

Tekrarlı Gözlem ve Sağkalım Verilerinin Bileşik Modellenmesi

Joint Modelling of Repeated Measurement and Survival Data: Review

Ege ÜLGEN,^a
Özgür ASAR^a

^aBiyostatistik ve Tıp Bilişimi Bölümü,
Acıbadem Üniversitesi Tıp Fakültesi,
İstanbul

Geliş Tarihi/Received: 30.05.2017
Kabul Tarihi/Accepted: 08.07.2014

Yazışma Adresi/Correspondence:
Özgür ASAR
Acıbadem Üniversitesi Tıp Fakültesi,
Biyostatistik ve Tıp Bilişimi Bölümü,
İstanbul,
TÜRKİYE/TURKEY
ozgurasarstat@gmail.com

ÖZET Sağlık durumu ve sağkalım süreçleri arasındaki ilişkinin anlaşılması ve bu süreçler hakkında ileriye dönük kestirimler yapılması güncel tıbbi araştırma konularındandır. Bu süreçlerden ilki tekrarlı biyobelirteçler ile takip edilmekte iken, ikincisi sağkalım süreleri ve sağkalım olayının gerçekleşip gerçekleşmediği değişkenleri ile takip edilmektedir. Tekrarlı gözlem ve sağkalım verilerinin birlikte analiz edilmesine olanak sağlayan bileşik modeller, bahsi geçen konuların araştırılmasına olanak sağlayan güncel (biyo)istatistiksel yöntemler arasında yer almaktadır. Bu modeller, boylamsal veri analizi literatüründe sıkça kullanılan lineer karma etkili modeller (diğer adıyla rastgele etkili modeller) ile sağkalım veri analizi literatüründe sıkça kullanılan genişletilmiş Cox modellerini birleştirmektedir. Parametre tahmini için beklenti-en çoklama ve direkt en çoklama algoritmalarına dayalı en çok olabilirlik yöntemi ve Markov Zinciri Monte Carlo örnekleme algoritmalarına dayalı Bayesçi yöntemler sıkça kullanılmaktadır. Bu makalenin amacı bileşik modeller üzerine bir derleme ve eğitsel metin sunmaktır. Yöntemler, 65 yaş üzeri kişilerde demans riskini araştırmayı hedefleyen PAQUID çalışmasından elde edilen bir veri setine uygulanmıştır. Kognitif fonksiyon ile demans riski arasındaki ilişkinin anlaşılması ve bu süreçler için ileriye dönük kestirimlerin elde edilmesi bilimsel ilgiler arasındadır. Bileşik modeller ile birlikte lineer karma etkili model ve genişletilmiş Cox modeli PAQUID verisine uygulanmış ve sonuçlar kıyaslanmıştır. Seçilen bir kişi için ileriye dönük kognitif fonksiyon ve demans olasılıkları dinamik olarak elde edilmiştir.

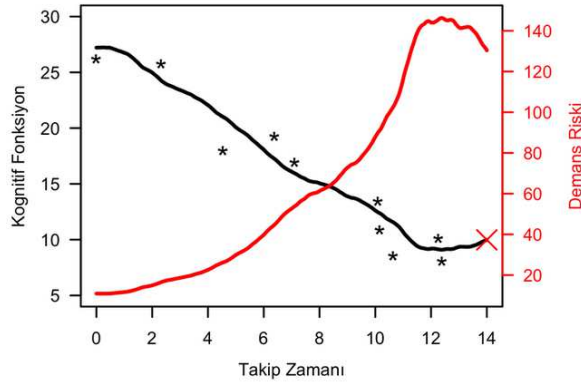
Anahtar Kelimeler: Boylamsal veri analizi; ortak parametrelili modeller; rastgele etkiler; yaşam çözümlemesi

ABSTRACT Understanding the relationship between health status and survival processes and obtaining forecasts regarding these processes are amongst the recent areas in medical research. Whilst the first process is measured by repeated biomarkers, the second is measured by survival times and event indicators. Joint models for simultaneous analysis of repeated measures and survival data enable researchers to investigate the aforementioned topics and are amongst the current (bio)statistical methods. These models combine linear mixed effects models (also known as random effects models) that are widely used in longitudinal data analysis literature, and extended Cox models that are widely used in survival data analysis literature. Direct-maximisation and expectation-maximisation based maximum likelihood and Markov Chain Monte Carlo based Bayesian methods are widely used for parameter estimation. The aim of the current article is to present a review on joint models. The methods are applied to a data-set obtained from the PAQUID study. The study aims to investigate the dementia incidence amongst people over 65. Understanding the relationship between cognitive function and risk of dementia and obtaining forecasts for these processes are amongst the scientific interests. Together with joint models, linear mixed effects models and extended Cox models are applied to the PAQUID data-set, and results are compared. Forecasts of cognitive function and dementia probabilities are obtained for a subject.

Keywords: Longitudinal data analysis; shared-parameter models; random effects; survival analysis

Bu makale, sağlık durumu ve sağkalım riski süreçleri arasındaki ilişkinin belirlenmesi ve bu süreçler için ileriye dönük kestirimlerin hesaplanması üzerinedir. Bu amaçlar için, tekrarlı gözlem ve sağkalım verilerinin birlikte analiz edilmesine olanak sağlayan bileşik modeller kullanılmaktadır. Sağlık durumu, bir ya da birden çok sürekli ve/veya kesikli biyobelirteçler ile takip edilebilmektedir. Sağkalım süreci ise, yine bir ya da birden çok soldan, aralık ve sağdan sansürlü, yinelenen olay ve/veya yarışan riskler ile takip edilebilmektedir. Bu çalışmada sağlık durumu için bir sürekli biyobelirteci, sağkalım süreci için sağdan sansürlü bir terminal olayı göz önünde bulunduracağız.

Bahsi geçen süreçler, zaman içinde sürekli olarak değişen (continuous-time) süreçlerdir. Bu süreçlerin gerçek hayatta doğrudan gözlemlenmeleri, ekonomik ve pratik nedenlerden ötürü mümkün olmamaktadır. Gerçekteyse, sağlık durumu, tekrarlı gözlemler şeklinde toplanmış biyobelirteçler ile ölçülmektedir. Toplanan tekrarlı gözlemler zaman içinde tipik olarak birbirinden eşit olmayan aralıklarla ayrılmakta ve ölçüm hatasına maruz kalmaktadır. Sağkalım süreci ise, sağkalım süresi ve sağkalım olayının gerçekleşip gerçekleşmediği bilgisi ile ölçülmektedir. Süreçlerin gözlemlenebilen ve gizli (latent) özellikleri simüle edilmiş bir hastaya ait veri ile gösterilmiştir (Şekil 1). Burada siyah ve kırmızı sürekli çizgiler, sırası ile sağlık durumu ve sağkalım riski süreçlerini göstermektedir. Yıldız ile gösterilen değerler tekrarlı biyobelirteç gözlemlerini göstermektedir. Bu değerlerin siyah düz çizgiye dikey uzaklıkları ölçüm hatasına tekabül etmektedir. Kırmızı çarpı sağkalım olayını göstermektedir.



ŞEKİL 1: Simüle hasta verisi. Siyah sürekli çizgi: (gizli) kognitif fonksiyon.

Yıldızlar (*): gözlemlenen biyobelirteç değerleri. Kırmızı devamlı çizgi: (gizli) demans riski.

Kırmızı çarpı: sağkalım olayı.

Cox oransal risk modeli sağkalım analizi literatüründe sıkça kullanılmaktadır.¹ Bu model, bilimsel ilgi tekrarlı biyobelirteç gözlemlerinden bir tanesi, örneğin ilk veya son değer, ile sağkalım olayı riski arasındaki ilişki üzerine olduğunda kullanılabilen bir yöntemdir. Eğer ilgi tekrarlı gözlemler ile risk arasındaki ilişki üzerine ise, genişletilmiş Cox modeli kullanılabilir bir yöntemdir.²⁻⁴ Bu yöntemlerin iki büyük dezavantajı bulunmaktadır. Bunlardan ilki, biyobelirteç gözlemlerindeki ölçüm hatasının göz ardı edilmesidir. Ölçüm hatasının, sağlık durumu ve sağkalım süreçleri arasındaki ilişkiyi ifade eden parametrenin sıfıra doğru yanlı tahmin edilmesine neden olduğu bilinmektedir.⁵⁻⁷ İkinci dezavantaj ise, istatistiksel çıkarım aşamasında gerekli olan, biyobelirteçlerin zaman içindeki sürekli değerleri için yapılan varsayımlardır. Cox oransal risk modeli, biyobelirteç değerlerinin tek bir seviyede olduğu varsayımı ile çalışmakta iken, genişletilmiş Cox modeli bu değerlerin birbirini takip eden tekrarlı

biyobelirteç seviyeleri arasında sabit kaldığı varsayımı ile çalışmaktadır. Bileşik modeller, hem biyobelirteçlerdeki ölçüm hatasının düzeltilmesini hem de biyobelirteçlerin zaman içinde sürekli olarak tahmin edilerek istatistiksel çıkarım aşamasında kullanılmasını olanaklı kılmaktadır. Bu modellerin sıkça kullanılan bir versiyonu, lineer karma etkili model ile zamana bağlı rastgele etkili terim içeren genişletilmiş Cox modelini birleştiren, iki seviyeli ortak parametrelili modellerdir.⁷⁻¹² Bileşik modellerin, bahsi geçen Cox modellerine kıyasla, bir diğer avantajı, sağlık durumu sürecinin de modellenmesine olanak sağlamasıdır. Bu sayede, bu süreç hakkında göz ardı edilemez kayıp veri (non-ignorable missing data) varlığında çıkarımlar yapılabilir.¹³ Bileşik modeller üzerine şimdiye kadar yazılmış derlemeler için şu kaynaklara başvurulabilir: Davidian ve Tsiatis (2004), McCrink, Marshall ve Cairns (2013), Asar, Ritchie, Kalra ve Diggle (2015), Gould ve diğ. (2015), Sudell ve diğ (2016).¹⁴⁻¹⁸ Ayrıca, bu modeller üzerine yazılmış iki adet kitap bulunmaktadır: Rizopoulos (2012), Elashoff, Li ve Li (2017).^{12,19}

Makalenin yapısı şu şekildedir. 2. bölümde yöntemler tanıtılacaktır. 3. bölümde yöntemlerin, R paketi lmm'de açık olarak mevcut olan PAQUID verisi-seti üzerine uygulamaları sunulacaktır. Makale 4. bölümde tartışma ile sonuçlandırılacaktır.

■ BİLEŞİK MODELLER

NOTASYON

i ($i = 1, \dots, n$) kişisinden toplanan veri; p adet zaman içinde değişmeyen değişken, $a_i = \{a_{i,1}, \dots, a_{i,p}\}$, m_i adet tekrarlı biyobelirteç gözlemi, $Y_i = \{Y_{i1}, \dots, Y_{im_i}\}$, biyobelirteçlerin toplandığı tarihler, $t_i = \{t_{i1}, \dots, t_{im_i}\}$, sağkalım süresi, T_i , ve sağkalım olayının gerçekleşip gerçekleşmediği gösterge değişkenidir, E_i . Burada gözlenen sağkalım süresi, gerçek sağkalım süresi, T_i^* , ve i kişisi için sağdan sansürlenme süresinin, C_i , minimumu olarak tanımlanmaktadır: $T_i = \min(T_i^*, C_i)$. Sansürlenme süresi, i kişisi için çalışmadan erken ayrılma (drop-out) süresi, D_i , ve çalışmanın planlanmış bitiş zamanının, C , minimumu olarak tanımlanmaktadır: $C_i = \min(D_i, C)$. Olay göstergesi değişkeni ise $E_i = I(T_i^* \leq C_i)$ olarak tanımlanmaktadır. Burada, $I(\cdot)$ gösterge fonksiyonudur. Zaman içinde değişen açıklayıcı değişkenleri, örneğin diyabet başlangıcı, kardiyovasküler olayların olması gibi durumlar, olası içsel değişkenlikten (endogeneity) ötürü göz ardı etmekteyiz.⁴

MODEL

Bileşik modeller için iki seviyeli ortak parametrelili formülasyon şu şekildedir:

$$Y_{ij} = Y_{ij}^* + Z_{ij} = x'_{ij}\alpha + d'_{ij}U_i + Z_{ij}, \quad (1)$$

$$h_i(t) = h_0(t) \exp\left(c'_i\beta + f(Y_i^*(t), \gamma)\right). \quad (2)$$

Burada, (1) numaralı eşitlik lineer karma etkili model iken, (2) numaralı eşitlik zamana bağlı rastgele etki terim içeren genişletilmiş Cox modelidir; $f(Y_i^*(t), \gamma) = Y_i(t)\gamma$ durumu standart genişletilmiş Cox modeline tekabül etmektedir.

Önceki notasyona ek olarak, (1) numaralı eşitlikte, Y_{ij}^* , i kişisi için j ($j = 1, \dots, m_i$) takibindeki ölçüm hatasından arındırılmış biyobelirteç seviyesini göstermekte iken, Z_{ij} ölçüm hatasını göstermektedir. Y_{ij}^* , sabit etkili terim, $x'_{ij}\alpha$, ve rastgele etkili terimin, $d'_{ij}U_i$, toplamı şeklinde ifade edilmektedir. Burada, x_{ij} ve d_{ij} , a_i ve t_i değişkenlerinden oluşan bir kolon matrisleridir; α sabit etkili regresyon parametrelerinden oluşan bir kolon matrisidir; U_i rastgele etki katsayılarından oluşan kolon matrisidir. d_{ij} genel olarak x_{ij} 'nin elemanlarının alt kümesi olarak tanımlanmaktadır. $d'_{ij}U_i$ teriminin $U_{oi} + U_{1i}t_{ij}$

olarak tanımlanması sıkça kullanılan rastgele kesim ve rastgele eğim formülasyonu olarak bilinmektedir. U_i katsayılarının tipik olarak sıfır ortalamaya ve Σ varyans-covaryans matrisine sahip çok değişkenli Gaussian rastgele değişkenler olduğu varsayımı yapılmaktadır ($U_i \sim \text{MVN}(0, \Sigma)$). Z_{ij} terimleri için özdeş ve birbirinden bağımsız Gaussian rastgele değişkenler oldukları varsayımı yapılmaktadır: $Z_{ij} \perp Z_{ik} \sim N(0, \sigma^2)$ ($j \neq k$). Diğer bir varsayım, rastgele etki ve ölçüm hatası terimlerinin birbirinden bağımsız olduklarıdır: $U_i \perp Z_{ij}$.

(2) numaralı eşitlikte, $h_i(t)$ i kişisi için t zamanındaki sağkalım olayı riskini göstermekte ve şu şekilde tanımlanmaktadır:

$$h_i(t) = \lim_{\Delta t \rightarrow 0} \frac{P(t \leq T_i^* < t + \Delta t | T_i^* \geq t, D_i(t); \theta)}{\Delta t} \quad (3)$$

Burada, $D_i(t)$ terimi i kişisi için t zamanına kadar (t zamanı da dahil) kaydedilmiş verileri göstermekte iken, θ , (1) ve (2) numaralı eşitliklerdeki parametrelerin toplam kümesidir. $h_0(t)$ temel risk olarak adlandırılmakta ve (2) numaralı eşitlikte bulunan $\exp(\cdot)$ terimindeki toplam ifadesinin 0 değerini aldığı kişiler için t zamanındaki sağkalım olayı riskini göstermektedir. $h_0(t)$, Cox modellerinde olduğu gibi parametrik olmayan bir biçimde tanımlanabilir.¹ Parametrik olmayan temel risk varsayımı altında bileşik model parametre tahminleri için Fisher matrisi ile edilen standart hata tahminlerinin sıfıra doğru yanlış olduğu bilinmektedir.²⁰ Bu durumda standart hataların bootstrap metodu ile elde edilmesi önerilmiştir. $h_0(t)$ için, sıkça kullanılan parametrik dağılımların risk fonksiyonları da bir seçenektir, örneğin Weibull hazard fonksiyonu: $h_0(t) = \delta \phi t^{\phi-1}$ (δ ölçek parametresi, ϕ şekil parametresidir). Diğer parametrik ve aynı zamanda esnek yöntemler ise parçalı sabit (piecewise constant) ve yiv (spline) temel risk fonksiyonlarıdır.^{12,21} c_i temel değişkenlerden oluşan kolon matrisi iken, β ilgili parametrelerden oluşan kolon matrisidir. $f(\cdot)$, ölçüm hatasından arındırılmış sürecin t zamanına kadar (t zamanı da dahil) olan geçmiş değerleri, $\mathcal{Y}_i^*(t) = \{Y_i^*(s) : s \leq t\}$, ile $h_i(t)$ arasındaki ilişkiyi belirleyen, matris sonuçlu bilinen bir fonksiyondur. γ ilişki parametrelerini göstermektedir. $f(\cdot)$ 'in en sık kullanılan versiyonları arasında, güncel değer formülasyonu, $f(\mathcal{Y}_i^*(t), \gamma) = Y_i^*(t)\gamma$, ve güncel değer ve değişim hızı formülasyonu, $f(\mathcal{Y}_i^*(t), \gamma) = Y_i^*(t)\gamma_1 + \frac{\partial Y_i^*(t)}{\partial t}\gamma_2$, gelmektedir. Diğer formülasyonlar için Henderson, Diggle ve Dobson (2000), Rizopoulos (2012), Hickey ve diğ. (2016) ve Elashoff, Li ve Li (2017) kaynakları incelenebilir.^{11,12,19,22}

PARAMETRE TAHMİNİ

Bileşik modellerde parametre tahmini için iki aşamalı yöntem kullanılabilir. Bu yöntemin ilk aşamasında boylamsal veri, (1) numaralı eşitlikteki lineer karma etkili model ile modellenmekte ve tekrarlı gözlemlerin ölçüm hatasından arındırılmış versiyonları, $\tilde{Y}_{ij} = x_{ij}'\hat{\alpha} + d_{ij}\tilde{U}_i$, elde edilmektedir. $\hat{\alpha}$ en çok olabilirlik, sınırlandırılmış en çok olabilirlik veya Bayesçi yöntemlerle elde edilebilir. \tilde{U}_i , parametre tahmini için en çok olabilirlik yöntemleri kullanıldığı durumlarda en iyi lineer yansız kestirimler, $\tilde{U}_i = E(U_i | Y_i; \hat{\theta})$, Bayesçi yöntemler durumunda ise sonsal dağılımın ortalaması, $\tilde{U}_i = E(U_i | Y_i)$, olarak elde edilebilir. \tilde{Y}_{ij} değerleri, $h_0(t)$ 'nin parametrik olmayan bir formda tanımlanması durumunda sansürlenmemiş sağkalım zamanlarında, parametrik bir formda tanımlanması durumunda ise 0 ile sağkalım süresinin küçük aralıklara bölünmesi ile elde edilen zamanlarda elde edilmektedir. Boylamsal verinin modellenmesi için, sıradan ve risk seti kalibrasyonları olarak bilinen iki popüler yöntem bulunmaktadır. Bunlardan ilki, mevcut boylamsal verinin tamamının tek seferde kullanılmasını esas almaktadır. İkinci yöntem ise, sansürlenmemiş sağkalım zamanları için ayrı ayrı karma etkili modellerin kurulmasını esas almaktadır. İki aşamalı yöntemler, genişletilmiş Cox modeline kıyasla, boylamsal ve

sağkalım süreçleri arasındaki ilişkiyi ölçen parametrenin daha yansız tahmin edilmesine olanak sağlamaktadırlar. Elde edilen sonuçlar yine de yansız değildir. Bunun nedenleri arasında, sağkalım verisinin sıradan kalibrasyonda göz ardı edilmesi, risk seti kalibrasyonunda kurulan karma etkili modellerden bazıları için az sayıda veri olması ve normal dağılım varsayımının geçersizliği, her iki yöntem için de \tilde{Y}_{ij} değerlerinin belirsizliklerinin göz ardı edilmesi sayılabilir. İki aşamalı yöntemin detayları için, Tsiatis, DeGruttola, Wulfson (1995), Sweeting ve Thompson (2011) ve Sweeting, Barrett, Thompson ve Wood (2016) makaleleri incelenebilir.^{6,23,24}

Bileşik olabilirlik fonksiyonuna, $[Y_i, T_i, E_i, U_i]$, dayalı yöntemler (en çok olabilirlik ve Bayesçi yöntemler) bileşik model parametre tahminlerinin yansız olarak elde edilmesini olanaklı kılmaktadır. Burada “[.]” sembolü olasılık dağılım fonksiyonunu göstermektedir. Bu yöntemler, rastgele etkiler biliniyorken, boylamsal ve sağkalım süreçlerinin bağımsız olduğu temel varsayımını yapmaktadır:

$$[Y_i, T_i, E_i, U_i] = [Y_i|U_i][T_i, E_i|U_i][U_i]. \quad (4)$$

En çok olabilirlik yöntemi için sıkça kullanılan tekrarlı temelli (iterative) iki algoritma bulunmaktadır: doğrudan en çoklama ve beklenti en çoklama algoritmaları. Doğrudan en çoklama algoritması bileşik olabilirlik fonksiyonunun rastgele etkiler üzerinden integrali ile çalışmaktadır.²⁵ Bu integral direkt olarak çözülememekte, nümerik yöntemler ile yaklaştırılması gerekmektedir. Düşük boyutlu integraller için uyarlayıcı ve yalancı uyarlayıcı (pseudo-adaptive) Guass-Hermite kare yapma yöntemi iyi sonuçlar vermekte iken, yüksek boyutlu integraller için Monte Carlo yöntemleri önerilmektedir.^{11,26} Beklenti-en çoklama algoritması bütün verinin, $\{Y_i, T_i, E_i, U_i\}$, gözlenen veriye, $\{Y_i, T_i, E_i\}$, koşullu beklentisinin en çoklanmasını esas almaktadır.^{7,11,12,19} Beklenti aşamasında, rastgele etkilerin bilinen birtakım fonksiyonlarının gözlenen veriye koşullu beklentileri, bir önceki tekrarlardan elde edilen parametre tahminleri kullanılarak hesaplanır. Bu aşamada nümerik yöntemlerle integral hesaplamaları yapılması gerekmektedir. Elde edilen beklenti değerleri en çoklama aşamasında yerine konularak parametre tahminlerinin yeni tekrarlardaki değerleri elde edilir. Bayesçi yöntemler, parametre ve rastgele etkiler için çıkarımların sonsal dağılım, $[U_i, \theta|Y_i, T_i, E_i,]$, kullanılarak elde edilmesine dayanır.²⁷⁻³⁰ İlgili sonsal dağılım bileşik modeller için bilinen bir olasılık dağılım fonksiyonuna benzememektedir. Bunun için bu dağılımdan Markov Chain Monte Carlo metotları, örneğin Gibbs örnekleme veya Metropolis-Hastings algoritması, kullanılarak sözde değerler (random numbers) türetilmektedir.

KESTİRİM

Bileşik modeller ile kişiselleştirilmiş, ileriye dönük boylamsal süreç ve sağkalım olasılıkları için kestirimler elde edilebilmektedir. t zamanına kadar verisi ($D_i(t)$) kaydedilen bir kişiye ait s ($s > t$) zamanındaki boylamsal süreç ve sağkalım olasılıkları için nokta tahmini ve güven aralıkları sırasıyla $[Y_i(s)|D_i(t)]$ ve $[T_i, E_i|D_i(t)]$ dağılımlarından elde edilebilir. Bu dağılımlar, bilinen bir dağılımın olasılık dağılım fonksiyonuna sahip değildir. Rizopoulos (2011, 2012) bu dağılımlardan Monte Carlo metotları ile sözde değerler üretmeyi önermiştir.^{12,31}

■ UYGULAMA

PAQUID ÇALIŞMASI

PAQUID çalışmasında güney-batı Fransa bölgesinde yaşayan 65 yaşın üzerinde kişiler, 1988’den 2004’e kadar takip edilmiştir.³² Çalışmanın amacı demans ve Alzheimer insidansını belirlemektir. Bu makalede PAQUID çalışmasından elde edilmiş, lcmm paketinde açık kaynaklı olarak mevcut bir veri seti

kullanılmaktadır. Amaç, demans riski ile demografik faktörler ve kognitif fonksiyon arasındaki ilişkilerin belirlenmesi ve kognitif fonksiyon ve demans olasılıkları için ileriye dönük kişiselleştirilmiş kestirimler yapmaktır. Bileşik modelleme ile cevaplanabilecek soruların bir kısmı şunlardır:

- Demografik değişkenler ile demans riski arasında ilişki var mıdır?
- Demografik değişkenler ile kognitif fonksiyon arasında ilişki var mıdır?
- Kognitif fonksiyon ile demans riski arasında ilişki var mıdır?
- Kognitif fonksiyon zaman içinde nasıl değişmektedir?

Değişken listesi Tablo 1’de sunulmuştur. Kognitif fonksiyonun anlaşılabilmesi için birden fazla biyobelirteç toplanmıştır. İçlerinden MMSE kognitif fonksiyon hakkında global bir fikir verdiğinden kullanılacak biyobelirteç olarak seçilmiştir. Özet istatistikler Tablo 2’de sunulmuştur. Tekrarlı MMSE gözlemleri toplandıkları zamana göre serpmme grafikte gösterilmiştir (Şekil 2). MMSE değerleri gri renkte gösterilmiş ve seçilen 8 kişi için zaman içinde değişim çizgiler olarak belirtilmiştir. Sağkalım sürecinin özetlemek için Kaplan-Meier sağkalım olasılık tahminleri hesaplanmıştır (Şekil 3). PAQUID verisi üzerinde bileşik model kullanılarak saptanabilecek ilişkiler bir diyagram ile ifade edilmiştir (Şekil 4).

SONUÇLAR

Parametre Tahminleri

PAQUID verisinin analizi için kurulan bileşik model aşağıdaki gibidir:

$$MMSE_{ij} = MMSE_{ij}^* + Z_{ij} = \alpha_0 + \alpha_1 CEP_i + \alpha_2 t_{ij} + \alpha_3 (age0_i - 65.25) + \alpha_4 male_i + \alpha_5 CEP_i t_{ij} + U_{oi} + U_{1i} t_{ij} + Z_{ij}, \quad (5)$$

$$h_i(t) = h_0(t) \exp(\beta_1 (age0_i - 65.25) + \beta_2 CEP_i + \beta_3 male_i + MMSE_i^*(t)\gamma). \quad (6)$$

Burada, $h_0(t)$ için parçalı sabit varsayımı yapılmıştır. Rastgele etkiler ve ölçüm hatası terimleri için Gaussian dağılım varsayımı yapılmıştır. Parametre tahmini için en çok olabilirlik yöntemi kullanılmıştır. En çoklama için beklenti-en çoklama algoritması kullanılmıştır. Beklenti adımıdaki integraller yalnızca uyarlayıcı Gauss-Hermite kare yapma yöntemi ile yakınlştırılmıştır.

Bileşik model ile birlikte, (5) numaralı eşitlikteki lineer karma etkili model ve (6) numaralı eşitlikteki genişletilmiş Cox modeli ($MMSE_i^*(t)\gamma$ terimi $MMSE_i(t)\gamma$ ile yer değiştirilerek kurulmuştur) de kurulmuştur. Bileşik modeller için parametre tahmini için JM paketi kullanılmıştır.³³ Lineer karma etkili için nlme paketi, genişletilmiş Cox modeli için survival paketi kullanılmıştır.

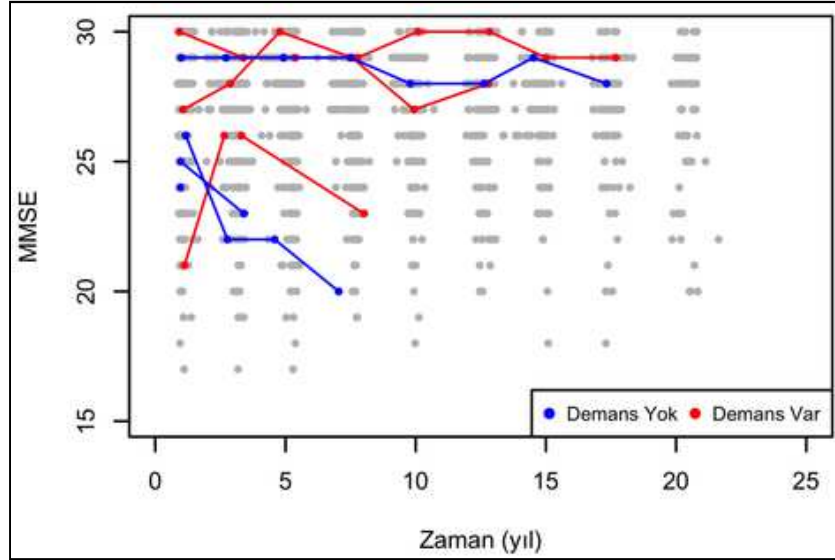
TABLO 1: PAQUID veri seti değişken listesi ve açıklamalar.

Değişken	Açıklama
CEP_i	i kişinin eğitim seviyesini bildiren ikili belirteç (CEP: Certificat d'études primaires); CEP = 1 ise kişinin ilköğretim diploması var, CEP = 0 ilköğretim diploması yok
$male_i$	i kişinin cinsiyetini bildiren ikili belirteç, male = 1 ise kişi erkek, male = 0 ise kadın
$age0_i$	i kişinin çalışmaya dahil olduğu zamandaki yaşı, minimum = 65.25
t_{ij}	tekrarlı gözlemlerin toplanma zamanları
$MMSE_i$	i kişinin global kognitif fonksiyonunu değerlendiren MMSE (Mini-Mental State Examination) test skoru
E_i	i kişinin demans tanısı için belirteç
T_i	i kişisi için demans zamanı; demans yaşı ile başlangıç yaşının farkı olarak hesaplanmıştır

TABLO 2: Özet istatistikler.

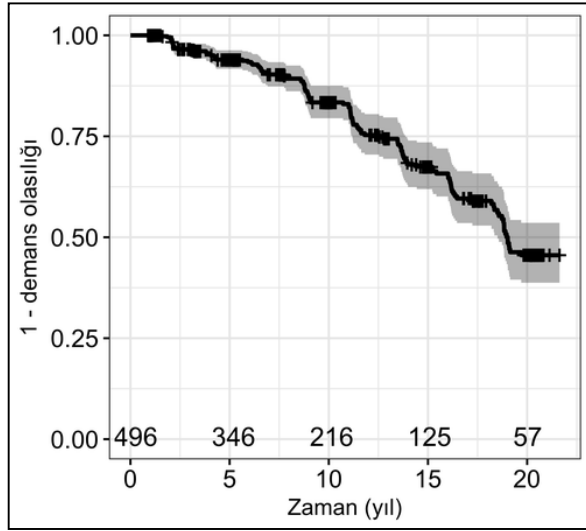
	Tüm Hastalar	Demans Yok	Demans Var
n	496	371 (%74.8)	125 (%25.2)
Tekrarlı Gözlem Sayısı min.-maks. Ortalama (std. sapma)	1-9 4.01 (2.54)	1-9 3.98 (2.62)	1-8 4.11 (2.29)
Başlangıç Yaşı minimum-maksimum Ortalama (std. sapma)	65.25-92.33 74.19 (6.39)	65.25-92.33 73.92 (6.43)	65.75-91.92 74.98 (6.2)
n (%) Kadın	286 (%57.66)	195 (%52.56)	91 (%72.8)
n (%) Erkek	210 (%42.34)	176 (%47.44)	34 (%27.2)
n (%) Eğitimsiz	143 (28.83%)	99 (26.68%)	44 (35.2%)
n (%) Eğitilmiş	353 (71.17%)	272 (73.32%)	81 (64.8%)
İlk MMSE skoru min.-maks. Ortalama (std. sapma)	17-30 27.01 (2.53)	17-30 27.18 (2.52)	20-30 26.51 (2.47)
Son MMSE min.-maks. Ortalama (std. sapma)	17-30 26.05 (2.72)	17-30 26.41 (2.69)	19-30 24.98 (2.55)

Min.: minimum, maks.: maksimum, std. sapma: standart sapma.

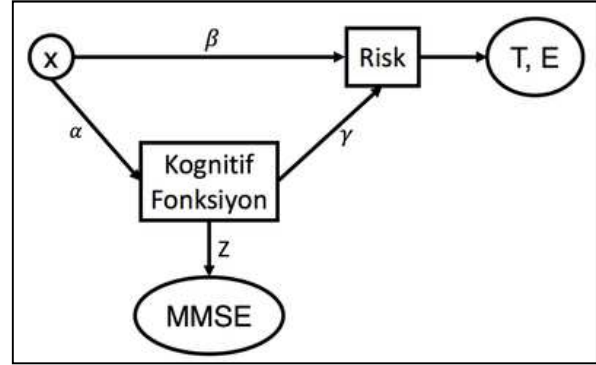


ŞEKİL 2: Tekrarlı MMSE skorları (gri). Kırmızı çizgiler: demans olan 4 kişi için MMSE skorlarının zamana göre değişimi. Mavi çizgiler: demans olmayan 4 kişi için MMSE skorlarının zamana göre değişimi.

Bileşik model, lineer karma etkili model ve genişletilmiş Cox modelinden elde edilen sonuçlar Tablo 3'te sunulmuştur. Bileşik modelden elde edilen boylamsal sürece ait sonuçların yorumu şu şekildedir. $\hat{\alpha}_1$, eğitilmiş bir kişiye ait başlangıçtaki MMSE skorunun eğitimsiz bir kişiye kıyasla yaklaşık 1.74 birim daha yüksek olduğunu belirtmektedir. $\hat{\alpha}_2$, eğitimsiz bir kişi için ortalama MMSE skorunun, bir yılda yaklaşık 0.17 birim azaldığını gösterirken, $\hat{\alpha}_5$, eğitilmiş bir kişi için bu azalmanın yaklaşık 0.10 birim ($\hat{\alpha}_2 + \hat{\alpha}_5 = -0.10$) olduğunu göstermektedir. $\hat{\alpha}_3$, başlangıç yaşındaki her bir yıllık artışın, ortalama MMSE skorundaki yaklaşık 0.13 birimlik azalma ile ilişkili olduğunu ifade etmektedir. Cinsiyet ile kognitif fonksiyon arasında anlamlı bir ilişki bulunmamıştır ($p = 0.96$). PAQUID verisi için lineer karma etkili model sonuçları bileşik modelden elde edilen boylamsal süreç sonuçlarıyla benzerdir. Bu durum, demans olayının MMSE değerleri için bilgi içermeyen kayıp veriye¹³ neden olduğu şeklinde yorumlanabilir.



ŞEKİL 3: Kaplan-Meier sağkalım tahminleri. Gri alan güven bandını göstermektedir. x ekseninin üzerinde bulunan sayılar, risk setinde bulunan kişi sayısını ifade etmektedir.



ŞEKİL 4: İlişki diyagramı. x temel değişkenleri ifade etmektedir. Ovaler: gözlemlenen süreçler. Dikdörtgenler: gizli süreçler.

TABLO 3: PAQUID verisi için bileşik model, lineer karma etkili model ve genişletilmiş Cox modeli sonuçları.

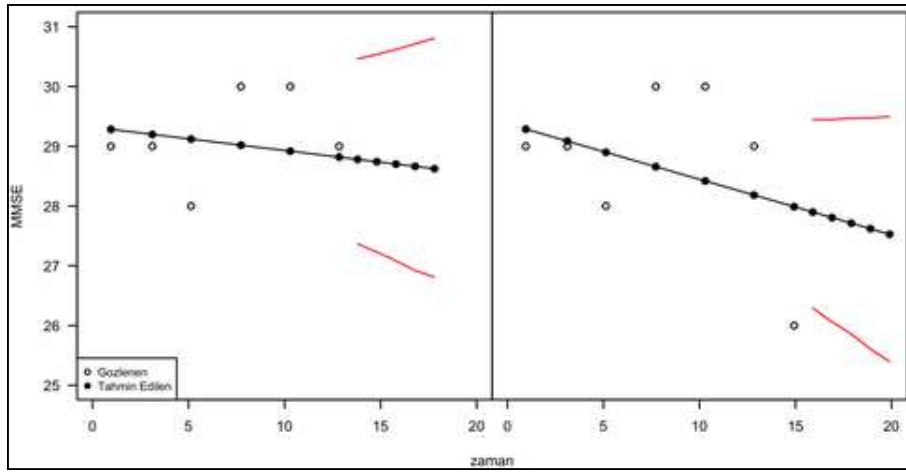
Boylamsal Süreç									
		Bileşik Model			Lineer Karma Etkili Model				
Değişken	Par.	Değer	Std. Hata	P	Değer	Std. Hata	P		
(Reg. sabiti)	α_0	27.21	0.24	<0.001	27.16	0.24	<0.001		
CEP == 1	α_1	1.74	0.22	<0.001	1.75	0.22	<0.001		
t_{ij}	α_2	-0.17	0.02	<0.001	-0.15	0.02	<0.001		
age0 - min(age0)	α_3	-0.13	0.01	<0.001	-0.13	0.01	<0.001		
male == 1	α_4	0.01	0.18	0.96	0.01	0.18	0.97		
(CEP == 1)* t_{ij}	α_5	0.07	0.02	<0.001	0.07	0.02	0.001		
Sağkalım Süreci									
		Bileşik Model			Genişletilmiş Cox Modeli				
Değişken	Par.	Değer	Std. Hata	p	RO	Değer	Std. Hata	p	RO
Kognitif Fonksiyon	γ	-0.29	0.06	<0.001	0.75	-0.26	0.04	<0.001	0.77
age0 - min(age0)	β_1	0.09	0.02	<0.001	1.09	0.09	0.02	<0.001	1.09
CEP == 1	β_2	0.54	0.25	0.03	1.72	0.42	0.22	0.06	1.52
male == 1	β_3	-0.16	0.21	0.45	0.85	-0.16	0.21	0.45	0.85

Reg. sabiti: Regresyon sabiti, Par.: Parametre, Değer: tahmin değeri, Std. Hata: Standart hata, p: p-değeri, RO: risk oranı (=exp(Değer)) Kognitif fonksiyon için bileşik model MMSE terimini kullanmaktayken genişletilmiş Cox modeli MMSE terimini kullanmaktadır. Rastgele etkiler, ölçüm hataları ve temel risk için sonuçlar ek dosyada sunulmuştur.

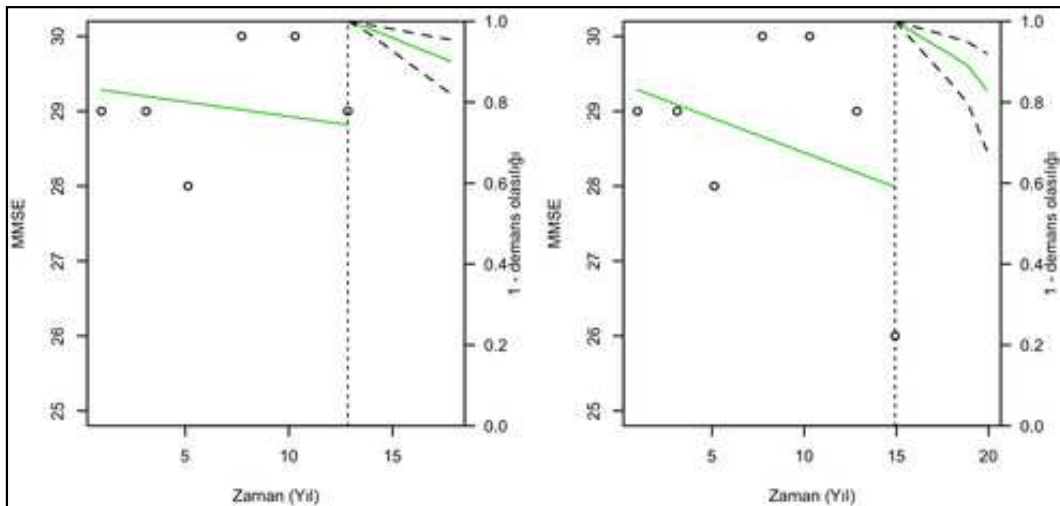
Bileşik model kognitif fonksiyondaki bir birimlik düşüşün, demans riskindeki yaklaşık %34'lük $(=(e^{0.29}-1)*100)$ artış ile ilişkilendirilebileceğini ifade etmektedir. Genişletilmiş Cox modeli bu ilişkinin yaklaşık %30 $(=(e^{0.26}-1)*100)$ olduğunu ifade etmektedir. Bu durum biyobelirteçteki ölçüm hatasının γ parametresinin 0'a doğru yanlı olarak tahmin edilmesine bir örnektir. Başlangıç yaşındaki bir yıllık artış, demans riskindeki yaklaşık %9'luk artış ile ilişkili bulunmuştur ($RO = 1.10$). β_2 , aynı başlangıç kognitif fonksiyon değerine, aynı başlangıç yaşına ve aynı cinsiyete sahip iki kişiden eğitilmiş olanın, eğitimsiz olana kıyasla yaklaşık %72 daha yüksek demans riskine sahip olduğunu belirtmektedir. Bu durum, eğitilmiş kişinin, eğitimsiz kişiyle aynı kognitif fonksiyon seviyesine sahip olmasına yol açabilecek ve demans riskini artırabilecek başka risk faktörlerine (örneğin genetik yatkınlık) sahip olabileceği şeklinde yorumlanabilir. CEP ile demans riski arasındaki doğrudan ilişki genişletilmiş Cox modelinde anlamlı değildir ($p = 0.06$). Cinsiyet ile demans riski arasındaki doğrudan ilişki anlamlı değildir ($p = 0.45$).

Kestirimler

Kestirim amacıyla, PAQUID verisinden ID değeri 61 olan kişi çıkarılıp, bileşik model tekrar kurulmuştur. Yeni sonuçlar temelinde kognitif fonksiyon değerleri ve sağkalım olasılıkları için ileriye dönük kestirimler JM paketindeki `survfitJM()` ve `predict()` fonksiyonları kullanılarak hesaplanmıştır (Şekil 5, 6). Kognitif fonksiyon için kestirim değerleri dinamik olarak, 6 ve 7 gözleme koşullu olarak hesaplanmıştır. Eklenen 7. gözlem genel eğimin daha büyük olarak kestirilmesine neden olmaktadır. Benzer olarak, sağkalım olasılıkları dinamik olarak, 6 ve 7 gözleme koşullu hesaplanmıştır. Eklenen 7. gözlem önceki 6 gözleme kıyasla daha düşük kognitif fonksiyon seviyesindedir. Bu durum daha düşük sağkalım olasılıkları hesaplanmasına neden olmaktadır.



ŞEKİL 5: Kognitif fonksiyon için 5 yıllık kestirim değerleri. Soldaki grafik 6 gözleme, sağdaki ise 7 gözleme koşullu olasılıkların hesaplandığı durumları göstermektedir. İçi boş yuvarlaklar gözlenen MMSE değerlerini, içi dolu siyah yuvarlaklar kestirim değerlerini, kırmızı çizgiler ise %95'lik güven aralıklarını göstermektedir.



ŞEKİL 6: Sağkalım olasılıkları için 5 yıllık kestirim değerleri. Soldaki grafik 6 MMSE gözlemine, sağdaki ise 7 MMSE gözlemine koşullu olasılıkların hesaplandığı durumları göstermektedir.

TARTIŞMA

Bu makalede, tekrarlı gözlem ve sağkalım verilerinin birlikte analiz edilmesine olanak sağlayan bileşik modeller üzerine bir derleme sunulmuştur. Derlemede bir sürekli biyobelirteç ve sağdan sansürlü terminal sağkalım olayı hesaba katılmıştır. Model formülasyonu için rastgele etkiler içeren ortak parametrelili formülasyon üzerinde durulmuştur. PAQUID verisi üzerine bir uygulamada, bileşik modeller ile parametre tahminleri ve ileriye dönük kestirimler sunulmuştur.

Birden çok belirteç ve sağkalım olayını hesaba katan bileşik modeller güncel olarak üzerinde çalışılmakta olan konulardır. Bu konudaki güncel bir derleme çalışması için Hickey ve diğ (2016) makalesi incelenebilir.²² Kesikli biyobelirteç için genelleştirilmiş karma etkili model içeren bileşik model için Viviani, Alfo ve Rizopoulos (2014) makalesi incelenebilir.³⁴ Soldan sansürlü ve aralık sansürlü sağkalım verilerini hesaba katan bileşik modeller için sırasıyla Crowther ve diğ (2015) ve Gueorguieva, Rosenheck ve Lin (2011) çalışmaları incelenebilir.^{35,36} Yinelemeli olay ve yarışan riskler varlığında bileşik modeller için sırasıyla Liu ve Huang (2009) ve Williamson ve diğ. (2008) makaleleri incelenebilir.^{37,38} Rastgele etkiler ve ölçüm hataları için Gaussian olmayan parametrik dağılımlar varsayımı yapan bileşik modeller için Baghfalaki, Ganjali ve Berridge (2013) çalışması, parametrik olmayan rastgele etkiler için Li, Hu ve Greene (2009) çalışması incelenebilir.^{39,40} Ortak parametrelili formülasyona alternatif olarak gizli sınıf formülasyonu kullanılabilir. Bu konuda Proust-Lima ve diğ (2014) makalesi incelenebilir.⁴¹ Tekrarlı gözlemlerin toplandığı zamanların rastgeleliliğini hesaba katan bileşik modeller için Sun, Sun ve Liu (2007) makalesi incelenebilir.⁴² Boylamsal ve sağkalım süreçleri arasındaki ilişkiyi hesaba katan parametrenin zaman içinde değişen parametre olarak tanımlandığı bileşik modeller için Song ve Wang (2008) makalesi incelenebilir.⁴³ Bileşik modellerde değişken seçimi için He, Tu, Wang, Fu ve Yu (2015) çalışması, model seçimi için Zhang ve diğ. (2014) çalışması, örneklem büyüklüğü hesaplaması için Chen, Ibrahim ve Chu (2011) çalışması incelenebilir.⁴⁴⁻⁴⁶ Regresyon tanımları için Dobson ve Henderson (2003), Rizopoulos, Verbeke ve Molenberghs (2010) makaleleri incelenebilir.^{47,48} Sağkalım süreci kestirimlerinin doğruluğunu test etmek için C-indeksi, Brier skoru ve zamana bağlı ROC eğrisi kullanılabilir. Bileşik modeller için birçok istatistiksel yazılım platformunda paketler mevcuttur: R dilinde JM, joineR, JMBayes, joineRML ve frailtypack paketleri, STATA'da stjm paketi, SAS'ta JMFit makrosu.^{33,30,51-53}

Teşekkür

Dr. Cecile Proust-Lima'ya PAQUID verisini bu çalışmada kullanmamıza izin verdiği için teşekkür ederiz.

Ek

PAUQID veri seti R paketi lcmm'de açık olarak mevcuttur. Makaledeki sonuçları tekrarlanabilir kılan R kodları için ozgurasarstat@gmail.com adresi ile iletişime geçiniz.

Çıkar Çatışması

Yazarlar herhangi bir çıkar çatışması veya finansal destek bildirmemiştir.

Yazar Katkıları

Fikir/Kavram: Özgür Asar

Tasarım: Ege Ülgen, Özgür Asar

Denetleme/Danışmanlık: Ege Ülgen, Özgür Asar

Analiz ve/veya Yorum: Ege Ülgen, Özgür Asar

Kaynak Taraması: Özgür Asar

Makalenin Yazımı: Ege Ülgen, Özgür Asar

Eleştirel İnceleme: Ege Ülgen, Özgür Asar

KAYNAKLAR

1. Cox DR. Regression models and life-tables. *J R Stat Soc Series B Stat Methodol* 1972;34(2):187-220.
2. Andersen PK, Gill RD. Cox regression model for counting processes: a large sample study. *Ann Stat* 1982;10(4):1100-20.
3. Kleinbaum DG, Klein M. *Survival Analysis: A Self-Learning Text*. 3rd ed. New York, NY: Springer; 2012. p.700.
4. Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. 2nd ed. New York, NY: John Wiley & Sons; 2002. p.462.
5. Prentice RL. Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika* 1982;69(2):331-42.
6. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error. Applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995;90(429):27-37.
7. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997;53(1):330-9.
8. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38(4):963-74.
9. Diggle PJ, Heagerty PJ, Liang KY, Zeger SL. *Analysis of Longitudinal Data*. 2nd ed. Oxford: Oxford University Press; 2002. p.379.
10. Fitzmaurice GM, Laird NM, Ware JH. *Applied Longitudinal Analysis*. 2nd ed. Hoboken, NJ: John Wiley & Sons; 2011. p.701.
11. Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics* 2000;1(4):465-80.
12. Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data: with Applications in R*. Boca Raton, FL: CRC Press; 2012. p.275.
13. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. 2nd ed. New York, NY: John Wiley & Sons; 2002. p.408.
14. Tsiatis AA, Davidian M. Joint modeling of longitudinal and time-to-event data: an overview. *Stat Sin* 2004;14(3):809-34.
15. Mccrink LM, Marshall AH, Cairns KJ. Advances in joint modelling: a review of recent developments with application to the survival of end stage renal disease patients. *Int Stat Rev* 2013;81(2):249-69.
16. Asar Ö, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. *Int J Epidemiol* 2015;44(1):334-44.
17. Lawrence Gould A, Boye ME, Crowther MJ, Ibrahim JG, Quartey G, Micallef S, et al. Joint modeling of survival and longitudinal non-survival data: current methods and issues. Report of the DIA Bayesian joint modeling working group. *Stat Med* 2015;34(14):2181-95.
18. Sudell M, Kolamunnage-Dona R, Tudur-Smith C. Joint models for longitudinal and time-to-event data: a review of reporting quality with a view to meta-analysis. *BMC Med Res Methodol* 2016;16(1):168.
19. Elashoff RM, Li G, Li N. *Joint Modelling of Longitudinal and Time-to-Event Data*. Boca Raton, FL: CRC Press; 2017. p.241.
20. Hsieh F, Tseng YK, Wang JL. Joint modeling of survival and longitudinal data: likelihood approach revisited. *Biometrics* 2006;62(4):1037-43.
21. Crowther MJ, Lambert PC. A general framework for parametric survival analysis. *Stat Med* 2014;33(30):5280-97.
22. Hickey GL, Philipson P, Jorgensen A, Kolamunnage-Dona R. Joint modelling of time-to-event and multivariate longitudinal outcomes: recent developments and issues. *BMC Med Res Methodol* 2016;16(1):117.
23. Sweeting MJ, Thompson SG. Joint modelling of longitudinal and time-to-event data with application to predicting abdominal aortic aneurysm growth and rupture. *Biom J* 2011;53(5):750-63.
24. Sweeting MJ, Barrett JK, Thompson SG, Wood AM. The use of repeated blood pressure measures for cardiovascular risk prediction: a comparison of statistical models in the ARIC study. *Stat Med* 2016. Doi: 10.1002/sim.7144.
25. Crowther MJ, Look MP, Riley RD. Multilevel mixed effects parametric survival models using adaptive Gauss-Hermite quadrature with application to recurrent events and individual participant data meta-analysis. *Stat Med* 2014;33(22):3844-58.
26. Rizopoulos D. Fast fitting of joint models for longitudinal and event time data using a pseudo-adaptive Gaussian quadrature rule. *Comput Stat Data Anal* 2012;56(3):491-501.
27. Berzuini C, Larizza C. A unified approach for modeling longitudinal and failure time data, with application in medical monitoring. *IEEE Trans Pattern Anal Mach Intell* 1996;18(2):109-23.
28. Ibrahim JG, Chen MH, Sinha D. *Bayesian Survival Analysis*. New York, NY: Springer; 2010. p.480.
29. Andrinopoulou ER, Rizopoulos D, Takkenberg JJ, Lesaffre E. Joint modeling of two longitudinal outcomes and competing risk data. *Stat Med* 2014;33(18):3167-78.
30. Rizopoulos D. The R package JMbayes for fitting joint models for longitudinal and time-to-event data using MCMC. *J Stat Softw* 2016;72(7).
31. Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011;67(3):819-29.
32. Letenneur L, Commenges D, Dartigues JF, Barberger-Gateau P. Incidence of dementia and Alzheimer's disease in elderly community residents of South-Western France. *Int J Epidemiol* 1994;23(6):1256-61.
33. Rizopoulos D. JM: an R package for the joint modelling of longitudinal and time-to-event data. *J Stat Softw* 2010;35(9):1-33.
34. Viviani S, Alfó M, Rizopoulos D. Generalized linear mixed joint model for longitudinal and survival outcomes. *Stat Comput* 2013;24(3):417-27.
35. Crowther MJ, Andersson TM, Lambert PC, Abrams KR, Humphreys K. Joint modelling of longitudinal and survival data: incorporating delayed entry and an assessment of model misspecification. *Stat Med* 2015;35(7):1193-209.
36. Gueorguieva R, Rosenheck R, Lin H. Joint modelling of longitudinal outcome and interval-censored competing risk dropout in a schizophrenia clinical trial. *J R Stat Soc Ser A Stat Soc* 2011;175(2):417-33.
37. Liu L, Huang X. Joint analysis of correlated repeated measures and recurrent events processes in the presence of death, with application to a study on acquired immune deficiency syndrome. *J R Stat Soc Ser C Appl Stat* 2009;58(1):65-81.
38. Williamson PR, Kolamunnage-Dona R, Philipson P, Marson AG. Joint modelling of longitudinal and competing risks data. *Stat Med* 2008;27(30):6426-38.
39. Baghfalaki T, Ganjali M, Berridge D. Robust joint modeling of longitudinal measurements and time to event data using normal/independent distributions: a Bayesian approach. *Biom J* 2013;55(6):844-65.
40. Li L, Hu B, Greene T. A semiparametric joint model for longitudinal and survival data with application to hemodialysis study. *Biometrics* 2009;65(3):737-45.
41. Proust-Lima C, Séne M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Stat Methods Med Res* 2014;23(1):74-90.
42. Sun J, Sun L, Liu D. Regression analysis of longitudinal data in the presence of informative observation and censoring times. *J Am Stat Assoc* 2007;102(480):1397-406.
43. Song X, Wang CY. Semiparametric approaches for joint modeling of longitudinal and survival data with time-varying coefficients. *Biometrics* 2008;64(2):557-66.
44. He Z, Tu W, Wang S, Fu H, Yu Z. Simultaneous variable selection for joint models of longitudinal and survival outcomes. *Biometrics* 2015;71(1):178-87.
45. Zhang D, Chen MH, Ibrahim JG, Boye ME, Wang P, Shen W. Assessing model fit in joint models of longitudinal and survival data with applications to cancer clinical trials. *Stat Med* 2014;33(27):4715-33.
46. Chen LM, Ibrahim JG, Chu H. Sample size and power determination in joint modeling of longitudinal and survival data. *Stat Med* 2011;30(18):2295-309.
47. Dobson A, Henderson R. Diagnostics for joint longitudinal and dropout time modeling. *Biometrics* 2003;59(4):741-51.
48. Rizopoulos D, Verbeke G, Molenberghs G. Multiple-imputation-based residuals and diagnostic plots for joint models of longitudinal and survival outcomes. *Biometrics* 2009;66(1):20-9.
49. Blanche P, Dartigues JF, Jacqmin-Gadda H. Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring. *Biom J* 2013;55(5):687-704.
50. Kolamunnage-Dona R, Williamson PR. Time-dependent efficacy of longitudinal biomarker for clinical endpoint. *Stat Methods Med Res* 2016. Doi:10.1177/0962280216673084.
51. Rondeau V, Mazroui Y, Gonzalez JR. Frailtypack: an R package for the analysis of correlated survival data with frailty models using penalized likelihood estimation or parametrical estimation. *J Stat Softw* 2012;47(4):1-28.
52. Crowther MJ, Abrams KR, Lambert PC. Joint modeling of longitudinal and survival data. *Stata J* 2013;13(1):165-84.
53. Zhang D, Chen MH, Ibrahim JG, Boye ME, Shen W. JMFit: a SAS macro for joint models of longitudinal and survival data. *J Stat Softw* 2016;71(3).