

# A Központi Statisztikai Hivatal halálozási adatainak összevetése a Nemzeti Rákregiszter adatbázisával

*Egy adat-összekapcsolás tanulságai*

Wéber András dr.<sup>1, 2, 3</sup> ■ Szatmári István<sup>1, 3</sup> ■ Dobozi Mária<sup>1, 3</sup>  
Hilbert Lászlóné<sup>4</sup> ■ Branyiczkiné Géczy Gabriella<sup>4</sup> ■ Nagy Péter dr.<sup>1, 3, 5, 6</sup>  
Kásler Miklós dr.<sup>1, 7</sup> ■ Polgár Csaba dr.<sup>1, 3, 8\*</sup> ■ Kenessey István dr.<sup>1, 3, 9\*</sup>

<sup>1</sup>Országos Onkológiai Intézet, Budapest

<sup>2</sup>International Agency for Research on Cancer (IARC/WHO), Cancer Surveillance Branch, Lyon, France

<sup>3</sup>Nemzeti Tumorsejtbiológiai Laboratórium, Budapest

<sup>4</sup>Központi Statisztikai Hivatal, Budapest

<sup>5</sup>Állatorvostudományi Egyetem, Anatómiai és Szövettani Tanszék, Budapest

<sup>6</sup>Debreceni Egyetem, Onkokémiai Intézet, Debrecen

<sup>7</sup>Emberi Erőforrások Minisztériuma, Budapest

<sup>8</sup>Semmelweis Egyetem, Általános Orvostudományi Kar, Onkológiai Tanszék, Budapest

<sup>9</sup>Semmelweis Egyetem, Általános Orvostudományi Kar, Patológiai, Igazságügyi és Biztosítási Orvostani Intézet, Budapest

**Bevezetés:** Nemzetközi összehasonlításban a rosszindulatú daganatos incidencia és mortalitás tekintetében Magyarország az élvonalba tartozik. A halálozási statisztikát a Központi Statisztikai Hivatal a hatpéldányos Halottvizsgálati bizonyítványok feldolgozása alapján készíti el, míg az új daganatos betegségek előfordulásának populációalapú mérését a Nemzeti Rákregiszter végzi.

**Célkitűzés:** A Központi Statisztikai Hivatal és a Nemzeti Rákregiszter független adatbázisainak összekapcsolása rávilágíthat egymás gyengeségeire, emellett lehetőséget teremt az adatok verifikációjára, pontosítására, kiegészítésére, valamint a jelenlegi adatcsere bővítésének szükségességére.

**Módszer:** A Központi Statisztikai Hivatal 2012 és 2020 közötti halálozási adatait a társadalombiztosítási azonosító jel alapján kötöttük össze a Nemzeti Rákregiszter 2001 és 2020 közötti adatbázisával. A 2018-ra vonatkozó – főképpen tüdőrákos – halálozások mélyebb vizsgálatnak vetettük alá, mellyel az adatbevitel hiányosságai mellett a két állomány közötti eltéréseket is demonstráltuk.

**Eredmények:** A Központi Statisztikai Hivatal 2018-ra vonatkozó halálozási adatbázisában 32 586 esetben rosszindulatú daganat volt a statisztikai közlésre kiválasztott elsődleges halálok, melyből a Nemzeti Rákregiszterben 29 970-et azonosítottunk. A Központi Statisztikai Hivatal adatai között 8716, statisztikai közlésre kiválasztott tüdőrákos halálozástól 7957 személyt találtunk meg a Regiszterben. A 7957 egyezésből a Nemzeti Rákregiszterben 7381-hez tartozott tüdőrákos diagnózis. A fennmaradó 576 esetet a Regiszter más daganattal rögzítette, a leggyakrabban, 69 esetben tüdőrák szerepelt.

**Megbeszélés:** A két adatbázis közötti eltérés adódhat az adatfelvételek metodikai különbségeiből, a jelentési fegyelem problémáiból, a hiányos, pontatlan kitöltésből és a feldolgozási algoritmusok különbözőségéből. Mindazonáltal a vizsgált adatbázisok tartalmának döntő hányada értékes információt tartalmaz, ezért alkalmasak adattudományi vizsgálatokra.

**Következtetés:** A jelen elemzés tapasztalatai alapján a két intézmény közötti adatátadás felülvizsgálata várható. Emellett az elektronikus Halottvizsgálati bizonyítvány bevezetése vélhetően javítani fogja a társadalombiztosítási azonosító jel kitöltöttség megbízhatóságát, ráadásul a rendszerbe épített ellenőrzéseknek köszönhetően a kitöltés minősége javulhat, a feldolgozás ideje lerövidülhet.

Orv Hetil. 2022; 163(37): 1481–1489.

**Kulcsszavak:** daganatos incidencia, halandóság, 'big data', adatbázis-összekapcsolás

A \*-gal megjelölt szerzők egyenlő mértékben járultak hozzá a közleményhez

# Comparison of Hungarian Central Statistical Office's causes of death data with the database of the Hungarian National Cancer Registry

## *Lessons from a record linkage*

**Introduction:** In international comparison, Hungary is in the forefront of cancer incidence and mortality statistics. Based on paper-based death certificates, mortality statistics are compiled by the Hungarian Central Statistical Office, while population-based measures of cancer incidences are performed by the Hungarian National Cancer Registry.

**Objective:** Linking the records of these two independent databases can highlight their weaknesses and provide an opportunity to reconcile and verify collected data, which may emphasize the need to expand current data exchange protocols.

**Method:** Based on the Hungarian unique health care insurance ID, the mortality database of the Hungarian Central Statistical Office between 2012 and 2020 was compared with the data of the Hungarian National Cancer Registry from 2001 to 2020. Deaths in 2018, in particular those related to lung cancer, were examined in more depth to demonstrate the biases resulting from erroneous data collection.

**Results:** The mortality database of the Hungarian Central Statistical Office contained 32 586 cases with an underlying cause of death of malignant neoplasm for 2018, of which 29 970 were identified in the Hungarian National Cancer Registry. Out of the 8716 deaths coded to lung cancer, 7957 corresponding individuals were also found in the Registry. From the matches, 7381 cases were marked with lung cancer in the Hungarian National Cancer Registry. For the remaining 576 cases, the Registry recorded different types of cancers, of which in 69 cases with lung metastasis.

**Discussion:** The differences between the two databases may be caused by methodological differences in data collection, incomplete, inaccurate reporting and differences in processing algorithms. Nevertheless, the majority of the data in the examined databases were found to be appropriate for epidemiological studies.

**Conclusion:** Based on the outcomes of the present analysis, a revision of the data transfer between the two institutions is in order. The introduction of electronic Death certificate recording and validity checks are expected to improve the reliability of ID numbers and may shorten data processing times.

**Keywords:** cancer incidence, mortality, big data, record linkage

Wéber A, Szatmári I, Dobozi M, Hilbert L, Branyiczkiné Géczy G, Nagy P, Kásler M, Polgár Cs, Kenessey I. [Comparison of Hungarian Central Statistical Office's causes of death data with the database of the Hungarian National Cancer Registry. Lessons from a record linkage]. *Orv Hetil.* 2022; 163(37): 1481–1489.

(Beérkezett: 2022. május 5.; elfogadva: 2022. június 15.)

## Rövidítések

BM = Belügyminisztérium; BNO = Betegségek Nemzetközi Osztályozása; eHVB = elektronikus Halottvizsgálati bizonyítvány; EMMI = Emberi Erőforrások Minisztériuma; EüM = Egészségügyi Minisztérium; GLOBOCAN = Global Cancer Observatory; HVB = Halottvizsgálati bizonyítvány; IM = Igazságügyminisztérium; KSH = Központi Statisztikai Hivatal; NEAK = Nemzeti Egészségbiztosítási Alapkezelő; NRR = Nemzeti Rákregiszter; taj = társadalombiztosítási azonosító jel; WHO = Egészségügyi Világszervezet

A rosszindulatú daganatos betegségcsoport a cardiovascularis betegségek után a világon a második leg súlyosabb népegészségügyi problémát jelenti: a Global Cancer Observatory (GLOBOCAN) becslései szerint 2020-ban 19,3 millió új esetet diagnosztizáltak, emellett 9,9 millióan haltak meg rosszindulatú daganatos betegség következtében [1]. Magyarország hagyományosan a daganatosbetegség-listák élmezőnyébe tartozik, Európán belül a negyedik legmagasabb incidenciával és a második legnagyobb mortalitással [2].

Az Egészségügyi Világszervezet (WHO) által felállított GLOBOCAN különböző paraméterek figyelembevételével megbecsüli az egy évben előforduló új daganatos betegségek és halálozások számát, és arról adatot szolgáltat daganattípus, nem és ország szerinti bontásban. Mindazonáltal a populációalapú mérés jóval pontosabb adatokat szolgáltat – azonban nagyságrendekkel munkaigényesebb feladatot jelent. Magyarországon a daganatos betegségek előfordulását az Országos Onkológiai Intézet gondozásában működő Nemzeti Rákregiszter (NRR) és Biostatistikai Központ végzi [3], míg a halálozási statisztikát a Központi Statisztikai Hivatal (KSH) készíti el [4].

Az NRR működését a 49/2018. (XII. 28.) EMMI rendelet (betegségregiszter-rendelet) szabályozza (jogelődje: 24/1999. [VII. 6.] EüM rendelet); e szerint minden egészségügyi szolgáltató számára kötelező az adatjelentés, akinek tevékenységi körében a Betegségek Nemzetközi Osztályozásának 10. revíziója (BNO-10) szerinti C00–C97, D00–D09, D30.3, D33 tételeinek bármelyike esetén a diagnózist felállítják, kezelését végzik, továbbá szövettani vizsgálatot, boncolást végeznek

vagy véleményeznek. A gyakorlat szerint először a jelentő intézményben működő informatikai rendszerből leválogatják az adatokat, és az elektronikus jelentéseket az NRR számára egy webalapú programon keresztül továbbítják.

A KSH halálozási statisztikájának forrása a Halottvizsgálati bizonyítvány (HVB), melynek kitöltését és továbbítását a 351/2013. (X. 4.) Korm. rendelet szabályozza (jogelődje: 34/1999. BM–EüM–IM rendelet). Ezen a hatpéldányos, papíralapú dokumentumon a személyi adatok mellett a halál tényét és okát megállapító orvos tünteti fel a halálhoz vezető betegségek (események) sorozatát, illetve a kísérő betegségeket vagy állapotokat. A dokumentum első példányát a KSH részére kell megküldeni, amelynek feldolgozása a szöveges diagnózisok rögzítésével kezdődik. A halálóki kódolás és osztályozás során először BNO-kódot kell kapcsolni a HVB-n szereplő összes diagnózishoz, ezt követően kerül sor az elsődleges halálok kiválasztására. Ha több diagnózis szerepel a bizonyítványon, a kiválasztást meghatározza, hogy a kitöltő orvos azokat a nyomtatvány melyik részén tünteti fel. Ezt követően a KSH a HVB halálóki részének kitöltési helyességét a BNO-10 II. kötetének szabályai alapján ellenőrzi, és a hibásnak ítélt bizonyítványok közül a legkritikusabbakat visszaküldi a területileg illetékes kormányhivatal Népegészségügyi Főosztályára. A Népegészségügyi Főosztályok HVB-validálással megbízott

munkatársai vagy felveszik a kapcsolatot a kitöltő orvossal, vagy saját hatáskörben javítják a kitöltési hibákat. A statisztikai közlésre kiválasztott elsődleges halálok meghatározása – az ún. előztetés – a WHO szabályai [5] szerint történik, szoftver segítségével.

A betegségegyesítő-rendelet alapján a KSH minden évben átadja az NRR számára az előző évben elhunytak adatait, ennek részleteit a felek együttműködési megállapodásban határozzák meg. A kétféle adatbázis összekapcsolása és elemzése lehetőséget teremt új tudományos következtetések levonására. Ugyanakkor az NRR a regisztrációs évet követő június 30-ig a KSH részére aggregált formában átadja a felfedezett daganatos esetszámokat daganattípus, nem, korcsoport szerinti és megyei bontásban.

Mindazonáltal a 'big data' módszer validitását alapvetően meghatározza a bemenő adattartalom, vagyis a torzítás elkerülése végett elengedhetetlen az adatok tisztítása és a minőség ellenőrzése [6]. Ez utóbbi célból elemeztük a KSH halálóki adatbázisát, és a daganatos elhunytak adatait összevetettük az NRR adatbázisával. Figyelembe véve, hogy mind az incidenciát, mind a mortalitást tekintve a leggyakoribb rosszindulatú daganat a tüdőrák, részletesebb elemzéseink a populációra terjedtek ki [7]. A tapasztalatok alapján elemeztük, hogy a kérdésesként maradt esetek besorolására a halálozási adatbázis milyen további adataira lenne szükség.

**1. táblázat** | A Központi Statisztikai Hivatal halálozási adatbázisából származó éves esetszámok a 2012 és 2020 közötti időszakban, azonosíthatóság szerint (\*: a statisztikai közlésre kiválasztott elsődleges halálok BNO-10 szerinti C-s kódok; \*\*: a statisztikai közlésre kiválasztott elsődleges halálok BNO-10 szerinti C33/C34-es kódok) (Forrás: Központi Statisztikai Hivatal – Halálóki adatbázis)

Év	2012	2013	2014	2015	2016	2017	2018	2019	2020
<b>Összes rekord</b>	129 440	126 778	126 308	131 697	127 053	131 674	131 045	129 603	141 002
Tajjal	64 865	89 602	96 849	107 379	107 617	113 484	114 083	113 143	127 336
Taj nélkül	64 575	37 176	29 459	24 318	19 436	18 190	16 962	16 460	13 666
Ismeretlen tajok (900000007)	9	8	9	7	13	23	12	13	14
Jelentett tajarány	50,11%	70,68%	76,68%	81,53%	84,70%	86,19%	87,06%	87,30%	90,31%
Hibás tajok	597	647	684	760	730	579	566	577	593
Hibás tajok aránya	0,92%	0,72%	0,71%	0,71%	0,68%	0,51%	0,50%	0,51%	0,47%
<b>Rosszindulatú daganatos rekordok*</b>	33 224	32 748	32 748	32 792	32 987	32 844	32 586	32 012	31 623
Tajjal	16 829	23 049	25 096	26 723	27 999	28 436	28 537	28 132	28 760
Taj nélkül	16 395	9 699	7 652	6 069	4 988	4 408	4 049	3 880	2 863
Ismeretlen tajok (900000007)	3	4	1	1	2	1	1	2	2
Jelentett tajarány	50,65%	70,38%	76,63%	81,49%	84,88%	86,58%	87,57%	87,88%	90,95%
Hibás tajok	125	149	160	181	176	130	114	106	107
Hibás tajok aránya	0,74%	0,65%	0,64%	0,68%	0,63%	0,46%	0,40%	0,38%	0,37%
<b>Tüdő rosszindulatú daganatos rekordok**</b>	8 896	8 591	8 733	8 753	8 883	8 840	8 716	8 447	8 159
Tajjal	4 780	6 499	7 112	7 386	7 782	7 876	7 825	7 576	7 550
Taj nélkül	4 116	2 092	1 621	1 367	1 101	964	891	871	609
Ismeretlen tajok (900000007)	0	1	0	0	1	1	0	1	1
Jelentett tajarány	53,73%	75,65%	81,44%	84,38%	87,61%	89,10%	89,78%	89,69%	92,54%
Hibás tajok	32	42	48	52	42	38	28	22	23
Hibás tajok aránya	0,67%	0,65%	0,67%	0,70%	0,54%	0,48%	0,36%	0,29%	0,30%

taj = társadalombiztosítási azonosító jel; BNO = Betegségek Nemzetközi Osztályozása

## Módszer

Az együttműködési megállapodás szerint a KSH évente átadja az NRR számára az előző évben történt halálozások rekordszintű adatait: az elhunyt tajsámát (társadalombiztosítási azonosító jel), születésének és halálozásának dátumát, nemét, lakóhelyének és tartózkodási helyének KSH-kódját és a statisztikai közlésre kiválasztott elsődleges haláloknak BNO-10 szerinti kódját. A jelen elemzésbe a 2012 és 2020 közötti halálozási adatokat vontuk be. Elemeztük a tajhiányos, illetve ismeretlen tajos esetek arányát, valamint egy validálóprogram segítségével meghatároztuk a hibás tajok mennyiségét [8].

Részletesebb elemzés céljából a KSH 2018. évi haláloki állományát vetettük össze az NRR 2001 és 2018 közötti időszakra vonatkozó adatbázisával. A két adatbázis összekapcsolásának alapja a tajsám volt. Azokban az esetekben, amikor a taj hiányzott a KSH adatbázisából, valószínűségi rekordösszekapcsolást alkalmaztunk születési dátum, nem és a lakóhely irányítószáma szerint. A lakóhelyre vonatkozó KSH-kód és a való irányítószám közötti átváltást a KSH idevonatkozó kódtáblája alapján végeztük el. Az összekapcsolást követően elemeztük azon betegek arányát, akik a KSH-ban szerepeltek, az NRR részére azonban nem voltak jelentve, illetve összevetettük a két adatbázisban rögzített diagnózisokat, fókuszálva az elsődlegesen légcső-, hörgő-, tüdő- (BNO-10 szerinti C33–34) halálokra kódolt esetekre.

Ismertetett műveleteinket az R statisztikai programcsomag 4.1.2-es verziójával végeztük [9].

## Eredmények

A 2012 és 2020 közötti időszakban a KSH évente 130–140 ezer körüli halálesetet regisztrált (1. táblázat). A HVB-re 2012-ben került rá a tajsám, melyet az első évben a rekordok csupán felénél töltöttek ki. A tajsámmal ellátott rekordok mennyisége évről évre fokozatosan emelkedett, 2020-ra már 90%-os volt a kitöltöttség. Az ellenőrző algoritmus lefuttatása alapján a jelentett tajok 0,5–1%-a nem volt valós, ugyanakkor ez az érték az idővel szintén csökkenő tendenciát mutatott. Az ismeretlenként jelentett tajok száma elenyészőnek bizonyult.

A rosszindulatú daganatos halálozásokat tekintve (BNO-10 szerinti C-s kódok) a KSH alapján 2012 és 2020 között évi mintegy 32–33 ezer körüli esettel kell számolni. A tajtöltöttség aránya szinte teljes egészében megegyezett az összes elhunytnál tapasztalt arányokkal. A tüdő elsődleges rosszindulatú daganatait tekintve a KSH adatbázisa évi 8–9 ezer körüli esettel számol. A taj nélkül, illetve hibás tajjal jelentettek aránya megközelítőleg azonos volt az összes elhunyt, illetve az összes rosszindulatú daganatos elhunyt körében mért arányokkal. A tendencia ezekben az esetekben is érvényesült, miszerint a tajkitöltöttség évről évre nőtt, emellett a nem valós tajok száma csökkent.

Mivel az NRR kezdettől fogva taj alapján dolgozik, a taj töltöttsége már 2001-ben is meghaladta a 99%-ot (2. táblázat). Mivel a regisztrációs szoftverbe 2017-től egy tajellenőrző alkalmazást is bevezettek, ettől kezdve nem kell taj nélküli és hibás tajú esetekkel számolni. Ugyanakkor ebben az időszakban némileg megszorodtak az ismeretlen tajjal (900000007) jelentett esetek.

A KSH-nak a 2018-as évre jelentett 131 045 halálesetéből taj és valószínűségi összekapcsolás alapján 49 217 elhunytat találtunk meg az NRR adatbázisában. Ebből 19 247 eset a KSH adatbázisában nem daganatos halálokkal szerepelt. A daganatos halálokkal regisztrált esetek közül 26 561-et taj alapján, 3409-et valószínűségi összekapcsolással azonosítottunk. Az összes halálozásból 2616 daganatos elhunyt nem szerepelt az NRR-ben; ebből 1862 esethez érvényes taj tartozott, míg 754 esetben nem szerepelt vagy érvénytelen volt a taj.

A KSH adatbázisából a 2018. évi tüdőrákos elhunytakat részletesen elemezve megállapítottuk, hogy a 131 045 összes regisztrált eset 6,7%-a, 8716 halálozás történt elsődlegesen tüdőrák miatt (1. ábra). Ezek közül 891 elhalt esetén hiányzott a tajsám, illetve 28-nál hibásan szerepelt. A hiányos 919 rekord az összes tüdőrákos eset mintegy 10%-át jelentette. A 7797 érvényes azonosítójú esethez taj alapján kerestünk egyezéseket az NRR által rögzített, 2001 és 2018 között felfedezett 202 401 tüdőrákos esetből: 6788 egyezést találtunk, vagyis ezeket az eseteket jelentették az NRR-be tüdőrákosként, amely a halál okának bizonyult. A megmaradt 1009, érvényes tajsámmal rendelkező, tüdőtumoros halálozási rekordból 418 esethez az NRR-ből valamilyen más felfedezett primer daganat tartozott – vélhetően a tüdőáttét volt rosszul jelentve a HVB-n. A KSH 591 tüdőrákos halálozásával az NRR-ben taj alapján nem találtunk egyezést – ezen eseteknél az NRR irányába jelentési mulasztás történt.

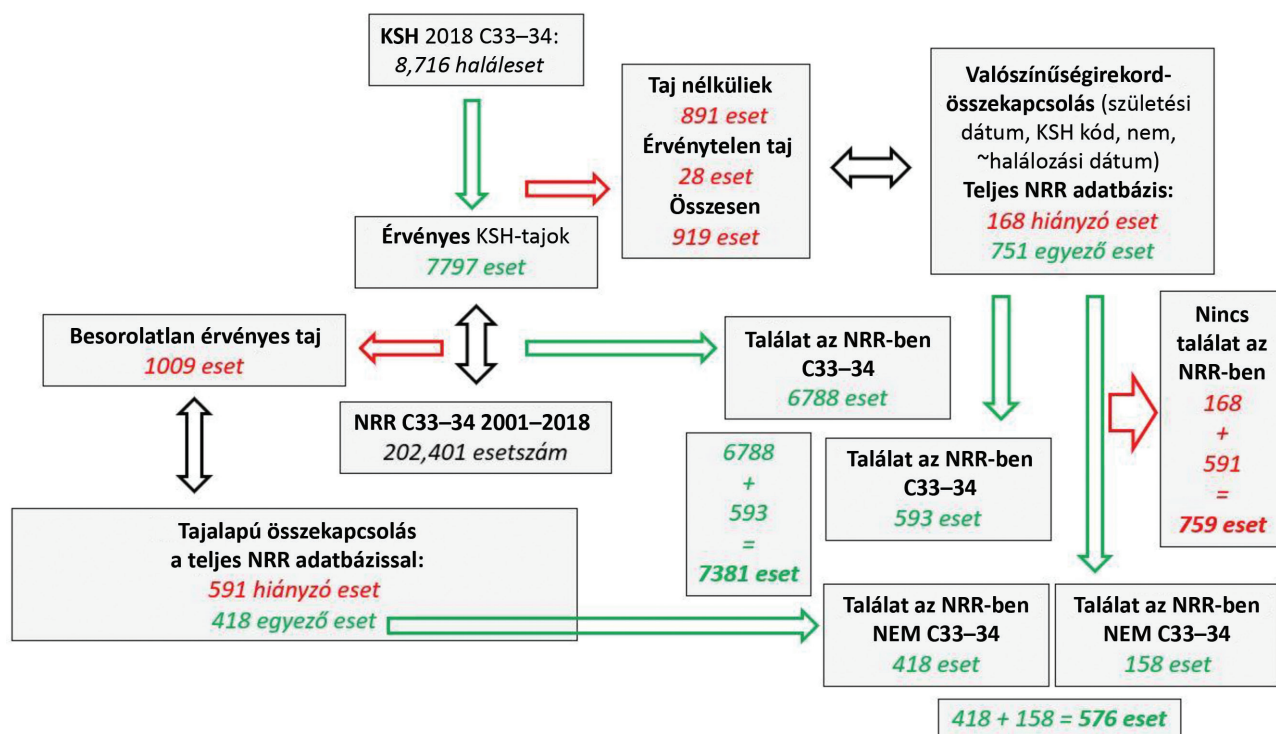
A 919 nem létező, illetve érvénytelen tajsámú esetet valószínűségi rekord-összekapcsolással próbáltuk párosítani az NRR adatbázisával felvett születési dátum, KSH területi kód nem és halálozási dátum szerint. Eszerint az NRR-ben 751 rekordot találtunk meg (80%-os találati arány), viszont 168 KSH-s tüdőtumoros halálozáshoz egyáltalán nem találtunk további információt. A 751 egyezés közül 593 esetben fordult elő tüdőrák az NRR-ben, míg 158-nál más tumoros diagnózis volt rögzítve.

Összegezve, a 2018-ra vonatkozó, KSH által közölt 8716, tüdőrákra kódolt halálesetből 7957 elhunytat találtunk meg az NRR-ben, 759-et nem. A 7957 egyezésből 7381 az NRR szerint is tüdőrákos volt. Ezek közül 492 esetről 1, 490-ről 2, 350-ről 3, 413-ról 4 és 5636-ról 5 vagy annál több jelentés érkezett. A fennmaradó 576 eset az NRR-ben nem primer tüdőrákként jelent meg: 69 eset légző- és emésztőszervi áttétként (primer nélkül), 60 eset a bőr nem melanocytar rosszindulatú daganataként, 57 eset egyéb áttétként (primer nélkül),

**2. táblázat** A Nemzeti Rákregiszter adatbázisának tájrlórttsége a 2001 és 2020 közötti időszakban (a Rákregiszter folyamatos adattisztítás alatt áll, emiatt az egyes számok változhatnak, minél régebbi az adat, annál kisebb mértékben) (Forrás: Nemzeti Rákregiszter)

	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
Összes személy	73 116	75 041	85 749	81 676	82 421	82 852	80 953	88 412	84 573	81 626	86 728	92 478	94 677	96 370	98 371	100 985	100 685	96 194	99 648	85 124
Tajjal	73 010	74 996	85 736	81 656	82 392	82 827	80 920	88 393	84 561	81 606	86 691	92 440	94 612	96 308	98 328	100 951	100 685	96 194	99 648	85 124
Taj nélkül	106	45	13	20	29	25	33	19	12	20	37	38	65	62	43	34	0	0	0	0
Ismeretlen tajok (900000007)	20	22	25	16	8	1	11	10	4	9	3	3	0	1	1	4	10	27	27	31
Jelentett tajarány	99,86%	99,94%	99,98%	99,98%	99,96%	99,97%	99,96%	99,98%	99,99%	99,98%	99,96%	99,96%	99,93%	99,94%	99,96%	99,97%	100,00%	100,00%	100,00%	100,00%
Hibás tajok	130	97	67	58	50	29	31	32	38	47	21	27	35	37	27	34	2	0	0	0
Hibás tajok aránya	0,18%	0,13%	0,08%	0,07%	0,06%	0,04%	0,04%	0,04%	0,04%	0,06%	0,02%	0,03%	0,04%	0,04%	0,03%	0,03%	0,00%	0,00%	0,00%	0,00%
Összes tűdórakos személy	11 176	11 042	11 637	10 943	11 034	11 169	11 174	11 757	11 075	10 410	10 977	11 240	11 133	11 437	11 748	12 037	12 217	9 571	11 297	9 839
Tajjal	11 173	11 041	11 637	10 942	11 033	11 167	11 170	11 753	11 074	10 410	10 977	11 239	11 133	11 429	11 748	12 035	12 217	9 571	11 297	9 839
Taj nélkül	3	1	0	1	1	2	4	4	1	0	0	1	0	8	0	2	0	0	0	0
Ismeretlen tajok (900000007)	5	4	4	3	0	0	1	4	1	5	1	0	0	0	0	1	0	1	0	0
Jelentett tajarány	99,97%	99,99%	100,00%	99,99%	99,99%	99,98%	99,96%	99,97%	99,99%	100,00%	100,00%	99,99%	100,00%	99,93%	100,00%	99,98%	100,00%	100,00%	100,00%	100,00%
Hibás tajok	9	3	2		2		2	4	1		1	2	1	1	3	3	0	0	0	0
Hibás tajok aránya	0,08%	0,03%	0,02%	0,00%	0,02%	0,00%	0,02%	0,03%	0,01%	0,00%	0,01%	0,02%	0,01%	0,01%	0,03%	0,02%	0,00%	0,00%	0,00%	0,00%

taj = társadalombiztosítási azonosító jel



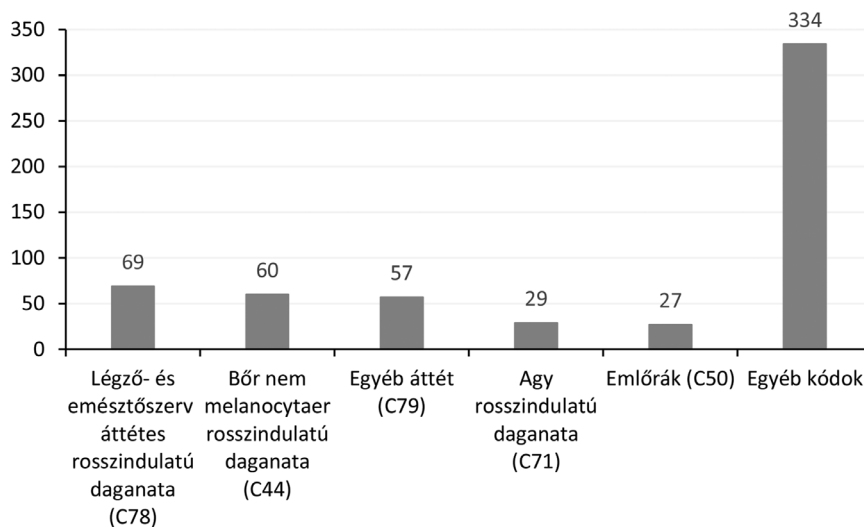
1. ábra A Központi Statisztikai Hivatal 2018. évi, statisztikai közlésre kiválasztott elsődleges tüdőrákos halálozásainak rekord-összekapcsolási diagramja a Nemzeti Rákregiszterrel

29 eset az agy rosszindulatú daganataként és 27 eset emlőrákként volt jelölve, míg további 334 esetben egyéb kódot találtunk (2. ábra).

## Megbeszélés

A megbízható halálozási statisztika segíthet felmérni egy-egy népegészségügyi probléma jelentőségét, így szolgáltatva adatokat a döntéshozók számára. Magyarországon a haláloki statisztika forrása a HVB. A taj HVB-n

való rögzítése a haláloki statisztikai munkának nem feltétele, az egyedi azonosító a HVB nyomtatványra 2012-ben került, az NRR kérésére. Elemzésünk alapján látható, hogy 2012 óta, ahogy egyre kevesebb régi nyomtatvány maradt forgalomban, a tajtöltöttség folyamatosan nőtt. Tekintve, hogy a taj egyedi személyazonosító, alkalmas lehet az adatbázisok összekapcsolására, ami segíthet az adattudományi elemzések kivitelezésében. Ugyanakkor a KSH-nak sem a validálásra, sem a pótlásra nincs lehetősége, mivel nem áll rendelkezésükre tajtörzsállo-



2. ábra A Központi Statisztikai Hivatal által statisztikai közlésre kiválasztott elsődleges tüdőrákos esetek száma, amelyek a Nemzeti Rákregiszterben nem primer tüdődaganatosként jelentek meg, besorolás szerint (Forrás: Nemzeti Rákregiszter)

mány. Ráadásul a betegségregistereknek való adatátadást követően a KSH törli a személyazonosításra alkalmas adatokat, a népmozgalommal összefüggő egészségügyi adatokat csak személyazonosításra alkalmatlan módon szabad kezelni [10, 11].

Ezzel szemben az NRR szempontjából alapvető fontosságú a taj, ugyanis a különböző jelentési sorok egymáshoz való felfűzése csak ennek ismeretében lehetséges. Emiatt a tajtöltöttség már a vizsgált időszak legelején is meghaladta a 99%-ot. 2017-től a jelentési gyakorlat teljes mértékben elektronikus, a webalapú rendszer nem is engedi a taj nélküli vagy hibás tajt tartalmazó sorok feltöltését. Ezzel szemben korábban az NRR papíralapú jelentéseket is elfogadott, illetve az adatbázis-építés során ellenőrző alkalmazás nem állt rendelkezésre. Ugyanakkor, mivel az üres és a hibás tajmezők jelentése lehetetlenné vált, 2017-től kezdve ezek az ismeretlen tajjal rögzített csoportban jelentek meg. Az NRR szoftvere későbbi jelentés alapján képes beazonosítani a korábban 900000007-tel jelentett eseteket. A teljes adatbázist nézve azonban a fennmaradó ismeretlen tajszerű esetek aránya elenyésző.

További eredményeink a két egymástól független adatfelvétel lényeges diszkrepanciáját fedték fel. Ezt támasztja alá, hogy a KSH által 2018-ban rögzített daganatos halálozások kb. 8%-ának nem volt nyoma az NRR-ben, vélhetően az eltérő metodika és jelentési fegyelem miatt. A tüdőrákos eseteket illetően a két adatbázis hasonló mértékű eltérése mutatkozott. Ugyanakkor a nem daganatos halálozások mintegy ötödénél találtunk tumoros jelentést is, vagyis a meglévő vagy gyógyult daganatos betegség a KSH statisztikájában közvetlen halálóki szereppel nem bír.

Ismert, hogy a világon a tüdőrák az egyik leggyakoribb rosszindulatú daganat, ráadásul ez a tumor felelős nemcsak a legtöbb korai halálozásért, hanem 5 éves túlélése még a magas jövedelmű országokban is igen alacsony [12]. Annak ellenére, hogy a dohányzás továbbra is jól meghatározható rizikófaktor a tüdőráknak, más tényezők – csakúgy, mint a légszennyezettség – is jelentős szerepet játszhatnak a kialakulásában [13]. Mindezek miatt társadalmi terhe az egyik legnagyobb az összes betegcsoport között [14], ennél fogva minden, tüdődaganattal kapcsolatos pontos és részletes statisztika fokozott jelentőséggel bír az egészségügyi költségtervezés szempontjából.

A GLOBOCAN becslései szerint 2020-ban 2,2 millió tüdődaganatot fedeztek fel a világon, ami az összes tumor 11,7%-át jelentette. Ugyanakkor a tüdőrákhoz 1,8 millió halálozás köthető, az összes tumoros halálozást 18,2%-a [2]. Nemzetközi összevetésben Magyarország mind az előfordulás, mind a halandóság tekintetében világelsőnek számít: a KSH statisztikái alapján Magyarországon évi 8–9 ezren halnak meg tüdőrák miatt, ami 2018-ban az összes daganatos halálozáson belül mintegy 27%-os részarányt jelentett, vagyis általában meghaladja a nemzetközi arányszámokat [15]. Tekintve, hogy

a GLOBOCAN-becsléseknél a mortalitás/incidencia arányszámokat veszik alapul, hazánk tüdődaganatos szempontból rendre a világon a legrosszabb kategóriában végez [2]. Ezeket az irreálisan magas számokat az utóbbi időben több hazai tanulmány igyekezett helyreigazítani a NEAK adatbázisának szisztematikus szűréseivel [16, 17]. Ezek az új megközelítések azonban nem képesek helyettesíteni a megbízható patológiai diagnózisokat figyelembe vevő és populációalapú adatgyűjtés eredményeit, amelyek kizárólag az egyes résztvevők (patológusok, klinikusok, adminisztrátorok, informatikai szolgáltatók, szakmai kollégiumok, NRR, KSH, EMMI) együttműködésével valósulhatnak meg [18].

Mind a halálozási, mind az incidenciát érintő statisztikákat meghatározza a bemeneti adatok minősége. Az NRR-be kerülő számos fals diagnózis egyik okát a BNO-kódolás hibái jelentik. Egyrészt az adatfelvételt végző személyzet nem tesz különbséget a daganatgyanú és a valós entitás között, másrészt más daganatos betegségeket is tüdőrákosnak kódolnak, halmozottan akkor, ha tüdőáttét is jelen van [17, 18]. A KSH halálozási statisztikáját kisebb mértékben érinti a primer tüdőrák tüdőáttéttel való összetévesztése, ugyanis az előztetési szabályok sok esetben korrigálják a helytelen kitöltést. A halálóki adatfelvétel bizonytalanságainak forrása lehet a halál okát megállapító orvos szubjektuma is. Eszerint, ha ismeretes, hogy az elhunyt tüdőrákja volt, akkor a kitöltő ezt gyakran járulékos halálókként tünteti fel, amennyiben az elhunyt nem volt dohányos, ha azonban az volt, akkor inkább elsődlegesnek. Ez a hazai és nemzetközi kutatások szerint is bizonyítottan alulbecsli a nem dohányzók és felül a dohányosok tüdődaganatos halálozási számát [19, 20].

Saját vizsgálataink során a 8716 tüdőrákos halálozásként nyilvántartott esetből csupán 7381-et igazolt vissza az NRR. A maradék esetek közül 576 az NRR-ben másik daganatos BNO-kóddal szerepelt, ez mintegy 7%-os eltérést jelentett. Ezen esetek egy része a HVB-ken hibásan szerepelhetett. Más részében a *post mortem* vizsgálat tisztázhatta a daganat valódi természetét, ám ezek jelentése az NRR felé való kötelezettség ellenére nem történt meg. Megjegyezzük, hogy a két intézmény közötti eredeti megállapítás alapján az NRR-nek eljuttatott, rekordszintű KSH-adatbázis a halál okát megállapító orvos személyét nem tartalmazta, tehát nem volt megállapítható, hogy csak klinikai vizsgálatra vagy kórboncolásra is sor került, vagyis milyen szinten történt meg a tüdőrákos diagnózis megerősítése. A jelen tapasztalatok alapján a szóban forgó megállapodás jövőbeli felülvizsgálata várható, a boncolásra, illetve a kitöltő orvosra vonatkozó információk kiegészítésével. Emellett a bizonytalan esetek (eltér a halálozási besorolás és az NRR-ben rögzített diagnózis) tisztázása többes halálóki vizsgálattal lenne lehetséges.

A 759, a KSH-ban tüdőrákos elhunytként, ám az NRR állományában egyáltalán nem szereplő beteget az idevonatkozó szakirodalom „csak Halottvizsgálati bizo-

nyítványból származó” eseteknek nevezi [21]. A potenciális *post mortem* diagnosztizált esetek ebből, illetve az NRR-ben más daganattal szereplő elhunytak összegéből (1335) kerülnek ki. A jövőbeli adatátadás annak eldöntését is megkönnyítheti, hogy ebben milyen arányban szerepelnek az NRR-be nem jelentett, hibásan kódolt, illetve a kórbonctani vizsgálattal felfedezett esetek.

Ugyan a nemzetközi trend alapján a kórboncolás szerepe egyre csökken, Magyarországon a boncolási arányszám az Európai Unió átlagosan 12%-os szintjéhez képest kiugróan magas: a vizsgált 2018. évben a 131 045 elhunyt közül 46 978 esetben a halál okát kórboncolás állapította meg (35,8%) [22]. A viszonylag nagy boncolási arány hozzájárul olyan esetek felfedezéséhez, amelyek más gyakorlattal rendelkező országok esetén rejtve maradnának. Vagyis a torzítás okát a hazai pontosabb metodika jelenti, ami viszont a többi országhoz képest magasabb esetszámokban nyilvánul meg. Mindazonáltal *Egerváry és mtsai* kutatása szerint hiába a magas arány, igen magas hazánkban a fals negatív és pozitív tüdődaganatos esetek aránya, azonban felvetették, hogy nemzetközi szinten is jelentős pontatlanság tapasztalható a tüdőrákos halálozásban [23].

## Következtetés

Megállapítottuk, hogy az NRR és a KSH halálozási adatbázisának összekapcsolása korlátozottan, de alkalmas adattudományi vizsgálatokra. Az eltérések az adatfelvételek különbségeiből, a jelentési és kitöltési fegyelem problémáiból, az adatrögzítés hibáiból és a feldolgozási algoritmusok különbözőségéből adódnak. A jelenlegi korlátok egy része a két intézmény közötti adatszere jövőbeli bővítésével, a többes halálokokra, boncolásra vonatkozó adatok átadásával csökkenthető. Tekintve, hogy a hatlapos, papíralapú HVB-t teljes körű funkcionalitás mellett hamarosan leváltja az elektronikus Halottvizsgálati bizonyítvány (eHVB), várhatóan javulni fog a kitöltés minősége, csökkenni fog a hiányosan, pontatlanul kitöltött bizonyítvány, azonban az orvosok szerepe, szakmai hozzáértése továbbra is alapvető fontosságú lesz [24]. Mindazonáltal az eHVB hatékonyságának értékelése már egy jövőbeli vizsgálat tárgyát fogja képezni.

**Anyagi támogatás:** A jelen munkát a Tématerületi Kiválósági Program (TKP2020-NKA-26, TKP2021-EGA-44), a Nemzeti Laboratóriumok Program (Nemzeti Tumorbiológia Laboratórium-2022-2.1.1-NL-2022-000 10), valamint a Nemzeti Népegészségügyi Stratégiával Összefüggő Feladatok (IV/4925/2021/EKF) támogatta.

**Szerzői munkamegosztás:** A vizsgálat megtervezése: W. A., K. I., N. P. Adatfeldolgozás és előkészítés: Sz. I., D. M., H. L. Statisztikai elemzés: W. A., K. I. A kézirat

szövegezése: W. A., K. I., N. P., K. M., P. Cs., H. L., B. G. G. A szerzők a cikk végleges változatát elolvasták és jóváhagyták.

**Érdekltségek:** A szerzőknek nincsenek érdekltségeik.

## Irodalom

- [1] Ferlay J, Colombet M, Soerjomataram I, et al. Cancer statistics for the year 2020: an overview. *Int J Cancer* 2021; 149: 778–789.
- [2] International Agency for Research on Cancer. Cancer today. Data visualization tools for exploring the global cancer burden in 2020. Available from: <http://gco.iarc.fr/today/home> [accessed: 2 March 2022].
- [3] National Institute of Oncology, Hungary – Hungarian National Cancer Registry. [Országos Onkológiai Intézet – Nemzeti Rákregiszter.] Available from: <https://onkol.hu/nemzeti-rakregiszter/> [accessed: 2 March 2022]. [Hungarian]
- [4] Hungarian Central Statistical Office. Data registration on vital events enacted by the law. [Központi Statisztikai Hivatal. Törvény által elrendelt adatgyűjtések.] Available from: [https://www.ksh.hu/2020\\_torveny\\_alta\\_elrendelt\\_adatgyujtesek](https://www.ksh.hu/2020_torveny_alta_elrendelt_adatgyujtesek) [accessed: 11 Febr 2022]. [Hungarian]
- [5] World Health Organization. International statistical classification of diseases and related health problems. 10th revision, fifth edition. WHO, Geneva, 2016.
- [6] Binder H, Blettner M. Big data in medical science – a biostatistical view. *Dtsch Arztebl Int.* 2015; 112: 137–142.
- [7] Kásler M, Ottó S, Kenessey I. The current status of cancer morbidity and mortality in light of the National Cancer Registry. [A rákmorbiditás és -mortalitás jelenlegi helyzete a Nemzeti Rákregiszter tükrében.] *Orv Hetil.* 2017; 158: 84–89. [Hungarian]
- [8] Acts and parliamentary regulations – Identification methods replacing the personal identification number and the use of identification codes, Act XX, 1996. [1996. évi XX. törvény a személyazonosító jel helyébe lépő azonosítási módokról és az azonosító kódok használatáról – Törvények és országgyűlési határozatok.] Available from: <https://mkogy.jogtar.hu/jogszabaly?docid=99600020.TV> [accessed: 2 March 2022]. [Hungarian]
- [9] The R project for statistical computing. Available from: <https://www.r-project.org/> [accessed: 2 March 2022].
- [10] Repertory of Legislations in Operation – management and protection of health and related personal data, Act XLVII, 1997. [1997. évi XLVII. törvény az egészségügyi és a hozzájuk kapcsolódó személyes adatok kezeléséről és védelméről – Hatályos Jogszabályok Gyűjteménye.] Available from: <https://net.jogtar.hu/jogszabaly?docid=99700047.tv> [accessed: 12 April 2022]. [Hungarian]
- [11] Repertory of Legislations in Operation – official statistics and data registration, Act CLV, 2016. [2016. évi CLV. törvény a hivatalos statisztikáról – Hatályos Jogszabályok Gyűjteménye.] Available from: <https://net.jogtar.hu/jogszabaly?docid=a1600155.tv> [accessed: 12 April 2022]. [Hungarian]
- [12] Arnold M, Rutherford MJ, Bardot A, et al. Progress in cancer survival, mortality, and incidence in seven high-income countries 1995–2014 (ICBP SURVMARK-2): a population-based study. *Lancet Oncol.* 2019; 20: 1493–1505.
- [13] The Cancer Atlas. Available from: <http://canceratlas.cancer.org/9WS> [accessed: 2 March 2022].
- [14] Fitzmaurice C, Abate D, Abbasi N, et al. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 29 cancer groups, 1990 to 2017. A systematic analysis for the global bur-

- den of disease study. JAMA Oncol. 2019; 5: 1749–1768. Erratum: JAMA Oncol. 2020; 6: 444. Erratum: JAMA Oncol. 2020; 6: 789. Erratum: JAMA Oncol. 2021; 7: 466.
- [15] Statistics | Eurostat. Data browser. Causes of death – standardised death rate by NUTS 2 region of residence. Available from: [https://ec.europa.eu/eurostat/databrowser/view/hlth\\_cd\\_asdr2/default/table?lang=en](https://ec.europa.eu/eurostat/databrowser/view/hlth_cd_asdr2/default/table?lang=en) [accessed: 11 Febr 2022].
- [16] Inotai A, Abonyi-Tóth Z, Rokszin G, et al. Prognosis, cost, and occurrence of colorectal, lung, breast, and prostate cancer in Hungary. Value Health Reg Issues 2015; 7: 1–8.
- [17] Bogos K, Kiss Z, Gálffy G, et al. Revising incidence and mortality of lung cancer in Central Europe: an epidemiology review from Hungary. Front Oncol. 2019; 9: 1051.
- [18] Kenessey I, Wéber A, Szilágyi I, et al. The application of medical coding systems in oncology – professional guideline according to the experiences of Hungarian National Cancer Registry. [Az orvosi kódtárak gyakorlati alkalmazása az onkológiában – szakmai útmutató a Nemzeti Rákregiszter tapasztalatai alapján.] Magy Onkol. 2022; 66: 4–10. [Hungarian]
- [19] Flanders WD. Inaccuracies of death certificate information. Epidemiology 1992; 3: 3–5.
- [20] Kendrey G, Szende B, Lapis K, et al. Misdiagnosis of lung cancer in a 2000 consecutive autopsy study in Budapest. Gen Diagn Pathol. 1996; 141: 169–178.
- [21] Bray F, Parkin DM. Evaluation of data quality in the cancer registry: principles and methods. Part I: Comparability, validity and timeliness. Eur J Cancer 2009; 45: 747–755.
- [22] World Health Organization. European health information gateway. A wealth of information at your fingertips. Autopsy rate (%) for all deaths. Last updated: 22 November 2021. Available from: [https://gateway.euro.who.int/en/indicators/hfa\\_545-6410-autopsy-rate-for-all-deaths/](https://gateway.euro.who.int/en/indicators/hfa_545-6410-autopsy-rate-for-all-deaths/) [accessed: 11 Febr 2022].
- [23] Egerváry M, Szende B, Roe FJ, et al. Accuracy of clinical diagnosis of lung cancer in Budapest in an institute specializing in chest diseases. Pathol Res Pract. 2000; 196: 761–766.
- [24] Repertory of Legislations in Operation – method related to post-mortem examination and on procedure related to the deceased, Decree 351/2013 (X. 4.). [351/2013. (X. 4.) Korm. rendelet a halottvizsgálatról és a halottakkal kapcsolatos eljárásról – Hatályos Jogszabályok Gyűjteménye.] Available from: <https://net.jogtar.hu/jogszabaly?docid=a1300351.kor> [accessed: 2 March 2022]. [Hungarian]

(Wéber András dr.,  
Budapest, Ráth György u. 7–9., 1122  
e-mail: weber.andras@oncol.hu)

„*lactum tacendo crimen facias acrius.*”  
(Ha már megesett, még nagyobb a bűn, ha titkolod.)