ORİJİNAL ARAŞTIRMA ORIGINAL RESEARCH

Bilgi İçeren Durdurma Varlığında Yinelemeli Olay Verisi İçin Yeni Bir Modelleme

A New Modelling for Recurrent Event Data Under Informative Censoring

Hande ÜNLÜ,[a]
Ayten YİĞİTER[b]

[a]Hacettepe Üniversitesi
Halk Sağlığı Enstitüsü,
[b]İstatistik Bölümü,
Hacettepe Üniversitesi Fen Fakültesi,
Ankara

Geliş Tarihi/*Received*: 08.02.2016
Kabul Tarihi/*Accepted*: 18.04.2016

Yazışma Adresi/*Correspondence:*
Hande ÜNLÜ
Hacettepe Üniversitesi
Halk Sağlığı Enstitüsü, Ankara,
TÜRKİYE/TURKEY
hkonsuk@hacettepe.edu.tr

ÖZET Amaç: Bilgi içeren durdurma, yinelemeli olay verisinin modellemesinde dikkat edilmesi gereken önemli bir konudur. Bilgi içeren durdurma; ölüm gibi bir başarısızlık olayının gerçekleşmesi ya da bireylerin çalışmadan çekilmesi sonucu meydana gelir. Bu nedenle yinelemeli olay zamanları ile bilgi içeren durdurma zamanı ilişkilidir. Yinelemeli olay verisinde bu ilişkinin göz ardı edilmesi model parametrelerinin yanlı tahmin edilmesine yol açmaktadır. Bu çalışmanın amacı, bilgi içeren durdurma varlığında yinelemeli olay verisini modellemek ve parametre tahminlerini elde etmektir. **Gereç ve Yöntemler:** Bu çalışmada, bilgi içeren durdurmanın ölüm olayından kaynaklandığı durum incelenmiş ve ölüm olayının varlığında yinelemeli olay süreci homojen Poisson süreci ile modellenmiştir. Yinelemeli olay sürecine ilişkin yoğunluk fonksiyonu için yeni bir model önerilmiş ve yineleme zamanları ile ölüm zamanı arasındaki ilişki yapısı, paylaşılmış zayıflık modeli kullanılarak oluşturulmuştur. Önerilen model, farklı durdurma oranları ve örneklem büyüklüklerini içeren senaryolar altında bir benzetim çalışması ile incelenmiştir. Bilgi içeren durdurmanın dikkate alınmadığı indirgenmiş modelle karşılaştırmalar yapılmıştır. **Bulgular:** Benzetim çalışması sonucunda önerilen modelin Akaike Bilgi Kriteri (AIC) değerleri indirgenmiş modelden daha düşük bulunmuştur. **Sonuç:** Yapılan çalışmada ölüm gibi bir bilgi içeren durdurmanın söz konusu olduğu yinelemeli olay verisinin, ölüm olayı göz ardı edilerek modellenmesi parametre tahminlerini olumsuz etkilediği görülmüştür.

Anahtar Kelimeler: Yinelemeli olay verisi; bilgi içeren durdurma; Poisson süreci; zayıflık; sağkalım analizi; mesane kanseri

ABSTRACT Objective: Informative censoring is an important issue taken into account while modeling recurrent event data. Informative censoring might happen as a result of terminal event such as death or drop out of individuals in the study. Therefore, there is correlation between recurrent event times and terminal event times. Ignoring this correlation in the recurrent event data, biased estimates of model parameters could be obtained. The aim of this study is to model recurrent event data in the presence of informative censoring and obtain the estimate of parameters. **Material and Methods:** In this study, we focused on the situation in which informative censoring happens as a result of death and modeled recurrent event data using homogeneous Poisson Process. We proposed the new model based on intensity function of recurrent event process and constructed the structure of correlations between recurrent event time and death time via shared frailty. Simulation study was conducted for different scenarios including different censoring rates and sample sizes, and the proposed model was investigated via a simulation study. The proposed model was compared with the reduced model in which informative censoring wasn't taken into account. **Results:** As a result of the simulation study, Akaike Information Criteria (AIC) values of the proposed model were found to be lower than those of the reduced model. **Conclusion:** In this study, it was found that when recurrent event data has informative censoring such as death, parameter estimates was adversely affected by ignoring the informative censoring.

Key Words: Recurrent event data; informative censoring; Poisson process; frailty; survival analysis; bladder cancer

doi: 10.5336/biostatic.2016-50770

Turkiye Klinikleri J Biostat 2016;8(2):125-32

Yinelemeli olay bireylerin aynı olayı birden fazla kez deneyimlemesi olarak tanımlanır. Yinelemeli olaylara örnek olarak epilepsi hastalarındaki nöbetler, kanser vakalarında tekrarlayan tümörler ya da üretim donanımlarında tekrarlayan arızalar verilebilir. Yinelemeli olay analizinde amaç, açıklayıcı değişkenler ile ilgilenilen olayın tekrarlama hızı arasındaki ilişkiyi değerlendirmektir.[1]

Yinelemeli olaylara ilişkin verilerin istatistiksel modellemesini anlamak için veri yapısının karakteristiklerini bilmek önemlidir. Yinelemeli olay verisinin üç önemli özelliği bulunmaktadır. Birincisi, yinelemeli olay verilerinde bireyler bağımsız olarak örneklenmiştir, ancak aynı bireylerden elde edilen yineleme zamanları ilişkilidir. Bu ilişki, parametre tahminlerine ilişkin varyansın olduğundan daha düşük tahmin edilmesine neden olur. İkincisi, yinelemeli olayların sayısı, yinelemeli olay süreci hakkında bilgi vericidir. Örneğin, bir birey için belli bir zaman periyodu içinde az sayıda yinelemeli olayın gözlenmesi, olayın oluş hızının yavaş olduğunu, olaylar arasında geçen zamanın uzun olduğunu belirtir. Üçüncüsü, yinelemeli olaylarda, birey ya da birim aynı olayı değişik zamanlarda tekrar yaşayabilir. Bu nedenle, yinelemeli olay analizinde ilgilenilen olay hiçbir zaman ölüm olamaz.

Yinelemeli olay verileri, Cox[2-3] tipi regresyon modellerinin geliştirilmesine dayanmakla birlikte; bu verilerin karakteristiklerini yansıtacak derecede tam bir modelleme yapılması olanaksızdır. Aynı zamanda bu verilerin modellenmesi ve analizi istatistik biliminin gündeminde olan konularından biridir. Literatürde, yinelemeli olayların modellenmesi ve bu modellerin çözümüne ilişkin çok sayıda çalışma vardır: Örneğin, Andersen ve Gill[4], yinelemeli olay sürecini bağımsız artmalı sayma süreci kabul ederek, açıklayıcı değişkenler yardımıyla Cox-tipi modeli kullanmışlardır. Lin ve Wei, her yineleme zamanını, hızlandırılmış başarısızlık zamanı modelleri kullanarak modellemiştir.[5] Pepe ve Cai, Poisson tipi süreçlerde bağımsız artma varsayımını gevşetip koşullu oran fonksiyonunu modelleyerek yeni bir yaklaşım önermişlerdir.[6] Balshaw, çalışmasında yinelemeli olay sürecini homojen olmayan Poisson süreci olarak ele almış ve yarı olabilirlik tahmin yöntemini kullanarak parametre tahminlerini elde etmiştir.[7] Lin, Wei ve Ying, sayma sürecini kullanarak ortalama fonksiyonlarını hızlandırılmış başarısızlık zamanı modelleri ile modellemişlerdir.[8] Lin, Wei, Yang ve Ying, yinelemeli olaylarda ortalama ve oran fonksiyonlarını Poisson varsayımı olmaksızın yarı parametrik olarak modellemişler ve modelin yeterliliğini test etmek için sayısal yöntemler önermişlerdir.[9] Pan, hızlandırılmış başarısızlık zamanı modellerine zayıflık terimini ekleyerek modellemiştir.[10] Nielsen, çalışmasında, yinelemeli olay sürecini karışık etkili homojen olmayan Poisson süreci ile modellemiştir. Bireye özgü heterojenliği ise zayıflık terimini modele dahil ederek açıklamıştır. Sayma sürecine ilişkin yoğunluk fonksiyonunun, eğrisel çizgili düzleştirme fonksiyonu olduğunu varsaymıştır. Açıklayıcı değişken etkilerini de eğrisel çizgileri kullanarak modellemiştir. Bu sayede model, zamana bağlı açıklayıcı değişken kullanımına olanak sağlamaktadır.[11]

Yinelemeli olayların stokastik sıralı yapısı ile birlikte durdurma (censored) olaylarının varlığı istatistiksel modellerde çeşitli güçlüklere neden olur. Yinelemeli olay verisinin modellenmesinde durdurma olayının yapısı çok iyi irdelenmelidir. Çoğu araştırmada, başarısızlık zamanı; ölüm gibi bir başarısızlık olayı, çalışmanın sona ermesi ya da izlenme döneminde bireyin çalışmadan çekilmesi sebepleriyle durdurmaya maruz kalabilir. Eğer durdurma, çalışmanın sona ermesinden kaynaklanıyorsa, durdurma zamanı bağımsız ya da bilgi içermeyen şeklinde kabul edilir. Birçok uygulamada yinelemeli olaylar, bireyin çalışmadan çekilmesi ya da bireyin ölümü ile sonlandırılır. Bu şekilde meydana gelen durdurma zamanı bilgi içerir ve yinelemeli olay zamanları ile ilişkilidir. Söz konusu ilişkiyi göz ardı ederek modelleme yapmak yanlı sonuçlar elde edilmesine neden olmaktadır.

Literatürde bilgi içeren durdurma varlığı altında yinelemeli olayları modelleyen çok sayıda çalışma bulunmaktadır. Cook ve Lawless, yinelemeli olayları, bilgi içeren durdurma varlığı altında ele almışlar; yinelemeli olaylar ile bilgi içeren durdurma olayı arasındaki bağımlılık yapısını bileşik modeller kullanarak modellemişlerdir. Parametre tahminleri için yarı parametrik yöntemleri kullanmışlardır.[12] Ghosh ve Lin, durdurma olayının varlığı altında birikimli yinelemeli olay sayısının marjinal ortalamalarına dayanan tek ve iki örnekleme ilişkin parametrik olmayan bir yöntem geliştirmişlerdir.[13] Ghosh ve Lin, yaşam verisinde ölüm olayı gibi bilgi içeren durdurma olduğu zaman yinelemeli olayların ortaya çıkış zamanlarını modellemede marjinal ortalama fonksiyonunu ele almışlar; olayların stokastik yapısını dikkate almadan açıklayıcı değişkenlere ilişkin parametrelerin tahminlerini parametrik olmayan yöntemler kullanarak elde etmişlerdir.[14] Huang ve Wolfe, yinelemeli olay verisinde, başarısızlık zamanı ile bilgi içeren durdurma olayı arasındaki ilişkiyi zayıflık terimi ile modellemişlerdir. Zayıflık teriminin dağılımının normal dağılıma sahip olduğu varsayılmıştır. Parametre tahminlerini EM ve Metropolis-Hastings algoritmaları kullanarak elde etmişlerdir.[15] Ghosh ve Lin, daha sonra yinelemeli olay ve bağımlı durdurma süreçlerinin marjinal modellerini hızlandırılmış başarısızlık zamanı modelleriyle formüle eden yarı parametrik bileşik bir model önermişlerdir.[16] Liu ve arkadaşları, yinelemeli olay verisini bilgi içeren durdurma varsayımı altında Poisson süreçlerini kullanarak modellemişlerdir. Çalışmalarında ölüm olayını, bilgi içeren durdurma olarak ele almışlar; durdurma olayı ile yinelemeli olaylara ilişkin bileşik olabilirlik fonksiyonunu paylaşılmış zayıflık modellerinden yararlanarak elde etmişlerdir.[17] Huang ve Wang çalışmalarında yinelemeli olay verisini homojen olmayan Poisson süreciyle modellemiş, yineleme süreci ile durdurma süreci arasındaki bağımlılık yapısını paylaşılmış zayıflık terimi kullanarak bağlamışlardır.[18] Huang, Qin ve Wang, yinelemeli olay verisinde, başarısızlık zamanı ile durdurma olayı arasındaki ilişkiyi zayıflık terimi ile modellemişlerdir. Modellerinde zayıflık teriminin dağılımına ilişkin bir varsayımda bulunmamışlardır. Huang ve ark., model parametrelerini, yineleme zamanları ve zayıflık terimi için herhangi bir dağılım varsayımında bulunmadan yarı parametrik yöntemler kullanarak tahmin etmişlerdir.[19]

Bu çalışmada, yineleme zamanlarının sağdan durdurmaya maruz kaldığı durum düşünülmüş, bilgi içeren durdurmanın ölüm olayından kaynaklandığı durum incelenmiş ve ölüm olayı varlığında yinelemeli olaylar homojen Poisson süreci ile modellenmiştir. Yinelemeli olay sürecine ilişkin yoğunluk fonksiyonu için bir model önerilmiş ve yineleme zamanları ile ölüm zamanı arasındaki ilişki yapısı, paylaşılmış zayıflık modeli kullanılarak oluşturulmuştur. Paylaşılmış zayıflık parametresinin Gamma dağılımından geldiği varsayılmıştır. Parametrelerin en çok olabilirlik tahminleri Newton-Raphson algoritması kullanılarak elde edilmiştir. Önerilen model, Byar'ın[20] yapmış olduğu çalışmada yer alan mesane kanseri verilerine uygulanmıştır. Önerilen modelin geçerliliğinin gösterilmesi için benzetim çalışması yapılmıştır.

GEREÇ VE YÖNTEMLER

C_i, *i.* bireye ait bilgi içermeyen durdurma zamanını, D_i, *i.* bireye ait bilgi içeren durdurma zamanını göstersin. $\tau_i = \min(D_i, C_i)$, *i.* bireye ilişkin takip süresi ve $\Delta_i = I(D_i \le C_i)$, *i.* bireyin ölüm olayını deneyimleme durumunu gösteren gösterge değişkenidir. $Y_i(t) = I(\tau_i \ge t)$, *i.* bireyin *t* zamanında hayatta olup olmadığını gösteren risk sürecidir. $N_i^D(t) = I(\tau_i \le t, \Delta_i = 1)$, *i.* bireyin *t* zamanına kadar ölüm olayını yaşayıp yaşamadığını gösteren sayma sürecidir. $dN_i^R(t) = \lim_{\Delta t \to 0} N^R(t+\Delta t^-) - N^R(t^-)$, *i.* bireyin $(0, \Delta t]$ zaman aralığında deneyimlediği yineleme sayısı olmak üzere, *i.* bireyin *t* zamanına kadar yaşadığı yinelemeleri sayan sayma süreci $N_i^R(t) = \int_0^t Y_i(t) dN_i^R(t)$ olarak ifade edilir. *i.* bireye ait geçmiş bilgisi $\mathcal{H}_i(t^-) = \{N_i^R(u), \mathbf{Z}_i(u), Y_i(u), 0 \le u \le t^-\}$ ve $\mathbf{Z} = (Z_1, Z_2, ..., Z_p)$ zamandan bağımsız açıklayıcı değişken vektörüdür. O_i, *i.* bireye ait gözlenen veriyi; ν_i, ise *i.* bireye ilişkin zayıflık terimini göstermektedir.

Önerilen model için yapılan varsayımlar şunlardır.

1. Yinelemeli olay, ölüm ve durdurma süreçleri sürekli dağılıma sahiptir; ölüm, durdurma ve yinelemeli olaylar aynı anda meydana gelemez.

2. Yinelemeli olay sürecine ilişkin yoğunluk fonksiyonu,

$$P\left(dN^R(t)=1 \middle| \mathcal{H}_i(t^-), D\geq t\right)=Y_i(t)\,dR_i(t)\equiv Y_i(t)\,r_i(t)\,dt$$

$$dR_i(t)=P\left(dN_i^{R_*}(t)=1 \middle| Z_i, v_i, D_i\geq t\right)$$

biçimindedir.

3. i. bireye ait sayma süreci $N_i(t)$, $\lambda_i(.)$ yoğunluğuyla homojen Poisson sürecine sahiptir.

4. Durdurma, zayıflık terimi v_i ' ye ilişkin bilgi içermemektedir.

5. Zayıflık teriminin dağılımı, $v_i \sim Gamma(\frac{1}{\theta}, \theta)$ olarak alınmıştır.

Yukarıda verilen varsayımlar altında, ölüm olayının varlığında yinelemeli olay sürecine ilişkin önerilen modelin yoğunluk fonksiyonu,

$$\lambda_i\left(t \mid \mathcal{H}_i(t^-)\right)=r_0(t)\,v_i\exp(\boldsymbol{\beta}_1^T\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i}) \quad (1)$$

biçiminde tanımlanmıştır.[‡] Eşitlik (1)'de $r_0(t)$, ölüm olayı varlığı altında yinelemeli olay sürecine ilişkin temel yoğunluk fonksiyonunu, $\mathbf{Z}_{1i}$, zamandan bağımsız açıklayıcı değişken vektörünü, $\mathbf{Z}_{2i}$, i. bireyin ölüm olayını deneyimleyip deneyimlemediği bilgisini gösteren açıklayıcı değişken değerini, $\boldsymbol{\beta}_1$, $\mathbf{Z}_{1i}$ açıklayıcı değişkenine karşılık gelen parametre vektörünü, ve β_2 , $\mathbf{Z}_{2i}$ açıklayıcı değişkenine karşılık gelen parametreyi göstermektedir.

Eşitlik (1)'de verilen önerilen modele ait yoğunluk fonksiyonu kullanılarak, i. bireye ait olabilirlik fonksiyonu,

$$\begin{aligned} L_i(O_i, v_i \mid z_i) &= L_i(O_i \mid v_i, Z_i)\, f_\theta(v_i) \\ &= \exp\left\{-\int_0^\infty Y_i(t)\,v_i\exp\left(\boldsymbol{\beta}_1^T\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i}\right)dR_0(t)\right\} \quad (2) \\ &\times\prod_k\left\{v_i\exp\left(\boldsymbol{\beta}_1^T\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i}\right)dR_0(t)\right\}^{\delta_{ik}} \\ &\times\frac{1}{\Gamma(1/\theta)\,\theta^{1/\theta}}\,v_i^{\frac{1}{\theta}-1}\,e^{-v_i/\theta} \end{aligned}$$

olarak yazılabilir.[21] Eşitlik (2)'de verilen olabilirlik fonksiyonunun çıkarsamasında Kalbfleisch ve Prentice çalışmasından yararlanılmıştır.[21] Burada δ_{ik} , i. bireyin $T_{i,k}$ anında yineleme yaşamasını ifade eden gösterge değişkenidir. Eşitlik (1)'de verilen yoğunluk fonksiyonunda, temel yoğunluk fonksiyonu $r_0(t)=\lambda$ ve β_1 zamandan bağımsız açıklayıcı değişkenine karşılık gelen parametre olarak alınırsa, Eşitlik (2)'de verilen olabilirlik fonksiyonu aşağıdaki biçimde yazılabilir:

$$\begin{aligned} L_i(O_i, v_i \mid \mathbf{Z}_i) &= \left(\prod_{k=1}^{N_i(\tau_i)}\lambda v_i\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})\right)\exp \\ &\left(-\int_0^{\tau_i}\lambda v_i\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})du\right) \quad (3) \\ &\times\frac{1}{\Gamma(1/\theta)\,\theta^{1/\theta}}\,v_i^{\frac{1}{\theta}-1}\,e^{-v_i/\theta}. \end{aligned}$$

Eşitlik (3)'te $1/\theta=\theta^*$ olarak alınırsa,

$$\begin{aligned} L_i(O_i, v_i \mid \mathbf{Z}_i) &= [\lambda\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})]^{N_i(\tau_i)} \\ &\exp[-\lambda v_i\tau_i\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})] \quad (4) \\ &\times\frac{(\theta^*)^{\theta^*}}{\Gamma(\theta^*)}\,v_i^{\theta^*-1}e^{-\theta^* v_i} \end{aligned}$$

biçiminde yeniden ifade edilebilir. Eşitlik (4)' te verilen olabilirlik fonksiyonu, yinelemeli olay sürecinin ve gözlenemeyen zayıflık parametresinin bileşik olabilirlik fonksiyonudur. Bu yüzden zayıflık parametresine göre integrali alınarak n birey için marjinal olabilirlik fonksiyonu;

$$\begin{aligned} L_{\mathrm{marg}}(O \mid v, \mathbf{Z}) &= \prod_{i=1}^n L_{i,\mathrm{marg}} = \prod_{i=1}^n\int_0^\infty L_i(O_i \mid v_i, \mathbf{Z}_i)\, f_\theta(v_i)\,dv_i \quad (5) \\ &= \prod_{i=1}^n\frac{(\theta^*)^{\theta^*}\times[\lambda\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})]^{N_i(\tau_i)}}{\Gamma(\theta^*)} \\ &\times\frac{\Gamma(N_i(\tau_i)+\theta^*)}{[\lambda\tau_i\,\exp(\beta_1\,\mathbf{Z}_{1i}+\beta_2\,\mathbf{Z}_{2i})+\theta^*]^{N_i(\tau_i)+\theta^*}} \end{aligned}$$

[‡] Eşitlik (1)'de verilen model ve çıkarsamalar, *"Ünlü H., Bilgi içeren durdurma varlığında yinelemeli olay süreci, Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Doktora Tezi, 2013, p. 1-118."* tez çalışmasının bir parçasını teşkil etmektedir.

biçiminde elde edilir. Eşitlik (5)' deki $\Gamma(.)$ fonksiyonu için,

$$\Gamma(N_i(\tau_i)+\theta^*) = \prod_{j=1}^{N_i(\tau_i)} \left\{ N_i(\tau_i) - j + \theta^* \right\} \Gamma(\theta^*)$$

eşitliğinden yararlanılarak marjinal olabilirlik fonksiyonu,

$$L_{m\arg}(O\,|\,\nu,\mathbf{Z}) = \prod_{i=1}^{n} \left\{ \prod_{j=1}^{N_i(\tau_i)} \left\{ N_i(\tau_i) - j + \theta^* \right\} \right\} \frac{\left(\theta^*\right)^{\theta^*} [\lambda \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})]^{N_i(\tau_i)}}{[\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i}) + \theta^*]^{N_i(\tau_i)+\theta^*}} \tag{6}$$

ve log olabilirlik fonksiyonu,

$$\begin{aligned} \log L_{m\arg}(O\,|\,\nu,Z) = {} & \sum_{i=1}^{n} \left\{ \sum_{j=1}^{N_i(\tau_i)} \log(N_i(\tau_i) - j + \theta^*) \right\} + N_i(\tau_i)\{\log(\lambda) + \beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i}\} \\ & + \theta^* \log(\theta^*) - \left\{ N_i(\tau_i) + \theta^* \right\} \log\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i}) + \theta^* \right) \end{aligned} \tag{7}$$

olarak elde edilir. Eşitlik (7)' de verilen log-olabilirlik fonksiyonunun bilinmeyen λ, β_1, β_2 ve θ^* parametrelerine göre türevleri aşağıda verilmiştir:

$$\begin{aligned} \frac{\partial L_{m\arg}(O\,|\,\nu,\mathbf{Z})}{\partial \lambda} &= \sum_{i=1}^{n} \frac{N_i(\tau_i)}{\lambda} - \frac{\left(N_i(\tau_i)+\theta^*\right)\left(\tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right)}{\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right) + \theta^*}, \\ \frac{\partial L_{m\arg}(O\,|\,\nu,\mathbf{Z})}{\partial \beta_1} &= \sum_{i=1}^{n} N_i(\tau_i)\,\mathbf{Z}_{1i} - \frac{\left(N_i(\tau_i)+\theta^*\right)\left(\lambda \tau_i \mathbf{Z}_{1i} \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right)}{\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right) + \theta^*}, \\ \frac{\partial L_{m\arg}(O\,|\,\nu,Z)}{\partial \beta_2} &= \sum_{i=1}^{n} N_i(\tau_i)\,\mathbf{Z}_{2i} - \frac{\left(N_i(\tau_i)+\theta^*\right)\left(\lambda \tau_i \mathbf{Z}_{2i} \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right)}{\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right) + \theta^*} \end{aligned} \tag{8}$$

$$\begin{aligned} \frac{\partial L_{m\arg}(O\,|\,\nu,\mathbf{Z})}{\partial \theta^*} = {} & \sum_{i=1}^{n} \left\{ \sum_{k=1}^{N_i(\tau_i)} \frac{1}{N_i(\tau_i) - j + \theta^*} \right\} + \log\left(\theta^*\right) + 1 - \\ & \left\{ \log\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i}) + \theta^* \right) + \frac{\left(N_i(\tau_i)+\theta^*\right)}{\left(\lambda \tau_i \exp(\beta_1 \mathbf{Z}_{1i} + \beta_2 \mathbf{Z}_{2i})\right) + \theta^*} \right\}. \end{aligned}$$

Eşitlik (8)'de verilen skor fonksiyonlarını sıfır yapan değerler parametre kestirimlerini vermektedir. Bu skor fonksiyonlarının eşanlı çözümü ancak denklemler doğrusal ise olasıdır. Bu nedenle doğrusal olmayan denklem sisteminin çözümü için iteratif yöntemlere başvurulur. Yukarıda yer alan denklem sistemlerinin çözümü için Newton-Raphson algoritması kullanılmıştır.

BENZETİM ÇALIŞMASI

Benzetim çalışmasında, bilgi içeren durdurma varlığında yinelemeli olay verisi homojen Poisson süreci kullanılarak farklı örneklem büyüklüklerine ve durdurma oranlarına göre türetilmiştir. Örneklem büyüklükleri 40, 60 ve 100 olarak alınmıştır. Modeldeki açıklayıcı değişken, 0 ya da 1 değerini alan kategorik bir değişkendir ve açıklayıcı değişken değerleri $p=0.50$ olasılıkla Bernoulli dağılımdan türetilmiştir. Regresyon parametreleri $\beta_1=-1$, $\beta_2=1$ ve modeldeki temel yoğunluk fonksiyonunun değeri $\lambda=2$ olarak alınmıştır.

Zayıflık teriminin dağılımı için $\theta^*=2$ ile $Gamma(\theta^*, \frac{1}{\theta^*})$ dağılımı düşünülmüş ve bireylere özgü zayıflık değerleri bu dağılımdan türetilmiştir. Çalışmadaki bireylerin her biri en az bir kez, en çok dört kez yineleme yaşamıştır. Bireylerin durdurma ya da ölüm zamanları birbirinden farklı olacak şekilde türetilmiştir. Ölüm oranı olarak 0.10, her yinelemeden sonra durdurma oranı ise 0.01, 0.05 ya da 0.10 olarak alınmıştır. Her bir örneklem büyüklüğü ve ölüm oranları için 500 tekrar yapılmıştır. Benzetim çalışması için R programında kod yazılarak her bir senaryoya ilişkin parametre tahminleri, hata kareler ortalamaları (HKO) ve AIC değerleri elde edilmiştir. Parametrelerin en çok olabilirlik tahminleri Newton-Raphson algoritması kullanılarak elde edilmiştir ve ε=0.001 durdurma kuralına göre iterasyonlara son verilmiştir.

Farklı örneklem büyüklükleri ve durdurma oranları için elde edilen parametrelerin tahmin değerleri, HKO ve AIC değerleri verilmiştir (Tablo 1, 2, 3). Ayrıca tablolarda yer alan indirgenmiş model, önerilen modelde $\beta_2=0$ ve $\theta^*=0$ durumuna karşılık gelmektedir. İndirgenmiş model, veride bilgi içeren durdurmanın ve zayıflık teriminin yer almadığı modele karşılık gelmektedir.

Tablo 1, 2 ve 3'ten görüldüğü gibi önerilen modelde λ ve β_1 parametrelerine ilişkin hata kareler ortalamaları, indirgenmiş modele göre daha düşük çıkmıştır. Bu sonuç, veride bilgi içeren durdurma söz konusu olduğunda bu bilginin dikkate alınarak modellenmesi gerektiği bilgisini desteklemektedir. β_2 ve θ^* parametreleri indirgenmiş modelde yer almadığından, bu parametreler bakımından iki model karşılaştırılamaz. Bu yüzden model karşılaştırması

TABLO 1: n=40 ve durdurma oranı=0.01, 0.05 ve 0.10 alındığında önerilen model için benzetim sonuçları.

			Önerilen Model			İndirgenmiş Model		
n	**Durdurma Oranı**	**Parametre**	**TAHMİN**	**HKO**	**AIC**	**TAHMİN**	**HKO**	**AIC**
		$\lambda = 2$	1.5665	0.5159	827.4091	0.3134	4.4458	3951.1422
40	0.01	$\beta_1 = -1$	-1.1831	0.0968		0.153	1.3396	
		$\beta_2 = 1$	1.0389	0.0648		--	--	
		$\theta^* = 2$	1.552	0.6887		--	--	
		$\lambda = 2$	1.5959	0.5164	788.5604	0.3176	4.4496	3491.9802
40	0.05	$\beta_1 = -1$	-1.1826	0.0977		0.1421	1.3143	
		$\beta_2 = 1$	1.0352	0.0689		--	--	
		$\theta^* = 2$	1.5264	0.7455		--	--	
		$\lambda = 2$	1.5821	0.5268	745.2377	0.2198	7.2018	3255.1475
40	0.1	$\beta_1 = -1$	-1.179	0.1056		0.1445	1.3191	
		$\beta_2 = 1$	1.0118	0.1002		--	--	
		$\theta^* = 2$	1.5605	0.7487		--	--	

TABLO 2: n=60 ve durdurma oranı=0.01, 0.05 ve 0.10 alındığında önerilen model için benzetim sonuçları

			Önerilen Model			İndirgenmiş Model		
n	**Durdurma Oranı**	**Parametre**	**TAHMİN**	**HKO**	**AIC**	**TAHMİN**	**HKO**	**AIC**
		$\lambda = 2$	1.5584	0.3925	1141.1738	1.6415	1.615	3309.6809
60	0.01	$\beta_1 = -1$	-1.1696	0.075		-0.0049	0.9934	
		$\beta_2 = 1$	1.0524	0.049		--	--	
		$\theta^* = 2$	1.7537	0.5221		--	--	
		$\lambda = 2$	1.5282	0.3969	1052.5197	1.7317	1.8445	2951.4873
60	0.05	$\beta_1 = -1$	-1.1567	0.0686		-0.0109	0.9808	
		$\beta_2 = 1$	1.0653	0.0484		--	--	
		$\theta^* = 2$	1.8359	0.5693		--	--	
		$\lambda = 2$	1.494	0.4511	1006.8226	1.6502	2.0832	2880.4654
60	0.1	$\beta_1 = -1$	-1.1414	0.0649		-0.0085	0.9857	
		$\beta_2 = 1$	1.0806	0.0513		--	--	
		$\theta^* = 2$	1.7745	0.4772		--	--	

TABLO 3: n=100 ve durdurma oranı=0.01, 0.05 ve 0.10 alındığında önerilen model için benzetim sonuçları.

			Önerilen Model			İndirgenmiş Model		
n	**Durdurma Oranı**	**Parametre**	**TAHMİN**	**HKO**	**AIC**	**TAHMİN**	**HKO**	**AIC**
		$\lambda = 2$	1.425	0.4169		2.0797	0.6594	
100	0.01	$\beta_1 = -1$	-1.1323	0.0412	1903.9224	-0.0211	0.9584	5522.9343
		$\beta_2 = 1$	1.0897	0.0318		--	--	
		$\theta^* = 2$	1.6431	0.2986		--	--	
		$\lambda = 2$	1.4006	0.4434		2.124	0.9605	
100	0.05	$\beta_1 = -1$	-1.1405	0.0426	1808.9962	-0.0214	0.9577	5108.5693
		$\beta_2 = 1$	1.0815	0.0295		--	--	
		$\theta^* = 2$	1.7478	0.3159		--	--	
		$\lambda = 2$	1.4067	0.4404		2.1168	0.829	
100	0.1	$\beta_1 = -1$	-1.1399	0.0437	1692.4875	-0.0214	0.9577	4716.6303
		$\beta_2 = 1$	1.0821	0.0309		--	--	
		$\theta^* = 2$	1.7513	0.3037		--	--	

AIC değerlerine bakılarak yapılmıştır. AIC değerleri modeldeki parametre sayısına duyarlı olmasına rağmen önerilen modelde daha düşük çıkmıştır. Her bir yineleme için durdurma oranlarının artması sonuçlar üzerinde belirgin bir etki yaratmamıştır (Tablo 1-3).

UYGULAMA

Uygulamada, Üroloji Araştırma Grubu tarafından elde edilen mesane kanseri çalışmasındaki veriler kullanılmıştır. Bu çalışmada, çalışmaya giren tüm hastalarda mesane tümörü vardır. Bu tümörler alınıp hastalar rasgele olarak tedavi (piridoksin ve tiyotepa) ve kontrol grubu olmak üzere üç gruba atanmış ve hastalar izlenmeye başlanmıştır. Birçok hastada, çalışma boyunca tümör yeniden oluşmuştur. Bu tümörler tekrar alınmış ve hastalar izlenmeye devam edilmiştir. Burada, her hastanın tümör yineleme zamanları, tedavinin başlangıcından itibaren ölçülmüştür.

Çalışmanın amacı, kontrol grubuna (plaseboya) karşı nükseden mesane tümörlerinde piridoksin (Vitamin B_6) ve tiyotepa etkisini belirlemektir. Bu veri kümesi literatürde birçok çalışmada kullanılmıştır.[13,14,22] Yapılan çoğu çalışmada kontrol ile tiyotepa grupları karşılaştırıldığı için bu çalışmada da bu iki grup alınmıştır. Çalışma, kontrol grubunda 48 hasta, tedavi (tiyotepa) grubunda 38 hasta olmak üzere toplamda 86 hastadan oluşmaktadır. Çalışmada yer alan hastalarda ortalama takip süresi 31.5 hafta ve ortalama durdurma oranı 0.35 olup, en fazla yineleme (tümör) sayısı 9'dur. Kontrol grubundaki hastaların 19'unda, tedavi grubundaki hastaların 20'sinde yineleme gözlenmemiştir. Yineleme gerçekleşen hastalarda ortalama yineleme sayısı kontrol grubu için 3 ve tedavi grubu için 2,5'dir. Çalışmanın sonunda kontrol grubundaki hastalardan 11 kişi, tedavi grubundaki hastalardan ise 10 kişinin öldüğü gözlenmiştir. Dolayısıyla bu çalışmada bilgi içeren durdurma söz konusudur.

Bu verinin modellenmesinde, bu çalışmada önerilen model kullanılmış ve modellerde yer alan parametrelerin tahmini R programında kod yazılarak elde edilmiştir. Benzetim çalışmasında olduğu gibi parametrelerin en çok olabilirlik tahminleri Newton-Raphson algoritması kullanılarak elde edilmiştir ve aynı ε değeri kullanılarak iterasyonlara son verilmiştir. Önerilen modele ilişkin parametrelerin tahmin değerleri, tahminlerin standart hataları (SH) ve AIC değeri elde edilmiştir (Tablo 4).

Önerilen modelde Eşitlik (1) de yer alan zayıflık terimine göre integral alınabildiği için parametre yorumları Cox orantılı hazard modelindekine benzer şekilde yapılabilir. Tablo 4'e göre çalışmanın sonunda ölüm olayını yaşamayan hastalar için kontrol grubunun tedavi grubuna göre kanser yineleme riski $\exp(\beta_1) = 1.41$ kat daha fazladır. Bu durum çalışmanın sonunda ölüm olayını yaşayan hastalar için tedavi grubunun kontrol grubuna göre kanser yineleme riski $\exp(\beta_1 + \beta_2) = 1.28$ kat fazladır (Tablo 4).

SONUÇ VE TARTIŞMA

Yinelemeli olay verisinin modellenmesinde bilgi içeren durdurma olayının göz önünde bulundurulması gerekmektedir. Çünkü veride ölüm gibi bir durdurma söz konusu olduğunda yineleme zamanları ile ölüm zamanları ilişkilidir; bu ilişkinin modelde dikkate alınmaması parametrelere ilişkin yanlı tahminler elde edilmesine yol açmaktadır. Gerçekte bu ilişkinin tam olarak modellenmesi mümkün değildir. Ancak, çalışmanın giriş bölümünde de verildiği gibi, literatürde bilgi içeren durdurma varlığı altında yinelemeli olay verisinin modellenmesi üzerine yapılan çok sayıda çalışma vardır.

Bu çalışmada, yinelemeli olay süreci ile ölüm süreci arasında güçlü bir ilişkinin olduğu düşünülerek bu ilişkinin yapısını da dikkate alan yeni bir model önerilmiştir. Önerilen model Cox orantılı hazard modelinin, yinelemeli olay süreci için genişletilmiş biçimi olarak düşünülebilir. Bilgi içeren durdurma varlığında, yinelemeli olay süreci, homojen Poisson süreci ile modellenmiş; bilgi içeren durdurma olarak ise ölüm olayı alınmıştır. Önerilen modelde ölüm olayının, yinelemeli olay sürecine ilişkin yoğunluk fonksiyonuna açıklayıcı değişken olarak eklen-

TABLO 4. Önerilen model için mesane kanseri verisi sonuçları.

		λ	β_1	β_2	θ^*
Önerilen Model	**TAHMİN**	0.0489	-0.3454	0.5968	0.5452
	SH	0.0119	0.3643	0.4230	0.0690
	AIC	980.2474			

mesi düşünülmüştür. Aynı biçimde yinelemeli olaylar arasındaki ilişki ise gözlenemeyen rasgele etkiyi ifade eden zayıflık teriminin modele eklenmesiyle oluşturulmuştur.

Bu çalışmada önerilen modeldeki parametrelerin tahminleri en çok olabilirlik yöntemi kullanılarak elde edilmiş; farklı senaryolar için bir benzetim çalışması yapılmıştır. Benzetim çalışması R programında kod yazılarak yürütülmüş ve her bir senaryoya ilişkin sonuçlar yaklaşık 20 gün sonunda elde edilebilmiştir.

Benzetim çalışmasında önerilen modelin geçerliliğini göstermek için, ölüm olayının dikkate alınmadığı ve bilgi içermeyen durdurma olarak düşünüldüğü indirgenmiş model ile karşılaştırma yapılmış, bunun için AIC değerleri kullanılmıştır. Çalışmada önerilen modelin, AIC değerlerine göre indirgenmiş model ile yapılan karşılaştırılmasında genel olarak daha iyi sonuçlar verdiği görülmüştür. Bu sonuç, ölüm olayı bilindiğinde ve durdurma varlığında yinelemeli olay sürecinin modele etkisinin önemli olduğunu düşündürmektedir.

Ayrıca önerilen model, gerçek bir veri kümesi kullanılarak modellenmiş ve sürece ilişkin parametre tahminleri elde edilmiştir. Ancak önerilen model, literatürde yer alan modellerle karşılaştırılamamıştır. Bunun nedeni literatürde yer alan çalışmalarda geçen modellerin yapısının ve modellerdeki parametrelerin anlamsal olarak birbirinden farklılık göstermesidir.

Bu çalışmada, önerilen modelde sadece açıklayıcı değişkenin 0 ya da 1 değerini alan kategorik bir değişken olduğu durum ele alınmıştır. Bundan sonraki çalışmalarda modele daha fazla sayıda zamandan bağımsız ya da zamana bağlı açıklayıcı değişken eklenmesi düşünülebilir. Parametre sayısının artması model yapısını ve tahmin sürecini daha karmaşık bir hale getireceğinden tahminler farklı algoritmalar (Genetik algoritma vb.) kullanılarak elde edilebilir.

KAYNAKLAR

1. Kleinbaum D, Klein M. Survival Analysis: A Self Learning Text. 2nded. New York: Springer; 2005. p.590.
2. Cox DR. Regression models and life tables. J R Statist Soc Series B 1972;34(2):187-34.
3. Ata N, Karasoy D, Sözer MT. [Stratified cox regression model for non-proportional hazards and an application on breast cancer patients]. Turkiye Klinikleri J Med Sci 2008;28(3):327-32.
4. Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. Ann Statist 1982;10(4):1100-20.
5. Lin JS, Wei LJ. Linear regression analysis for multivariate failure time observations. Journal of the American Statistical Association (JASA) 1992;87(420):1091-7.
6. Pepe MS, Cai J. Some graphical displays and marginal regression analyses for recurrent failure times and time dependent covariates. Journal of the American Statistical Association (JASA) 1993;88(423):811-10.
7. Balshaw RF, Dean CB. A semiparametric model for the analysis of recurrent-event panel data. Biometrics 2002;58(2):324-8.
8. Lin DY, Wei LJ, Ying Z. Accelerated failure time models for counting processes. Biometrika 1998;85(3):605-14.
9. Lin DY, Wei LJ, Yang I, Ying Z. Semiparametric regression for the mean and rate functions of recurrent events. J R Statist Soc Series B 2000;62(4):711-20.
10. Pan W. Using frailties in the accelerated failure time model. Lifetime Data Analysis 2001;7(1):55-64.
11. Nielsen JD, Dean CB. Clustered mixed nonhomogeneous Poisson process spline models for the analysis of recurrent event panel data. Biometrics 2008;64(3):751-61.
12. Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. Stat Med 1997;16(8):911-4.
13. Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. Biometrics 2000;56(2):554-62.
14. Ghosh D, Lin DY. Marginal regression models for recurrent and terminal events. Statistica Sinica 2002;12(3):663-88.
15. Huang X, Wolfe RA. A frailty model for informative censoring. Biometrics 2002;58(3):510-20.
16. Ghosh D, Lin DY. Semiparemetric analysis of recurrent events data in the presence of dependent censoring. Biometrics 2003;59(4):877-85.
17. Liu L, Wolfe RA, Huang X. Shared frailty models for recurrent events and a terminal event. Biometrics 2004;60(3):747-56.
18. Huang CY, Wang MC. Joint modelling and estimation for recurrent event processes and failure time data. J Am Stat Assoc 2004;99(468):1153-65.
19. Huang CY, Qin J, Wang MC. Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. Biometrics 2010;66(1):39-49.
20. Byar D, Blackard C. Comparisons of placebo, pyridoxine, and topical thiotepa in preventing recurrence of stage 1 bladder cancer. Urology 1976;10(6):556-61.
21. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. 2nded. Hoboken, NJ: Wiley-Interscience; 2002. p.1-447.
22. Zhao X, Zhou J, Sun L. Semiparametric transformation models with time-varying coefficients for recurrent and terminal events. Biometrics 2010;67(2):404-14.