KASPAR
Yeditepe Üniversitesi Tıp Fakültesi,
Biyostatistik BD, İstanbul,
TÜRKİYE/TURKEY
ecaltunok@yeditepe.edu.tr

ÖZET Amaç: Gözlemsel çalışmalarda ve/veya retrospektif çalışmalarda, araştırmacının olgu ve kontrol gruplarındaki birimleri rastlantısal olarak gruplara atama işleminde kontrolü yoktur. Bu nedenle, olgu-kontrol gruplarına düşen bireylerin ortak değişkenlerinde bazı farklılıklar gözlemlenebilir ve bu farklılıklarda tedavi etkisinin yanlı (biased) tahminlerine neden olabilmektedir. Bir dengeleme skoru olarak tanımlayabileceğimiz eğilim skoru (ES), tedavinin gözlenen ortak değişkenlere göre koşullu olasılığı olarak ifade edilir. ES tahmin ettikten sonra eşleştirme, tabakalara ayırma ve regresyon düzeltmesi yöntemleriyle yeniden örnekleme yapılarak retrospektif gözlemsel çalışmalarda örneklemedeki yanlılığın azaltılmasında, prospektif çalışmalardaki gibi kesinliğin artmasında, kanıt seviyesi bakımından görece düşük olan olgu-kontrol çalışmalarının, prospektif bir çalışmanın kanıt seviyesine çıkartılması ve bazı ortak değişkenlerin etkilerini ortaya koymak amacıyla kullanılır. Çalışmamızda ES öncesi ve sonrası örneklem için olgu ve kontrol bireylerine lojistik regresyon uygulanarak risk faktörlerinin anlamlılıklarındaki farkları ortaya koymak amaçlanmıştır. **Gereç ve Yöntemler:** Çalışmamızda, göğüs cerrahisi ameliyatı geçirmiş n=478 hasta kullanılmıştır. Ameliyat sonrası delirium tanısı alan ve almayan hastalara ait, 24 risk faktörüne lojistik regresyon uygulanmış ve sonuçlar elde edilmiştir. ES lojistik regresyon ile hesaplanmış, tabakalara ayırma yöntemi kullanılarak oluşan yeni örneğe istatistik analiz uygulanmış ve değerlendirilmiştir. **Bulgular:** ES ile oluşturulmuş yeni örneklemedeki olgu ve kontrol grupları hemen hemen aynı karakteristiklere sahip olmuş ve yanlılık azalmıştır. **Sonuç:** ES öncesi ve sonrası örneklem için olgu ve kontrol bireylerine uygulanan lojistik regresyon sonucunda risk faktörlerinin anlamlılıklarında değişiklik gözlemlenmiştir.

Anahtar Kelimeler: Eğilim skoru, geriye dönük (retrospektif) çalışmalar, yanlılığın azaltılması, olgu-kontrol çalışmaları, gözleme dayalı çalışmalar, tabakalara ayırma

ABSTRACT Objective: In observational and/or nonrandomized studies, investigators have no control over the treatment assignment. Therefore, some differences may exist on observed covariates in case-control groups, and these differences could lead to biased estimates of treatment effects. Propensity score (PS), which can be determined as a balancing score, is the conditional probability of being treated given covariates. New sample is generated by the estimated PS through matching, stratification, regression adjustment methods. In retrospective studies, it is aimed to reduce bias, to increase the precision of the estimates as in the prospective studies, to increase the relatively low evidence level of case-control studies to an evidence level of prospective study and to determine the effects of some covariates. In this study, we aimed to observe the difference between the significance in risk factors, when logistic regression was used for the sample of case-control groups' patients before and after adjusting the PS. **Material and Methods:** In this study, there are 478 patients who had thoracic surgery. Logistic regression was performed with regard to 24 risk factors, regarding patients with postoperative delirium (POD) and without POD after surgery. PS was calculated by logistic regression and statistical tests were performed for the new sample which is generated by the estimated PS through stratification. **Results:** After resampling, case-control groups had similar characteristics in risk factors and bias was reduced. **Conclusion:** The significance in risk factors, when logistic regression was used for the sample of case-control groups' patients before and after adjusting the PS, was observed to be different.

Key Words: Propensity score, retrospective studies, reduction bias, case-control studies, observational studies, stratification

İstatistik yöntemleri günümüzde ölçmeye ve verilerin değerlendirilmesine dayanan çok çeşitli bilim dallarında kullanılmaktadır. İstatistiğin kullanıldığı en önemli alanlardan birini de Tıp Bilimleri oluşturmaktadır. Tıp Bilimlerinde araştırmalar genelde gözlem sonucu elde edilen verilerin değerlendirilmesi ile ilerleme kaydetmektedir. Eğilim skoru (propensity skor) son on beş yılda geriye dönük gözleme dayalı çalışmalarda kullanılan önemli bir yöntemdir ve örneklemedeki yanlılığı (bias) azaltma özelliğine sahiptir.

Eğilim skoru, geriye dönük gözleme dayalı bir veri kümesinden hareket ederek tedavi etkilerinin tahminindeki (varsa) biası azaltmak için Rosenbaum ve Rubin (1983) tarafından bir yöntem olarak önerilmiştir. Gittikçe artarak kullanılan bu yöntem, daha ziyade tıbbi denemelerde ve ekonomik politika müdahalelerinin değerlendirilmesinde popüler hale gelmiştir.

Geriye dönük gözleme dayalı çalışmalarda rastlantısallık pratik veya etik nedenlerden dolayı sağlanamadığında rastlantısal olarak atanmayan olgu ve kontrol gruplarının gerek demografik özelliklerinde gerekse gözlenen ortak değişkenlerinde büyük farklar olabilir ve bu durumda olgu gruplarında çıktılar doğrudan karşılaştırılması yanıltıcı olur. Bu farklar olgu etkisinin yanlı (biased) tahminlerine sebep olur. Çalışma tasarımı gözlenen ortak değişkenlerdeki bilgilerin birleşmesi ile örneğin eşleştirilmiş örnekleme yaparak veya tedavi etkisinin tahmininde tabakalara ayırma veya kovaryans düzeltmesi yaparak bu farklardan kaçınılabılır. Ortak değişken sayısı arttığında, aynı veya benzer eşleştirilmiş çiftler bulmak zorlaşacaktır ve geleneksel düzeltme metotları (eşleştirme, tabakalara ayırma, kovaryans düzeltmesi) düzeltme için sadece sınırlı sayıda ortak değişken kullanabildiklerinden genellikle sınırlıdır. Eğilim skorunda ise böyle bir sınırlama yoktur. Eğilim skoru çok sayıda ortak değişken içerdiğinde eşleştirilmiş setlerin ve tabakaların oluşturulması için bir düzenektir, ortak değişken bilgilerinin özet ölçüsünü içerir.

Eğilim skoru geriye dönük gözleme dayalı çalışmalarda bir kez tahmin edildikten sonra eşleştirme, tabakalara ayırma (sınıflara ayırma), reg-

resyon düzeltmesi (kovaryans düzeltmesi) veya bu üçünün bazı birleşiminin kullanılması yöntemleriyle yeniden örnekleme yapılarak, başlıca örneklemedeki yanlılığın azaltılmasında, ileriye dönük çalışmalardaki gibi kesinliğin artmasında, kanıt seviyesi bakımından görece düşük olan olgu-kontrol çalışmalarının, ileriye dönük bir çalışmanın kanıt seviyesine çıkartılması ve bazı ortak değişkenlerin etkilerini ortaya koymak amacıyla kullanılır.

Gerçek eğilim skoru değeri bilindiğinde, eğilim skorunda eşleştirme veya tabakalara ayırma yapılarak tabakalar ve eşleştirilmiş kümeler elde edilir. Bu kümeler gerçekte heterojen olsalar bile eğilim skorunda homojen olarak şekillenirler yani olgu ve kontrol gruplarındaki eşit (veya eşite yakın) eğilim skoruna sahip birimler, kendi geçmiş ortak değişkenlerine bağlı olarak aynı (veya hemen hemen aynı) dağılıma sahip olma eğiliminde olacaktır. Bu nedenle örneklemedeki yanlılığı ortadan kaldırmak için yapılan düzeltmeler, tek tek tüm ortak değişkenleri kullanmak yerine eğilim skoru kullanılarak yapılabilir.¹

GEREÇ VE YÖNTEMLER

Çalışmamızda, Marmara Üniversitesi Hastanesi göğüs cerrahisi bölümünde 1996–2003 yılları arasında aynı doktor tarafından göğüs cerrahisi ameliyatı geçirmiş N=478 hasta kullanılmıştır. Ameliyat sonrası delirium tanısı alan (vaka) 18 hasta ile delirium tanısı almayan (kontrol) 460 hastaya ait ameliyat öncesi 10 risk faktörü ve ameliyat sonrası 14 risk faktörüne istatistik analiz uygulanmıştır. Olgu ve kontrol grupları arasında ortak değişkenler bakımından fark olup olmadığını görmek için nicel değişkenler için iki bağımsız örnek t testi, nitel değişkenler için Ki-kare testi ve Fisher's kesin olasılık testi uygulanmıştır. Eğilim skoruları tahmin etmek için lojistik regresyon uygulanmış ve delirium tanısı için önemli risk faktörleri belirlenmiştir. Eğilim skoru öncesi ve sonrası grupları arasındaki dengesizliği ortaya koymak açısından ise iki yönlü varyans analizi uygulanmıştır. İstatistiksel analizler için SPSS 13.0 programı kullanılmış ve $p < 0.05$ değeri istatistiksel olarak anlamlı kabul edilmiştir.

PROPENSITY SKOR TANIMI

Eğilim skoru, tam bir veri setine sahip ($i=1,2,\dots,N$) deneklerde χ_i olarak tanımlanır ve χ_i olgunun ($Z_i=1$), kontrole ($Z_i=0$) göre gözlenen ortak değişkenler vektörünün koşullu olasılığı olarak;

$$e(x_i) = pr(Z_i = 1 | X_i = x_i) \quad (1)$$

ifade edilir. Burada verilen X 'ler için Z_i 'lerin bağımsız olduğu varsayılır;

$$pr(Z_i = z_1, \dots, Z_N = z_N | X_1 = x_1, \dots, X_N = x_N) = \prod_{i=1}^N e(x_i)^{z_i} \{1 - e(x_i)\}^{1-z_i} \quad (2)$$

Eğilim skorları elde edildiğinde olgu öncesi karakteristiklerin koşullu dağılımları olgu ve kontrol grupları için aynı olması nedeniyle eğilim skoru bir dengeleme skorudur.² Açıkça ifade etmek gerekirse, eğilim skoru bir kişinin sadece ortak değişken skoru kullanılarak tedavi edilmesinin bir olasılık ölçüsüdür.

Ortak değişkenlerde kayıp veri durumu söz konusu değilse, eğilim skoru diskriminant analizi veya lojistik regresyon yardımıyla tahmin edilir. Ortak değişkenlerin çok değişkenli normal dağılıma sahip olduğu varsayıldığında diskriminant analizi kullanılır, lojistik regresyon için ise bu varsayımına gerek yoktur.¹

Eğilim skoru genellikle lojistik regresyon kullanılarak aşağıda ifade edilen yöntem ile tahmin edilmiştir.

$$\ln\left(\frac{PS}{1-PS}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p \quad (3)$$

$$PS = \frac{\text{beklenen}(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)}{1 + \text{beklenen}(\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p)} \quad (4)$$

$$\text{Odds oranı} = \frac{\text{beklenen}(\text{sabit} + (\text{katsayı} \times \text{var1}) + \dots + (\text{katsayı} \times \text{varN}))}{1 + \text{beklenen}(\text{sabit} + (\text{katsayı} \times \text{var1}) + \dots + (\text{katsayı} \times \text{varN}))} \quad (5)$$

$$\text{PPSkor} = \text{Oddsoranı} / (1 + \text{Oddsoranı}) \quad (6)$$

Eğilim skoru hesaplamasının amacı, geriye dönük rassal olmayan gözleme dayalı çalışmalarda olgu alacak birimin olasılığının yardımıyla (eğilim skoru) olgu etkisinin tahmininin düzeltilerek hemen hemen rassal bir deneyden elde edilen tahminin kesinliğini yaratmaktır. Aynı eğilim skoru ile biri olgu grubunda biri de kontrol grubunda iki birim bulunursa, bu durumda iki birimin her gruba rassal olarak atandığını, bir anlamda eşit benzerlikte olgu veya kontrolde olduğu düşünülebilir.

EĞİLİM SKORU İLE YANLILIK (BİAS) AZALTMA YÖNTEMLERİ

EŞLEŞTİRME

Araştırmacılar çalışmalarda, genellikle sınırlı sayıda olgu birimi ve buna karşılık fazla sayıda kontrol birimi ile karşılaşmaktadır. Mümkün tüm birimlerden çıktı ölçülerini sağlamak ise finansal açıdan oldukça belirsiz bir durumdur. Bu gibi durumlarda bazı örnekleme biçimlerinin uygulanması gerekmektedir. Eşleştirmede bu örnekleme biçimlerinden biridir ve araştırmacılar tarafından sıklıkla araştırmada kontrol edilmesine gerek duyulan ortak değişkenlerde, olgu bireyleri ile eşleşmiş kontrol birimleri seçmek için kullanılmaktadır. Eğilim skoru eşleştirmesi ile araştırmacı tek bir skaler değişken yardımıyla eşleştirme yaparak çok sayıda ortak değişken üzerinde kontrol sağlamaktadır.

Rosenbaum ve Rubin eğilim skoru kullanarak eşleşmiş örnek oluşturmak için üç teknik önermişlerdir.³

i. Tahmin edilen eğilim skora bağlı olarak mümkün olan en yakın eşleştirme: Bu yöntemde olgu ve kontrol bireyleri rassal şekilde dizilir. İlk olgu bireyi ile eğilim skoru mümkün olan en yakın kontrol bireyi ile eşleştirilir. Her iki bireyde eşleştirme yapılacak örneklemeden çıkartılır. Bir sonraki olgu bireyi seçilir. Eşleştirme yapmak için tahmini eğilim skorunun logitlerinin kullanılması tavsiye edilir.³ $\hat{q}(X)$ in dağılımı genellikle yaklaşık olarak normal olduğundan $\hat{q}(X) = \log[1 - \hat{e}(X) / \hat{e}(X)]$ kullanılır.

ii. Eğilim skoru içeren Mahalanobis metrik eşleştirmesi: Ortak değişkenler çok değişkenli normal dağılıyorsa ve olgu ve kontrol gruplarının ortak kovaryans matrisleri var ise bu durumda Mahalanobis metrik eşleştirmesi eşit yüzdeli yanlılık azaltma tekniğidir. Burada yanlılık olguya sahip olanların ortalamasından kontrollerin ortalamasından çıkartılması ile bulunur. Diğer bir deyişle, tüm ortak değişkenlerdeki yanlılığın azalma yüzdesi eşittir ve eşleştirme ile yanlılığı yükselecek hiçbir ortak değişken yoktur.⁴ Mahalanobis metrik eşleştirmesinde, olgu ve kontroller rassal şekilde dizilirler. Olgu birimi i ile kontrol birimi j arasındaki Mahalanobis farkı;

$$d(i, j) = (u - v)^T C_{OR}^{-1} (u - v) \quad (7)$$

şeklinde dir.¹

İlk olgu birimi Mahalanobis farkın en yakın olduğu kontrol birimi ile eşleşir. Burada u ve v , $\{x^r, \hat{q}(x)\}^r$ nin olgu birimi i ve kontrol birimi j için eşleşen değişkenlerin değerleridir ve C_{OR} kontrol grubundaki $\{x^r, \hat{q}(x)\}$ in örnek kovaryans matrisidir. Eşleşen bu iki birim örneklemeden çıkarılır ve bu süreç tüm olgu birimleri için eşler bulunana kadar tekrar edilir.

iii. Eğilim skorları tarafından tanımlanan aralıklar arasında mümkün olan en yakın Mahalanobis metrik eşleştirmesi: Bu yöntem yukarıda anlatılan iki yöntemi birleştirmektedir. Tüm olgu birimleri rassal şekilde sıralanır ve ilk olgu birimi seçilir. Olgu birimlerinin tahmini eğilim skorlarının değer aralığındaki tüm kontrol birimleri seçilir ve az sayıda ortak değişkene bağlı olarak bu birimler ile olgu birimi arasındaki Mahalanobis mesafesi hesap edilir. En yakın kontrol birimi ve olgu birimi örneklemeden çıkarılır, geriye kalan tüm kontrol birimleri bir olgu birimi ile bir sonraki eşleşme için hazır bulunurlar ve bu süreç bu şekilde tekrarlanır.

Bu üç yöntem de örneklemedeki yanlılığın azaltılması için kullanışlı tekniklerdir. Tahmin edilen eğilim skoruna bağlı olarak mümkün olan en yakın eşleştirme hesaplaması en kolay eşleştirme tekniğidir.³ İkinci yöntem, eğilim skoru içeren Mahalanobis metrik eşleştirmesi yalnız değişkenler için daha küçük standart biçime dönüştürülmüş farklar üretirken, eğilim skorunda önemli farklar kalmaktadır. Üçüncü yöntem, eğilim skorlar tarafından tanımlanan aralıklar arasında mümkün olan en yakın Mahalanobis metrik eşleştirmesi bu üç yöntem arasındaki en iyi tekniktir. Olgu ve kontrol gruplarındaki ortak değişkenler arasındaki en iyi dengeyi üretmektedir. Bu üçüncü yöntem eğilim skoruna bağlı değer aralığı oluşturularak, araştırmacı rassal bir deney oluşturmaya çalışmaktığı şeklinde düşünülebilir.^{1,3}

TABAKALARA AYIRMA

Tabakalara ayırma yöntemi de gözleme dayalı çalışmalarda olgu ve kontrol grupları arasındaki yanlılığı kontrol etmek için kullanılır. Bu teknik gözlenen geçmiş karakteristikleri tarafından bireyleri tabakalara gruplar. Eğilim skoruna bağlı olarak tam tabakalara ayırma yapıldığında, tabakalar ara-

sında ortalama olgu etkisi, gerçek olgu etkisinin yansız tahmincisini üretir.² Eğilim skoruna bağlı olarak tabakalara ayırma, eğilim skorunu tahmin etmek için kullanılan k ortak değişkeni dengeler ve genellikle eğilim skoruna bağlı beş tabaka, bu ortak değişkenlerin her birindeki biasın %90 nını ortadan kaldırır.^{5,6} Eğilim skoru lojistik regresyon veya diskriminant analizi ile tahmin edildikten sonra araştırmacı, her iki grubun birlikte veya olgu grubunda ve kontrol grubunda ayrı olarak eğilim skoru değerlerine bağlı tabaka sınırlarına karar vermeli-dir.

REGRESYON DÜZELTMESİ

Regresyon düzeltmesinde, olgu etkisi τ şu şekilde tahmin edilir. Burada t olguyu, c kontrolü belirtir.

$$\hat{\tau} = (\bar{Y}_t - \bar{Y}_c) - \beta(\bar{X}_t - \bar{X}_c) \quad (8)$$

Geçmiş ortak değişkenlerin etkisi, yukarıdaki denklemin sağ tarafındaki ikinci terim çıkartılarak düzeltilir. Burada β geçmiş ortak değişkenlere bağlı olarak olgu ve kontrol grupları için bir regresyon tahminidir. Bu nedenden ötürü, olgu ve kontrol gruplarında eğilim skoruna bağlı yanıtların regresyonunu bulmak ve bunu kullanarak olgu etkisinin en son tahminini düzeltmek için eğilim skoru değişkeninin kullanımı oldukça kullanışlıdır.¹

BULGULAR

Tablo 1 ve Tablo 2 her iki grup için ayrı ayrı 24 risk faktörünün tanımlayıcı istatistiklerini içermektedir. Örneğin delirium tanısı almayan 460 hastada solunum yetmezliği geçirenlerin sayısı %18,5 iken, delirium tanısı alan 18 hastanın solunum yetmezliği geçirenlerinin sayısı %27,8 dir.

Burada olgu ve kontrol grupları arasında ortak değişkenler bakımından nicel değişkenlerde ortalamalar arasında fark olup olmadığı, nitel değişkenlerde ise oranlar arası fark olup olmadığı test edilmiştir. Buna göre ameliyat öncesi yaş, alkol kullanımı, tümör tipine göre acil operasyon türü, diyabet risk faktörleri ile ameliyat sonrası serumdaki kimyasal değerlerin anormallliği, operasyon süresi, uyku bozukluğu risk faktörlerinin p değerleri 0.05 altında çıkmış, yani iki grup arasında ortak değişkenler bakımından fark olduğu sonucuna varılmış-

TABLO 1: Ameliyat öncesi risk faktörlerinin tek değişkenli analizleri

Ameliyat Öncesi	Delirium tanısı almayan hasta (n=460)	Delirium tanısı alan hasta (n=18)	p değeri †
Yaş			0,005**
(18–59)	307 (%66,7)	6 (%33,3)	
(59>)	153 (%33,3)	12 (66,7)	
Cinsiyet			0,38
Erkek	310 (%67,4)	11 (%61,1)	
Bayan	150 (32,6)	7 (38,9)	
Kronik Hastalık ^a			0,55
Var	15 (%3,3)	0 (%0)	
Yok	445 (%96,7)	18(%100)	
Alkol Kullanımı			0,017**
Var	90 (%19,6)	8 (%44,4)	
Yok	370 (%80,4)	10 (%55,6)	
Psikiyatri sorunu			0,171
Var	55 (%12)	4 (%22,2)	
Yok	405 (%88)	14 (%77,8)	
Serebrovasküler hastalık			0,891
Var	3 (%0,7)	0 (%0)	
Yok	457 (%99,3)	18 (%100)	
Diyabet			0,046**
Var	34 (%7,4)	4 (%22,2)	
Yok	426 (%92,6)	14 (%77,8)	
Kemoterapi			0,536
Var	16 (%3,5)	0 (%0)	
Yok	444 (%96,5)	18 (%100)	
Tümöre göre operasyon			0,514
Var	244 (%53)	8 (%44,4)	
Yok	216 (%47)	10 (%55,6)	
Operasyon türü acil (tümörün tipine göre)			0,000**
1. derece acil	79 (%17,2)	10 (%55,6)	
2. derece acil	135 (%29,3)	2 (%11,1)	
3. derece acil	59 (%12,8)	4 (%22,2)	
4. derece acil	187 (%40,7)	2 (%11,1)	

† Ki-kare testi p değerleri, **p<0,005; ^a Herhangi bir akciğer, kalp, böbrek ve karaciğer hastalığının mevcudiyeti

tır. Eğilim skoruları tahmin etmek için lojistik regresyon uygulanmış ve delirium tanısı için önemli risk faktörleri belirlenmiştir.

Tablo 3’de lojistik regresyon sonucu yaş, psikiyatri sorunu, operasyon süresi, serumdaki kimyasal değerlerin anormalliği risk faktörlerinin p değerleri 0.05 altında çıkmış yani bu risk faktörlerinin parametre tahminleri anlamlı kabul edilmiştir. Modelimizin başlangıç olabirlik oranı (-2 LL₀) değeri 78,43 olup Nagelkerke R² değeri 0.528’dir.

Hosmer-Lemeshow goodness of fit testi ki-kare değeri 1,081 olup serbestlik derecesi 8 dir ve p değeri 0,998 çıkmış yani oluşturduğumuz modelin uygun bir model olduğu sonucuna varılmıştır.

Eğilim skoru genellikle lojistik regresyon kullanılarak (6) da ifade edilen yöntem ile tahmin edilmiştir ve ilk 10 delirium tanısı konmayan hasta için eğilim skorları şöyledir: 0.01644, 0.00339, 0.01951, 0.02112, 0.00000, 0.00335, 0.00635, 0.00000, 0.01095, 0.00768.

TABLO 2: Ameliyat sonrası risk faktörlerinin tek değişkenli analizleri.

Ameliyat Sonrası	Delirium tanısı almayan hasta (n=460)	Delirium tanısı alan hasta (n=18)	p değeri †
Solunum yetmezliği ^b			0,237
Var	85 (%18,5)	5 (%27,8)	
Yok	375 (%81,5)	13 (%72,2)	
Serumdaki kimyasal değerlerin anormalliği ^c			0,004**
Var	26 (%5,7)	5 (%27,8)	
Yok	434 (%94,3)	13 (%27,8)	
Operasyon süresi			0,004**
5 saat ve üzeri	47 (%10,2)	13 (%72,2)	
5 saatten az	413 (%89,8)	5 (%27,8)	
Yoğun bakım süresi	2,56 ± 7,89	3,05 ± 1,92	0,79 ^d
Uyku bozukluğu			0,001**
Var	0 (%0)	2 (%11,1)	
Yok	100 (%100)	16 (%88,9)	
Hipertansiyon			0,272
Var	58 (%12,8)	4 (%27,2)	
Yok	402 (%87,4)	14 (%77,8)	
Enfeksiyon			0,760
Var	91 (%19,8)	4 (%27,2)	
Yok	369 (%80,2)	14 (%77,8)	
Kan transfüzyonu			0,606
Var	136 (%29,6)	4 (%27,2)	
Yok	324 (%70,4)	14 (%77,8)	
Aminofilin kullanımı			0,464
Var	15 (%3,3)	1 (%5,6)	
Yok	445 (%96,7)	17 (%94,4)	
Antiaritmi kullanımı			0,371
Var	32 (%7)	2 (%11,1)	
Yok	428 (%93)	16 (%88,9)	
Antibiyotik kullanımı			0,564
Var	110 (%110)	4 (%22,2)	
Yok	350 (%76,1)	14 (%77,8)	
Streoid kullanımı			0,285
Var	26 (%5,7)	2 (%11,1)	
Yok	434 (%94,3)	16 (%88,9)	
Antihipertansif kullanımı			0,576
Var	74 (%16,1)	3 (%16,7)	
Yok	386 (%83,9)	15 (%83,3)	
Hareket edememe			0,288
Var	47 (%19,2)	3 (%16,7)	
Yok	413 (%89,8)	15 (%83,3)	

† Ki-kare testi p değerleri; ** p<0,005; ^bOda havasındaki P_{O_2} <55 mm Hg kandaki oksijen miktarının yetersizliği ve P_{CO_2} >44 mm Hg kandaki anormal karbondioksit seviyesi; ^c Kanda olması gerekli limit değerlerine sahip (sodyum 130-150 meq/L, potasyum 3.0-6 meq/L, glukoz 60-300 mg/dL) serum değerlerinden herhangi biri limit değerlerinden farklı ise serum değerlerinde anormallik vardır ^dt istatistiği p değeri (süreklili veri, ortalama±standart sapma şeklinde gösterilmiştir.)

TABLO 3: Eğilim skoru düzeltmesi öncesi lojistik regresyon çıktısı.

Değişkenler	B	Standart Hata	Wald	s.d.	Anlamlılık	Beklenen(B)	(B) için %95 Güven Aralıkları Beklenen	
							Alt	Üst
Tümör türü	0.515	0.966	0.284	1	0.594	1.673	0.252	11.112
Yaş	-1.907	0.884	4.654	1	0.031**	0.149	0.026	0.840
Kronik Hastalık	20.568	7836.056	0.000	1	0.998	8.6E+08	0.000	b
Alkol Kullanımı	-0.787	0.920	0.733	1	0.392	0.455	0.075	2.760
Psikiyatri sorunu	-2.095	0.926	5.118	1	0.024**	0.123	0.020	0.756
Serebrovasküler hast.	-1.360	22029.444	0.000	1	1.000	0.257	0.000	.
Ameliyat süresi	-2.732	0.968	7.960	1	0.005**	0.065	0.010	0.434
Hareket edememe	0.183	1.269	0.021	1	0.886	1.200	0.100	14.422
Solunum yetmezliği	-0.300	0.775	0.149	1	0.699	0.741	0.162	3.386
Uyku Bozukluğu	-43.443	28943.294	0.000	1	0.999	0.000	0.000	.
Hipertansiyon	-0.032	1.058	0.001	1	0.976	0.968	0.122	7.708
Enfeksiyon	-0.309	0.964	0.102	1	0.749	0.735	0.111	4.862
Kan Transfüzyonu	1.256	0.810	2.407	1	0.121	3.512	0.718	17.176
Diyabet	-1.198	0.955	1.574	1	0.210	0.302	0.046	1.961
Aminofilin kullanımı	-1.987	1.508	1.736	1	0.188	0.137	0.007	2.634
Antiaritmi kullanımı	18.993	5811.123	0.000	1	0.997	1.8E+08	0.000	.
Antibiyotik kullanımı	0.533	0.977	0.298	1	0.585	1.704	0.251	11.568
Antihipertansif Kul.	2.078	1.437	2.090	1	0.148	7.988	0.477	133.623
Kemoterapi	16.539	8692.541	0.000	1	0.998	1.5E+07	0.000	.
Streoid Kullanımı	-1.688	1.366	1.527	1	0.217	0.185	0.013	2.688
Serumdaki Kim. Deg. anormallliği	-1.944	0.939	4.290	1	0.038**	0.143	0.023	0.901
Cinsiyet	1.149	0.797	2.078	1	0.149	3.156	0.661	15.058
			1.432	4	0.839			
Tümör tipine göre Acil Op. türü	-42.552	25344.458	0.000	1	0.999	0.000	0.000	.
1. derece	-24.325	25075.102	0.000	1	0.999	0.000	0.000	.
2. derece	-24.169	25075.102	0.000	1	0.999	0.000	0.000	.
3. derece	-25.559	25075.102	0.000	1	0.999	0.000	0.000	.
4. derece	0.016	0.055	0.085	1	0.771	1.016	0.913	1.131
Hastanede yatış süresi	17.271	45148.061	0.000	1	1.000	3.2E+07		
Sabit								

** p<0,005.

Tüm delirium tanısı alan ve almayan hastalar için eğilim skorları tahmin edildikten sonra her iki grubun birlikte bu değerlerine bağlı olarak eşit örneklem sayısına sahip tabaka sınırlarına karar verilmiştir. Tablo 4 bu çalışmamız için yapılan tabakaları ve sınırlarını göstermektedir.

Tabakalara ayırdıktan sonra çalışmamızda delirium tanısı alan 12 hasta ile almayan 163 hasta kalmış ve böylece birbiri ile benzer dağılıma sahip olgu ve kontrol bireyleri seçilmiştir. Tabakalara ayırmadan önceki ve sonraki F oranları Tablo 5 ve 6'da verilmiştir.

Yukarıdaki tablolarda görülebileceği gibi tabakalara ayırdıktan sonra birbirleri ile benzer ortala-

TABLO 4: 5'li tabakalara ayrıldığında eğilim skoru düzeltmesi sonrası kontrol ve olgu gruplarına düşen örneklem sayısı.

Tabakalar	Sınırlar	Delirium tanısı alan	Delirium tanısı almayan
1	0.5–0.115	7	28
2	0.115–0.05	2	33
3	0.05–0.022	1	34
4	0.022–0.0121	1	34
5	0.0121–0.006	1	34

malara sahip olgu ve kontrol gruplarının elde edildiğine F oranı sonucunda karar verilmiştir. Eğilim skoruna bağlı olarak, tabakalara ayırmadan önce

TABLO 5: Tabakalara ayırmadan önce ve sonra iki grup arasındaki dengesizliği ölçmek amacıyla hesaplanan F-oranları.

Ameliyat Öncesi	Tabakalara ayırmadan önce	Tabakalara ayırdıktan sonra
	F oranı †	F oranı†
Yaş	8,55 **	0,021
Cinsiyet	0,267	0,043
Kronik Hastalık	0,606	-
Alkol Kullanımı	6,578**	0,055
Psikiyatri sorunu	1,687	0,662
Serebrovasküler hastalık	0,118	-
Diyabet	5,207**	0,349
Kemoterapi	0,648	-
Tümör türü	0,044	0,021
Operasyon türü acil	14,9**	0,458

† F oranı: $\chi^2 = F_{(r,\infty)} \cdot r$ (r= serbestlik derecesi); †F oranı= t².

** p<0,05

TABLO 6: Tabakalara ayırmadan önce ve sonra iki grup arasındaki dengesizliği ölçmek amacıyla hesaplanan F-oranları.

Ameliyat Sonrası	Tabakalara ayırmadan önce	Tabakalar ayırdıktan sonra
	F oranı †	F oranı
Solunum yetmezliği	0,980	1,284
Serumdaki kimyasal değerlerin anormallığı	13,982**	1,077
Operasyon süresi	8,357**	0,744
Yoğun bakım süresi	0,068*	0,082
Uyku bozukluğu	51,32**	-
Hipertansiyon	1,418	0,197
Enfeksiyon	0,065	0,459
Kan transfüzyonu	0,451	0,058
Aminofilin kullanımı	0,282	0,936
antiaritmi kullanımı	0,453	-
Antibiyotik kullanımı	0,027	0,705
Streoid kullanımı	0,936	0,002
Antihipertansif kullanımı	0,004	0,010
Hareket edememe	0,769	0,315

†F oranı: $\chi^2 = F_{(r,\infty)} \cdot r$ (r= serbestlik derecesi); †F oranı= t².

** p<0,05

yaş, alkol kullanımı, diyabet, tümör tipine göre acil operasyon türü, serumdaki kimyasal değerlerin anormallığı, operasyon süresi ve uyku bozukluğu değişkenlerinde delirium tanısı alanlar ile alma-

yanlar arasında farklılık olduğu görülmüş fakat tabakalara ayırma yapıldıktan sonra bu farkların düzeldiğine F oranı sonucunda karar verilmiştir. Kronik hastalık, serebrovasküler hastalık, uyku bozukluğu, antiaritmi kullanımı, kemoterapi değişkenlerinin var veya yok şeklindeki iki yanıtında, tabakalara ayırma yapıldıktan sonra yanıt değerlerinde sabit değer gözlenildiğinden bu değişkenlere istatistik analiz uygulanamamıştır. Elde ettiğimiz yeni örneklem ile lojistik regresyon yardımıyla risk faktörlerinin anlamlılıkları test edilmiş ve sonuçlar Tablo 7’de verilmiştir.

Tablo 7’deki çıktıya göre hiçbir risk faktörünün parametre tahmini p değerleri 0,05 altında çıkmamış yani anlamlı kabul edilmemiştir. Ancak psikiyatri sorunu ve ameliyat süresi risk faktörlerinin parametre tahmini p değerleri 0,10 dan küçük çıkmıştır. Modelimizin başlangıç olabilirlik oranı (-2 LL₀) değeri 76,229 olup Nagelkerke R² değeri 0.159, ROC eğrisi altındaki alan 0,77’dir. Hosmer-Lemeshow goodness of fit testi ki-kare değeri 9,969 olup serbestlik derecesi 8 dir ve p değeri 0,273 çıkmış yani oluşturduğumuz modelin uygun bir model olduğu sonucuna varılmıştır.

SONUÇ VE TARTIŞMA

Özellikle, eğilim skoru uygulayarak rassal olmayan (retrospective) geriye dönük çalışmalarla elde edilen sonuçlar, hemen hemen (prospective) ileriye dönük rassal çalışmalar gibi çoğu kez doğru sonuçlar vermektedir. Eğilim skoru, istatistik analizlerde yoğun olarak özellikle uygulamalı tıp alanında kullanılmaktadır. Rassal klinik denemelerin maliyeti artıkça ve çoğu araştırmacının daha az pahalı olan araştırmalara yönelmesi sonucunda eğilim skorumun bir yöntem olarak kullanımı artış göstermektedir. Çalışmanın tasarım aşamaları (eşleştirme veya tabakalara ayırma) göz önüne alındığında eğilim skoru yöntemi azami faydayı ürettiği görülmüştür. Bu fayda aynı zamanda zaman ve maliyetten de tasarruf ederek gerçek olgu etkisinin daha kesin tahminlerini sağlar. Ayrıca bu tasarruf ile de çalışma için uygun olmayacak bireylerin çalışmaya alınmasından kaçınılmış olacaktır.¹ Çalışmanın en önemli tarafı ise eğilim skor uygulanarak örneklemedeki yanlılığın (bias) azaltılmasıyla, (retrospec-

TABLO 7: Eğilim skoru düzeltilmesi sonrası lojistik regresyon çıktısı.

	B	Standart Hata	Wald	s.d.	Anlamlılık	Beklenen(B)	Beklenen(B) için %95 Güven Aralıkları	
							Alt	Üst
Tümör türü	0,662	1,012	0,428	1	0,513	1,939	0,267	14,084
Yaş	-1,354	0,959	1,993	1	0,158	0,258	0,039	1,692
Alkol Kullanımı	-0,387	0,999	0,150	1	0,699	0,679	0,096	4,810
Psikiyatri sorunu	-1,734	0,958	3,278	1	0,070	0,177	0,027	1,154
Ameliyat süresi	-2,091	1,069	3,824	1	0,051	0,124	0,015	1,005
Hareket edememe	0,226	1,281	0,031	1	0,860	1,254	0,1012	15,444
Solunum yetmezliği	-0,223	0,790	0,080	1	0,778	0,800	0,170	3,763
Hipertansiyon	-0,170	1,067	0,025	1	0,873	0,844	0,104	6,826
Enfeksiyon	-0,164	0,960	0,029	1	0,864	0,849	0,129	5,566
Kan Transfüzyonu	0,902	0,826	1,193	1	0,275	2,465	0,489	12,436
Diyabet	-1,046	0,946	1,222	1	0,269	0,351	0,055	2,244
Aminofilin kullanımı	-1,777	1,479	1,443	1	0,230	0,169	0,009	3,072
Antibiyotik kullan.	0,224	0,977	0,053	1	0,819	1,251	0,184	8,490
Antihipertansif Kul.	1,519	1,485	1,046	1	0,306	4,570	0,249	84,001
Stteroid Kullanımı	-1,374	1,400	0,963	1	0,326	0,253	0,016	3,937
Serumdaki Kim. Deg. anormalliği	-1,346	1,061	1,611	1	0,204	0,260	0,033	2,080
Cinsiyet	0,974	0,794	1,505	1	0,220	2,649	0,559	12,565
	-		1,200	1	0,753			
Tümör tipine göre Acil Op. türü	15,931	40192,970	0,000	1	1,000	0,000	0,000	.
1. derece	1,190	1,113	1,143	3	0,285	3,287	0,371	29,119
2. derece	1,316	1,388	0,899	1	0,343	3,727	0,246	56,582
3. derece	0,012	0,054	0,050	1	0,824	1,012	0,911	1,125
Hastanede yatış süresi	1,874	3,488	0,288	1	0,591	6,511		
Sabit								

** p<0,005.

tive) rassal olmayan geriye dönük çalışmaların, ileriye dönük (prospective) rassal çalışmalar kadar yakın sonuçlar vermesidir.^{7,8} Böylelikle kanıt seviyesi bakımından görece düşük olan olgu-kontrol çalışmalarının, ileriye dönük bir çalışmanın kanıt seviyesine çıkartılması sağlanmış olacaktır.

Çalışmamızda eğilim skor yöntemiyle yanlılığı azaltılmış örnekleme uygulanan istatistiksel analiz sonucu, Nagelkerke değeri düşük çıkmış olsa bile klinik olarak olası tüm risk faktörlerinin incelenmiş olması çalışmamızın diğer bir kuvvetli tarafıdır. Bu nedenle delirium tanısına klinik risk faktörlerinin dışında başka faktörlerinde etki etmiş olabileceği düşünülmektedir.

Delirium tanısı az rastlanır bir durum olduğundan çalışmamızda olgu sayısının oldukça düşük olması çalışma sonuçlarını olumsuz yönde etkilemiştir. Bu nedenle de eğilim skoru tahmin ettikten

sonra eşleştirme yöntemi yerine daha büyük örneklem ile çalışmak için tabakalara ayırma yöntemi kullanılmıştır. Olgu sayısı daha fazla olsaydı eğilim skoru öncesi ve sonrası yapılan istatistik analizler farklı sonuçlar verebilirdi. Bu durum çalışmanın en önemli kısıtıdır.

Cochran (1968), ortak değişken veya değişkenin alt sınıflamasına bağlı olarak yapılan beş tabakaya ayırmanın yanlılığı (biası) %90 olarak azalttığını göstermiştir.⁶ Bu nedenle çalışma için eğilim skorunu beş tabakaya ayırma yöntemi tercih edilmiştir. Tabakalara ayırma sonucunda tam eşleştirme sağlanamamış, denek sayısı 478 den 175 düşmüştür çünkü olgu ve kontrol grupları eğilim skorlarının birbirini tutmayan değişim aralıklarına sahiptir.⁹ Çalışmamızda delirium tanısı alan hastaların minimum ve maksimum eğilim skorları 0,00687 ve 1,00, delirium tanısı almayan hastalar

için ise minimum ve maksimum eğilim skorları 0,509 ve 0,00'dır. En yüksek eğilim skoruna sahip delirium tanısı alanlar ($>0,509$) ve en düşük eğilim skoruna sahip delirium tanısı almayanlar ($<0,00687$) çalışmadan çıkarılmıştır.

Eğilim skorun tanımında sözü edilen Rosenbaum ve Rubin (1983) eğilim skoru hesaplanması

ve sonrasında yapılan uygulamaları tam veri kümesi için geliştirmiştir. Ancak günümüzde eğilim skorunun kayıp veriler için hesaplanması istatistik programlarda henüz yer almamakta, kayıp verilere sahip veri kümeleri için çalışmalar sürdürülmektedir. Bu konudaki en önemli çalışmalar olarak ta D'Agostino'nun yapıtlarına değinilmektedir.¹⁰

KAYNAKLAR

1. D'Agostino RB Jr. Tutorial in biostatistics: propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat Med* 1998;17(19): 2265-81.
2. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* 1983 70(1):41-55.
3. Rosenbaum PR, Rubin DB. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *Am Stat* 1985;39(1): 33-8.
4. Rubin DB. Multivariate matching methods that are equal percent bias reducing: some examples. *Biometrics* 1976; 32(1):109-20.
5. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc* 1984;79(387):516-24.
6. Cochran WG. The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* 1968;24(2): 295-313.
7. Boening A, Friedrich C, Hedderich J, Schoettler J, Fraund S, Cremer JT. Early and medium-term results after on-pump and off-pump coronary artery surgery: a propensity score analysis. *Ann Thorac Surg* 2003;76(6): 2000-6.
8. Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. *Pharmacoepidemiol Drug Saf* 2004;13(12):841-53.
9. Rosenbaum PR. Constructing matched sets and strata. *Observational Studies*. 1st ed. New York: Springer-Verlag; 1995. p. 200-21.
10. D'Agostino RB Jr, Rubin DB. Estimating and using propensity scores with partially missing data. *J Am Stat Assoc* 2000;95(451):749-59.
11. Cameron E, Pauling L. Supplemental ascorbate in the supportive treatment of cancer: Prolongation of survival times in terminal human cancer. *Proc Natl Acad Sci USA* 1976;73(10):3685-9.
12. Cochran WG. The planning of observational studies of human populations. *Journal of the Royal Statistical Society, Series A* 1965; 128:134-55.
13. Doll R, Hill AB. Mortality of British doctors in relation to smoking: observations on coronary thrombosis. *Natl Cancer Inst Monogr* 1966; 19:205-68.
14. Hays WL. The general linear model and analysis of variance. *Statistics*. 5th ed. New York: Harcourt Brace College Publishers; 1994. p.376-8.
15. Rubin DB. Assignment to the treatment group on the basis of a covariate. *JEBS* 1977;2(1):1-26.
16. Stone RA, Obrosky DS, Singer DE, Kapoor WN, Fine MJ. Propensity score adjustment for pretreatment differences between hospitalized and ambulatory patients with community-acquired pneumonia. *Pneumonia Patient Outcomes Research Team (PORT) Investigators. Med Care* 1995;33(4 Suppl):AS56-66.
17. Ury HK. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics* 1975;31(3):643-9.