



Betül DAĞOĞLU HARK<sup>1, a</sup>  
Hüseyin Refik BURGUT<sup>2, b</sup>

<sup>1</sup> Fırat Üniversitesi,  
Tıp Fakültesi,  
Biyostatistik ve Tıbbi Bilişim  
Ana Bilim Dalı,  
Elazığ, TÜRKİYE

<sup>2</sup> Maltepe Üniversitesi,  
Tıp Fakültesi,  
Biyostatistik Ana Bilim Dalı,  
İstanbul, TÜRKİYE

<sup>a</sup> ORCID: 0000-0002-5189-1929

<sup>b</sup> ORCID: 0000-0002-8335-1927

Geliş Tarihi : 21.02.2022  
Kabul Tarihi : 16.08.2022

#### Yazışma Adresi Correspondence

Betül DAĞOĞLU HARK  
Fırat Üniversitesi,  
Tıp Fakültesi,  
Biyostatistik ve Tıbbi  
Bilişim Ana Bilim Dalı,  
Elazığ - TÜRKİYE

bdagoglu@firat.edu.tr

## SNAP-II VE SNAPPE-II Risk Tahmin Modellerinin Performansının Değerlendirilmesi \*

**Amaç:** Risk tahmin modellerinin performanslarının değerlendirilmesi için farklı istatistiksel yöntemler kullanılmaktadır. Modelin genel performansının değerlendirilmesinde Brier skoru ve R<sup>2</sup> yöntemleri kullanılırken ayırım yeteneğinin değerlendirilmesi için alıcı-işletim karakteristikleri (Receiver-operating characteristic-ROC) eğrisi ve ayırım eğrileri kullanılmaktadır. Bunlara ek olarak kalibrasyon için kalibrasyon eğrileri ve Hosmer-Lemeshow test istatistiği ile modelin performansı değerlendirilmektedir. Model performansını değerlendirmek için kullanılan istatistiksel yöntemlerle, SNAP-II (Score for Neonatal Acute Physiology) ve SNAPPE-II (Score for Neonatal Acute Physiology- Perinatal Extension-II) risk tahmin modellerinin performanslarının değerlendirilmesi amaçlanmaktadır.

**Gereç ve Yöntem:** Çukurova Üniversitesi Tıp Fakültesi Balcalı Hastanesi'nin Yeni Doğan Yoğun Bakım Ünitesine, 2013-2015 Temmuz yılları arasında yatırılan 864 bebeğin ölüm olasılıklarını tahmin etmek için SNAP-II ve SNAPPE-II risk tahmin modelleri kullanılmıştır.

**Bulgular:** SNAP-II risk tahmin modelinin genel performansını değerlendirmek için kullanılan yöntemlerden Brier Skoru 0.076 ve Nagelkerke R<sup>2</sup> 0.479 olarak bulunmuştur. Ayırım yeteneğini değerlendirmek için kullanılan ROC-AUC değeri 0.885 (GA: 0.851-0.919)'dir (p<0.001). Kesim noktası 14.50 olarak belirlenmiş ve bu noktadaki seçicilik 0.725, duyarlılık 0.897 ve doğruluk 0.750 olarak elde edilmiştir. Kalibrasyon için kullanılan Hosmer-Lemeshow testinin p değeri 0.619 ( $\chi^2=6.249$ ) olarak bulunmuştur. SNAPPE-II risk tahmin modeline ait Brier skoru 0.056 iken Nagelkerke R<sup>2</sup> değeri 0.619 olarak bulunmuştur. ROC-AUC değeri 0.921 (GA: 0.890-0.953)'dir (p<0.001). Kesim noktası 26.50 olarak belirlenmiş ve bu noktadaki seçicilik 0.877, duyarlılık 0.857 ve doğruluk 0.874 olarak elde edilmiştir. Hosmer-Lemeshow testinin p değeri 0.968 ( $\chi^2=2.338$ ) olarak bulunmuştur.

**Sonuç:** SNAP-II ve SNAPPE-II risk tahmin modellerinin her ikisi de yüksek model performansına sahiptir. Ayrıca SNAPPE-II risk tahmin modelinin genel performans, ayırım yeteneği ve kalibrasyon bakımından SNAP-II'den daha iyi olduğu elde edilmiştir.

**Anahtar Kelimeler:** Risk tahmin modeli, lojistik regresyon, performans ölçütleri, SNAP-II, SNAPPE-II

### Evaluation of Performance of SNAP-II and SNAPPE-II Risk Prediction Models

**Objective:** Different statistical methods are used to evaluate the performance of risk prediction models. While Brier score and R<sup>2</sup> methods are used to evaluate the overall performance of the model, ROC curve and discrimination curves are used to evaluate discrimination ability. Additionally, calibration curves and the Hosmer-Lemeshow test statistic are used to evaluate the calibration.

It is aimed to evaluate the performance of SNAP-II and SNAPPE-II risk prediction models with the statistical methods used to evaluate model performance.

**Materials and Methods:** SNAP-II and SNAPPE-II risk prediction models were used to estimate the mortality probabilities of 864 infants hospitalized in the Neonatal Intensive Care Unit of Çukurova University Medical Faculty Balcalı Hospital between July 2013 and 2015.

**Results:** The Brier Score of the SNAP-II risk prediction model was found to be 0.076 and the Nagelkerke R<sup>2</sup> to be 0.479. The ROC-AUC value was 0.885 (CI: 0.851-0.919). The cutoff point was determined as 14.50. The p value of the Hosmer-Lemeshow test was found to be 0.619 ( $\chi^2=6.249$ ). While the Brier score of the SNAPPE-II risk prediction model was 0.056, the Nagelkerke R<sup>2</sup> value was 0.619. The ROC-AUC value was 0.921 (CI: 0.890-0.953). The cutoff point was determined as 26.50. The p value of the Hosmer-Lemeshow test was found to be 0.968 ( $\chi^2=2.338$ ).

**Conclusion:** Both SNAP-II and SNAPPE-II risk prediction models have high model performance. In addition, it was found that the SNAPPE-II risk prediction model was superior to SNAP-II in terms of overall performance, discrimination ability and calibration.

**Key Words:** Risk prediction model, logistic regression, performance measures, SNAP-II, SNAPPE-II

### Giriş

Risk değerlendirmesi, modern klinik ve koruyucu hekimlik uygulamalarında önemli bir rol oynar. Yaşam şekli, genetik yatkınlık ve yaş gibi faktörler bir hastalığın gelişme riskini etkilemektedir. Bu risk faktörlerini içeren risk tahmin modellerinin

\* Bu çalışma, Çukurova Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Ana Bilim Dalı'nda yürütülen Betül DAĞOĞLU HARK'a ait yüksek lisans tezinin bir kısmıdır.

geliştirilmesi istatistiksel bir süreçtir (1). Tanı ve prognoz için en sık kullanılan lojistik risk tahmin modellerinin performansının değerlendirilmesinde kullanılan yöntemler önem kazanmıştır. Bir tahmin modeli kurulduktan sonra bu kurulan modelin tahminlerde ne kadar iyi olduğunun ölçülmesi modelin performans ölçütlerine bağlıdır. Performans değerlendirmesinde kullanılan geleneksel ölçüm yöntemleri; genel model performansının değerlendirilmesinde açıklanan varyasyon ( $R^2$ ) katsayısı ve Brier skor, ayırım yeteneğinin testi için ROC eğrisi altında kalan alan değeri ve son olarak da kalibrasyon için uyum iyiliği istatistikleri kullanılmaktadır (2).

**Genel Performansın Değerlendirilmesi:** Modelin genel performansının değerlendirilmesinde tahmin edilen değer ile gerçek değer arasındaki farkın incelenmesi merkezi bir role sahiptir. Sürekli ölçümlü çıktılarına sahip modellerde bu değerlendirme  $(Y - \hat{Y})$  farkına dayanmaktadır. İkili çıktıları modeller için  $Y$ , (0,1) aralığında tanımlanmaktadır ve tahmin edilen olasılığa yani  $p$ 'ye eşittir. Bu gözlenen ve tahmin edilen değerler arasındaki fark, "modelin uyum iyiliği" prosedürü ile sınırlanmaktadır. İyi bir model için bu fark minimum seviyelerde olması gerekmektedir (3).

Açıklanan varyasyon ( $R^2$ ), sürekli ölçüme sahip sonuç değişkenli modellerin performanslarının değerlendirilmesi için kullanılmaktadır (4). Genelleştirilmiş doğrusal modellerde artık varyanslarını belirlemek oldukça zor bir süreçtir ve bu modeller maksimum olabilirlik yaklaşımına dayanmaktadır (5). Bu nedenlerle, Cox-Snell ve Nagelkerke tarafından iki farklı  $R^2$  yaklaşımı geliştirilmiştir. Nagelkerke  $R^2$ , Cox ve Snell tarafından geliştirilen  $R^2$  yaklaşımının düzeltilmiş halidir.  $R^2$  istatistiği, bir uyum iyiliği ölçütü değildir. Bu istatistik "Açıklayıcı değişkenlerin, cevap değişkenini ne kadar açıklayabildiğinin" bir katsayısıdır ve etki büyüklüğünün ölçüsü olarak da tanımlanabilmektedir (6).

Brier skoru, kuadratik skorlama yöntemine dayanan bir yaklaşımdır. Bu istatistik değeri, ikili sonuç değişkeni olan  $Y$  ve tahmin edilen  $p$  arasındaki farkın karesi  $((Y - p)^2)$  esasına dayanır. Brier skor 0 ile 0.25 aralığında değer aldığı iyi bir model çıkarsaması yapılırken 0.50 dolaylarında değer olması modelin bilgi içermeyen bir model olduğu sonucuna varılmasını sağlar (7).

**Ayırım Yeteneğinin Değerlendirilmesi:** Vaka ve kontrol dağılımları arasındaki farkın belirlenmesine yönelik olarak kullanılan yaklaşım ROC eğrisi altında kalan alanın (AUC) belirlenmesidir (8). ROC, duyarlılık ve (1-seçicilik) eksenlerine sahip ve muhtemel tüm c kesim noktaları için çizilen bir eğridir (9). AUC değerinin hesaplanabilmesi için parametrik, semi-parametrik ve parametrik olmayan yaklaşımlar olmak üzere 3 yaklaşım söz konusudur. Parametrik metotta, "hastalığı olan ve olmayan bireyler kendi içlerinde çok değişkenli normal dağılım gösteriyor" varsayımı aranır. Semi-parametrik yaklaşımda çok değişkenli normal dağılım varsayımı aranmaksızın risk skorlarının monoton dönüşümleri üzerinden AUC değeri hesaplanmaktadır. Yapılan monoton dönüşümle hasta ve sağlam kategorilerine ait risk skorları normal dağılıma dönüştürülür. Bu nedenle

semi-parametrik yaklaşımla AUC değerinin hesaplanması "bi-normal AUC" olarak da adlandırılmaktadır. Parametrik olmayan yaklaşımla ise AUC değerinin yansız tahmin edicisi Mann-Whitney istatistiğine dayanır (10).

**Kalibrasyonun Değerlendirilmesi:** Kalibrasyon genel bir tanım ile gözlem değerleri ile tahmin değerleri arasındaki farkın ölçümüdür. Kalibrasyonu grafiksel olarak değerlendirmek mümkündür. Bu grafik, x ekseninde tahmin değerleri y ekseninde gözlem değerleri yer alan ve serpm (scatter) grafiği ile aynı görsele sahip olan bir grafikdir. İyi bir tahmin modelinde kalibrasyon eğrisi 45 derecelik bir eğime sahip olmalıdır. İkili sonuç değişkenine sahip modellerde bu eğrinin y ekseninde sadece 0 ve 1'lerden oluşacaktır. Bu durumda, gözlem değerlerinin tahmininde Loess algoritmasına dayanan Smoothing tekniği kullanılır (11).

Kalibrasyonun test prosedürü için Hosmer-Lemeshow istatistiği kullanılmaktadır. Hosmer-Lemeshow istatistiği  $H_0$  hipotezi altında grup sayısı eksi 2 serbestlik dereceli ki-kare dağılımı göstermektedir ( $\chi^2(g-2)$ ). Yapılan simülasyon çalışmalarıyla grup sayısının 10 olarak alınmasının uygun olacağı gösterilmiştir. Ancak bu test yöntemi birey sayısına ( $n$ ) bağlı olduğu için bazı sınırlamalar içermektedir. Yani  $n$  yeteri kadar büyük ise sonuç istatistiksel olarak anlamlı  $n$  küçük ise anlamsız olarak bulunacaktır (8).

Bu çalışmada, ikili yanıt değişkenine sahip risk tahmin modellerinin performansını değerlendirmek için kullanılan performans ölçütlerinin, SNAP-II ve SNAPPE-II risk skorları için değerlendirilmesi amaçlanmaktadır.

## Gereç ve Yöntem

**Araştırma ve Yayın Etiği:** Çalışmaya Çukurova Üniversitesi Tıp Fakültesi Girişimsel Olmayan Klinik Araştırmalar Etik Kurulu onayı (04.09.2015-2015/45 nolu) alındıktan sonra başlandı.

Veri seti, Çukurova Üniversitesi Tıp Fakültesi Balcalı Hastanesi'nin Yeni Doğan Yoğun Bakım Ünitesine, 2013-2015 Temmuz yılları arasında yatırılan tüm bebekleri içermektedir. Çalışma toplam 1281 yeni doğanı içermektedir. Yirmi dört saat içinde ölen veya taburcu edilen, yaşla bağdaşmayan konjenital malformasyonu olan, çalışma protokolü eksik bilgiler nedeniyle tamamlanamayan ve hasta dosyalarına ulaşılmayan yeni doğanlar çalışmaya alınmamıştır. Bu dışlama kriterleri nedeniyle çalışmaya 864'lük veri seti ile devam edilmiştir. Çalışmaya alınan gözlem sayısı; I. Tıp hata miktarı (alfa) 0.05, testin gücü (1-beta) 0.8, fark değeri 0.07 ve alternatif hipotez ( $H_1$ ) iki yönlü iken ROC-AUC analizi kullanılarak anlamlı bir fark bulunabilmesi için gerekli olan minimum örneklem büyüklüğü yatış sonlanma durumunun hayatta olduğu grup için 588, ölü grubu için 98 olmak üzere toplamda 686 gözlem olmalıdır (12). Minimum örnek büyüklüğünün sağlanması ve 2013-2015 yılları arasındaki tüm dahil etme kriterlerine uyan yeni doğanların çalışmaya dahil edilebilmesi adına toplam 864 hasta çalışmaya alınmıştır. Risk tahmin modeli olarak SNAP-II ve SNAPPE-II kullanılmıştır. SNAP-II ve SNAPPE-II yeni doğan yoğun bakım ünitesine yatırılan bebeklerin ölüm

olasılıklarını hesaplamak için kullanılmaktadır. SNAP-II risk tahmin modeli; kan basıncı, vücut ısısı, PO<sub>2</sub>/FiO<sub>2</sub>, serum pH'ı, çoklu nöbetler ve idrar çıkışı bilgilerini içermektedir. SNAPPE-II, SNAP-II modeline doğum ağırlığı, gestasyonel yaşa göre küçük olma durumu ve 5 dakikalık Apgar skoru dahil olmak üzere 3 parametre daha ekler. Yüksek SNAP-II veya SNAPPE-II skorları, yenidoğanın daha ağır hasta olduğunu gösterir (13).

İstatistiksel değerlendirme, R 3.2.1 paket programı kullanılarak yapıldı. Kategorik ölçümler sayı ve yüzde olarak, sürekli ölçümlerse ortalama ve standart sapma olarak özetlendi. Kategorik ölçümlerin yatış sonlanma grupları ile karşılaştırılmasında ki-kare test istatistiği kullanıldı. Sürekli ölçümlerin normal dağılım varsayımını sağlayıp sağlamadığı Kolmogrov Smirnov testi ile test edildi. Yatış sonlanma grupları arasındaki sayısal ölçümlerin karşılaştırılmasında bağımsız gruplarda t-testi kullanıldı. Tüm testlerde istatistiksel önem düzeyi 0.05 olarak alındı. SNAP-II ve SNAPPE-II risk skorlarının genel performansının değerlendirilmesi için Nagelkerke R<sup>2</sup> ve Brier Skoru kullanılırken ayırım gücünün değerlendirilmesinde ROC eğrisi ve bununla ilişkili olan

AUC değeri ve güven aralıkları kullanılmıştır. Youden indeksi kullanılarak optimum kesim noktası belirlenmiş ve seçicilik, duyarlılık değerleri bu kesim noktasına göre saptanmıştır. Ayrıca uyum iyiliği ölçütü olan Hosmer-Lemeshow testi ile kalibrasyon değerlendirilmiştir.

### Bulgular

İlk olarak yeni doğan yoğun bakım ünitesine yatırılan bebeklerin yatış sonlanma durumuna göre SNAP-II ve SNAPPE-II risk tahmin modellerini oluşturan faktörlere ve demografik özelliklerine bağlı tanımlayıcılar Tablo 1'de verilmiştir. 2013-2015 yılları arasında yatırılan bebeklerde ölüm olasılığı %14.6 olarak bulunmuştur.

**SNAP-II Risk Tahmin Modelinin Performansının Değerlendirilmesi:** Tek değişkenli analizde SNAP-II risk tahmin modelini oluşturan tüm bağımsız değişkenler ile yatış sonlanma durumu arasındaki ilişki istatistiksel olarak anlamlı bulunmuştur. Buradan hareketle çoklu lojistik regresyon analizi uygulanmış ve Tablo 2'deki sonuçlar elde edilmiştir.

**Tablo 1.** Risk tahmin modellerini oluşturan faktörlere ve demografik özelliklerine bağlı tanımlayıcılar

Değişkenler		Yatış Sonlanma Durumu		P değeri
		Hayatta n(%)	Ölü n(%)	
Cinsiyet	Erkek	413(56.0)	68(54.0)	0.699
	Kız	325(44.0)	58(46.0)	
Doğum Şekli	NVYS	189(25.6)	26(20.6)	0.265
	C/S	549(74.4)	100(79.4)	
Celestone Kullanımı	Var	79(10.7)	44(34.9)	<0.001
	Yok	659(89.3)	82(65.1)	
Süpfaktan Kullanımı	Var	39(5.3)	68(54.0)	<0.001
	Yok	699(94.7)	58(46.0)	
Ortalama Kan Basıncı (mm Hg)	>=30	717(97.2)	98(77.8)	<0.001
	20-29	17(2.3)	9(7.1)	
	<20	4(0.5)	19(15.1)	
En Düşük Vücut Isısı (°C)	>35.6	651(88.2)	63(50.0)	<0.001
	35-35.6 °C	65(8.8)	30(23.8)	
	<35 °C	22(3.0)	33(26.2)	
PO <sub>2</sub> /FiO <sub>2</sub> oranı	>2.49	120(16.3)	10(7.9)	<0.001
	1.0-2.49	574(77.8)	64(50.8)	
	0.3-0.99	42(5.7)	37(29.4)	
	<0.3	2(0.3)	15(11.9)	
En Düşük Serum pH	>=7.20	634(85.9)	53(42.1)	<0.001
	7.10-7.19	76(10.3)	28(22.2)	
	<7.10	28(3.8)	45(35.7)	
Çoklu Nöbet	Var	11(1.5)	13(10.3)	<0.001
	Yok	727(98.5)	113(89.7)	
Diürez Miktarı (mL/kg/s)	>0.9	678(91.9)	76(60.3)	<0.001
	0.1-0.9	52(7.0)	30(23.8)	
	<0.1	8(1.1)	20(15.9)	
SGA	Var	14(1.9)	16(12.7)	<0.001
	Yok	724(98.1)	110(87.3)	
Doğum Ağırlığı	>=1000	733(99.3)	76(60.3)	<0.001
	750-999	5(0.7)	20(15.9)	
	<750	0(0.0)	30(23.8)	
Apgar (5. dk)	>=7	653(88.5)	50(39.7)	<0.001
	<7	85(11.5)	76(60.3)	
Anne Yaşı†		29.83±6.360	28.79±6.34	0.613
Doğum Haftası†		35.60±3.32	30.96±5.47	0.502

† Değişkenler için ortalama ± standart sapma değerleri verilmiştir.

**Tablo 2.** SNAP-II risk tahmin modelini oluřturan bađımsız deđiřkenlerin katsayıları ve risk skorları

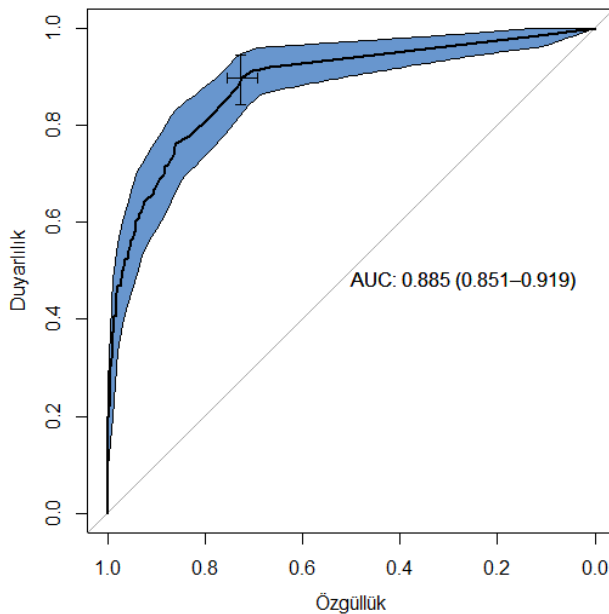
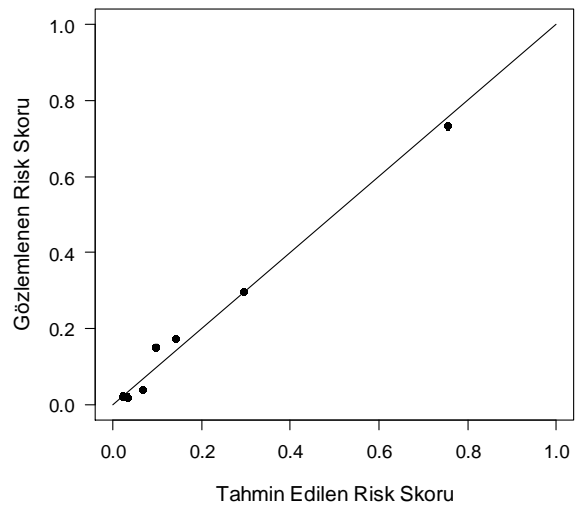
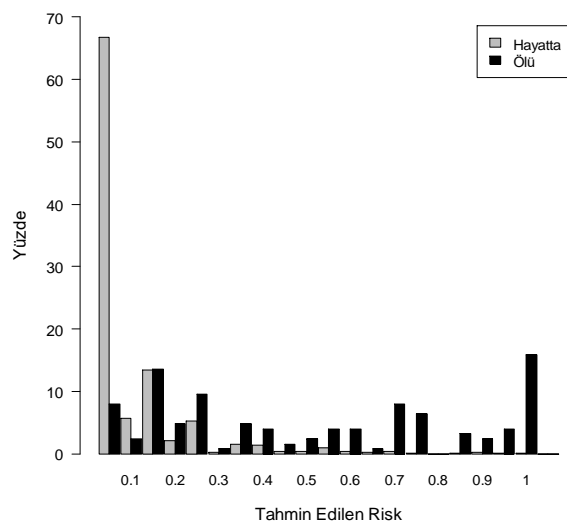
Deđiřkenler		$\beta$	Odds Oranları	%95 GA	SNAP-II Skor Deđerleri
Ortalama Kan Basıncı	20-29 mm Hg	1.241	3.460	1.198-9.993	12
	<20 mm Hg	2.026	7.587	1.682-34.216	20
En Düşük Vücut Isısı	35-35.6 °C	1.301	3.675	1.952-6.919	13
	<35 °C	2.109	8.242	3.749-18.119	21
PO <sub>2</sub> /FiO <sub>2</sub> oranı	1.0-2.49	0.394	1.482	0.649-3.383	4
	0.3-0.99	1.891	6.625	2.539-17.288	19
	<0.3	3.183	24.113	3.323-174.995	32
En Düşük Serum pH	7.10-7.19	1.270	3.561	1.907-6.650	13
	<7.10	2.202	9.040	4.525-18.058	22
Çoklu Nöbet	Var	1.932	6.904	2.273-20.965	19
Diürez Miktarı	0.1-0.9 mL/kg/s	0.762	2.142	1.118-4.105	8
	<0.1 mL/kg/s	1.736	5.674	1.610-19.992	17
Sabit		-3.801			

SNAP-II risk tahmin modelinin genel performansını deđerlendirmek için kullanılan yöntemlerden Brier Skoru 0.076 ve Nagelkerke R<sup>2</sup> 0.479 olarak bulunmuřtur.

Ayırım yeteneđini deđerlendirmek için kullanılan ROC-AUC deđeri 0.885 (GA: 0.851-0.919)'dir (p<0.001) (řekil 1). Kesim noktası 14.50 olarak belirlenmiř ve bu noktadaki seđicilik 0.725, duyarlılık 0.897 ve dođruluk 0.750 olarak elde edilmiřtir.

Kalibrasyonu deđerlendirmek için kullanılan Hosmer-Lemeshow testinin p deđeri 0.619 ( $\chi^2=6.249$ ) olarak bulunmuřtur. Yani gözlem deđerleri ile tahmin deđerleri arasındaki fark istatistiksel olarak anlamlı deđildir. Elde edilen kalibrasyon eđrisi řekil 2'de verildiđi gibidir.

Tahmin edilen risk deđerlerinin hayatta kalan ve kalamayan bebeklerdeki risk skorlarının dađılım yüzdelerini gösteren risk dađılım grafiđi řekil 3'te verildiđi gibidir.

**řekil 1.** SNAP-II risk tahmin modelinin ROC**řekil 2.** SNAP-II risk tahmin modelinin kalibrasyon eđrisi**řekil 3.** SNAP-II için risk deđerlerinin yatıř sonlanma durumuna göre dađılım grafiđi

**SNAPPE-II Risk Tahmin Modelinin Performansının Değerlendirilmesi:** SNAP-II risk tahmin modeline eklenen 3 adet perinatal risk faktörleri ile SNAPPE-II risk tahmin modeli elde edilir. SNAP-II risk skorları ve bu 3 adet perinatal risk faktörünü içeren lojistik regresyonla tahmin değerleri elde edilmektedir. Buradan hareketle, doğrusal regresyona dayanarak elde edilen 3 değişkene ait katsayı tahminleri ve risk skorları Tablo 3'te verildiği gibidir.

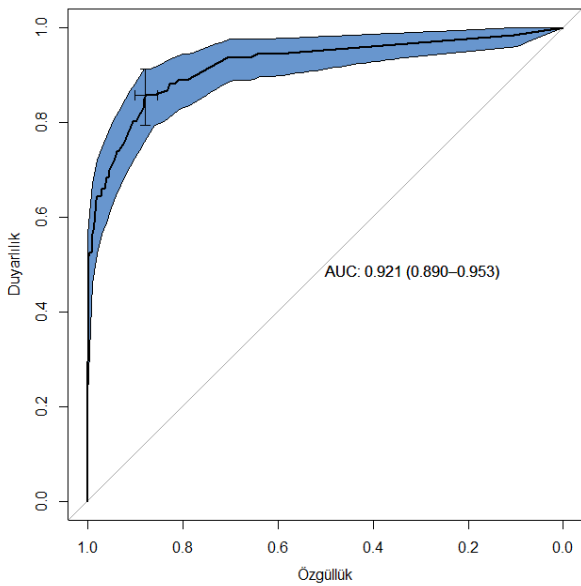
SNAPPE-II risk tahmin modeline ait Brier skoru 0.056 iken Nagelkerke  $R^2$  değeri 0.619 olarak bulunmuştur. Ayırım yeteneğini değerlendirmek için kullanılan ROC-AUC değeri 0.921 (GA: 0.890-0.953)'dir ( $p < 0.001$ ) (Şekil 4). Kesim noktası 26.50 olarak belirlenmiş ve bu noktadaki seçicilik 0.877, duyarlılık 0.857 ve doğruluk 0.874 olarak elde edilmiştir.

Hosmer-Lemeshow testinin p değeri 0.968 ( $\chi^2=2.338$ ) olarak bulunmuştur. Yani gözlem değerleri ile tahmin değerleri arasındaki fark istatistiksel olarak anlamlı değildir. Elde edilen kalibrasyon eğrisi Şekil 5'te verildiği gibidir.

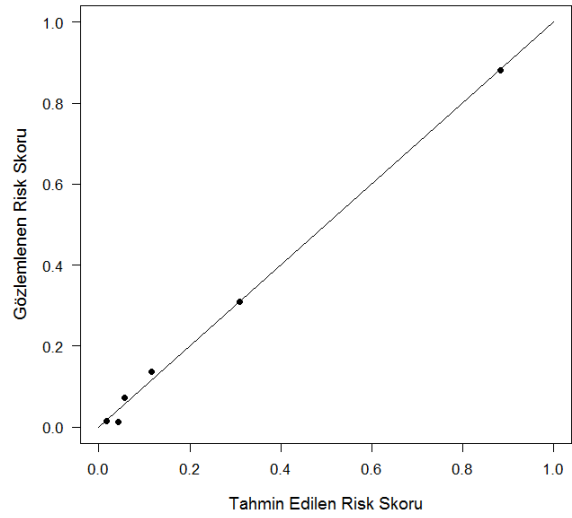
Tahmin edilen risk değerlerinin hayatta kalan ve kalamayan bebeklerdeki risk skorlarının dağılım yüzdelere gösteren risk dağılım grafiği Şekil 6'da verildiği gibidir.

**Tablo 3.** SNAP-II risk tahmin modeline eklenen 3 adet risk faktörünün katsayıları ve risk skorları

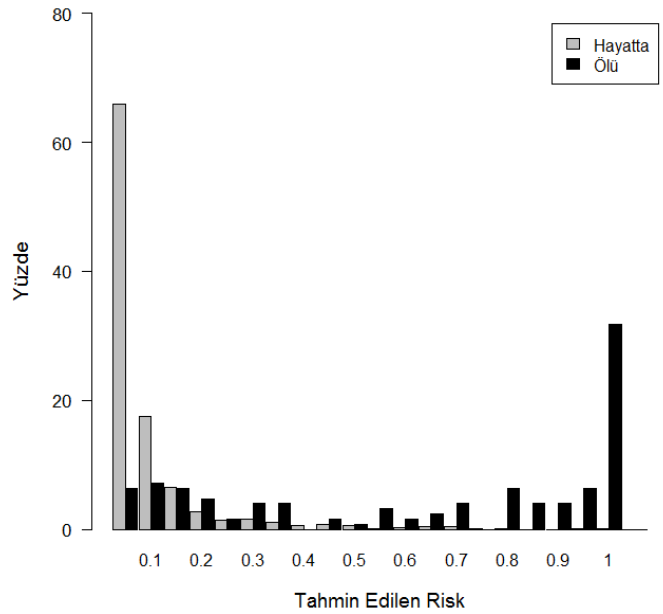
Değişkenler	Tahmin Değeri	P değeri	Risk Skoru	
Doğum Ağırlığı	750-999	3.102	<0.001	31
	<750	3.248	<0.001	32
SGA	Var	1.638	0.004	16
5.dakikadaki Apgar Skoru	<7	1.265	<0.001	13



**Şekil 4.** SNAPPE-II risk skoruna bağlı ROC eğrisi



**Şekil 5.** SNAPPE-II risk skoruna bağlı kalibrasyon eğrisi



**Şekil 6.** SNAPPE-II için risk değerlerinin yatış sonlanma durumuna göre dağılım grafiği

### Tartışma

Risk tahmin modelleri, klinik olarak karar verme sürecinde sıklıkla kullanılan bir istatistiksel modelleme tekniğidir (14). Tahmin modelleri, çeşitli belirteçlerden gelen bilgileri birleştiren istatistiksel bir süreçtir. Yaygın model türleri arasında lojistik regresyon modelleri, Cox orantılı tehlike modelleri ve sınıflandırma ağaçları bulunur. Her model türü, her bir birey için modeldeki bilgileri kullanarak tahmini bir risk üretir (15). Karar verme sürecinde kullanılan risk tahmin modellerinin tahminlerde ne kadar iyi olduğunun ölçülmesi ise modelin performans ölçütlerine bağlıdır. Performans

deđerlendirmesinde kullanılan geleneksel ölçüm yöntemleri; genel model performansının deđerlendirilmesinde açıklanan varyasyon ( $R^2$ ) katsayısı ve Brier skoru, ayırım yeteneđinin testi için ROC eğrisi altında kalan alan deđeri ve son olarak da kalibrasyon için uyum iyiliđi istatistikleri kullanılmaktadır (4). Bu performans ölçütleri (Brier skoru hariç) ne kadar yüksek ise modelin performansı o kadar iyidir.

İç içe olmayan risk tahmin modellerinde skor tahminlerinin uygun puanlama kuralları ile elde edilmesi literatürde kullanılan yöntemler arasındadır. Örneđin, Breiman (16) çalışmasında lojistik regresyondan türetilen bir prognostik modeli rastgele ormandan türetilen başka bir modelle karşılaştırmıştır. Başka bir çalışmada Bartfay ve ark. (17) sinir ađı modellerini lojistik regresyon modelleriyle karşılaştırmak için Brier skorunu ve AUC'yi kullanmıştır. Son olarak, Bernardo ve Smith (18) makalelerinde Bayesci çıkarsama yöntemleri ile skor puanlamaları hesaplamıştır. Makine öğrenme algoritmaları, model karşılaştırmadan ziyade model oluşturma sürecidir (19).

Bu çalışmada, risk tahmin modeli olarak SNAP-II ve SNAPPE-II modelleri kullanılmıştır. Bu iki model, klinik karar vermede halihazırda kullanılır ve lojistik regresyon modellemesine dayanır. Bu çalışmada amaç; model oluşturma sürecinden ziyade model performans karşılaştırma yöntemlerini ortaya koymak ve SNAP-II, SNAPPE-II modellerinin birbirleri ile karşılaştırmaktır. SNAP-II ve SNAPPE-II risk tahmin modellerinin kalibrasyonunu deđerlendirmek için kullanılan Hosmer Lemeshow uyum iyiliđi testinin p deđeri literatüre paralel olarak anlamsız bulunmuştur. Yani gözlem deđerleri ile beklenen deđerler arasındaki fark istatistiksel olarak anlamlı deđildir. Bununla beraber bu modellerin genel performansını deđerlendirmek için Nagelkerke  $R^2$  ve Brier skoru kullanılmıştır.

SNAP-II için literatürde bulunan ROC eğrisi altında kalan alan deđeri 0.82 (GA:0.68-0.95) (20) olarak bulunurken çalışmada bu deđer 0.885 (GA: 0.851-0.919) olarak elde edilmiştir. Ayrıca kesim noktası 14.50 olarak belirlenmiş ve bu noktadaki seçicilik 0.725, duyarlılık

0.897 ve doğruluk 0.750 olarak elde edilmiştir. SNAPPE-II risk tahmin modeli için ise 0.84 ile 0.92 arasında deđişen bir ROC eğrisi altında kalan alan deđeri söz konusu iken çalışmada bu deđer 0.921 (GA: 0.890-0.953)'dir (13). Richardson ve ark. (21) çalışmalarında AUC deđerini 0.91, Zupanic ve ark.(22) 0.90, Mia ve ark.(23) ise 0.863 olarak elde etmiştir. Ayrıca SNAPPE-II için kesim noktası 26.50 olarak belirlenmiş ve bu noktadaki seçicilik 0.877, duyarlılık 0.857 ve doğruluk 0.874 olarak elde edilmiştir. SNAPPE-II'nin kesim noktası literatürde 13.5 iken seçicilik 0.63 ve duyarlılık 0.67 şeklindedir (24). SNAPPE-II risk tahmin modelinin tüm performans ölçütlerine göre daha iyi bir performansa sahip olduđu elde edilmiştir.

İleri çalışmalarda, modelleme tekniđi olarak lojistik regresyon modellemesi ile birlikte makine öğrenme algoritmalarının karşılaştırılması önerilebilir. Bu sayede model skorları farklı yöntemlerle elde edilir ve karşılaştırmalar yapılabilir. Ayrıca risk tahmin modellerinin performans deđerlendirmesine ilişkin literatürde sağkalım sürelerini dikkate alacak şekilde geliştirilen Cox orantılı hazard modelleri ve makine öğrenme yöntemleri mevcuttur.

Sonuç olarak lojistik regresyon, risk faktörlerine bađlı tahmin modellerinde en sık karşılaşılan modellemelerden biridir. Bunun nedeni risk tahmin modellerinin genel amacı hastalığı, ölümü vs. gibi durumları etkileyen faktörleri ortaya koymak ve gerekli koruyucu önlemleri almaktır. Bu nedenle en uygun modelin belirlenmesi ve kurulan modelin performansının deđerlendirilmesi gerekmektedir. Modelin iyi bir performansa sahip olabilmesi için açıklanan varyasyonun, ayırım gücünün ve kalibrasyon yeteneđinin yüksek olması ve bu özelliklerin istatistiksel olarak geçerli olması gerekmektedir.

Model performansının deđerlendirilmesi için kullanılan yöntemlere göre elde edilen bulgular ışığında, SNAPPE-II risk tahmin modelinin genel performans, ayırım yeteneđi ve kalibrasyon bakımından SNAP-II'den daha iyi olduđu elde edilmiştir.

## Kaynaklar

- Demler OV. Impact of new variables on discrimination of risk prediction models. PhD. thesis, Boston: Boston University, Boston University Theses & Dissertations, 2012.
- Bedogni G, TSYBAKOV A, Berlin S. Clinical prediction models-a practical approach to development, validation and updating. Development 2009; 18: 53.
- Gu W. Evaluating risk prediction markers and models: University of Washington; 2009.
- Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: A framework for some traditional and novel measures. Epidemiology (Cambridge, Mass) 2010; 21: 128.
- Nagelkerke NJ. A note on a general definition of the coefficient of determination. Biometrika 1991; 78: 691-692.
- Bewick V, Cheek L, Ball J. Statistics review 14: Logistic regression. Critical Care 2005; 9:1-7.
- Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. Circulation 2007; 115: 654-657.
- Pepe, M. S., & Janes, H. (2013). Methods for Evaluating Prediction Performance of Biomarkers and Tests. Springer New York eBooks.
- Pepe MS. An interpretation for the ROC curve and inference using GLM procedures. Biometrics 2000; 56: 352-359.
- DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. Biometrics 1988: 837-845.
- Alonzo TA. Clinical prediction models: A practical approach to development, validation, and updating: By Ewout W. Steyerberg. Oxford University Press 2009.

12. Arslan AK, Yaşar Ş, Çolak C, Yoloğlu, S. WSSPAS: An interactive web application for sample size and power analysis with R using shiny. *Türkiye Klinikleri Journal of Biostatistics* 2008; 10: 224-246.
13. Richardson DK, Corcoran JD, Escobar GJ, Lee SK. SNAP-II and SNAPPE-II: Simplified newborn illness severity and mortality risk scores. *The Journal of Pediatrics* 2001;138: 92-100.
14. Mühlenbruch K, Heraclides A, Steyerberg EW, et al. Assessing improvement in disease prediction using net reclassification improvement: Impact of risk cut-offs and number of risk categories. *European Journal of Epidemiology* 2013; 28: 25-33.
15. Janes H, Pepe MS, Gu W. Assessing the value of risk predictions by using risk stratification tables. *Annals of Internal Medicine* 2008; 149: 751-760.
16. Breiman L. Statistical modeling: The two cultures (with comments and a rejoinder by the author). *Statistical Science* 2001;16:199-231.
17. Bartfay E, Mackillop W, Pater J. Comparing the predictive value of neural network models to logistic regression models on the risk of death for small-cell lung cancer patients. *European Journal of Cancer Care* 2006;15:115-124.
18. Bernardo JM, Smith AF. *Bayesian theory*: John Wiley & Sons; 2000.
19. Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 2008; 50: 457-479.
20. Sundaram V, Dutta S, Ahluwalia J, Narang A. Score for neonatal acute physiology II predicts mortality and persistent organ dysfunction in neonates with severe septicemia. *Indian Pediatrics* 2009; 46: 775-780.
21. Richardson DK, Phibbs CS, Gray JE, et al. Birth weight and illness severity: independent predictors of neonatal mortality. *Pediatrics* 1993; 91: 969-975.
22. Zupancic JA, Richardson DK, Horbar JD, et al. Revalidation of the score for neonatal acute physiology in the Vermont Oxford Network. *Pediatrics* 2007; 119: e156-e163.
23. Mia R, Etika R, Harianto A, et al. The use of score for neonatal acute physiology perinatal extension II (SNAPPE II) in predicting neonatal outcome in neonatal intensive care unit. *Paediatrica Indonesiana* 2005; 45: 241-245.
24. Asker H, Satar M, Yapicioglu Yildizdas H, et al. Evaluation of score for neonatal acute physiology and perinatal extension II and clinical risk index for babies with additional parameters. *Pediatrics International* 2016; 58: 984-987.