

Költség eloszlások statisztikai modellezése cenzorált adatokon

Rakonczai Pál, Lang Zsolt, Csiha Julianna, Balázs Tamás, Dr. Bacskai Miklós
Healthware Tanácsadó Kft.

Betegutakhoz kapcsolódó fekvő- és járóbeteg ellátási költségek modellezését vizsgáltuk számítógépes szimulációs kísérlettel és az OEP betegkövető adatbázisának hepatitisz C betegkörén. A követési idő adminisztratív és informatív cenzorálásának torzító hatását túlélés elemzéssel, általános lineáris kevert modellekkel és összekapcsolt modellekkel mértük meg és hasonlítottuk egymással össze. Megállapítottuk, hogy a cenzorált adatok arányának és típusának megfelelően különböző statisztikai modellek alkalmasak a költségtrendek és az informatív cenzorálást okozó események kockázatának becslésére.

In this study the performance of different healthcare cost models have been investigated. In the first step a simulation study has been performed and then statistical models have been fitted to real cost data related to hepatitis C patients available in the NHIFA database. The biasing effect of administrative and informative censoring of follow-up time has been measured in several models such as survival/generalized linear mixed effect/joint models. The outcome of these models have been compared and lead to the conclusion that depending on the ratio and type of the censored data, there are different statistical models appropriate to estimate the trend of cost and risk of events related to informative censoring.

BEVEZETÉS

Az egészség-gazdaságtani és biztosításmatematikai elemzések kulcsfontosságú részét képezik a költségmodellek. Különösen így van ez a két terület határán elhelyezkedő kassza-hatás és kockázat-elemzések eseteiben, illetve az ezekben megfogalmazott előrejelzések vissza ellenőrzésére. Az egyszerű kiadásbecslések gyakran a becsült átlagos betegszám és erőforrás felhasználás trendjének figyelembevételével születnek, azonban az ellátási események és betegek heterogenitása jelentős szórást eredményeznek mind egyéni, mind a betegpopulációra aggregált kiadás értékek tekintetében. Az előrejelzések pontosításához a kezdeti paraméterek kontrollja, valamint az esetleges eltéréseket magyarázó újabb paraméterek beemelése szükséges. Mindezek rámutatnak az empirikus adatok feldolgozásának és az adatokra illesztett teoretikus, paraméteres költségmodellek szerepére. Kiemelten, ha kockázatmegosztó, biztosítás technikai eszközökkel élünk a finanszírozás kialakítása során és ezeket eltérően alkalmazzuk a szóban forgó betegpopuláció egyes alcsoportjaiban.

Mind a klasszikus ár-, mind az ár-teljesítmény alapú beszerzés és szolgáltatásvásárlás alkalmával szükség van az egyéni és csoportos kiadások tervezésére, s óhatatlanul szembesülhetünk a szezonaritás, akut-krónikus betegséglefolyás, a kezelést váratlanul befolyásoló események átmeneti vagy fatális hatásának problematikájával. További kihívást jelent, hogy mindebből csak néhány kezelhető egyszerűen az általában naptári évre szóló előrejelzésekben, a megfigyelések száma sokszor kevés beteghez tartozik, valamint a nem egy éves időszakra vonatkozó megfigyelés hatásainak éves tendenciáivá alakításakor figyelembe kell venni a követési idő adminisztratív és informatív cenzorálásának torzító hatását is.

Jelen tanulmány célja a hepatitisz C betegek egészségügyi ellátással kapcsolatos ápolási költségeinek időbeli tendenciájának modellezése. A költség alatt a fekvő- és járóbeteg kasszában keletkezett, az idő előrehaladtával kumulálódó összköltséget értjük, a jelentős nagyságrendi különbség miatt kihagyott gyógyszerterápia költsége nélkül. A betegutak követését meghatározott időablakban, 2005.01.01.-2009.12.31. között egy 387 betegből álló minta esetén vizsgáltuk. Ebben az időintervallumban a betegutakat balról is, jobbról is adminisztratív, betegúttól független módon cenzoráltuk. A betegek megfigyelését, költségeinek halmozódását az időablakon belül az elhalálozás is megszakíthatja. A halálozás időpontja és a költségek nagysága, növekedésének intenzitása függ a beteg állapotától, a betegút állomásaitól, ez a cenzorálás informatív. A cikkben összehasonlítottunk számos releváns statisztikai módszert, amelyek alkalmasak lehetnek a költség modellezésére a fenti típusú adatszerkezet esetén. Emellett vizsgáljuk, hogy a cenzorálás figyelembe vétele mekkora hatással lehet a modellek becslésére.

STATISZTIKAI MÓDSZEREK

A költségek időbeli alakulását adminisztratív cenzorálás mellett regressziós modellekkel vizsgálhatjuk [1, 2, 3]. Ezeknek a módszereknek az alkalmazásához fel kell tenni, hogy a vizsgált időszak megválasztása, a betegek időbeli követésének cenzorálása nincs kapcsolatban, statisztikailag független a költségek időbeli alakulásától. Ha a követés megállítása, a cenzorálás függhet a betegút eseményeitől, azaz informatív cenzorálásról van szó, akkor az idézett költségelemző módszerek torzított eredményeket adhatnak. Ilyen esetben a követést megszakító eseményig (pl. halálozásig) tartó idő hagyományos túlélés elemzéssel, pl. Cox-regresszióval történő vizsgálata szintén torzított becslésekre vezethet.

Költségszkalás módszerek

Egy adott időszak költségeinek modellezésére kézenfekvő megoldás az ún. általánosított lineáris modell (GLM) használata, amely számos eloszlást és skála transzformációt kínál fel alkalmazható opcióként, univerzális módszert szolgáltatva különböző betegség faktorok hatásának minél pontosabb feltárásához. Cenzorált költségeket is tartalmazó adat esetén viszont használata különös körülményeket igényel. A GLM mellett, a túlélés elemzés területén gyakran alkalmazott, ún. accelerated failure time (AFT) regressziós modell [4] is alkalmazható, költség skálán. Ez tulajdonképpen az időfüggvény monoton növekvő transzformációval való átszkalázása költség skálára, hiszen jogosan feltételezhetjük, hogy hosszabb idő alatt nagyobb költség termelődhet. Az AFT modell egy olyan paraméteres modell, amely a feltételes túlélési függvényt írja le $S(x|z) = S_0(x \exp(\beta'z))$ alakban, ahol β a paraméterek és z a kovariánsok vektora. A modell azt feltételezi, hogy a költség logaritmusai lineáris kapcsolatban van a várható értékkel (μ), a kovariánsokkal ($\beta'z$) és egy hibataggal (σW), adott W eloszlást feltételezve $\log(X) = \mu + \beta'z + \sigma W$. A modellek közötti lényegi eltérés az eloszlás megválasztásában van. Az AFT modell szemiparaméteres változata az ún. Cox-féle arányos kockázat (COX) modell [3, 4], amely az AFT módszertől eltérően az $S(x) = P(X \geq x)$ túlélés függvényt nem közvetlenül, hanem a $h(x) = -S'(x)/S(x)$ kockázati függvényen keresztül modellezi.

A költségszkalás modellek hátránya lehet, hogy a cenzorálás torzító hatását nem minden esetben szüntetik meg [5]. A torzítás nagyságának a cenzorálás mértékétől való függését számítógépes szimulációs kísérlettel vizsgáltuk meg. A kísérlet során egy általunk előre meghatározott ismert struktúrájú mintát generálunk, majd a fent említett modelleket illesztjük a mesterséges mintára. Az eljárás segítségével megvizsgálhatjuk, hogy melyik módszer adja legjobban vissza az ismert paramétereket. A szimulált struktúra tehát a következő: lognormális eloszlásmodellből generáltunk ($\beta = [0; 1; 2; 3]$ együttható vektorral) cenzorált mintákat (független cenzorálással) olyan módon, hogy a cenzorált költségek az egész minta kb. 30 százalékát tegyék ki. A szimulált adatra 4 modellt illesztettünk: lognormális eloszlású GLM, lognormális és Weibull eloszlású AFT, illetve COX modelleket. A becült paramétereknek az ismert β valódi értékeivel való összehasonlítását az 1. táblázat tartalmazza.

Elméleti β	AFT(lognorm.)	COX	AFT(Weibull)	GLM(lognorm.)
1,00	0,99 (0,03)	1,04 (0,04)	0,95 (0,04)	0,86 (0,04)
2,00	2,00 (0,04)	2,05 (0,04)	1,92 (0,04)	1,61 (0,03)
3,00	3,00 (0,03)	3,07 (0,06)	2,90 (0,04)	2,16 (0,03)
Cenzorálás figyelembevételével	IGEN	IGEN	IGEN	NEM

1. táblázat

Statistikai modellek összehasonlítása ismert paraméterekkel cenzorált minta esetén

A táblázat mutatja, hogy a cenzorálást figyelmen kívül hagyó GLM modell paramétereiben a legjelentősebb a torzítás. Ennél még a téves eloszlást feltételező, de cenzorálást figyelembe vevő AFT modell is jobban teljesít, azaz pontosabban adja vissza az eredeti paramétereket. Ugyanez a modell a helyes eloszlást feltételezve a legpontosabb. Érdekes eredmény továbbá, hogy a COX modell paraméteres eloszlás családot feltételezése nélkül is, kismértékű szisztematikus felülbecsléstől eltekintve, elfogadható becsléshez vezet. További vizsgálatokkal bizonyítottuk, hogy 30%-nál akár jóval magasabb cenzorálási arány mellett is elfogadható becslésekhez jutunk, feltéve, hogy az adat nem cenzorált költségű része elegendő statisztikai információt tartalmaz a vizsgált paraméterekre nézve.

A szimulációs kísérlethez képest a valódi adatok szerkezete annyiban összetettebb, hogy a rendelkezésre álló minta költségei egyenként más-más ütemben kumulálódnak, ami pozitív korrelációt eredményez az időszakonkénti költségnövekményekben. Ez a probléma megoldható úgy, hogy az egyén hatását véletlen (random) komponensként beépítjük a modellbe.

A kísérlet alapján megállapítható, hogy a cenzorálás hatásának vizsgálata nagy jelentőséggel bír. A következő fejezetben betegadatokon bemutatjuk, milyen modellekkel vizsgálható a költségek alakulása, ha mindkét faktor: a cenzorálás és az egyének véletlen hatása is nehezítik a modellezést.

Összekapcsolt modellek

A betegkövetést megállító eseményig tartó idő és a költségek halmozódása közötti kapcsolat erőssége, pozitív vagy negatív iránya betegenként és időszakonként eltérő lehet. Annak érdekében, hogy a költségek időbeli alakulását és a követést megszakító eseményig tartó időt egyaránt torzítatlanul becsülhessük, ezek közös modellkeretben való vizsgálatára van szükség [6, 7, 8]. Ezekben az ún. összekapcsolt regressziós modellekben (jointmodel) a költségeket és az eseményig tartó időt közös magyarázó változókkal, kovariánsokkal jellemezzük, a betegeket jellemző, szubjektív kapcsolatokat pedig közös beteg specifikus random hatásokkal vesszük figyelembe. A két részmodell összekapcsolását, integrálását tovább erősíti, ha a költségek beteg specifikus regressziós trendjét, időtől való függését időben dinamikus magyarázó változóként beépítjük az eseményig tartó időtartam Cox-regressziójába [9, 10]. A hepatitisz C betegek halálozással cenzorált költség-idő kapcsolatot Rizopoulos eljárásával modelleztük, az R 3.0.2 statisztikai szoftver 'JM' programcsomagjával [9]. A cenzorálást és/vagy annak informatív jellegét figyelmen kívül hagyó statisztikai becslések akkor alkalmazhatóak, ha a cenzorálás torzító hatása nem jelentős. Ennek mértékét megállapíthatjuk, ha a részmodellek egymástól független illesztésével kapott becsléseket összehasonlítjuk az összekapcsolt modell becsléseivel. A költségek időtől való függése ebben az esetben olyan lineáris kevert, fix és random hatásos modellre redukálódik, melyben az eltelt idő, mint fix, minden betegre

ugyanolyan hatású és mint random, beteg specifikus magyarázó változó is szerepet kap [11]. A halálozásig tartó idő önálló modellje paraméteres vagy szemiparaméteres túlélés elemzési modell lehet [3, 4, 12, 13]. A statisztikai számításokat az R 3.0.2 statisztikai szoftverrel végeztük el, a lineáris kevert modell (LME) illesztéséhez az 'nlme' programcsomagot [14], a túlélés elemzéshez a 'survival' programcsomagot használtuk [15]. Az összegzés során az LME modell paramétereit összevetettük egy általánosított lineáris kevert modell (GLME) paramétereivel is, amelynek az illesztését az 'lme4' programcsomaggal végeztük [14, 16].

KÖLTSÉG MODELLEZÉS VALÓSZ ADATOKON

A költségek modellezéséhez számos kovariáns állt rendelkezésünkre. Az elemzéshez a vizsgált típusú költséget szignifikánsan befolyásoló, diagnózisokat azonosító változókat válogattuk ki (p-érték<0.01): cirrózis, anémia, egyéb májbetegség, májelégtelenség és májdaganat. Három modellt illesztettünk: egy gamma eloszlású GLME költség modellt (logaritmus skálán), egy normális eloszlású LME költség modellt (logaritmus skálán) és egy JM modellt, ami tekinthető az LME költség modell cenzorált változatának, melyet a halálozásig eltelt idő COX modellje szabályoz. Az 1. ábra bal oldali grafikonján jól látszik, hogy a költségek kezdetben hirtelen megnövekednek, később a növekedés üteme egyre csökken. Az időt és a költséget is logaritmizálva az összefüggés hozzávetőlegesen lineárisra tehető, ez látható a jobb oldali grafikonon.

Az adatbázisban egy beteg költsége az idő függvényében kumulálódva van jelen, ami jelentős korrelációt vinne egy lineáris modellbe. Ezt orvosoltuk egy olyan kevert modellel, amelyben az egyének különbözőségét véletlen hatásként figyelembe lehet venni. A véletlen hatás trendjét az idő logaritmusától tettük függővé, amit az 1. ábrán látottak alátámasztanak.

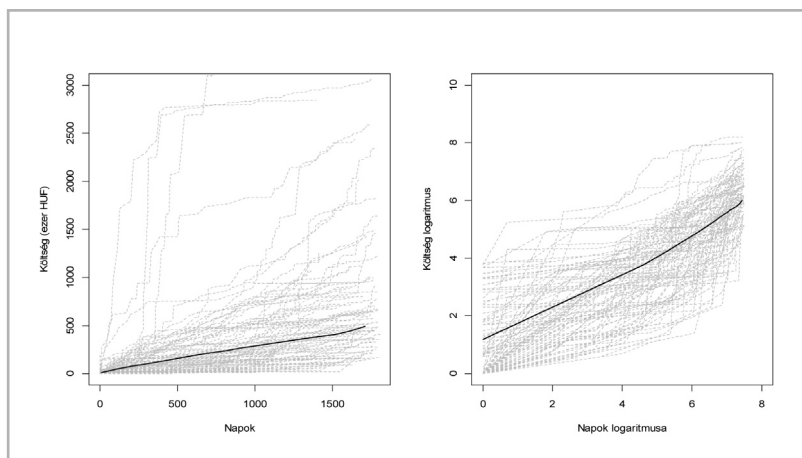
Az illesztett modellek paramétereit a 2. táblázatban foglaltuk össze. Látható, hogy a GLME modell paraméterei hasonlítanak az LME modell együtthatóihoz, bár a random

hatás gyengébben jelentkezik a GLME modellben. Mindkét modell alapján mondhatjuk, hogy a felsorolt betegségek diagnózisa és az eltelt napok száma növeli a várható költséget. A 2. ábrán 3 betegség diagnosztizálásának hatását vizsgáljuk. A bal oldali grafikon az LME modell által becsült költségeket ábrázolja. A legalacsonyabb költséggörbe azt az esetet illusztrálja, amikor nem diagnosztizálták egyik betegséget sem. Három magasabb görbe növekvő sorrendben a cirrózis, anémia és májdaganat diagnosztizálását követő időszak várható költségeit mutatja, végül a legmagasabb azt az esetet, amikor mindhárom betegséget diagnosztizálták. A GLME modell várható költségeinek az LME modelltől való

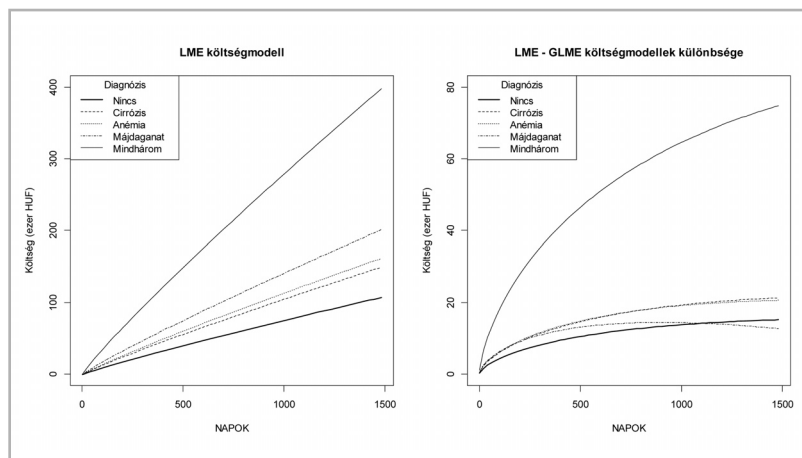
Költségmodell - Véletlen hatás	GLME	LME	JM
Konstans	0,6998	1,3395	1,3375
Napok logaritmusa	0,1097	0,2073	0,2069
Költségmodell - Fix hatás	GLME	LME	JM
Konstans	0,7923	0,5067	0,5067
Cirrózis	0,3317	0,3305	0,328
Anémia	0,3959	0,4080	0,4072
Egyéb májbetegség	0,5561	0,6266	0,6256
Májelégtelenség	0,7581	0,8319	0,8319
Májdaganat	0,6193	0,5677	0,5657
Napok logaritmusa	0,6023	0,6451	0,6453
Halálozás		COX	COX
Cirrózis		2,62	2,033
Költségtrend			0,4691
Cenzorálás figyelembe vétele	NEM	NEM	IGEN

2. táblázat
Hepatitisz C költség modellek becsült együtthatói

eltérése a jobb oldali panelon található, láthatjuk, hogy a lognormális eloszlást feltételező modell magasabb költségeket becsül, mint a gamma eloszlású GLME. A diagnózist követő egy évben 5-10 ezer forintos eltérés van a modellek között, ami mindhárom diagnózis meglétét követően már a 30 ezer forintos nagyságrendet is átlépi. A vizsgált populáció szintjén számolt becsült átlagos költségekben az eltérés



1. ábra
Kumulált költségek az idő függvényében



2. ábra
LME és GLME költségmodellek cirrózis, anémia és májdaganat diagnózist követően

mindössze 10 ezres nagyságrendű. Az LME modell populációra számított átlagos költségének idő trendjét az 1. ábrán vastag folytonos vonallal jelenítettük meg.

A halálig eltelt időt, mint cenzort figyelembe vevő JM költség modell paraméterei lényegében megegyeznek az LME modell paramétereivel, lényegesebb eltérést a COX modellben tapasztalhatunk. Itt új magyarázó változóként megjelenik a költségtrend, ennek pozitív együttműködője mutatja, hogy nagyobb kumulált költséghez nagyobb halálozási kockázat kapcsolódik.

MEGBESZÉLÉS, KÖVETKEZTETÉSEK

A számítógépes szimulációs kísérlet eredményéből megállapítható, hogy jelentősebb cenzorálási aránynál mindenképpen olyan modellt érdemes alkalmazni, amelyik a cenzorálás hatását csökkentő vagy azt megszüntető korrekciót tartalmaz. A valós adatokra épülő, hepatitisz C betegek időben kumulált költségeinek trendjét lényegében ugyanolyan hatékonyan tudtuk modellezni kevert (fix és random) hatásos line-

áris modellekkel (GLME és LME) és az adminisztratív cenzorálás és a halálozás kockázatát a költségtrenddel együtt közös modellkeretben vizsgáló ún. összekapcsolt modellel. Az összekapcsolt modell torzításmentesen távolítja el a cenzorálás hatását. Az ilyen korrekciót nem tartalmazó GLME és LME modellek eredményeivel való nagymérvű egyezés azzal magyarázható, hogy a vizsgált időszakban a halálozási arány viszonylag alacsony volt (8%). A költségtrend figyelembe vétele ugyanakkor lényegesen módosította, pontosabbá tette a halálozás kockázatának, házárdjának becslését. Ebben a vonatkozásban az összekapcsolt modell alkalmazása előnyösebb a többi módszerhez képest.

Az összekapcsolt modell használatával felállított költségmodellezés hasznos eszköznél bizonyulhat az ellátási területre kialakítható ápolásbiztosítási csomagok fenntarthatósági elemzésének vizsgálata során. Véleményünk szerint a krónikus betegségek több évre vonatkozó költség előrejelzésekor a cenzort figyelembe vevő költség modellek használata indokolt, különösen igaz lehet ez az alacsony prevalenciájú kórképek eseteiben.

IRODALOMJEGYZÉK

- [1] Lin DY (2000): Linear regression analysis of censored medical costs. *Biostatistics* 1(1), 35–47.
- [2] Lin DY (2003): Regression analysis of incomplete medical cost data. *Statist. Med.* 22, 1181–1200.
- [3] Lang Zs, Rakonczai P, Lazányi I, Lohn E, Bacskai M (2012): Eseményt követő kiadások eloszlásának és várható értékének modellezése Cox-féle arányos házárd modellel. *IME -Az egészségügyi vezetők szaklapja* XI:(1) 2012, 34–39.
- [4] Kalbfleisch JD, Prentice RL (2002): *The statistical Analysis of Failure time data*. 2nd Edition. Wiley.
- [5] Lin DY, Feuer EJ, Etzioni R, Wax Y (1997): Estimating Medical Costs from Incomplete Follow-UP Data. *Biometrics* 53, 419–434.
- [6] Liu L, Wolfe RA, Kalbfleisch JD (2007): A shared random effects model for censored medical costs and mortality. *Statist. Med.* 26: 139–155.
- [7] Liu L (2009): Joint modeling longitudinal semi-continuous data and survival, with application to longitudinal medical cost data. *Statist. Med.* 28, 972–986.
- [8] Xu H, Daggy J, Yu D, Craig BA (2013): Joint modeling of medical cost and survival in complex sample surveys. *Statist. Med.* 32, 1509–1523.
- [9] Rizopoulos D (2010): JM: An R Package for the Joint Modelling of Longitudinal and Time-to-Event Data. *Journal of Statistical Software*, 35(9), 1-33. URL <http://www.jstatsoft.org/v35/i09/>.

- [10] Rizopoulos D (2012): Joint Models for Longitudinal and Time-to-Event Data With Applications in R. Boca Raton: Chapman & Hall/CRC.
- [11] Brown H, Prescott R (2006): Applied Mixed Models in Medicine. Second edition. John Wiley & Sons.
- [12] Cox DR (1972): Regression models and life-tables (with discussion). J. Royal Stat. Soc. B, 34:187-220.
- [13] Therneau TM, Grambsch PM (2000): Modeling Survival Data. Extending the Cox Model. Springer.
- [14] Pinheiro J, Bates D, DebRoy S, Sarkar D and the R Development Core Team (2013): nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-111.
- [15] Therneau TM (2013): A Package for Survival Analysis in S. R package version 2.37-4, URL: <http://CRAN.R-project.org/package=survival>.
- [16] R Core Team (2013): R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>

A SZERZŐK BEMUTATÁSA



Rakonczi Pál alkalmazott matematikus, 2004-ben szerzett oklevelet az Eötvös Loránd Tudományegyetem Természettudományi Karán. 2004-től a Matematikai Intézet Valószínűségelmélet és Statisztika Tanszék doktorandusz hallgatója, 2007-től egyetemi tanársegéd. Főbb érdeklődési területei: többdimenziós összefüggési struktúrák, extrém érték elmélet, lineáris modellek. Jelenleg a Kutatás Üzletág vezetője.

Kutatás Üzletág vezetője.



Balázs Tamás a Szegedi Tudományegyetem Alkalmazott matematikus (BSc) alapszakán szerzett diplomát, később ugyanitt 2013-ban Pénzügyi alkalmazott matematikus (MSc) oklevelet. 2013 októberétől a Healthware Tanácsadó Kft kutatási elemzője. Kiemelt érdeklődési területe a matematikai szoftveres alkalmazások összetett modellezési eljárások során történő felhasználása.

szett modellezési eljárások során történő felhasználása.



Lang Zoltán jelenleg a Healthware Tanácsadó Kft. biostatistikusaként dolgozik. 1982-ben szerzett okleveles matematikus végzettséget az Eötvös Loránd Tudományegyetemen. 1983-1989 között az MTA Számítástechnikai és Automatizálási Kutató Intézet Geofizikai Osztályának, 1990-2008-ig az Abacom Kft., 1994-2008 között az

OEP Közgazdasági és Biztosításpolitikai Főosztály Statisztikai Osztályának munkatársa. 2001-2008 között a Semmelweis Egyetemen és az Országos Környezetegészségügyi Intézet Környezetepidemiológiai Osztályán dolgozik. 2007-2008-ban a Szent István Egyetem Állatorvos-tudományi Kar Biomatematikai és Számítástechnikai Tanszékének munkatársa. 1992-től a Nemzetközi Klinikai Biostatistikai Társaság Magyar Tagozatának tagja.

Csiha Julianna a Budapesti Corvinus Egyetem Gazdaságelemzés (BSc) alapszakán szerzett diplomát, majd 2013-ban az ELTE Társadalomtudományi karán Survey sta-

tisztikus (MSc) oklevelet. 2012. februártól a Healthware Tanácsadó Kft kutatási elemzője, ezt megelőzően 2009 és 2012 között piackutatóként dolgozott.



Dr. Bacskai Miklós a HealthWare Tanácsadó Kft. ügyvezető igazgatója. Diplomáit a Debreceni Orvostudományi Egyetem Általános Orvostudományi Karán (1999), és a Szegedi Tudományegyetem Közgazdaságtudományi Kará-

nak orvos-szakközgazdász szakán (2000) szerezte. 2000-től az Országos Egészségbiztosítási Pénztár Gyógyszerügyi Főosztályán dolgozott, 2001-2002 között elemzési osztályvezető, 2002-2004 között gyógyszerértékelési osztályvezető munkakörben. Szakmai gyakorlatait farmakoökonomia és információs rendszerszervezés területén szerezte.