

## ELKERÜLHETŐ HALÁLOZÁS VIZSGÁLATA MAGYAR ÉS ÍR ADATOKON PCLM, P-SPLINE ÉS LEE-CARTER MODELL ALKALMAZÁSA

Csiszár Dorottya (Deloitte Zrt. Aktuáriusi és Biztosítási Megoldások csapatában aktuárius tanácsadó),  
dcsiszar@deloitte.com

### ÖSSZEFOGLALÓ

Ezen tanulmány kiindulópontjaként a Magyar Nemzeti Bank 330 pontból álló versenyképességi programjának egészségügyre vonatkozó javaslatai szolgáltak, az ennek alapján felvázolt egészségügyi rendszert az Írországból bevezetett modellhez hasonlították. A tanulmányban az ír és magyar társadalom egészségi állapotára vonatkozó mutatók vizsgálatát követően előrejelzem 50 évre a megfelelő kezeléssel elkerülhető öt legjelentősebb halálok miatti halandóság alakulását Lee-Carter, valamint P-spline modellel. Az ötéves csoportosításban megtalálható halandósági adatok évekre történő szétbontásához a PCLM (penalized composite link model) modellt alkalmazom, melynek bemutatása a gyakorlatban történő hasznosíthatósága miatt előremutató lehet. A publikáció alapjául a Budapesti Corvinus Egyetem és az Eötvös Loránd Tudományegyetem közös Biztosítási és Pénzügyi Matematika mesterszakán írt diplomamunka szolgált, melyért a Magyar Aktuárius Társaság (MAT) a szerzőnek ítélte a Biztosításmatika Ifjú Mestere Díjat.

### SUMMARY

The Hungarian National Bank published a 330-point competitiveness program which contains recommendations regarding the health system as well. The recommended health system is considered to be similar to the Irish model. I examine indicators regarding the health status of the Irish and Hungarian society, then I forecast for 50 years the mortality caused by the five most significant treatable death causes using Lee-Carter and P-splines model. I use the PCLM (penalized composite link model) for decomposition of the five year-grouped mortality data. Demonstration of the usage of this model in practice can be forward-looking and advantageous. The publication is based on the thesis written in the Actuarial and Financial Mathematics joint Master's program of Corvinus University of Budapest and Eötvös Loránd University. Based on the thesis, the Hungarian Actuarial Society (MAT) awarded the author the Young Master of Actuarial Mathematics Award.

**Kulcsszavak:** Lee-Carter modell, P-spline modell, PCLM, elkerülhető halálozás, halandóság-előrejelzés

**Keywords:** Lee-Carter model, P-splines model, PCLM, avoidable mortality, mortality forecasting

JEL: C14, G22

DOI: 10.18530/BK.2022.1-2.12

<http://dx.doi.org/10.18530/BK.2022.1-2.12>

### Bevezetés

Az egészségügy helyzete, valamint az egészségügyi rendszer javításának lehetőségei igen aktuális téma napjainkban mind a világ, mind Magyarország vonatkozásában. A Magyar Nemzeti Bank (MNB) 330 pontból álló versenyképességi programjának 39 pontja foglalkozik az egészséges magyar társadalommal, és kínál megoldási javaslatokat a megfogalmazott problémákra (MNB, 2019). Kiss és Tókey (2019) véleménye szerint az MNB célkitűzései az Írországból működő modellre hasonlítanak, ahol a lakosságnak nagyjából a fele rendelkezik magánbiztosítással. Felmerülhet a kérdés, hogy az ír, valamint ahhoz képest a magyar egészségügyi rendszer mennyire hatékony? Endrédi és Constantinovits (2016) 25 mutató mentén összehasonlította Magyarországot és Írországot egészségügyi rendszerét, és azt találta, hogy mindössze 8 mutató tekintetében teljesített jobban hazánk Írországnál. Az egészségügyi rendszer hatékonyságának egyik indikátora lehet az elkerülhető halálozások alakulása (Allin és Grignon, 2014). Elkerülhető halálozáson a megelőzhető halálozás és a megfelelő kezeléssel elkerülhető halálozás összességét értjük (OECD, 2020). A megelőzhető halálozás olyan halálozás, amelyet hatékony népegészségügyi és primer prevenciók beavatkozásokkal (a betegségek diagnosztizálását megelőzően, az esetek előfordulásának mérséklésével) el lehetett volna kerülni. A megfelelő kezeléssel elkerülhető halálozás olyan halálozás, amely időben meghozott és hatékony egészségügyi, szekunder prevenciók beavatkozásokkal és kezeléssel (a betegségek diagnosztizálását követően, a következmények súlyosságának mérséklésével) elkerülhető lett volna (OECD 2019b).

AZ OECD által kiadott országprofilok (OECD, 2019a; OECD, 2019b) tartalmazzák mind Írországra, mind Magyarországra vonatkozóan a fő okait az elkerülhető halálozásnak, mely országprofilok alapján Írország jobban, Magyarország azonban rosszabbul szerepel az EU átlagánál.

Ezen fő, megfelelő kezeléssel elkerülhető halálozásokra készíték előrejelzést egyrészt a mortalitást előrejelző egyik legismertebb és legjobb modell, a Lee-Carter modell, másrészt a hazai szakirodalomban eddig még kisebb figyelemben részesített P-spline mortalitási modell segítségével.

Az adatbázisokban az egyes halálokok szerinti halálozási adatok ötéves csoportosításban, korcsoportokra vonatkozóan található meg. Simításra, a korcsoportonkénti halálozási ráták egyes életkorok szerinti mortalitási rátákra történő szétbontásához a PCLM modellt alkalmazom. A PCLM, mely egy P-spline interpoláció, a GLM (generalized

linear model) modellnek egy büntető likelihood módszerrel illesztett kiterjesztése, mely P-spline modell esetében a jövőbeli értékeket hiányzó értéknek tekintve, az előrejelzés egy természetes következménye a simításnak (Currie et al., 2004). Mivel nem ritkák azon esetek, amikor az adatok csoportba rendezve érhetőek el, úgy gondolom, hogy a PCLM modell mélyrehatóbb vizsgálata és egy konkrét alkalmazási lehetőségének áttekintése fontos és előremutató lehet.

### A magyar és ír társadalom egészségi állapota

Az 1. táblázat tartalmazza azt a 2013-as adatok alapján meghatározott 25 mutatót, amelyek mentén Endrédi és Constantinovits (2016) összehasonlította a két társadalom egészségi állapotát. Annak érdekében, hogy naprakészebb eredményeket lássunk, a légszennyezettségnek kitett lakosság, valamint a későbbiekben bizonyos halálokok tekintetében részletesebben vizsgált halálozások számának kivételével, Magyarországra és Írországra vonatkozóan megvizsgáltam ugyanezen mutatók 2017-es értékeit, melyeket szintén az 1. táblázat foglal össze. Az 1. táblázatban szürke háttérrel vannak feltüntetve a két ország közötti különbségek azon mutatók esetén, ahol Magyarország számára kedvezőbb eredmény adódott.

Endrédi és Constantinovits (2016) megállapítják, hogy a 25 mutatóból mindössze 8 esetben sikerült hazánknak jobb eredményt elérnie. Ugyanakkor véleményem szerint érdekes megfigyelés az, hogy MRI- és CT-gépek tekintetében Írország jobban ellátott, mégis Magyarországon egy lakosra jóval több MRI- és CT-vizsgálat jut. Valamint annak ellenére, hogy hazánkban az egy főre jutó orvosok száma 2013-ban csak 0,4-gyel volt magasabb, mégis jóval több a konzultáció orvos és páciens között. A 2017-es és 2018-as adatokat tekintve, OECD (2018) alapján Írország az MRI- és CT-vizsgálatok számát nem jelentette, valamint az átlagos várakozási időt illetően sem volt feltüntetve adat, így Írországra vonatkozóan ezen cellákban nem szerepel érték. Azt figyelhetjük meg, hogy Magyarország 2017–2018-ban ugyanazon mutatók esetén ért el jobb eredményt, mint 2013-ban, ugyanakkor a különbség a két ország között az egy főre jutó éves átlagos alkoholfogyasztás kivételével csökkent – például Írországból jelentősen növekedett az egy főre jutó orvosok száma. Fontosnak tartom továbbá megemlíteni, hogy az Írországból dolgozó orvosok 42,3%-a külföldön tanult (OECD, 2019c). Ugyanakkor néhány olyan mutató esetében, ahol Írország szerepelt jobban, sikerült csökkenteni a lemaradást, melyre példa az egészségben eltöltött évek száma. A napi zöldségfogyasztás esetében is kisebb lett az eltérés, ez azonban annak tulajdonítható, hogy Írországból nagyobb mértékben csökkent a napi zöldségfogyasztás, mint Magyarországon.

Endrédi és Constantinovits (2016) az egészségügy teljesítményének megítélése kapcsán hozta fel ezen mutatókat. A következőkben járjuk körül jobban, hogyan lehet mérni az egészségügyi rendszer teljesítményét.

### Elkerülhető halálozás mint az egészségügy hatékonyságának indikátora

Az egészségügy hatékonyságának mérése rendkívül fontos, ugyanakkor nagyon összetett feladat. Cylus et al. (2016) alapján az egészségügy hatékonyságának mérése során azt vizsgálják, hogy az egészségügyi rendszer inputjait (mint a kiadások és egyéb források) milyen mértékben használják fel, hogy azok a legjobb hatást érik el a rendszer outputjain és/vagy az egészségügyi rendszer céljain. Input minden olyan erőforrás, amit az egészségügy outputjainak az előállításánál felhasználnak, és melyek lehetnek pénzügyi vagy fizikai erőforrások (pl. tőke, munkaerő, gyógyszerek) vagy egészségügyi tevékenységek (pl. diagnosztikai vizsgálatok vagy műtéti beavatkozások). Az outputok tevékenységek, melyeket az inputok együttes felhasználásával állítottak elő, lehetnek például műtéti beavatkozások (melyeket tőke, munkaerő és egyéb erőforrások által hajtottak végre) vagy egészségügyi tevékenységek egy csoportja.

Hollingsworth (2016) alapján a data envelopment analysis (DEA) és a stochastic frontier analysis (SFA) a két leggyakrabban alkalmazott módszer az egészségügyi szolgáltatások hatékonyságának mérésére. A DEA lineáris programozási modellt alkalmaz, melynek keretében súlyokat rendel az inputokhoz és outputokhoz. Az SFA-t egyelőre jóval kevesebbszer alkalmazták egészségügyi hatékonyság mérésére, mint a DEA-t, ugyanakkor az erről szóló publikációk száma növekszik. Ez utóbbi módszer regresszióanalízist alkalmaz, melynek során egy adott kórház hatékonyságának egy kalkulált hatékonysági szinttől vett eltérését nézi.

Radojicic et al. (2020) 44 darab, 1998–2018 között íródott, DEA módszert alkalmazó tanulmányt tekintett át, és mindegyik esetében megvizsgálta a használt inputokat, outputokat és a környezeti indikátorokat. Azt találta, hogy az öt indikátor, mely a legtöbb tanulmányban előfordult, az inputok esetén az egészségügyi kiadás, orvosok száma, ápolók száma, ágyak száma és a GDP (melyeket Endrédi és Constantinovits (2016) alapján mi is vizsgáltunk), az outputok esetén a várható élettartam, csecsemőhalandóság, öt éves kor alatti halálozási ráta, az egészségkárosodással korrigált várható élettartam és az elkerülhető halálozás. A leggyakoribb környezeti inputok pedig az oktatás, GDP, táplálkozás/elhízottság, dohányzás, valamint a 65 éves kor feletti lakosság voltak. Tehát az egészségügyi hatékonyság méréséhez ezen indikátorokat használták a legtöbb alkalommal. Cylus et al. (2016) arról ír, hogy megfelelően alkalmazva a hatékonysági indikátorok fontos segítséget nyújthatnak a döntéshozóknak azzal kapcsolatban, hogy az erőforrások allokációja optimális volt-e, illetve rámutathatnak az egészségügyi rendszer azon részeire, amik nem úgy teljesítenek, ahogyan az elvárt lenne.

1. táblázat: Magyarország és Írország összehasonlítása makrogazdasági adatok alapján

Mutatók	Endrédi és Constantinovits (2013-as adatok)			Saját összeállítás (2017-es és *2018-as adatok)		
	Magyarország	Írország	Magyarország mutatói Írországhoz képest	Magyarország	Írország	Magyarország mutatói Írországhoz képest
Születéskor várható élettartam (év)	75,3	80,9	-5,6	76	82,2	-6,2
Egészségben eltöltött évek száma (év)	59,8	67,2	-7,4	65,7	70,4	-4,7
Elhalálozások kardiovaszkuláris megbetegedések miatt (100.000 lakosra)	639	373	266	-	-	-
Stroke okozta halálesetek (100.000 lakosra)	336	154	182	-	-	-
Rákos megbetegedések okozta halálesetek (100.000 lakosra)	767	629	138	-	-	-
Gyermekhalandóság (1000 élve születésre)	4,9	3,5	1,4	-	-	-
Dohányzás a felnőtt lakosság körében	26,5%	24%	2,5%	26%	17%	9%
Alkoholfogyasztás (liter/év)	11,4	11,6	-0,2	11,35*	12,88*	-1,53*
Napi gyümölcsfogyasztás a felnőtt lakosság körében	68,3%	73,8%	-5,5%	59,1%	73%	-13,9%
Napi zöldségfogyasztás a felnőtt lakosság körében	49%	95,4%	-46,4%	46,3%	63%	-16,7%
Elhízott lakosság aránya	28,5%	23%	5,5%	20%	18%	2%
Légszennyezettségnek kitett lakosság	33%	18%	15%	-	-	-

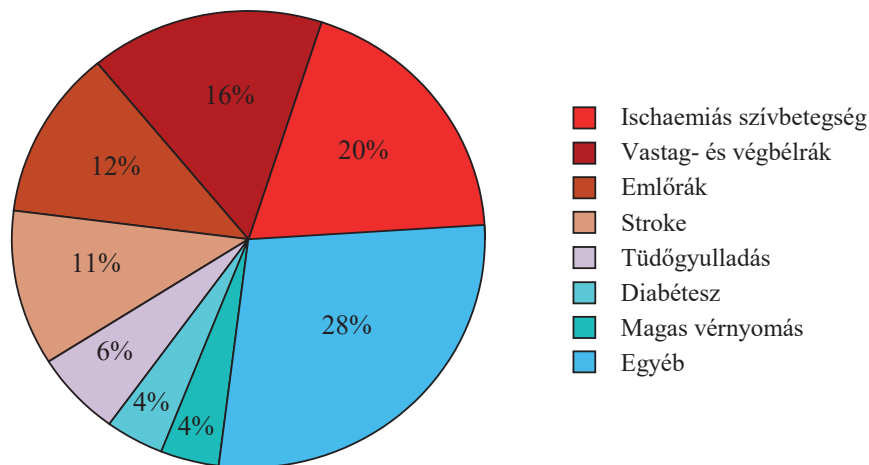
Egy főre jutó orvosok száma	3,1	2,7	0,4	3,3	3,1	0,2
Egy főre jutó éves orvosi konzultációk száma	11,8	3,8	8	10,9	5,8	5,1
Egy évben egy orvosra jutó konzultációk száma	3.820	1.224	2.596	3.278	1.976	1.302
Ápoló személyzet létszáma (1.000 lakosra)	6,3	12,8	-6,5	6,5	12,2	-5,7
MRI-gépek száma (1 millió lakosra)	2,8	12,4	-9,6	5	15	-10
CT-gépek száma (1 millió lakosra)	7,7	16,8	-9,1	9	19	-10
MRI-vizsgálatok (1.000 lakosra)	34,1	17,7	16,4	44	-	-
CT-vizsgálatok (1.000 lakosra)	86,2	71,3	14,9	123	-	-
Kórházi ágyak száma (1.000 lakosra)	7	2,8	4,2	7	3	4
Kórházban töltött átlagos idő (nap/fő)	9,5	6,2	3,3	9,6	6,1	3,5
Átlagos várakozási idő csípőműtétre (nap)	150	102	48	104		
GDP-arányos egészségügyi kiadások	8%	8,9%	-0,9%	6,6%*	7,0%*	-0,4%*
Egy főre eső egészségügyi kiadás (euró)	940	3.202	-2.262	2.047*	4.915*	-2.868*

Forrás: Endrédi és Constantinovits (2016), valamint saját összeállítás OECD (2019a), OECD (2019b), OECD (2019c), Worldbank (2021a) és Worldbank (2021b) alapján

Rutstein et al. (1975) már 1975-ben felvetette, hogy az elkerülhető halálózásra lehetne úgy tekinteni, mint az egészségügyi ellátás mutatószámára. Allin és Grignon (2014), valamint Vergara-Duarte et al. (2018) szerint az elkerülhető halálozás megfelelő indikátora az egészségügyi rendszer hatékonyságának, emellett Hone et al. (2017) is ezt az indikátort használta az egészségügyi szolgáltatások hatékonysága kapcsán.

A következőkben a megfelelő kezeléssel elkerülhető öt legjelentősebb halálokozást vizsgálom Írországra és Magyarországra vonatkozóan, és a továbbiakban elkerülhető halálozáson, illetve elkerülhető halálok alatt a megfelelő kezeléssel elkerülhető halálozást, illetve halálokozást értem. Az 1. ábrán azt láthatjuk, hogy az Európai Unióban az összes elkerülhető halálozásnak mekkora hányadát tették ki az egyes elkerülhető halálokozások miatti halálozások.

1. ábra: Fő elkerülhető halálokok az Európai Unióban, 2017



Forrás: OECD (2020)

## A halandósági modellezés szakirodalmi háttere

Vékás (2019) összefoglalta, hogy melyek a leginkább elterjedt halandósági modellek, azonban először szeretném bevezetni a halandósági ráta fogalmát, amely egy választott időszak és populáció vonatkozásában az adott időszak során az adott populációban bekövetkezett halálozások számának a populáció létszámahoz viszonyított arányaként határozható meg. Képlettel felírva:

$$m = \frac{D}{E}, \quad (1)$$

ahol  $m$  a halandósági ráta,  $D \in \mathbb{N}$  a vizsgált időszakban elhunytak száma és  $E > 0$  a vizsgált populáció létszáma, mely utóbbi alatt érthető az időszak kezdetén élő egyének létszáma (kezdeti kitétség, jele:  $E^0$ ) vagy az időszak alatt élő egyének átlagos létszáma (központi kitétség, jele:  $E^c$ ).

A halandóság-előrejelzés területén a Lee-Carter modell hatalmas áttörésnek számított, ugyanis egy életkortól és naptári időszaktól függő paraméterekkel rendelkező, viszonylag egyszerű log-bilineáris modell illesztésével meglepően jól írta le az 1900–1989. évi életkorfüggő halandósági ráták alakulását az Egyesült Államokban, és mely modell Deaton és Paxson (2001) szerint az ezredfordulóra a világ vezető halandóság-előrejelző módszere lett (Vékás, 2019). Ez a múltbeli trendek meghosszabbítására épülő, a változások háttérben álló mögöttes okokat figyelmen kívül hagyó extrapolatív eljárás.

A Lee-Carter modellnek több kiterjesztése is létezik. Az egyik ilyen a Poisson Lee-Carter modell, amely az eredeti modellben lévő normális eloszlású hibtagok alkalmazása helyett az egyes korcsoport-naptári év kombinációkhoz tartozó halálozások számainak Poisson-eloszlását feltételezi. Gogola és Vékás (2020) a modell ezen változatát alkalmazzák kohorsz halandósági táblák készítésére, Brouhns et al. (2002) alapján feltételezve, hogy a halálozások számai Poisson-eloszlású valószínűségi változók.

A halandóság-előrejelző modellek egyik továbbfejlesztett változata egységesíti az életkorfüggő halandóság simítását és a halandóság előrejelzését, melyhez p-spline interpolációs függvényeket alkalmaz (Currie et al., 2004). Simításra vagy kiegyenlítésre (angolul smoothing vagy graduation) azért van szükség, mert a becsült koréves halálozási valószínűségek sorozata szabálytalan, véletlen ingadozást tartalmaz, amelyet célszerű kiszűrni az adatokból. Ez történhet félpáraméteres, páraméteres (halandósági törvények) vagy nempáraméteres (pl. p-spline interpoláció) eljárásokkal (Vékás, 2019).

### Poisson Lee-Carter modell

Mivel a Poisson Lee-Carter modell az eredeti Lee-Carter modellnek egy továbbfejlesztett változata, először tekintsük az eredeti modellt. A Lee-Carter modell feltételezi, hogy minden  $x \in 1, 2, \dots, X$  korcsoportra és  $t \in 1, 2, \dots, T$  időszakra ismertek egy populáció  $m_{xt} > 0$  korcsoport- és időszakfüggő központi halandósági rátái, amelyek alakulását a következő egyenlettel írhatjuk fel:

$$(2) \quad \ln m_{xt} = a_x + b_x k_t + \varepsilon_{xt} \quad (x=1, 2, \dots, X, \quad t=1, 2, \dots, T),$$

ahol az  $X \geq 2$  és  $T \geq 2$  egész számok a vizsgált korcsoportok és időszakok számai. A korcsoportokat jellemzően korévenként, az időszakokat pedig többnyire naptári évenként szokták megadni (Vékás, 2019).

Az  $a_x$ ,  $b_x$  és  $k_t$  értékek a halandóság alakulásának életkortól, illetve időszaktól függő paraméterei, míg  $\varepsilon_{xt}$  a modell véletlen hibáit jelöli, mely utóbbiról általában felteszik, hogy független, 0 várható értékű, azonos  $\sigma^2 > 0$  varianciájú, normális eloszlású valószínűségi változók. Ekkor teljesül a következő összefüggés:

$$(3) \quad \ln m_{xt} \sim \mathcal{N}(a_x + b_x k_t, \sigma^2) \quad (x=1, 2, \dots, X, \quad t=1, 2, \dots, T).$$

Vékás (2019) alapján a modell paramétereinek értelmezése a következő:

$a_x$  – átlagos logaritmikus halandóság ( $x=1, 2, \dots, X$ ): Ezen paraméterek a halandóság életkor szerinti tipikus alakulását jelenítik meg.

$k_t$  – mortalitási index ( $t=1, 2, \dots, T$ ): A halandóság időbeli változását testesíti meg a modellben. Általában csökkenő sorozat, de gyakran tartalmaz növekvő szakaszokat is.

$b_x$  – életkorfüggő érzékenység ( $x=1, 2, \dots, X$ ): Ezen együtthatók a logaritmikus halandósági ráták érzékenységét adják meg a mortalitási index változására nézve:

$$(4) \quad \mathbb{E}(\ln m_{x,t+1} - \ln m_{xt}) = b_x (k_{t+1} - k_t) \quad (t=1, 2, \dots, T-1)$$

A paraméterek becslése maximum likelihood módszerrel történik. A modell illesztése a legkisebb négyzetek módszerével (angolul ordinary least squares, OLS) valósul meg, ahol független változó a megfigyelt logaritmikus halandóság (Currie et al., 2004).

A Poisson Lee-Carter modell az általánosított korcsoport-időszak-kohorsz (angolul Generalized Age-Period-Cohort, GAPC) modellcsalád tagja. A GAPC modellcsalád alkalmazása feltételezi, hogy minden egyes  $x \in 1, 2, \dots, X$  korcsoportra és  $t \in 1, 2, \dots, T$  időszakra ismert a bekövetkezett  $D_{xt} \in \mathbb{N}$  halálesetek száma, és az  $E_{xt}^c > 0$  központi vagy az  $E_{xt}^0 > 0$  kezdeti kitétségi értéke. A korcsoport- és időszakspecifikus halálozások  $D_{xt}$  számait a modell a (független)  $\tilde{D}_{xt}$  valószínűségi változók megvalósult értékeinek tekinti,

melyek peremeloszlására vonatkozó feltevés (a kitétségi adatok függvényében) Poisson- vagy binomiális eloszlás.

A modellben a kezdeti vagy központi halandósági ráták becslőegyenlete a következő:

$$g(m_{xt}) = \eta_{xt} \quad (x=1, 2, \dots, X, \quad t=1, 2, \dots, T), \quad (5)$$

ahol  $\eta_{xt}$  a modell szisztematikus komponense,  $g: \mathbb{R}_{>0} \rightarrow \mathbb{R}$  folytonosan differenciálható, szigorúan monoton növekvő függvény (kapocsfüggvény, angolul link function), mely esetében az általánosított lineáris modell szakirodalmára építve központi kitétségek és Poisson-eloszlás alkalmazása esetén a  $g(y) = \ln y$  ( $y > 0$ ) logaritmikus, míg kezdeti kitétségek és binomiális eloszlás esetén  $g(y) = \ln \frac{y}{1-y}$  ( $0 < y < 1$ ) logit kapocsfüggvény használata javasolt.

A Poisson Lee-Carter modellben  $\tilde{D}_{xt}$  valószínűségi változók megvalósult értékeinek peremeloszlása tehát Poisson-eloszlású, mely feltételezés D'Amato et al. (2011) alapján helytálló:

$$\tilde{D}_{xt} \sim \text{Poisson}(E_{xt}^c, m_{xt}^c) \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T). \quad (6)$$

Ezen modellkeretben a halandóságot leíró szisztematikus komponens a Lee-Carter-modell alapegyenletének hibatagok nélküli változata, illetve központi kitétségeket és logaritmikus kapocsfüggvényt feltételez a modell:

$$\eta_{xt} = \ln m_{xt}^c = a_x + b_x k_t \quad (x = 1, 2, \dots, X, \quad t = 1, 2, \dots, T). \quad (7)$$

A halandósági ráták becslési értékek körüli véletlen ingadozását a halálesetek számát leíró  $\tilde{D}_{xt}$  valószínűségi változók varianciája ragadja meg, ez indokolja a normális eloszlású  $\varepsilon_{xt}$  hibatagok hiányát. Ugyanakkor érvényesek továbbra is a következő paramétermegkötések:

$$\sum_{x=1}^X b_x = 1$$

$$\sum_{t=1}^T k_t = 0. \quad (8)$$

### P-spline és PCLM modell

Currie et al. (2004) a P-spline módszert munkájukban kétdimenziós halandósági tábla simítására és előrejelzésére alkalmazták. Ezen módszer esetén az előrejelzés a simító eljárás természetes következménye. Eilers és Marx (1996) egydimenziós regressziós

bázisként B-spline-okat alkalmazott, és Currie et al. (2004) ezt terjesztették ki úgy, hogy B-spline-okat használtak többdimenziós regresszió építéséhez, mint bázist. A regressziós megközelítés büntető likelihooddal (angolul penalized likelihood) illesztett általánosított lineáris modellhez (angolul generalized linear model, GLM) vezet. A P-spline módszer tömör összefoglalása a következő: 1) B-spline-ok használata a regresszió bázisaként és 2) a log-likelihood megváltoztatása regressziós koefficiensek eltérés-büntetésével. Ebben a büntető általánosított lineáris modellben (angolul penalized generalized linear model, PGLM) meghatározunk egy regressziós mátrixot, egy hibaeloszlást, egy link függvényt (GLM) és egy büntető mátrixot.

#### Simítás egydimenzióban

A következőkben a P-spline-okra vonatkozó ismereteket Currie et al. (2004) tanulmánya alapján mutatom be.

Tegyük fel, hogy ismertek a kárszámok (halálesetek) és a megélt életévek száma (kítetttség), melyek mátrixba rendezve az  $\mathbf{Y}$  és  $\mathbf{E}$  mátrixokban található, melyek sorai az életkorokkal, oszlopai a vizsgált évekkel indexelve. Bevezetjük az  $\mathbf{R}$  nyers házárdáták mátrixát, ahol  $r_{ij} = y_{ij}/e_{ij}$ . Tegyük fel, hogy a következő adatokkal rendelkezünk:  $(y_i, e_i, x_i), i=1, \dots, n$ , ahol  $y_i$  az  $x_i$  évi halálozások száma,  $e_i$  pedig a kockázati kítetttség.

Legyen  $\mathbf{y}' = (y_1, \dots, y_n)$ ,  $\mathbf{e}' = (e_1, \dots, e_n)$  és így tovább. Feltesszük, hogy az  $y_i$  halálozások száma Poisson-eloszlású  $\mu_i = e_i \theta_i$  várható értékkel. A nyers halandósági ráta  $\hat{\theta}_i = y_i/e_i$ , mely kapcsán keressük a  $\theta = \theta_i$  simított becslését. Klasszikus megközelítésként GLM modellt lehetne illeszteni, négyzetes regresszióval,  $\log \mu = \log e + \log \theta$ , ahol  $\log e$  a regressziós kifejezésben szereplő eltolás. Ebben a modellben  $\{1, x, x^2\}$  a bázisfüggvények. Azonban rugalmasabb bázist kapunk, ha B-spline-okat  $\{B_1(X), \dots, B_K(X)\}$  alkalmazunk. Minden B-spline harmadfokú polinomiális szakaszokból áll, melyek összeillesztési pontjait csomóknak hívjuk. Továbbra is a hagyományos regressziós modellkeretben vagyunk, ahol  $\mathbf{B}$  regressziós mátrix sorai a B-spline-ok értékei az egyes évek bázisán értékelve.

Az illesztést követően a regressziós koefficiensek ( $\hat{a}_i$ ) változékonyság viselkedése miatt egy eltérés-büntetés van beépítve a szomszédos koefficiensekhez:

$$(9) \quad (a_1 - 2a_2 + a_3)^2 + \dots + (a_{K-2} - 2a_{K-1} + a_K)^2 = \mathbf{a}' \mathbf{D}' \mathbf{D} \mathbf{a},$$

ahol  $\mathbf{D}$  egy másodrendű eltérés mátrix (angolul difference matrix), ami egy másodrendű büntetést definiál, azonban lineáris vagy harmadrendű függvények is lehetségesek. A büntető függvényt (angolul penalty function) beépítve a log-likelihood függvénybe kapjuk a büntető log-likelihood-ot (angolul penalized log likelihood):

$$l_p = l(\mathbf{a}; \mathbf{y}) - \frac{1}{2} \mathbf{a}' \mathbf{P} \mathbf{a} \quad (10)$$

$$l_p = l(\mathbf{a}; \mathbf{y}) - \frac{1}{2} \lambda \mathbf{a}' \mathbf{D}' \mathbf{D} \mathbf{a} \quad (11)$$

ahol  $l(\mathbf{a}; \mathbf{y})$  egy GLM esetén megszokott log-likelihood,  $\mathbf{P} = \lambda \mathbf{D}' \mathbf{D}$  a büntető mátrix és  $\lambda$  a simító paraméter.

A (10) függvény maximalizálása a következő büntető likelihood egyenleteket adja:

$$\mathbf{B}' (\mathbf{y} - \boldsymbol{\mu}) = \mathbf{P} \mathbf{a}, \quad (12)$$

amit a pontozó algoritmus büntető változatával is meg lehet oldani:

$$(\mathbf{B}' \tilde{\mathbf{W}} \mathbf{B} + \mathbf{P}) \hat{\mathbf{a}} = \mathbf{B}' \tilde{\mathbf{W}} \mathbf{B} \tilde{\mathbf{a}} + \mathbf{B}' (\mathbf{y} - \tilde{\boldsymbol{\mu}}), \quad (13)$$

ahol  $\mathbf{B}$  a regressziós mátrix,  $\mathbf{P}$  a büntető mátrix,  $\tilde{\mathbf{a}}, \tilde{\boldsymbol{\mu}}$  és  $\tilde{\mathbf{W}}$  a súlyok diagonálmátrixai, melyek a jelenlegi becsléseket jelölik, míg a  $\hat{\mathbf{a}}$  az  $\mathbf{a}$  új becslése.

Az algoritmust a következőképpen is fel lehet írni:

$$(\mathbf{B}' \tilde{\mathbf{W}} \mathbf{B} + \mathbf{P}) \hat{\mathbf{a}} = \mathbf{B}' \tilde{\mathbf{W}} \tilde{\mathbf{z}} \quad (14)$$

ahol  $\tilde{\mathbf{z}} = \tilde{\boldsymbol{\eta}} + \tilde{\mathbf{W}}^{-1} (\mathbf{y} - \tilde{\boldsymbol{\mu}})$ . A (14) egyenletből következik, hogy a  $\mathbf{H}$  mátrix, melyet a következőképpen adunk meg:

$$\mathbf{H} = \mathbf{B} (\mathbf{B}' \tilde{\mathbf{W}} \mathbf{B} + \mathbf{P})^{-1} \mathbf{B}' \tilde{\mathbf{W}} \quad (15)$$

és a  $\mathbf{H}$  mátrix nyoma,  $tr(\mathbf{H})$ , a mérőszáma a modell effektív dimenziójának vagy szabadságfokának.  $\mathbf{B} \hat{\mathbf{a}}$  hozzávetőleges szórásnégyzete:

$$Var(\mathbf{B} \hat{\mathbf{a}}) \approx \mathbf{B} (\mathbf{B}' \tilde{\mathbf{W}} \mathbf{B} + \mathbf{P})^{-1} \mathbf{B}' \quad (16)$$

A P-spline-okkal kapcsolatban többféle döntést kell meghoznunk: 1) mennyi legyen a csomók száma, hanyadfokú legyen a P-spline és a büntetés rendje, valamint 2) a simító paraméterrel kapcsolatban is határozunk kell. Az 1) pontban említett paramétereket P-spline paramétereknek nevezzük, és  $ndx$ ,  $bdeg$  és  $pord$  jelölésekkel látjuk el. P-spline-okkal való előrejelzéskor a becsült értékek nagyban függenek a büntetés rendjétől, míg a másik két P-spline paraméter kevésbé fontos, mert különböző választások hasonló simított függvényekhez vezetnek. A következő hüvelykujjszabály gyakran megfelelő: egyenletes eloszlású adatok esetén négy vagy öt megfigyelésenként egy csomó legyen, maximum 40 csomóig, használjunk harmadfokú spline-okat ( $bdeg=3$ ) és másodrendű

büntetést ( $pord=2$ ). Currie et al. (2004) a simító paraméter megválasztására a Schwarz-féle bayesi információs kritériumot (angolul Bayesian information criterion, BIC) használja:

$$(17) \quad BIC=2 \text{ Dev}+\log n \text{ Tr}$$

ahol  $Dev$  egy GLM-ben a deviancia,  $Tr=tr(\mathbf{H})$  az illesztett modell effektív dimenziója,  $n$  a megfigyelések száma. Akaike-féle információs kritérium (angolul Akaike information criterion, AIC) vagy általánosított keresztvalidáció (angolul generalised cross-validation, GCV) alapján is lehetne dönteni, azonban utóbbi két módszer hajlamos az alulillesztésre.

#### Kétdimenziós regressziós mátrix

Kétdimenziós regressziós feladat esetén a regressziós változóink legyenek  $x_1$  és  $x_2$ , ahol az  $\mathbf{Y}$  mátrix adatai rendelkezésre állnak egy  $m \times n$ -es rácson, és a sorai  $x_1$ -gyel (életkor), az oszlopai  $x_2$ -vel (naptári évek) vannak indexelve. Feltesszük, hogy az adatok oszlop szerinti sorrendbe vannak rendezve:  $\mathbf{y}=\text{vec}(\mathbf{Y})$ . Tekintsünk egy, a következő lineáris prediktorral meghatározott egyszerű polinomiális modellt:  $1 + x_1 + x_2 + x_2^2 + x_1x_2 + x_1x_2^2$ , mely modell regressziós mátrixa  $\mathbf{X}$ . Most tekintsünk két egydimenziós modellt, az egyik  $x_1$ -ben az  $\mathbf{y}_c$  oszlopa  $\mathbf{Y}$ -nak, a másik  $x_2$ -ben az  $\mathbf{y}_r$  sora az  $\mathbf{Y}$ -nak. Legyen  $\mathbf{X}_1$  az  $1+x_1$  lineáris prediktorú modell regressziós mátrixa,  $\mathbf{X}_2$  az  $1 + x_2 + x_2^2$  lineáris prediktorú modell regressziós mátrixa. Ekkor az  $1 + x_1 + x_2 + x_2^2 + x_1x_2 + x_1x_2^2$  formula a  $(1 + x_2 + x_2^2)(1 + x_1)$  tényezőkre bomlik, ebből következően

$$(18) \quad \mathbf{X}=\mathbf{X}_2 \otimes \mathbf{X}_1$$

ahol  $\otimes$  két mátrix Kronecker-szorzatát jelöli. A (18) egyenlet alapján készítjük el kétdimenziós modellünk regressziós mátrixát. Legyen  $\mathbf{B}_a=\mathbf{B}(\mathbf{x}_a)$ ,  $n_a \times c_a$ , az  $\mathbf{x}_a$  életkor magyarázó változón alapuló B-spline-ok regressziós mátrixa, ahol  $n_a$  a vizsgált életkorok száma,  $c_a$  pedig a  $\mathbf{B}_a$ -hoz tartozó csomók száma. Hasonlóan, legyen  $\mathbf{B}_y=\mathbf{B}(\mathbf{x}_y)$ ,  $n_y \times c_y$ , az  $\mathbf{x}_y$  naptári évek magyarázó változón alapuló B-spline-ok regressziós mátrixa, ahol  $n_y$  a vizsgált naptári évek száma,  $c_y$  a  $\mathbf{B}_y$ -hoz tartozó csomók száma. Ekkor a kétdimenziós modellünk regressziós mátrixa:

$$(19) \quad \mathbf{B}=\mathbf{B}_y \otimes \mathbf{B}_a$$

#### Kétdimenziós büntető mátrix

A (19) egyenlettel meghatározott  $\mathbf{B}$  regressziós mátrixhoz tartozik egy  $c_a c_y$  hosszúságú  $\mathbf{a}$  regressziós koefficiens vektor, melynek elemeit egy  $c_a \times c_y$  méretű  $\mathbf{A}$  mátrixba rendezzük, ahol  $\mathbf{a}=\text{vec}(\mathbf{A})$  és az  $\mathbf{A}$  mátrix oszlopai és sorai a következők:

$$\mathbf{A} = \left( \mathbf{a}_1, \dots, \mathbf{a}_{c_y} \right) \quad \mathbf{A}' = \left( \mathbf{a}_1^r, \dots, \mathbf{a}_{c_a}^r \right). \quad (20)$$

A Kronecker-szorzat definíciójából adódik, hogy az  $\mathbf{Y}$  mátrix  $j$ . oszlopához tartozó lineáris prediktor a  $\sum_{k=1}^{c_y} b_{jk}^y \mathbf{B}_a \mathbf{a}_k$  formában írható fel, ahol  $\mathbf{B}_y = (b_{ij}^y)$ . Tehát  $\mathbf{Y}$  mátrix oszlopainak lineáris prediktorai a  $c_y$ -ok lineáris kombinációjaként írhatók fel, mely az életkorok szerint simít. Ezért  $\mathbf{A}$  minden  $c_y$  oszlopára pontatlansági büntetést célszerű alkalmazni. A (21) egyenlet bal oldalát a megfelelő büntetés adja, ahol  $\mathbf{D}_a$  eltérés mátrix az  $\mathbf{A}$  mátrix oszlopaire vonatkozóan, míg az egyenlet jobb oldala a Kronecker-szorzat definíciójából következik:

$$\sum_{j=1}^{c_y} \mathbf{a}_j' \mathbf{D}_a' \mathbf{D}_a \mathbf{a}_j = \mathbf{a}' \left( \mathbf{I}_{c_y} \otimes \mathbf{D}_a' \mathbf{D}_a \right) \mathbf{a}. \quad (21)$$

A gondolatmenet hasonlóan belátható  $\mathbf{A}$  mátrix soraira is, figyelembe véve az  $\mathbf{Y}$  mátrix  $i$ . sorához tartozó lineáris prediktort:

$$\sum_{i=1}^{c_a} \mathbf{a}_i^r' \mathbf{D}_y' \mathbf{D}_y \mathbf{a}_i^r = \mathbf{a}' \left( \mathbf{D}_y' \mathbf{D}_y \otimes \mathbf{I}_{c_a} \right) \mathbf{a}, \quad (22)$$

ahol  $\mathbf{D}_y$  eltérés mátrix az  $\mathbf{A}$  mátrix soraira vonatkozóan. Az  $\mathbf{a}$  regressziós koefficiens becslése a (10) egyenletben meghatározott büntető log-likelihood maximalizálásával történik, ahol  $\mathbf{B}$  a (19) egyenlettel megadott B-spline, míg a  $\mathbf{P}$  büntető mátrix a következőképpen adódik:

$$\mathbf{P}=\lambda_a \mathbf{I}_{c_y} \otimes \mathbf{D}_a' \mathbf{D}_a + \lambda_y \mathbf{D}_y' \mathbf{D}_y \otimes \mathbf{I}_{c_a}, \quad (23)$$

ahol  $\lambda_a$  és  $\lambda_y$  az életkorhoz és évekhez tartozó simító paraméterek.

#### Előrejelzés P-spline-okkal

A jövőbeli értékek előrejelzését hiányzó érték problémaként kezeljük. Egydimenzióban rendelkezünk az  $y_1$  és  $e_1$  adatokkal  $n_1$  évre vonatkozóan valamely életkorra. Legyen  $\mathbf{B}_1$  a B-spline regressziós mátrix egy P-spline halálozási modellben. Tegyük fel, hogy  $n_2$  évre szeretnénk előrejelezni, ekkor az  $n_1+n_2$  évre vonatkozó  $\mathbf{B}$  regressziós mátrix a következő formába írható:

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{0} \\ \mathbf{B}_2 & \mathbf{B}_3 \end{bmatrix} \quad (24)$$

Az  $\mathbf{a}$  regressziós koefficiensek becslése a (10) egyenlet segítségével:

$$(25) \quad l_p = l(\mathbf{a}; \mathbf{y}_1) - \frac{1}{2} \mathbf{a}' \mathbf{P} \mathbf{a}.$$

Ha bevezetjük a  $\mathbf{V} = \text{blockdiag}(\mathbf{I}, \mathbf{0})$  súlymátrixot, ahol  $\mathbf{I}$   $n_1$ -es méretű identitás mátrix, és  $\mathbf{0}$   $n_2$ -es méretű, 0-kból álló négyzetes mátrix, akkor (24) alapján meghatározott  $\mathbf{B}$  regressziós mátrixszal a (13) büntető algoritmus a következő formába írható:

$$(26) \quad (\mathbf{B}' \mathbf{V} \tilde{\mathbf{W}} \mathbf{B} + \mathbf{P}) \hat{\mathbf{a}} = \mathbf{B}' \mathbf{V} \tilde{\mathbf{W}} \mathbf{B} \tilde{\mathbf{a}} + \mathbf{B}' \mathbf{V} (\mathbf{y} - \tilde{\boldsymbol{\mu}}),$$

ami megfelelő alak a kalkulációhoz, mert az illesztés és előrejelzés egyidejűleg elvégezhető.

Az  $\mathbf{a}$  koefficienseken lévő büntetés biztosítja a koefficiensek simaságát, és így simított előrejelzett értékek adódnak.

### PCLM modell

A korcsoportonkénti halálozási adatok szétbontásához Rizzi et al. (2019) a PCLM modellt alkalmazta, majd kétdimenziós simításra a P-spline-okat, módszere a következő.

Tegyük fel, hogy minden életkorra és adott naptári évekre meg tudunk figyelni halálozási számokat, melyeket a  $J \times N$  dimenziós  $\mathbf{Z} = (z_{jn})$  mátrix tartalmaz, ahol  $j=1, \dots, J$  az egyes évenkénti életkorokat és  $n=1, \dots, N$  az egyes naptári éveket jelöli. Feltesszük, hogy a halálozások száma Poisson-folyamatot követ  $E(z_{jn}) = \gamma_{jn}$  várható értékkel. A gyakorlatban megfigyelt, korcsoportonként csoportosított halálozási számokat az  $I \times N$  dimenziós  $\mathbf{Y} = (y_{in})$  mátrix tartalmazza, ahol  $i=1, \dots, I$  az egyes korcsoportokat,  $n=1, \dots, N$  pedig az egyes naptári éveket jelöli. Az  $y_{in}$  halálozások száma Poisson-eloszlású  $E(y_{in}) = \mu_{in}$  várható értékkel, ahol  $\mu_{in}$  tekinthető úgy, mint a látens, nem megfigyelt  $\gamma_{jn}$ -ekből álló várható érték. Tehát  $\gamma_{jn}$  az a várható mortalitási felszín, amelyet becsülni szeretnénk a  $\mu_{in}$  összetett várható értékekből. Ezt büntető Poisson likelihood maximalizálásával érjük el.

Az adatokat először is a következő oszlopvektorokba rendezzük:  $\mathbf{y} = (y_{11}, y_{21}, \dots, y_{IN})'$ ,  $\boldsymbol{\mu} = (\mu_{11}, \mu_{21}, \dots, \mu_{IN})'$  és  $\boldsymbol{\gamma} = (\gamma_{11}, \gamma_{21}, \dots, \gamma_{IN})'$ . Ekkor a PCLM a következő formában írható fel:

$$(27) \quad \boldsymbol{\mu} = \mathbf{C} \boldsymbol{\gamma} = \mathbf{C} \exp(\mathbf{B} \mathbf{a}),$$

ahol  $\boldsymbol{\gamma}$  azon várható halálozási számok sorozata, melyet becsülni szeretnénk,  $\mathbf{C}$  kompozíciós mátrix,  $\mathbf{B}$  pedig B-spline bázis a koefficienssel. A  $\mathbf{C}$  mátrix egy 0/1 blokkmátrix, ami  $\boldsymbol{\mu}$ -t a  $\boldsymbol{\gamma}$ -ból állítja elő, és azt írja le, hogy milyen volt a látens eloszlás. Valamint  $\mathbf{C} = \mathbf{C}_y \otimes \mathbf{C}_a$ , ahol  $\mathbf{C}_y$  marginális kompozíciós mátrix egy  $N \times N$ -es identitásmátrix, míg  $\mathbf{C}_a$  egy  $I \times J$

dimenziós mátrix, melynek elemei mind 0-k, a  $c_{ij}=1$ -ek kivételével. Annak érdekében, hogy halálozási számok helyett halandósági rátákat kapjunk, a  $\mathbf{C}$  mátrix minden oszlopa el van osztva a vonatkozó kitettséggel. A PCLM-ben a  $\gamma \mu$ -re vonatkozó túlparaméterezését a szerzők P-spline-ok segítségével oldották meg, ahol a  $\mathbf{B}_a$  marginális B-spline regressziós mátrixot  $J \times K_a$ , míg a  $\mathbf{B}_y$  mátrixot  $N \times K_y$  méretűnek definiálták, melynél  $K_a$  a  $\mathbf{B}_a$ -hoz tartozó csomók száma + a P-spline foka (azaz  $ndx + bdeg$ ), hasonlóan  $K_y$  a  $\mathbf{B}_y$ -hoz tartozó csomók száma + a P-spline foka.

A PCLM becslési eljárása, mely során az egyes naptári évekre vonatkozóan szétbontja a korcsoportokat egyes életkorok szerinti halandóságra, a büntető Poisson log-likelihood maximalizálásával történik:

$$l^* = l - \mathbf{P} = \sum_{i=1}^I \sum_{n=1}^N (y_{in} \ln \mu_{in} - \mu_{in}) - \mathbf{P}, \quad (28)$$

ahol  $\mathbf{P}$  a (23) egyenletben meghatározott büntető mátrix. A (28) egyenlet maximalizálásával egy egyenletrendszert kapunk, mely a következőre vezet:

$$(\check{\mathbf{B}}' \tilde{\mathbf{W}} \check{\mathbf{B}} + \mathbf{P}) \mathbf{a} = \check{\mathbf{B}}' \tilde{\mathbf{W}} [\tilde{\mathbf{W}}^{-1} (\mathbf{y} - \tilde{\boldsymbol{\mu}}) + \check{\mathbf{B}} \tilde{\mathbf{a}}] \quad (29)$$

ahol  $\check{\mathbf{B}} = \tilde{\mathbf{W}}^{-1} \mathbf{C} \tilde{\mathbf{B}}$ , melyben  $\tilde{\mathbf{I}} = \text{diag}(\tilde{\boldsymbol{\gamma}})$  és  $\tilde{\mathbf{W}} = \text{diag}(\tilde{\boldsymbol{\mu}})$ , melyek diagonális súlymátrixok, hullámmal jelölve egy adott iteráció közelítő megoldását, és  $\mathbf{P}$  a büntető mátrix. A simító paraméterek adott értékeire a rendszer megoldható. Az optimális simított értékekhez a szerzők az Akaike-féle információs kritériumot (AIC) számították ki a  $\lambda$  értékek logaritmikuskálán vett kétdimenziós rácsára, és annak minimális értéke alapján határozták meg az optimális simítást.

### Adatok és módszerek

Az elemzés elvégzéséhez az ír és magyar halálozások száma a WHO Mortality Database (WHO, 2021a) és az ír Központi Statisztikai Hivatal (CSO, 2021) adatbázisaiból származik, míg a kitettségi adatokat a Human Mortality Database (HMD, 2021) adatbázisából töltöttem le. A halálozások száma korcsoportok szerint csoportosítva van meg ezen adatbázisokban, mely csoportok a 0-1, 1-4, 5-9 éveseket, majd így folytatva 5 éves csoportosításban egészen 85 éves korig foglalják magukba, ahonnan egy összevont csoportban érhetők el a 85 éves kor felettiekre (85+) vonatkozó adatok, tehát összesen 19 korcsoporttal dolgoztam. A kitettségi adatok az elemzés elvégzésekor 2017-ig voltak elérhetők, így ennek megfelelően a halálozási adatokat is a 2017-es adatokkal bezárólag vettem figyelembe.



## Kezdeti lépések

Vizsgálataimat az R programnyelv segítségével végeztem el, melyben azon adattábla kialakításához, amellyel dolgoztam, több adattáblát kellett összefűzni: négy adattáblát a WHO (2021a) adatbázisából, és mivel ebben Írországra vonatkozóan csak 2015-ig voltak elérhetőek az adatok, így itt még egy ötödik adattáblát is hozzá kellett vennem a CSO (2021) oldaláról, hogy 2017-ig kipótoljam a halálozások számát. A különböző betegségek pontos beazonosításához az ICD (International Statistical Classification of Diseases and Related Health Problems) kódrendszerét alkalmazzák (WHO, 2021b), melyből az általam vizsgált évek alatt: 1955–2017 között négy verziót, az ICD-7, ICD-8, ICD-9 és ICD-10-et használtak, emiatt voltak a WHO (2021a) adatbázisában négy különböző táblába rendezve az adatok. A vizsgálat elvégzésekor még az ICD-10 volt érvényben, azonban 2022.01.01-től az ICD-11 került bevezetésre. Mivel a stroke esetében a különböző verzióknál nem voltak egyértelműen megfeleltethetők egymásnak a kódok, így ezen haláloknál az azt magába foglaló cerebrovaszkuláris betegségeket vizsgáltam.

## Korcsoportok felbontása PCLM modellel

Az öt fő elkerülhető halálokra PCLM modellel elvégeztem a korcsoportok felbontását, így valamennyi vizsgált évre vonatkozóan megkaptam az egyes életkorok szerinti halandósági ráták logaritmusát.

A PCLM modellt Írország és Magyarország esetén egyaránt 1955–2017 közötti adatokra használtam, az életkorok esetén pedig 0 éves kortól egyéves lépésközt alkalmaztam, ahol a maximális életkor 110 év volt, feltételezve, hogy ezen kor után már nem várható az adott halálok miatti halálozás. Az én alkalmazásomban a korcsoportok száma tehát  $I = 19$ , míg az egyéves lépésközű életkorok száma  $J = 111$  és a vizsgált naptári évek (1955–2017) száma  $N = 63$ . Az  $I \times N$  dimenziós  $\mathbf{Y}$  mátrix jelöli a csoportosított halálozási számokat  $\mu$  várható értékkel, míg a  $J \times N$  dimenziós  $\mathbf{Z}$  mátrix az egyéves lépésközű életvényekénti halálozási számokat  $\gamma$  várható értékkel. A célunk, hogy  $\gamma$ -t megbecsüljük a  $\mu$ -ból, ahol  $\mu = \mathbf{C}\gamma = \mathbf{C}\exp(\mathbf{B}\alpha)$ .

A  $\mathbf{B}$  regressziós bázis  $6993 \times 160$  ( $63 \times 111 = 6993, 10 \times 16 = 160$ ) dimenziós, mely a két marginális B-spline bázis; a  $63 \times 10$  dimenziós  $\mathbf{B}_y$  és a  $111 \times 16$  dimenziós  $\mathbf{B}_a$  Kronecker-szorzatából adódik. A (10)-es egyenletben volt, hogy  $\mathbf{B} = \mathbf{B}_y \otimes \mathbf{B}_a$ , ahol  $\mathbf{B}_y N \times K_y$  dimenziós marginális B-spline bázis a naptári év irányában és  $\mathbf{B}_a J \times K_a$  dimenziós marginális B-spline bázis az életkor irányában, ahol  $K_y$  és  $K_a$  megegyeznek az adott B-spline bázisokhoz tartozó csomók száma és a spline fokának összegével. Tehát  $\mathbf{B}_y$  mátrix 63 sorból áll, mely a vizsgált naptári évek száma,  $\mathbf{B}_a$  mátrix 111 sorból áll, mely a szétbontott, évenkénti életkorok száma.  $\mathbf{B}_y$  és  $\mathbf{B}_a$  oszlopainak száma úgy adódott, hogy a Rizzi et al. (2019) által alkalmazott hüvelykujjszabály alapján harmadfokú spline-ok használata esetén, az egyenletes lépésközű  $\gamma$  minden nyolc elemére egy csomót alkalmazva kaptuk, hogy  $K_y = \text{floor}(63 \setminus 8) + 3 = 10$  és

$K_a = \text{floor}(111 \setminus 8) + 3 = 16$ , ahol a csomók száma:  $ndx = \text{floor}(63 \setminus 8)$ , illetve  $ndx = \text{floor}(111 \setminus 8)$ , a spline foka:  $bdeg = 3$ . A kapott modell  $10 \times 16 = 160$  paraméterrel rendelkezik  $\alpha = (\alpha_1, \dots, \alpha_{160})$ , melyekkel megbecsülhetjük a mortalitási felületet. A  $\mathbf{C}$  kompozíciós mátrix  $1197 \times 6993$  ( $63 \times 19 = 1197, 63 \times 111 = 6993$ ) dimenziós, mely a  $63 \times 63$  dimenziós  $\mathbf{C}_y$  és a  $19 \times 111$  dimenziós  $\mathbf{C}_a$  marginális mátrixok Kronecker-szorzata:  $\mathbf{C} = \mathbf{C}_y \otimes \mathbf{C}_a$ .  $\mathbf{C}_y$  annyi sorral és oszloppal rendelkező identitás mátrix, ahány naptári évet vizsgálunk, míg  $\mathbf{C}_a$  az életkorcsoportok számával megegyező sorú és a szétbontott életkorok számával megegyező oszlopú mátrix.  $\mathbf{C}$  oszlopai továbbá el vannak osztva a vonatkozó kitétségekkel, így a becsült  $\gamma$  értékek halandósági ráták lesznek, így kapjuk meg a mortalitási felszínt. A Human Mortality Database (HMD, 2021) adatbázisából 1950–2017 között, 0 éves kortól 110 éves korig évenkénti lépésközzel az egyes életkorokra elérhetőek a kitétségi adatok. A simító paraméterek optimális értékeit,  $\widehat{\lambda}_a$ -t és  $\widehat{\lambda}_y$ -t Currie et al. (2004) javaslata szerint a BIC alapján határoztam meg. A BIC – mint azt a szerzők megjegyzik – jobban bünteti a modellkomplexitást, különösen nagy  $n$  megfigyelésszám esetén. Emellett a csoportbontás esetén végig harmadfokú spline-okat és másodrendű büntetést alkalmaztam, Currie et al. (2004) és Rizzi et al. (2019) alapján. Az illesztés eredményeként létrejött a  $111 \times 63$  dimenziós simított mortalitási felület.

A PCLM modell segítségével minden életévre vonatkozóan megkaptuk a halandósági rátákat, melyekből a kitétségi adatok ismeretében minden életévre visszaszámolhatók a halálozási számok is. A Poisson Lee-Carter és a P-spline modelleket ezen szétbontott, valamennyi életévre ismert halálozási számok felhasználásával illesztettem. Mindkét modell esetén 100 éves korig vettem figyelembe az adatokat, és annak megfelelően illesztettem a modellt azon feltételezés miatt, hogy afelett már nincsen kellő számú megfigyelés.

## Modellillesztés és előrejelzés

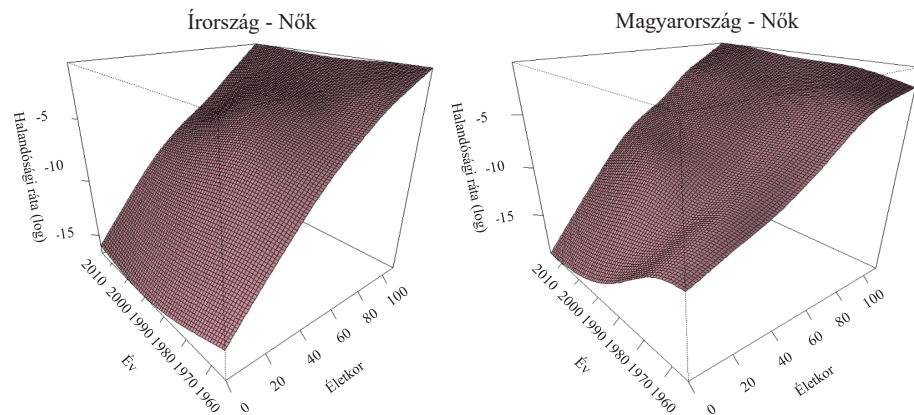
A Poisson Lee-Carter modell illesztéséhez az R-ben az StMoMo package-et használtam, míg a P-spline-okkal való előrejelzéshez a MortalitySmooth package nyújtott segítséget. A Lee-Carter modell esetén az illesztést követően kirajzoltattam a modell paramétereit  $(a_x, k_p, b_x)$ , majd előrejelzést készítettem, valamint az illesztés alapján 5000 szimulációt futtattam 2017-hez képest az elkövetkező 50 évre. Az eredmények szemléltetésénél mindkét modellnél az 1 év alatti, 35, 45, 55, 60, 65, 70, 75, 80, 85, 90, 95 és 100 évesekre nézve a tényleges, illetve az illesztett halandósági ráta logaritmusát mutatom be 95%-os konfidenciaintervállummal 1955–2067 között. Az előrejelzési horizontot Currie et al. (2004) alapján választottam 50 évre, azonban a szerzők megjegyzik, hogy ilyen távoli időpontra becsült értékek esetén már nagyobb a bizonytalanság, így a következtetések kapcsán is óvatosan kell eljárni. Az ábrázolás során azért vizsgáltam többet inkább az idősebb korosztályok halandóságából, mert az idősödés és élettartam-kockázat miatt úgy gondoltam, hogy érdemes az idősebb korosztályok halandóságára helyezni a hangsúlyt ezen halálokok esetében is, ugyanakkor a fiatalabb korosztályok halandósága is megjelenik (35, 45, 55 évesek, valamint 1 év alattiak).

## Mortalitási felületek

Először tekintsük a PCLM modellel létrehozott mortalitási felületeket. A korlátozott terjedelem miatt a nők halandóságát fogom bemutatni, ugyanis Írországnál és Magyarországnál is minden vizsgált haláloknál általánosságban elmondható, hogy hasonlóan változott a nők és férfiak halandósága, a harmadik fő elkerülhető halálokat, az emlőrákot kivéve. Ezen halálok a nőket nagyobb mértékben érinti, mint a férfiakat, és náluk a változás is nagyobb mértékű volt az évek során, így a női halandóság bemutatása szemléletesebb. A mortalitási felületeket bemutató ábrák függőleges tengelye a halandósági ráta logaritmusát mutatja.

Az ischaemiás szívbetegség miatti mortalitási felületet mutatja be a 2. ábra. Írországot tekintve azt láthatjuk, hogy a 40 és 80 éves kor közöttiek körében nagyjából 2000-től kezdődően van láthatóbb csökkenés a halandóságban, az elmúlt 20 évben ezen korosztályok körében javult a halandóság. Emellett 1970–2000 között 70 éves kortól laposabbá vált a görbe, kevésbé meredeken növekedett a halandóság, 1990-től kezdve pedig a fiatalabb korosztályokban is jelentősebb mértékben javult. Magyarországon nagyjából 1955–1970 között az ischaemiás szívbetegség miatti fiatalkori halandóság magasabb szinten volt, mint Írországnál, majd ez jelentősen lecsökkent.

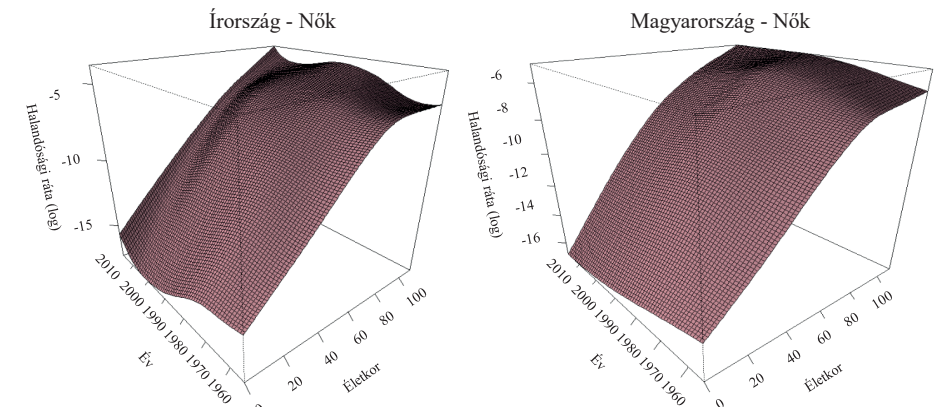
2. ábra: Az ischaemiás szívbetegség miatti halandóság mortalitási felülete



Forrás: Saját szerkesztés

A 3. ábra alapján Írországnál nagyjából 2000-től kezdett el erőteljesen csökkenni a vastag- és végbélrák miatti halálozás a 20 éves kor felettek körében, mely csökkenés leginkább az idősebb korosztályt érintette. 2010 után azonban növekedést figyelhetünk meg. Magyarországon 1980-tól enyhe növekedés, majd 1990-től kismértékű csökkenés látható a mortalitási ráta logaritmusában az 50 éves kor felettek körében. Emellett, elmondható még, hogy Írországnál és Magyarországon is 1955-ben és az azt követő néhány évben magasabb volt a fiatalkori halandóság ezen halálokhöz kapcsolódóan, ami az évek alatt mérséklődött.

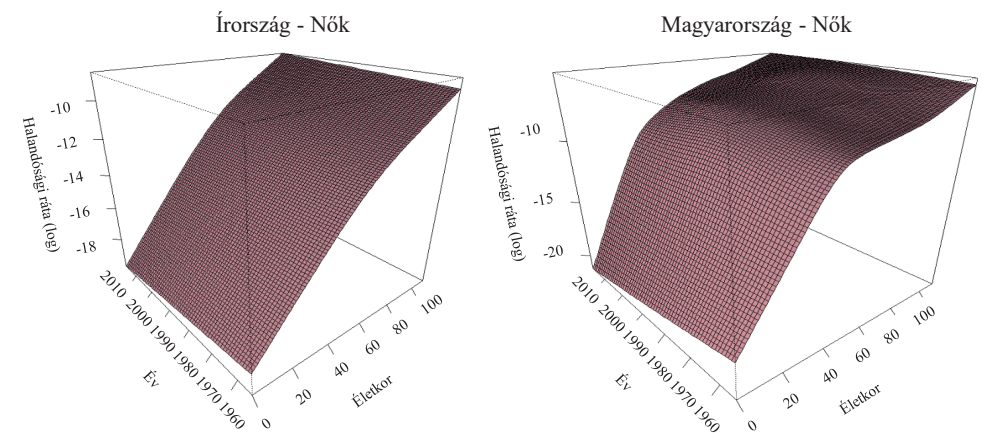
3. ábra: A vastag- és végbélrák miatti halandóság mortalitási felülete



Forrás: Saját szerkesztés

A 4. ábrán azt láthatjuk, hogy az emlőrák esetében nem voltak olyan időszakok, mint a korábban vizsgált betegségeknél, amikor bizonyos években jelentősen megnőtt vagy lecsökkent adott korosztályok mortalitása, ami hullámot keletkeztetett volna a mortalitási felületen. Írországnál 60 éves kor alatt kissé csökkent az emlőrák miatti halandóság, 60–70 éves kor között 2017-ben nagyjából az 1955-ös szinten volt, 70 év feletti életkorok esetén pedig enyhén emelkedett. Magyarországon valamennyi korosztályban leginkább a 2000-es évektől figyelhető meg kisebb csökkenés a halandóságban.

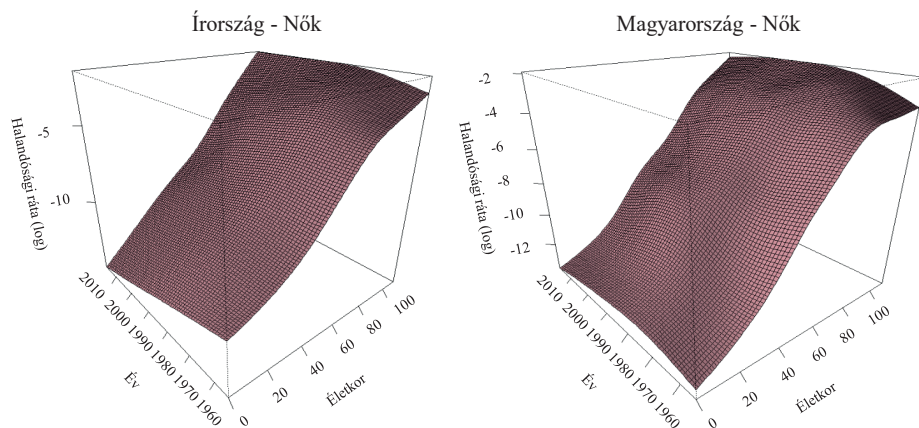
4. ábra: Az emlőrák miatti halandóság mortalitási felülete



Forrás: Saját szerkesztés

Az 5. ábra alapján Írorszában 90 éves kor alatt valamennyi korosztály esetében csökkent a cerebrovaszkuláris betegségek miatti halandóság 1955-höz képest 2017-re. Ez a csökkenés már 1980 körül elkezdődött, leginkább a 40 éves és idősebb korosztálynál jelentkezett, és tartott egészen 90 éves korig. A 90 év feletti körében ugyanis 1955-től növekedés történt, ami magyarázható a hosszabbodó élettartammal is, ugyanakkor 90 éves kor felett a halandósági adatok az alacsony megfigyelésszám miatt kevésbé megbízhatók.

5. ábra: A cerebrovaszkuláris betegségek miatti halandóság mortalitási felülete

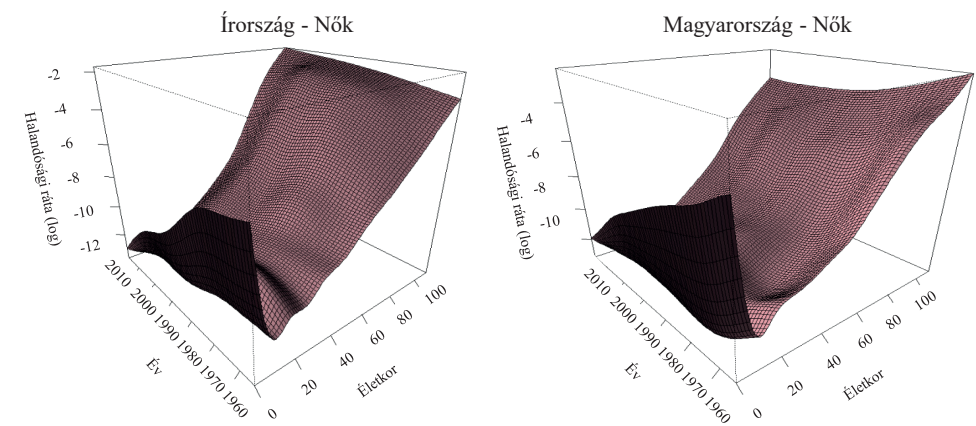


Forrás: Saját szerkesztés

Magyarországon az 5. ábrán 1955-től eleinte csökkenés látható a cerebrovaszkuláris betegségek miatti halandóságban a 80 év alattiak esetén, ami nem sokkal 1970 előtt növekedésbe fordult át, mely növekedés annál inkább jelentős volt, minél fiatalabb (30–50 év között) korúakról beszélünk. 1990-től hazánkban a mortalitás a 90 éven aluliaknál ismét csökkenni kezdett, ez esetben is a csökkenés annál jelentősebb volt, minél fiatalabbakról van szó (20 éves korig nézve). Magyarországon az 1 éves kor alattiak halandóságában nem következett be változás 2017-re, bár az ő halandóságuk már 1955-ben is alacsony volt.

A 6. ábrán Írország esetében azt láthatjuk, hogy az 1 éves kor alatti tüdőgyulladás miatti halandóság 1955-höz képest jelentősen lecsökkent 2017-re. 1970 körül főként az idősebb korosztályoknál, nagyjából 50 éves kortól elkezdődött egy növekedés, majd 1980-tól egy csökkenés a halandóságban, aztán 1985-től ismét növekedés indult el, mely 2000-ben csúcspontot ért el, és onnantól kezdve jelentősen mérséklődött. A 80 év feletti esetében vagy az 1955-ös szintre sikerült leszorítani a mortalitást, vagy (a 90 év feletti korosztálynál) annál magasabb szintet ért el, ami ugyancsak a hosszabbodó élettartam jelenségével lehet összefüggésben, bár mivel nagyon magas életkorról beszélünk, a következtetések levonásánál kell némi fenntartással élnünk.

6. ábra: A tüdőgyulladás miatti halandóság mortalitási felülete



Forrás: Saját szerkesztés

Magyarországon is megfigyelhető a 6. ábra alapján, hogy az 1 év alattiak halandósága 1955 és 2017 között nagymértékben csökkent. 1955–1970 között volt egy visszaesés a mortalitásban valamennyi korosztályban, majd ezután növekedés következett. 1980-tól újabb kismértékű csökkenés történt, melyet 1990-től ismét kisebb halandóság növekedés követett, azonban 2000-től jelentősebb mértékű csökkenés kezdődött. Ugyanakkor 2010-től a fiatalabb korosztályoknál (45 éves kor és alatta) megtorpanás volt, 45 éves kor felett pedig újból növekedés bontakozott ki a mortalitásban, amellyel kapcsolatban kérdésként merülhet fel, hogy mennyire tartós. A COVID-19 okozta tüdőgyulladásos halálos esetek még nem szerepelnek az általam vizsgált halálozási számokban, azonban Csépai és Kovács (2021) cikkéből egyértelműen látszik Magyarországon a járvány hatására bekövetkező növekedés az átlagos halálozásban, mely halálozási többlet elsősorban a 60 év feletti korcsoportokban mutatkozott meg.

### Átlagos négyzetes hiba

Hogy számszerűsíthető legyen, melyik modell illeszkedett jobban, és ezáltal pontosabb következtetéseket vonhassunk le, tekintsük az átlagos négyzetes hiba (angolul Mean Squared Error, MSE) értékeit a 2. táblázatban. A segítségével láthatjuk, hogy melyik modell illesztett értékei térnek el kevésbé a tényleges értékektől. Ehhez az adatok egy részét leválasztottam: az 1955–2007 közötti adatokon végeztem az illesztést, és a 2008–2017 közötti adatokon teszteltem az előrejelzés pontosságát úgy, hogy kigyűjtöttem a 2008–2017-re előrejelzett halandósági rátákat, valamint az ezen időszakra vonatkozó tényleges halandósági rátákat (melyek a PCLM modell révén minden korévre vonatkozóan rendelkezésemre álltak).

Ezt követően a becsült és tényleges halandósági ráták logaritmusainak néztem az átlagos négyzetes hibáját, mely kapcsán minél kisebb az érték, annál jobban reprodukálja a modell a tapasztalatokat (Vékás, 2019). A 2. táblázatban szürke háttérrel van kiemelve a két modell MSE értékei közül a kisebb, azaz kedvezőbb érték.

2. táblázat: Átlagos négyzetes hiba értékei

Halálozás modellezése	Valamennyi életkorban		Felnőttkorban	
	Lee-Carter	P-spline	Lee-Carter	P-spline
<b>Ischaemiás szívbetegség</b>				
Írország, nők	3,7840	0,1270	0,8214	0,0784
Magyarország, nők	2,4031	0,7713	0,6475	0,1007
<b>Vastag- és végbélrák</b>				
Írország, nők	2,0041	0,1718	0,2884	0,0608
Magyarország, nők	0,4101	0,1290	0,0384	0,0356
<b>Emlőrák</b>				
Írország, nők	0,0777	0,5807	0,1030	0,0181
Magyarország, nők	0,1462	0,7962	0,0689	0,0240
<b>Cerebrovaszkuláris betegségek</b>				
Írország, nők	0,1035	0,0090	0,0433	0,0177
Magyarország, nők	0,1974	0,1131	0,2404	0,0158
<b>Tüdőgyulladás</b>				
Írország, nők	2,3096	0,1920	2,1245	0,1664
Magyarország, nők	0,4036	0,1331	0,2370	0,2830

Forrás: Saját szerkesztés

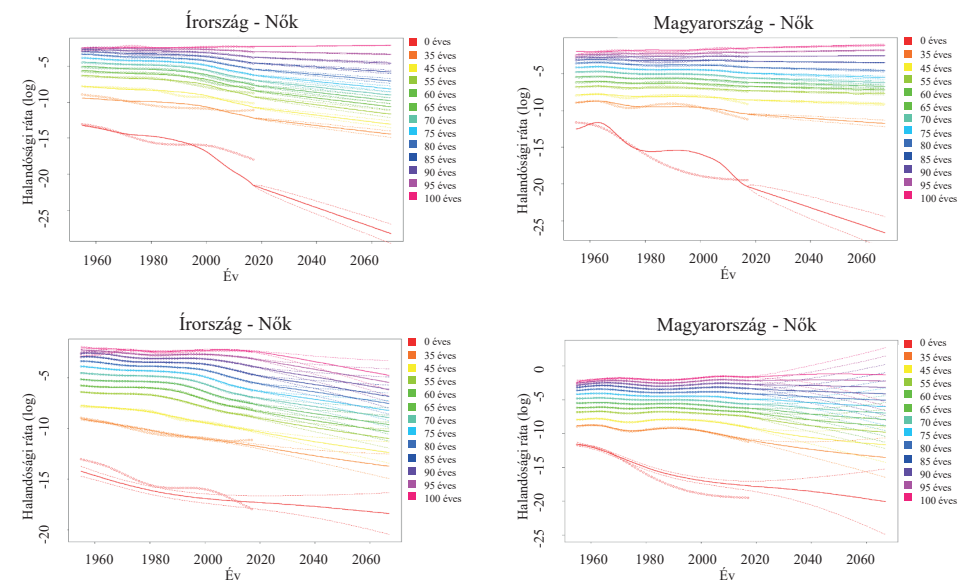
Megállapíthatjuk, hogy majdnem minden halál esetén a P-spline módszer pontosabb eredményt adott, mint a Lee-Carter modell, akár a valamennyi életkorra vett, akár kizárólag a felnőttkori (25 év feletti) halandóságot tekintjük. Egyedül az emlőrák esetében a valamennyi életkorban nézett halandóság esetén volt pontosabb a Lee-Carter modellel az illeszkedés, azonban ezen halálknál is a kizárólag felnőttkori értékeket tekintve a P-spline modell ad jobb értékeket. A tüdőgyulladás esetében a nőknél Magyarországon ugyancsak a Lee-Carter teljesített jobban, azonban az eltérés nagyon csekély a két modell között.

### Ischaemiás szívbetegség

A 7. ábrán látható a két modell illesztésének eredménye, mely alapján mindkét ország esetén azt állapíthatjuk meg, hogy a fiatalkori ischaemiás szívbetegség miatti halandóságot kevésbé

tudják megbecsülni a modellek; a legjobb illeszkedés a 45–80 év közötti korúak esetén látható, míg 80 éves kor felett már tapasztalható eltérések a tényadatokhoz képest.

7. ábra: Az ischaemiás szívbetegség miatti halandóság Lee-Carter modell (fent) és P-spline modell (lent) szerinti előrejelzése



Forrás: Saját szerkesztés

Megfigyelhetjük, hogy Írorszában szinte valamennyi korcsoportban nagyobb mértékben csökkent a halandóság, mint Magyarországon. Ugyanakkor érdekes, hogy a 35 évesek halandósága Írorszában 2000-től stagnált, majd növekedni kezdett. A Lee-Carter modell a magyar nőknek stagnáló, az ír nőknek enyhén csökkenő mortalitást vetít előre az ischaemiás szívbetegséggel összefüggésben. A P-spline modell, mely az MSE alapján jobban illeszkedik, egy fokkal kedvezőbb képet jelenít meg, ugyanis eszerint a magyar nők halandósága is csökkenni fog (igaz, a nagyon idős korosztályoké kevésbé), az ír nők halandósága pedig nagyobb mértékben csökken majd a jövőben, mint a Lee-Carter modell szerint.

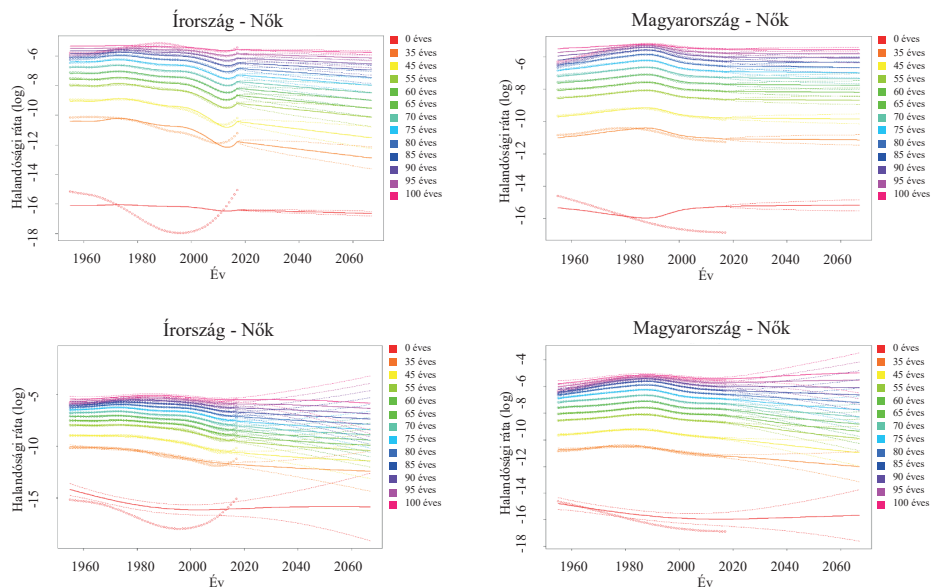
Összességében azt állapíthatjuk meg, hogy az ír nők ischaemiás szívbetegség miatti halandósága kedvezőbben alakult, és az előrejelzés alapján a jövőben is kedvezőbben fog alakulni, mint a magyar nőké. Az ischaemiás szívbetegség esetében rizikófaktornak számítanak a helytelen étkezési szokások, magas vérnyomás, magas koleszterinszint, a mozgáshiány és a magas alkoholfogyasztás (Dai et al., 2020; Scarborough et al., 2011). Az 1. táblázatban láttuk, hogy mind a zöldség- és gyümölcsfogyasztást, mind az elhízottak, illetve a dohányosok

arányát tekintve Magyarországon rosszabbak az adatok. Így, bár ezen halálok a megfelelő kezeléssel elkerülhető halálokok közül számít vezető haláloknak, mégis, primer prevenciók beavatkozásokkal is lehetne javítani az ebből eredő halálozáson.

### Vastag- és végbélrák

A 8. ábrán Magyarországon és Írországon mindkét nem esetén 45 és 75 éves kor között az illesztés nagyon jónak mondható, 75 éves kor felett azonban a Lee-Carter modell az 1980-át megelőző évek esetén felülbecsüli, 45 éves kor alatt alulbecsüli a tényleges halandóságot. Az átlagos négyzetes hiba alapján is megállapítottuk, hogy a vastag- és végbélrák esetében szintén a P-spline modell illeszkedett jobban.

8. ábra: A vastag- és végbélrák miatti halandóság Lee-Carter modell (fent) és P-spline modell (lent) szerinti előrejelzése



Forrás: Saját szerkesztés

A jövőre vonatkozó előrejelzést illetően hasonló tendenciát látunk, mint az ischaemiás szívbetegségnél, miszerint a Lee-Carter modell Magyarországra stagnálást jelez előre (minden korosztályra), míg Írországra enyhe csökkenést. A jobban illeszkedő P-spline modell ugyanakkor a magyar nőknél 90 éves kor alatt csökkenést projektál annál nagyobb mértékben, minél fiatalabbakról van szó, míg Írországra még az idősebb korosztályoknál is csökkenést jelez a vastag- és végbélrák miatti halandóságra.

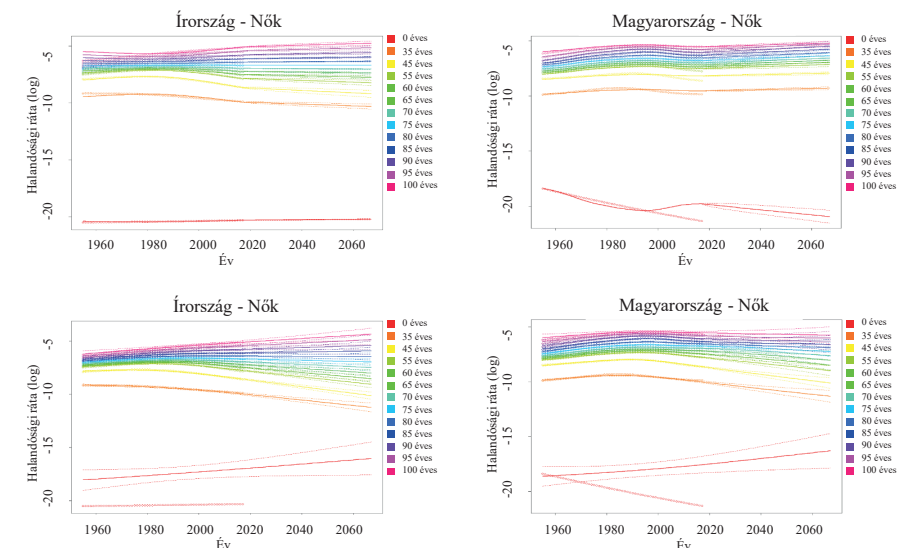
Ezen halálokat tekintve is azt láthatjuk, hogy Írországon kedvezőbben alakult a nők halandósága, mint Magyarországon, mert bár 1955-ben Írországon szinte valamennyi korcsoportban magasabb szinten volt a halandóság, mint Magyarországon, mégis azóta nagyobb mértékben csökkent.

Rawla et al. (2019) rávilágított, hogy 2018-ban világszinten Magyarországon volt a legmagasabb a vastag- és végbélrákból eredő új esetek száma és a halandósági ráta is. Menyhárt et al. (2018) a 2016-2030 közötti időszakra a 65–85 évesek esetén a rákos megbetegedések 35%-os, a halálozás 30%-os növekedését vetítette előre Magyarországon.

### Emlőrák

A 9. ábrán az emlőrák esetében mindkét országban összességében jó illeszkedést láthatunk. Ha csak az idősebb korosztály halandóságát tekintjük, akkor Magyarországon, ha azonban a fiatalabb korosztályt nézzük, akkor Írországon illeszkedik jobban a Lee-Carter modell. A P-spline modellnek 1 éves kor alatt nem jó az illesztése, hiszen nagyon felülbecsüli. 35 éves kornál, illetve felette azonban jó a P-spline modell becslése, amit abból is láthattunk, hogy az átlagos négyzetes hiba értékei alapján valamennyi életkort tekintve a Lee-Carter modell, míg kizárólag a 25 éves kor feletti halandóságot tekintve a P-spline modell illeszkedett jobban. Mivel az emlőrák esetében is a fiatalok halandósága nagyon alacsony, a felnőttkori illeszkedés jósága tekinthető inkább mérvadónak.

9. ábra: Az emlőrák miatti halandóság Lee-Carter modell (fent) és P-spline modell (lent) szerinti előrejelzése



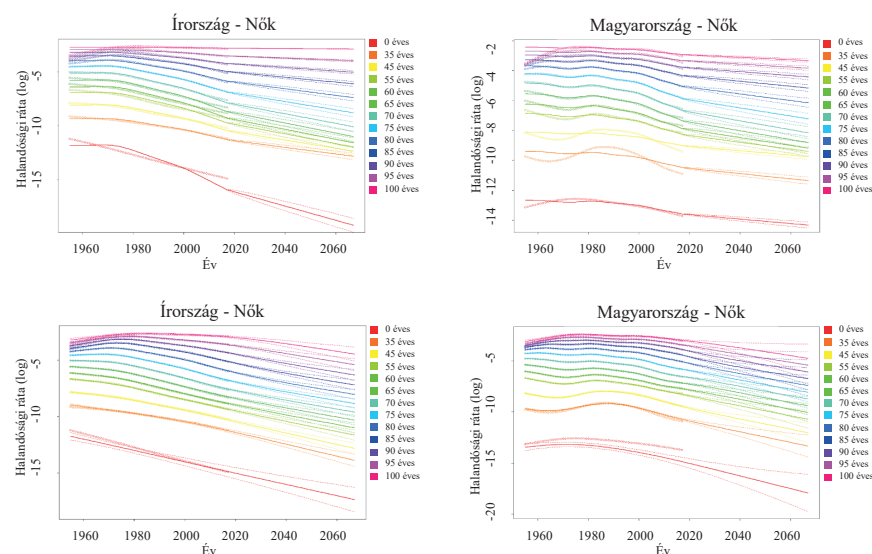
Forrás: Saját szerkesztés

Az emlőrák miatti halandóság előrejelzését vizsgálva megfigyelhetjük, hogy a Lee-Carter modell az ír nők halandóságára stagnálást, a magyar nők halandóságára növekedést vetít előre minden felnőtt korosztályra. Ezzel szemben a P-spline modell a felnőttek esetén Magyarországon minden korosztályra csökkenést, Írországon 80 éves kor alatt csökkenést, felette enyhe növekedést jelez előre. Ezen halálok esetén nincsen jelentősebb különbség a két országban tapasztalt, illetve előrejelzett halandóságban, Magyarországon a legidősebb korosztályok számára a mortalitás kedvezőbb alakulását vetíti előre a modell.

### Cerebrovaszkuláris betegségek

A 10. ábrán megfigyelhetjük, hogy a cerebrovaszkuláris betegségek miatti halandóságra a P-spline modell jobban illeszkedik, mint a Lee-Carter modell. Ezt a megállapítást az átlagos négyzetes hiba értékei is alátámasztják. Írországon valamennyi életkorban szinte tökéletesnek mondható az illeszkedés.

10. ábra: A cerebrovaszkuláris betegségek miatti halandóság Lee-Carter modell (fent) és P-spline modell (lent) szerinti előrejelzése



Forrás: Saját szerkesztés

Írországon valamennyi korosztályban jelentősen csökkent a cerebrovaszkuláris betegségek miatti halandóság, így a P-spline modell a későbbiekre is valamennyi korosztály számára csökkenést vetít itt előre. Magyarországon ugyanakkor nem olyan folyamatos a csökkenés, mint Írországon, sőt, az 1970-es éveket követően volt egy növekedés, ami leginkább a fiatalabb

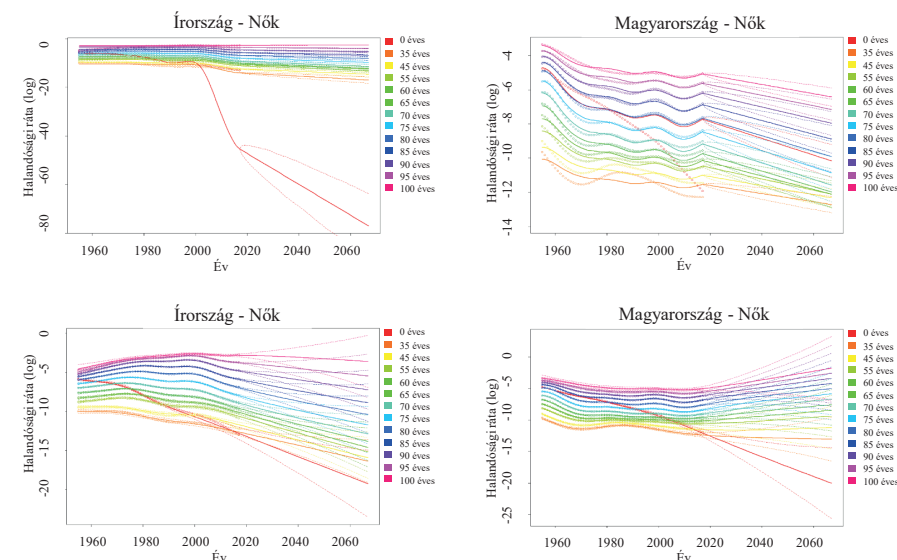
korosztályoknál mutatkozott, és ami az 1990-es években fordult csak vissza. Bonita et al. (1990) alapján 1970–1985 között Magyarországon volt a legmagasabb mindkét nem esetén a 40–69 évesek körében a stroke halandósági ráta éves növekedése a vizsgált 27 ország közül, míg Németh et al. (2021) arról ír, hogy 1990–2014 között Magyarországon 59,1%-kal csökkent a nők cerebrovaszkuláris betegség miatti halálozása, ami felülmúlta a nyugat-európai országok átlagát. A P-spline modell Magyarországra is csökkenést jelez előre valamennyi korosztályra, azonban a halandóság csökkenése Írországon nagyobb mértékű volt, mint hazánkban.

A cerebrovaszkuláris betegségek rizikófaktorai közé tartozik többek között a dohányzás, diabétesz mellitusz, a túlzott alkoholfogyasztás, mozgásszegény életmód és elhízás (Morys et al., 2021; Soto-Cámara et al., 2020). Az ischaemiás szívbetegségnél már megállapítottuk, hogy ezen faktorok tekintetében Magyarországon rosszabb a helyzet, mint Írországon.

### Tüdőgyulladás

A 11. ábra alapján megállapíthatjuk, hogy a Lee-Carter modell sem az ír, sem a magyar tüdőgyulladás miatti halandósági adatokra nem illeszkedik elég jól, mert az 1 éves kor alattiak halandóságát nem tudta leképezni, illetve a többi korosztálynál is láthatóak eltérések. A P-spline modellel való becslés alapján Írország és Magyarország esetében is valamennyi életkorra nagyon jó az illeszkedés. Még a többi halálok esetén általában rosszul becsült 1 éven aluliak tényleges halandóságát is rendkívül jól becsülte meg a modell.

11. ábra: A tüdőgyulladás miatti halandóság Lee-Carter modell (fent) és P-spline modell (lent) szerinti előrejelzése



Forrás: Saját szerkesztés

Az átlagos négyzetes hiba alapján, ha csak a felnőttkori halandóságot tekintjük, akkor a magyar nők halandóságára a P-spline modell illeszkedése romlik, még hozzá annyit, hogy a Lee-Carter modellre jobb illeszkedés adódik – bár a két modell MSE értéke közötti eltérés nagyon kicsi. Itt tehát kérdéses, hogy melyik modell előrejelzését fogadjuk el, mert a Lee-Carter modell a magyar nők esetén minden ábrázolt életkorra csökkenést, míg a P-spline modell az 1 éven aluliak kivételével stagnálást, illetve növekedést jelez előre. Az ír nők halandóságára a Lee-Carter modell enyhe csökkenést, és a P-spline modell is csökkenést vetít előre.

A tüdőgyulladás miatti női halandóság esetén Magyarországon a fiatalabb korosztályok halandósága 1955–2017 között kisebb mértékben csökkent, mint Írországból, azonban az idősebb korosztályoknál Írországból növekedett az ezen halálok miatti halandóság, aminek hátterében véleményem szerint a hosszabbodó élettartam is állhat.

## Összefoglalás

Tanulmányomban a megfelelő kezeléssel elkerülhető öt legjelentősebb halálok miatti halálozást vizsgáltam a nőkre vonatkozóan Írországból és Magyarországon, ennek során a PCLM, P-spline és Lee-Carter modelleket alkalmaztam. A modellek illesztése előtt röviden áttekintettem a magyar és ír társadalom egészségi állapotával kapcsolatos főbb mutatókat, melyek alapján általánosságban Írország esetén beszélhetünk kedvezőbb értékekről.

Az elkerülhető halálozás a szakirodalom alapján tekinthető az egészségügyi rendszer indikátorának. Arra a kérdésre, hogy az elkerülhető halálozás alapján hatékonyabb-e az ír egészségügyi rendszer, mint a magyar, a következőket mondhatjuk. A vizsgált halálokok többségében azt láttuk, hogy a tényidőszakban az ír halandóság jobban alakult, mint a magyar, és a modellek a jövőre vonatkozóan is jellemzően Írországra vetítették előre a mortalitás csökkenését. Ugyanakkor a tanulmányban bemutatott vizsgálatok nem elégségesek ahhoz, hogy a két rendszert a hatékonyságuk alapján rangsoroljuk. Azonban jó alapot jelenthetnek a hatékonyság vizsgálata során, és úgy gondolom, hogy érdemes lenne az egyéb halálokok miatti halandóság alakulását is megvizsgálni, ideértve a megelőzhető halálokokat, melyekkel ez a tanulmány nem foglalkozott. Nyitott kérdés továbbá az is, hogy a Magyar Nemzeti Bank versenyképességi programjának javaslataiból ténylegesen melyek, milyen formában és mélységben fognak megvalósulni.

Láttuk azt is, hogy az elemzett halálokok közül több esetében olyan rizikófaktorok is állnak a betegség kialakulásának a hátterében (pl. egészségtelen étrend, dohányzás, mozgásszegény életmód, magas alkoholfogyasztás), amelyek primer prevenciók beavatkozásokkal, szemléletformálással elkerülhetők vagy csökkenthetők lennének. Ezen rizikófaktorok nagy részénél láttuk a két társadalom egészségi állapotára vonatkozó mutatóknál, hogy Magyarországon rosszabbak az adatok, mint Írországból, így véleményem szerint az egészségtudatosság további erősítése mindenképpen nagyon fontos feladat továbbra is.

A tanulmányban bemutatott PCLM modell a csoportosított értékek szétbontásához

nyújthat számottevő segítséget. A P-spline modell alkalmazása a halandóság előrejelzésében bírhat jelentőséggel, hiszen a hosszabbodó várható élettartam jelenség következtében az időskori mortalitás minél pontosabb előrejelzése és ahhoz megbízható halandóság előrejelző modell alkalmazása egyre fontosabb lesz (gondolva itt különösen a járadékbiztosításokra).

## IRODALOMJEGYZÉK

- Allin, Sara – Grignon, Michel (2014): Examining the Role of Amenable Mortality as an Indicator of Health System Effectiveness. *Healthcare Policy*. 9(3): pp.12–19. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3999566/> Letöltés dátuma: 2022.02.14. <https://doi.org/10.12927/hcpol.2014.23733>
- Bonita, Ruth – Stewart, Alistair – Beaglehole, Robert (1990): International trends in stroke mortality: 1970–1985. *Stroke*. 21(7): pp. 989–992. <https://doi.org/10.1161/01.str.21.7.989> Letöltés dátuma: 2022.02.14.
- Brouhns, Natacha – Denuit, Michel – Vermunt, Jeroen K. (2002): A Poisson Log-Bilinear Regression Approach to the Construction of Projected Life Tables. *Insurance Mathematics and Economics* 31(3):373–393. [http://dx.doi.org/10.1016/S0167-6687\(02\)00185-3](http://dx.doi.org/10.1016/S0167-6687(02)00185-3) Letöltés dátuma: 2022.02.14.
- Currie, Iain D. – Durban, Maria – Eilers, Paul H.C. (2004): Smoothing and forecasting mortality rates. *Statistical Modelling*, 4(4):279–298. <https://doi.org/10.1191/1471082X04st0800a> Letöltés dátuma: 2022.02.14.
- Cylus, Jonathan – Papanicolas, Irene – Smith, Peter C. [editors]. (2016): *Health System Efficiency: How to make measurement matter for policy and management*. Health Policy Series, (46):243. European Observatory on Health Systems and Policies. ISBN: 978 92 890 5041 8. [https://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0004/324283/Health-System-Efficiency-How-make-measurement-matter-policy-management.pdf](https://www.euro.who.int/__data/assets/pdf_file/0004/324283/Health-System-Efficiency-How-make-measurement-matter-policy-management.pdf) Letöltés dátuma: 2022.02.14.
- Csépai, Orsolya – Kovács, Erzsébet (2021): Koronavírus-járvány adatok és biztosítási hatások elemzése. *Biztosítás és kockázat*, 8(3–4): pp. 24–43. <http://dx.doi.org/10.18530/BK.2021.3-4.24> Letöltés dátuma: 2022.02.14.
- CSO (2021): Deaths Occuring. Labels: Sex, Cause of Death, Age at Death, Year, 2007–2018. VSA08. [Central Statistics Office]. <https://data.cso.ie/> Letöltés dátuma: 2021.01.10.
- Dai, Haijiang – Much, Arsalan A. – Maor, Elad – Asher, Elad – Younis, Arwa – Xu, Yawen – Lu, Yao – Liu, Xinyao – Shu, Jingxian – Bragazzi, Nicola L. (2020): Global, regional, and national burden of ischemic heart disease and its attributable risk factors, 1990–2017: results from the global Burden of Disease Study 2017. *European Heart Journal – Quality of Care and Clinical Outcomes*. pp. 1–11. <https://doi.org/10.1093/ehjcco/qcaa076> Letöltés dátuma: 2022.02.14.
- D’Amato, Valeria – di Lorenzo, Emilia – Haberman, Steven – Russolillo, Maria – Sibillo, Marilena (2011): The Poisson Log-Bilinear Lee-Carter Model: Applications Of Efficient Bootstrap Methods to Annuity Analyses. *North American Actuarial Journal*. 15(2): pp. 315–333. <https://doi.org/10.1080/10920277.2011.10597623> Letöltés dátuma: 2022.02.14.
- Deaton, Angus – Paxson, Christina (2001): Mortality, Income, and Income Inequality Over Time in Britain and the United States. NBER Working Paper No. 8534. Cambridge. <http://dx.doi.org/10.3386/w8534> Letöltés dátuma: 2022.02.14.
- Eilers, Paul H.C. és Marx, Brian D. (1996): Flexible smoothing with B-splines and penalties. *Statistical Science*, 11(2):89–121. <https://doi.org/10.1214/ss/1038425655> Letöltés dátuma: 2022.02.14.
- Endrédi Balázs Szabolcs – Constantinovits Milán (2016): Magyarország és Írország egészségügyi rendszerének összehasonlítása, avagy a fogyasztói paradoxon az egészségügy területén. *Külügyi Szemle*, 15(2): pp. 71–92. <https://kki.hu/assets/upload/Endreedi-Constantinovits.pdf> Letöltés dátuma: 2022.02.14.
- Gogola, Ján – Vékás, Péter (2020): Élettartam-kockázat Csehországban és Magyarországon. *Biztosítás és kockázat*, 7(3–4):14–26. <http://dx.doi.org/10.18530/BK.2020.3-4.14> Letöltés dátuma: 2022.02.14.
- HMD (2021): The Human Mortality Database. [Human Mortality Database]. <https://www.mortality.org/> Letöltés dátuma: 2021.01.10.
- Hollingsworth, Bruce (2016): Health system efficiency: measurement and policy. In: *Health System Efficiency: How to make measurement matter for policy and management*. Health Policy Series, (46):243. European Observatory on Health Systems and Policies. ISBN: 978 92 890 5041 8. [https://www.euro.who.int/\\_\\_data/assets/pdf\\_file/0004/324283/Health-System-Efficiency-How-make-measurement-matter-policy-management.pdf](https://www.euro.who.int/__data/assets/pdf_file/0004/324283/Health-System-Efficiency-How-make-measurement-matter-policy-management.pdf) Letöltés dátuma: 2022.02.14.

- Hone, Thomas – Rasella, Davide – Barreto, Mauricio – Atun, Rifat – Majeed, Azeem – Millett, Christopher (2017): Large Reductions in Amenable Mortality Associated With Brazil's Primary Care Expansion And Strong Health Governance. *Health Affairs*. 36(1): pp. 149–158.  
<https://doi.org/10.1377/hlthaff.2016.0966> Letöltés dátuma: 2022.02.14.
- Kiss Norbert – Tókey Balázs (2019): Az MNB versenyképességi programjában szereplő, egészségügyi makrofinanszírozást érintő javaslatok értékelése. *Biztosítás és kockázat*, 6(4): pp. 12–27.  
<http://dx.doi.org/10.18530/BK.2019.4.12> Letöltés dátuma: 2022.02.14.
- Lee, Ronald D. – Carter, Lawrence R. (1992): Modeling and Forecasting U.S. Mortality. *Journal of the American Statistical Association*, 87(419): pp. 659–671.  
<https://doi.org/10.1080/01621459.1992.10475265> Letöltés dátuma: 2022.02.14.
- Menyhárt, Otilia – Fekete, János T. – Györfly, Balázs (2018): Demographic shift disproportionately increases cancer burden in an aging nation: current and expected incidence and mortality in Hungary up to 2030. *Clinical Epidemiology*. 2018(10): pp. 1093–1108.  
<https://doi.org/10.2147/CLEP.S155063> Letöltés dátuma: 2022.02.14.
- MNB (2019): Egészséges társadalom. In: Versenyképességi program 330 pontban. pp. 124–142. ISBN 978-615-5318-26-9. [Magyar Nemzeti Bank]  
<https://www.mnb.hu/letoltes/versenykepességi-program.pdf> Letöltés dátuma: 2022.02.14.
- Morys, Filip – Dadar, Mahsa – Dagher, Alain (2021): Association Between Midlife Obesity, Its Metabolic Consequences, Cerebrovascular Disease and Cognitive Decline. *The Journal of Clinical Endocrinology & Metabolism*. 106(10):e4260–e4274.  
<https://doi.org/10.1210/clinem/dgab135> Letöltés dátuma: 2022.02.14.
- Németh, Noémi – Endrei, Dóra – Horváth, Lilla – Elmer, Diána – Csákvári, Tímea – Pónusz, Róbert – Szapáry, László – Boncz, Imre (2021): A cerebrovasculáris betegségekből eredő, idő előtti halálozás egyenlőtlenségei Európában 1990 és 2014 között. *Orvosi Hetilap*. 162(4): pp. 144–152.  
<https://doi.org/10.1556/650.2021.31980> Letöltés dátuma: 2022.02.14.
- OECD (2018): Health at a Glance: Europe 2018: State of Health in the EU Cycle. [Organisation for Economic Co-operation and Development]. OECD Publishing. Paris. ISBN 978-92-64-30335-5 (PDF). ISSN 2305-6088 (online).  
[https://doi.org/10.1787/health\\_glance\\_eur-2018-en](https://doi.org/10.1787/health_glance_eur-2018-en) Letöltés dátuma: 2022.02.14.
- OECD (2019a): Ireland: Country Health Profile 2019. State of Health in the EU. [Organisation for Economic Co-operation and Development]. OECD Publishing. Paris/European Observatory on Health Systems and Policies. Brussels. ISBN 9789264684461. SSN 25227041 (online).  
<https://doi.org/10.1787/2393fd0a-en> Letöltés dátuma: 2022.02.14.
- OECD (2019b): Magyarország: egészségügyi országprofil 2019. State of Health in the EU. [Organisation for Economic Co-operation and Development]. OECD Publishing. Paris/European Observatory on Health Systems and Policies. Brussels. ISBN 9789264690851. SSN 25227041 (online).  
<https://doi.org/10.1787/2247f4e0-hu> Letöltés dátuma: 2022.02.14.
- OECD (2019c): Health at a Glance 2019: OECD Indicators. [Organisation for Economic Co-operation and Development]. OECD Publishing. Paris. ISBN 978-92-64-80766-2 (pdf). ISSN 1999-1312 (online).  
<https://doi.org/10.1787/4dd50c09-en> Letöltés dátuma: 2022.02.14.
- OECD (2020): Health at a Glance: Europe 2020. State of Health in the EU Cycle. [Organisation for Economic Co-operation and Development]. OECD Publishing. Paris. ISBN 978-92-64-81194-2 (pdf). ISSN 2305-6088 (online).  
<https://doi.org/10.1787/82129230-en> Letöltés dátuma: 2022.02.14.
- Radojicic, Milan – Jeremic, Veljko – Savic, Gordana (2020): Going beyond health efficiency: What really matters? *The International Journal of Health Planning and Management*. 35(1): pp. 318–338.  
<https://doi.org/10.1002/hpm.2914> Letöltés dátuma: 2022.02.14.
- Rawla, Prashanth – Sunkara, Tagore – Barsouk, Adam (2019): Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. *Gastroenterology Review*, 14(2): pp. 89–103.  
<https://doi.org/10.5114/pg.2018.81072> Letöltés dátuma: 2022.02.14.
- Rizzi, Silvia – Halekoh, Ulrich – Thinggaard, Mikael – Engholm, Gerda – Christensen, Niels – Johannesen, Tom B. – Lindahl-Jacobsen, Rune (2019): How to estimate mortality trends from grouped vital statistics. *International Journal of Epidemiology*, 48(2): pp. 571–582.  
<https://doi.org/10.1093/ije/dyy183> Letöltés dátuma: 2022.02.14.
- Rutstein, David D. – Berenberg, William – Chalmers, Thomas C. – Child, Charles G. 3rd. – Fishman, Alfred P. – Perrin, Edward B. – Feldman, Jacob J. – Leaverton, Paul E. – Lane, J. Michael – Sencer, David J. – Evans, Carleton C. (1975): Measuring the quality of medical care. A clinical method. *The New England Journal of Medicine*, 294(11): pp. 582–588.  
<https://doi.org/10.1056/nejm197603112941104> Letöltés dátuma: 2022.02.14.
- Scarborough, Peter – Wickramasinghe, Kremlin – Bhatnagar, Prachi – Rayner, Mike (2011): Trends in coronary heart disease 1961–2011. British Heart Foundation. London.  
<https://www.bhf.org.uk/informationsupport/publications/statistics/trends-in-coronary-heart-disease-1961-2011> Letöltés dátuma: 2022.02.14.
- Soto-Cámara, Raúl – González-Bernal, Jerónimo J. – González-Santos, Josefa – Aguilar-Parra, José M. – Trigueros, Rubén – López-Liria, Remedios (2020): Age-Related Risk Factors at the First Stroke Event. *Journal of Clinical Medicine*. 9(7): Article 2233.  
<https://doi.org/10.3390/jcm9072233> Letöltés dátuma: 2022.02.14.
- Vergara-Duarte, Montse – Borrell, Carme – Pérez, Glòria – Martín-Sánchez, Juan C. – Clèries, Ramon – Buxó, Maria – Martínez-Solanas, Èrica – Yasui, Yutaka – Muntaner, Carles – Benach, Joan (2018): Sentinel Amenable Mortality: A New Way to Assess the Quality of Healthcare by Examining Causes of Premature Death for Which Highly Efficacious Medical Interventions Are Available. *BioMed Research International*. 2018: pp. 1–15.  
<https://doi.org/10.1155/2018/5456074> Letöltés dátuma: 2022.02.14.
- Vékás, P. (2019): Az élettartam-kockázat modellezése. Budapesti Corvinus Egyetem, Budapest. ISBN: 9789635037681.  
<https://www.who.int/data/data-collection-tools/who-mortality-database> Letöltés dátuma: 2021.01.29.
- WHO (2021a): WHO Mortality Database. [World Health Organization].  
<https://www.who.int/standards/classifications/classification-of-diseases> Letöltés dátuma: 2021.01.10.
- Worldbank (2021a): Healthy Life Expectancy, Years.  
[https://tcdata360.worldbank.org/indicators/h7a6911d2?country=BRA&indicator=41390&viz=line\\_chart&years=2017,2019#table-link](https://tcdata360.worldbank.org/indicators/h7a6911d2?country=BRA&indicator=41390&viz=line_chart&years=2017,2019#table-link) Letöltés dátuma: 2021.03.02.
- Worldbank (2021b): World Development Indicators.  
<https://datacatalog.worldbank.org/dataset/world-development-indicatorss> Letöltés dátuma: 2021.03.02.