

Magyar klinikai vizsgálatok sajátosságai egy nemzetközi adatbázis elemzése alapján

Tóth Tamás¹ ■ Pollner Péter² ■ Palla Gergely² ■ Dinya Elek dr.¹

¹Semmelweis Egyetem, Egészségügyi Közzolgálati Kar, Digitális Egészségtudományi Intézet, Budapest

²MTA–ELTE Statisztikus és Biológiai Fizika Kutatócsoport, Budapest

Bevezetés: A ClinicalTrials.gov az Amerikai Egyesült Államok kormányzata által működtetett weboldal, amely a klinikai vizsgálatokkal kapcsolatos adatokat gyűjti. **Célkitűzés:** Vizsgálatunkban az oldalról XML fájlok formájában letölthető, magyar vonatkozású adatokat dolgoztuk fel. **Módszer:** Nagyobb mennyiségben a 2000-es évektől állnak rendelkezésre adatok, így jó áttekintő képet kaphatunk az elmúlt 10–15 év klinikai kutatásairól. Az adatmezők jelentős része szabad szöveges, ezért az adatok elemzése előtt jelentős munkát kellett fektetni az adattisztítás elvégzésébe. **Eredmények:** Az adatbázisban 2863 magyar vonatkozású klinikai vizsgálat szerepelt az elemzés elkészítésekor, amelyek 189 településen rendelkeznek vizsgálati hellyel. A konkrét intézményeket sokszor nem lehet azonosítani, a vizsgálati helyek alig ötödénél sikerült ez. A többi esetben csak valamilyen azonosító vagy általános megnevezés szerepel, így ezt a publikusnak szánt információt anonimá teszik. **Következtetés:** Vizsgálatunk az adatbázisból kinyerhető információk elemzése mellett rámutat arra, hogy milyen problémák befolyásolhatják a magyar klinikai kutatás helyzetéről kialakult nemzetközi képet. *Orv. Hetil., 2017, 158(9), 345–351.*

Kulcsszavak: adatbázis, klinikai vizsgálat, információkinyerés

Characteristics of clinical trials in Hungary based on the analysis of an international database

Intorduction: The ClinicalTrials.gov website, which is operated by the US government, collects data about clinical trials. **Aim:** We have processed data related to Hungary by downloading from the website as XML files. **Method:** Most of the data describe trials performed after 2000, so we got an overview about the clinical research of the last 10 to 15 years. As the majority of the data fields are collected as free text, significant data cleaning was needed. **Results:** The database contained 2863 trials related to Hungary from 189 settlements. Only 20 per cent of the actual research organizations could have been identified as many times only an “id” number or a general name was given, thus this information was anonymised in many cases. **Conclusion:** Besides the analysis of the information obtained from this database, our study points out the relevant issues that may influence the international view of the Hungarian clinical research.

Keywords: database, clinical trial, information extraction

Tóth, T., Pollner, P., Palla, G., Dinya, E. [Characteristics of clinical trials in Hungary based on the analysis of an international database]. *Orv. Hetil., 2017, 158(9), 345–351.*

(Beérkezett: 2016. november 14.; elfogadva: 2017. január 6.)

Rövidítések

FDAMA = Food and Drug Administration Modernization Act; MeSH = Medical Subject Headings; NIH = National Institutes of Health; NLM = National Library of Medicine

A nyílt hozzáférésű, interneten elérhető adatgyűjtemények egyre növekvő szerepet játszanak az egészségügyben és az orvosi kutatásban is. A legtöbb ilyen adatbázis a genetika, genomika, molekuláris biológia területéhez

kapcsolódik, de más tudományterületeken is jelen vannak. A szabadon elérhető információk fontosságát felismerve az Amerikai Egyesült Államok 1997-ben elfogadott, 105–115. számú törvénye (Food and Drug Administration Modernization Act – FDAMA) előírta egy olyan átfogó, nyilvánosan elérhető adatbázis létrehozását, amely információkat tartalmaz az országban folyó humán klinikai vizsgálatokról [1]. Ennek eredményeként jött létre a National Library of Medicine (NLM) által fejlesztett és üzemeltetett *ClinicalTrials.gov* weboldal, amely 2000. február 29-én indult el [2]. Az FDAMA által meghatározott kötelező adatszolgáltatás köre többször bővült, legutóbb 2016 szeptemberében, amely szabályozás 2017-ben lép hatályba [3]. Ezen jogszabályok hatálya alá tartozik a legtöbb 1-nél magasabb fázisú kontrollált klinikai vizsgálat, beleértve nemcsak a gyógyszervizsgálatokat, hanem egyes orvosi eszközök fejlesztését is. A jogszabályokon kívül különféle irányelvek írják elő az adatok feltöltését, például a National Institutes of Health (NIH) 2017-től megköveteli, hogy az általa részben vagy teljesen finanszírozott vizsgálatok adatai bekerüljenek az adatbázisba még akkor is, ha azt egyébként jogszabály nem teszi kötelezővé [4].

Az adatbázis elsősorban klinikai vizsgálatokat tartalmaz, de lehetőség van megfigyeléses vizsgálatok feltöltésére is. A kötelezően előírtakon kívül az önkéntes adatszolgáltatás is lehetséges, így a *ClinicalTrials.gov* mára az egész világra kiterjedő adatbázissá vált. A vizsgálati adatokat a kutatás szponzora vagy a kutatásvezető töltheti fel előzetes regisztrációt követően egy webes felületen keresztül [5].

Az adatbázis az alábbi témaköröket tartalmazza [6]:

- a vizsgálat címe és összefoglaló leírása,
- vizsgált betegség vagy állapot,
- beavatkozás (például a vizsgált gyógyszer, eszköz vagy módszer),
- a vizsgálat jellemzői (például vizsgálati fázis, típus),
- a résztvevőkkel szemben támasztott feltételek,
- a vizsgálatban részt vevő helyszínek adatai,
- releváns források (például publikációk).

Lehetőség – és bizonyos vizsgálatfajták esetén kötelezettség – van az eredmények feltöltésére is, amely magában foglalja:

- a résztvevők jellemzőit (a vizsgálatot elkezdő és befejező személyek száma, alapvető demográfiai jellemzőik),
- a vizsgálat fő eredményeit,
- a vizsgálat során észlelt mellékhatásokat.

Módszer

A *ClinicalTrials.gov* weboldalon különféle szempontok szerint lehet keresni és szűrni az adatbázisban tárolt vizsgálatokat. Lehetőség van a keresési eredmények letöltésére is, további feldolgozás, elemzés céljára. A letöltés egy tömörített állományformába történik, amely minden kiválasztott vizsgálatot különálló XML fájlban tartalmaz.

A magyar vonatkozású vizsgálatok letöltéséhez a legegyszerűbbnek a térképes nézet bizonyult, amelynek segítségével országonként (vagy akár régiónként, földrészenként) ki lehet listázni a kapcsolódó vizsgálatokat, majd ezeket le lehet tölteni.

Ahhoz, hogy statisztikai és informatikai módszerekkel feldolgozhatóvá váljanak az adatok, egy előfeldolgozást végeztünk. Ennek során egy erre a célra fejlesztett szoftver segítségével egyenként beolvastuk az XML fájlokat, a bennük talált információkat egy relációs adatbázisba töltöttük be. Az SQL nyelv használatával elvégzett lekérdezések eredményét pedig statisztikai szoftverbe importáltuk, és elvégeztük a kinyert adatok analízisét. Néhány adatmező esetében adattisztítást kellett végeznünk a pontosabb eredmény érdekében, amelyet az OpenRefine 2.6 ingyenes szoftver segítségével hajtottuk végre [7].

Jelen kutatás során vizsgáltuk:

- az adatok általános jellemzőit, minőségét, feldolgozhatóságát;
- a vizsgálatok időbeli és földrajzi megoszlását;
- a vizsgálatok témáját (betegségek, beavatkozások/gyógyszerek);
- a vizsgálatokban részt vevő települések közötti kapcsolatokat.

Eredmények

A vizsgálatok témáját illetően több adatelem is tartalmaz információkat, de ezek nem egyforma mértékben alkalmasak statisztikai kiértékelésre. Az egyik lehetőség, hogy kulcsszavakat lehet rendelni hozzájuk, azonban mivel ezeket szabad szöveges formában adhatja meg a feltöltő, a feldolgozásuk számos nehézségbe ütközik. Például ha egy vizsgálathoz több kulcsszó is tartozik, akkor az egyes kulcsszavakat külön adatelemként (úgynevezett XML tag-ként) kellene feltölteni, ám sok esetben egy tag tartalmazza az összes kulcsszót vesszővel vagy pontosvesszővel elválasztva. A kulcsszavak ezen elválasztó karakterek mentén történő feldarabolása sem jelent megoldást, mivel nem mindig ilyen célt szolgálnak: gyakori például a „Transplantation, renal” típusú kulcsszó megadása is, ahol a vessző után egy minősítő kifejezés szerepel, tehát nem két különálló kulcsszóról, hanem egy kifejezésről van szó. Több esetben komplett kifejezéseket, akár mondatokat adtak meg kulcsszó gyanánt, amelyek szintén nem felelnek meg a kulcsszavakkal szemben támasztott általános elvárásoknak, például egyértelműség.

Az XML fájlok szerkezetét leíró sémában létezik külön Condition (állapot) és Intervention (beavatkozás) mező is, de ezek is szabad szövegesek, így hasonló problémák állnak elő. Gyakori, hogy ugyanazon betegségnek különféle variánsait használják, például a 2-es típusú diabetes esetén legalább nyolc különféle változat jelenik meg az adatokban:

- diabetes mellitus type 2;
- diabetes mellitus, non-insulin-dependent;
- diabetes mellitus, type 2;

- diabetes mellitus, type II;
- diabetes, type 2;
- type 2 diabetes;
- type 2 diabetes mellitus;
- type II diabetes mellitus.

A harmadik szóba jöhető adatelem a vizsgálatokhoz rendelt, a Medical Subject Headings (MeSH) szótárból származó kulcsszavak, az úgynevezett MeSH term-ek, külön állapot és beavatkozás kategóriában. Mivel ezek egy előre meghatározott szótárból kerülnek kiválasztásra, ezért nem jelentkeznek az előzőekben bemutatott problémák. Ugyanakkor fontos megjegyezni, hogy ezek hozzárendelése automatikusan történik, egy számítógépes algoritmus segítségével, amely a vizsgálat többi adatmezőjében megjelenő kifejezéseket vizsgálja, és csak a legrelevánsabbnak ítélteteket rendeli hozzá az adatrekordhoz. Humán ellenőrzés hiányában előfordulhatnak ugyan hibák, de az algoritmust folyamatosan fejlesztik, finomhangolják, így ezen esetleges hibák hatását kisebbnek ítéltük, mint a szabad szöveges adatelemek fent említett problémáit, ezért a további elemzéseket a MeSH termeken végeztük el.

Összesen 2863 magyar vonatkozású (azaz legalább egy magyarországi vizsgálóhellyel rendelkező) vizsgálatot találtunk az adatbázisban.¹ Átlagosan két állapot- és két beavatkozásterm került hozzárendelésre egy-egy vizsgálathoz. A legmagasabb szám 12 állapot és 14 beavatkozás volt, azonban négyénél több kevés esetben szerepel. Figyelemre méltó, hogy közel 30% esetében egyetlen beavatkozásterm sem szerepel. Ha egy vizsgálathoz több kifejezés is hozzárendelésre került, akkor ezek legtöbbször vagy ugyanazon fogalom általánosabb és konkrét megnevezését jelentik (például cardiovascularis betegség és infarktus), vagy egy adott betegség különböző változatait jelölik (például hepatitis A és B), vagy pedig alapbetegség és szövődmény kapcsolatban állnak (például diabetes és diabeteses neuropathia).

708 különböző állapot- és 730 beavatkozásterm került hozzárendelésre a rekordokhoz. Az 1. táblázat a 20 leggyakoribb betegséget mutatja, ezek között megtalálható az arthritis, a diabetes, többféle daganatos betegség, valamint légzőszervi megbetegedések. A 2. táblázat a 20 leggyakoribb beavatkozást (többnyire gyógyszereket) mutatja. Nem meglepő módon az előbb felsorolt betegségek esetén alkalmazott gyógyszerek (például inzulin és egyéb, diabetes kezelésére szolgáló szerek) szerepelnek a lista élén.

Minden, az adatbázisban szereplő vizsgálathoz rögzítésre kellene, hogy kerüljön a tervezett kezdés és befejezés dátuma. Ez a legtöbb esetben meg is történik, de 14 vizsgálat esetében hiányzott a kezdő dátum, míg 34 esetben a befejezés dátuma. A legkorábbi, az adatbázisban szereplő vizsgálat 1993 májusában indult, de a legtöbb adat 2000 utánról származik. A vizsgálatok száma nö-

1. táblázat | A 20 leggyakoribb, betegséget leíró MeSH term

Sorszám	MeSH term	Vizsgálatok száma
1.	Arthritis	178
2.	Diabetes mellitus, type 2	170
3.	Arthritis, rheumatoid	151
4.	Diabetes mellitus	133
5.	Breast neoplasms	119
6.	Carcinoma, non-small-cell lung	116
7.	Lung neoplasms	100
8.	Pulmonary disease, chronic obstructive	86
9.	Asthma	75
10.	Lung diseases	68
11.	Lung diseases, obstructive	64
12.	Sclerosis	64
13.	Ulcer	62
14.	Colitis, ulcerative	57
15.	Prostatic neoplasms	56
16.	Colitis	56
17.	Hypertension	53
18.	Multiple sclerosis	53
19.	Crohn disease	46
20.	Kidney diseases	45

2. táblázat | A 20 leggyakoribb, hatóanyagot leíró MeSH term

Sorszám	MeSH term	Vizsgálatok száma
1.	Antibodies, monoclonal	106
2.	Methotrexate	74
3.	Insulin	70
4.	Metformin	64
5.	Insulin, globin zinc	63
6.	Paclitaxel	57
7.	Docetaxel	55
8.	Carboplatin	53
9.	Albumin-bound paclitaxel	52
10.	Cisplatin	48
11.	Insulin glargine	45
12.	Adalimumab	44
13.	Rituximab	42
14.	Tiotropium bromide	41
15.	Gemcitabine	37
16.	Trastuzumab	36
17.	Cyclophosphamide	35
18.	Etanercept	33
19.	Capecitabine	33
20.	Doxorubicin	32

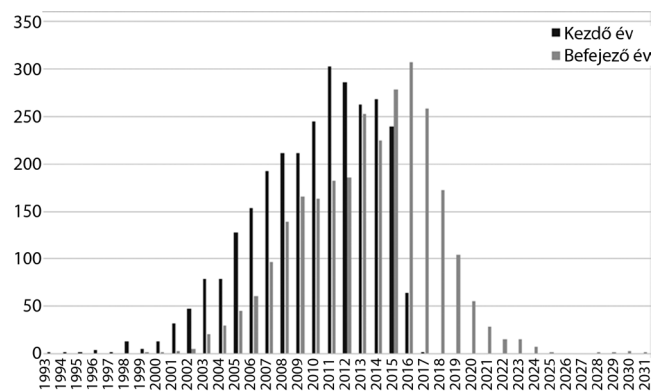
¹Az adatok letöltésének dátuma: 2016. július 29.

vekvő tendenciát mutat, ami azonban az utolsó néhány évben megállni, sőt kismértékben csökkenni látszik. (A 2016-os adatok még töredékesek, mivel a vizsgálatok jellemzően az indulás előtt nem sokkal kerülnek feltöltésre, és még egy előzetes jóváhagyáson is át kell esniük, mielőtt bekerülnek a nyilvános adatbázisba.)

A befejező dátumokat vizsgálva kétféle adatelem is rendelkezésre áll:

- „Primary completion date”: Az a dátum, amikor a vizsgálat elsődleges kimenetének mérése lezárul.
- „Completion date”: Az utolsó nyomon követési dátum vagy a vizsgálat befejezése (ez általában az előzőnél későbbi vagy azzal megegyező dátum, ha mindkét mező ki van töltve).

Amennyiben meg volt adva, úgy az elemzésnél a „completion date” értékét vettük figyelembe. Ha ez hiányzott, de a „primary completion date” ki volt töltve, úgy azzal számoltunk. A vizsgálatok befejezésének legkorábbi értéke 1999 áprilisa, a legkésőbbi 2031 júliusa volt az adatbázisunkban (1. ábra).



1. ábra | A vizsgálatok kezdetének és befejezésének évenkénti megoszlása

Mivel a kezdés és a befejezése dátuma is hónapban van megadva, a vizsgálatok hossza is hónap pontossággal számítható ki. 40 esetben nem állapítható meg a vizsgálat hossza, mivel vagy a kezdő, vagy a befejező dátum (esetleg mindkettő) hiányzik. Két esetben a két érték megegyezik, vagyis a vizsgálat egy hónapon belül zajlik le. A legmagasabb értéknek 443 adódik egy 1994. januárban kezdődő és 2030. decemberben záruló vizsgálat esetében. A vizsgálatok hosszának átlaga és szórása $41 \pm 32,5$ hónap, mediánja 33 hónap.

Az adatbázis tartalmazza az új gyógyszerek klinikai vizsgálatánál alkalmazott vizsgálati fázisra vonatkozó információt is. Az adatfeltöltésnél 1-től 4-es fázisig lehet kiválasztani, valamint megadható kombinált 1/2 és 2/3 fázis is. Amennyiben a fázis fogalma nem értelmezett az adott vizsgálat esetében (például nem gyógyszer, hanem eszközt vagy viselkedési beavatkozást vizsgálnak), akkor N/A jelölést kell megadni. Az adatbázisba feltöltött, magyarországi vizsgálóhellyel is rendelkező kutatók 55%-a 3-as fázisúnak, 23%-a 2-es fázisúnak, 7%-a

3. táblázat | Fázisvizsgálatok megoszlása

Fázisvizsgálatok típusa	Vizsgálatok száma
N/A	272
Fázis I	59
Fázis I/II	39
Fázis II	654
Fázis II/III	59
Fázis III	1573
Fázis IV	207

4-es fázisúnak van jelölve. A pontos vizsgálati számokat a 3. táblázat mutatja.

Amint az a 3-as fázisú vizsgálatok nagy arányából is látszik, a magyarországi intézmények elsősorban nagy, multicentrikus vizsgálatokban vesznek részt, amelyek számos magyar és külföldi helyszínen zajlanak egy időben. A vizsgálatokhoz összesen 13 924 magyarországi vizsgálati hely² tartozik. (Ez természetesen nem ennyi egyedi intézményt jelent, hiszen egy-egy egyetemen, kórházban számos klinikai vizsgálat zajlott a vizsgált időszakban.) A vizsgálati helyeket leíró adatmező tartalmazza az intézmény nevét, valamint címét (külön mezőkkel az ország, régió, város, cím számára). Ezen mezők tartalmát vizsgálva, a következő tipikus problémákat észleltük:

- Az intézmény nevének különféle változatait használják, időnként részben vagy egészben angolra fordítva (például Uzsoki utcai Kórház, Uzsoki utcai Hospital, Uzsoki Street Hospital, Uzsoki Hospital).
- Egyes vizsgálatoknál megadják az érintett szervezeti egység (osztály, egyetemi klinika) nevét is, máskor csak az intézményét.
- Egyes adatfeltöltők, főleg nagy gyógyszergyárak által végzett vizsgálatok esetén, csak egy vizsgálatihely-azonosítót adnak meg, vagy esetleg egy kontaktszemélyt, telefonszámot, ahol további felvilágosítás kérhető, és legfeljebb a település neve derül ki az adatokból, a konkrét intézmény nem.
- A városneveknél elütések, valamint az ékezetes karakterek eltérő változatai jelentik a legjellemzőbb hibákat.

Ezen problémák miatt az adatok elemzése előtt adattisztítást végeztünk a városneveket, valamint az intézményneveket tartalmazó mezőn. A települések többségét be lehetett azonosítani, itt elsősorban elírásokat tapasztaltunk, valamint néhány esetben városrészt is megadtak (például budapesti kerület számát vagy nevét). Az adattisztítás során az utóbbiakat eltávolítottuk, és csak a település megnevezését hagytuk meg. 65 rekord esetében nem tudtuk beazonosítani a települést sem, mivel ezeknél „Unknown”, „Many Locations”, „Various

²Vizsgálati hely alatt egy adott vizsgálatnak az adatbázisban szereplő helyszínét értjük. Az összesített értékben minden egyes intézmény annyiszor szerepel, ahányszor az adatbázisban előfordult a vizsgált időszakban.

Cities” és hasonló megnevezések szerepeltek. Összesen 189 település szerepel az adatbázisban, ezek térképes nézetét a 2. ábra mutatja. A helyszínek közül messze kiemelkedik Budapest 4117 vizsgálati hellyel (29,5%), ezt követik az egyetemi városok és a többi megyeszékhely. A 4. táblázat a 20 legtöbb hellyel rendelkező települést mutatja.

A konkrét intézmények tekintetében sokkal rosszabb a helyzet: a vizsgálati helyek felénél egyáltalán nem volt azonosítható az intézmény (például csak annyi szerepelt, hogy „Investigation site”), és további számos esetben csak a vizsgálatot végző gyógyszergyár neve derült ki az értékből. Összességében a vizsgálati helyek alig ötödénél volt egyértelműen azonosítható az intézmény. Az adat tisztítás során ezekben az esetekben csak az intézmény nevét hagytuk meg, az esetleges részleg, osztály, klinika nevét töröltük, így 299 különböző intézményt találtunk. A legtöbbször azonosítható intézmény a Semmelweis Egyetem volt, itt 304 vizsgálat zajlott, de ez is csak az összes rekord 2,2%-át jelenti. Sorrendben a három vidéki orvosegyetem következett 150–230 azonosítható rekorddal.

Az egy-egy intézményben végzett vizsgálatokhoz rendelt betegségeket és beavatkozásokat megvizsgálva képet kaphatunk az intézményben folyó tudományos kutatás profiljáról. A Semmelweis Egyetemre szűrve az adatokat például megállapíthatjuk, hogy a leggyakrabban előfor-



2. ábra | A vizsgálóhelyek földrajzi megoszlása. A pontok mérete arányos az adott településen lévő vizsgálóhelyek számával

duló kórképek egyes daganatos, valamint szív- és érrendszeri betegségek. A gyógyszerek közül viszont egyértelműen a daganatellenes szerek szerepelnek legnagyobb számban. Nem szabad azonban figyelmen kívül hagyni, hogy nagyszámú anonim intézmény van, amelyek között nagy valószínűséggel előfordul a Semmelweis Egyetem is, így a fenti adatok nem tekinthetők teljesen pontosnak.

A nemzetközi együttműködések vizsgálva megállapítottuk, hogy a magyar kutatóhelyek 113 másik országgal vesznek részt közös vizsgálatban. A vizsgálatok közül 114-hez mindössze egyetlen vizsgálóhely van megadva, míg a legmagasabb érték 1704 vizsgálóhely volt. Az egy vizsgálatra jutó helyszínek átlaga 105, mediánja 67.

A vizsgálatok közül 186 (6,5%) rendelkezik kizárólag magyarországi vizsgálóhellyel. A legtöbb országra kiterjedő vizsgálat 60 különböző országban rendelkezik vizsgálóhellyel. Az országok számának átlaga 14, mediánja 12. A leggyakrabban előforduló partner országok: Amerikai Egyesült Államok, Németország, Spanyolország, Lengyelország, Franciaország. Érdemes megemlíteni, hogy annak ellenére, hogy amerikai adatbázisról van szó, a vizsgálatok 40%-a nem rendelkezik amerikai vizsgálóhellyel. Ez azt mutatja, hogy az adatbázis nemzetközileg is meghatározóvá vált, és a kutatók fontosnak tartják a benne való megjelenést.

A települések közötti kapcsolatok vizsgálatára hálózat-elemzési módszereket alkalmaztunk. Első lépésként egy gráfot készítettünk, amelynek csúcsai a települések. Két csúcs között akkor fut él, ha van olyan vizsgálat az adatbázisban, amely mindkét településen rendelkezik vizsgálóhellyel. A gráf 7361 élt tartalmaz, tehát az eredmény egy meglehetősen sűrű gráf lett. Az élekhez súlyszámot rendeltünk, amely a közös vizsgálatok számát mutatja. Az élek több mint fele (3760 él) 1-es súlyszámot kapott, és 90%-uk súlyszáma legfeljebb 10. A legmagasabb érték 882, amely Budapest és Debrecen, tehát a klinikai kutatásban leginkább élenjáró két város között húzódik. A 10 legnagyobb súlyszámú él, azaz a legtöbb közös projektben részt vevő várospárok listáját az 5. táblázat mutatja.

4. táblázat | A legtöbb vizsgálóhellyel rendelkező települések

Sorszám	Város neve	Vizsgálóhelyek száma
1.	Budapest	4117
2.	Debrecen	1131
3.	Szeged	728
4.	Pécs	628
5.	Győr	485
6.	Miskolc	483
7.	Nyíregyháza	436
8.	Gyula	363
9.	Székesfehérvár	344
10.	Zalaegerszeg	299
11.	Szombathely	294
12.	Kecskemét	283
13.	Kaposvár	267
14.	Veszprém	261
15.	Szolnok	257
16.	Szekszárd	185
17.	Balatonfüred	183
18.	Eger	153
19.	Békéscsaba	141
20.	Törökbálint	128

5. táblázat | A gráf legmagasabb súlyszámú élei (a közös vizsgálatok száma)

1. város	2. város	Súlyszám
Budapest	Debrecen	882
Budapest	Szeged	591
Budapest	Pécs	507
Budapest	Győr	407
Budapest	Miskolc	382
Budapest	Nyíregyháza	374
Debrecen	Szeged	372
Budapest	Gyula	322
Debrecen	Pécs	321
Budapest	Székesfehérvár	299

Ugyanezt az eredményt adta a *k-core*-elemzés is. Ezt a módszert gyakran alkalmazzák sűrű, sok éllel rendelkező gráfok elemzésére, magok, összetartozó csoportok keresésére. Egy gráf *k*-magja egy olyan maximális részgráf, amelynek minden csúcsából legalább *k* számú él fut a részgráfon belüli csúcsokhoz [8]. A teljes gráfot vizsgálva, a legmagasabb *k*-érték 74-nek adódott, amely mag 75 csúcsot (39,7%) és 2775 élt (37,7%) jelent.

A fenti vizsgálatot úgy is elvégeztük, hogy súlyszám szerint szűrtük az éleket. Először a legalább 5 súllyal rendelkező élekre szűkítettük a gráfot, ekkor 1147 él maradt, amelyek 88 csúcsot kötnek össze. Ebben a részgráfban a legmagasabb *k*-értéknek 31 adódott, amely mag 34 csúcsot és 555 élt tartalmaz. Legalább 10 súlyszámmal rendelkező élekre történő szűrés esetén a gráf 69 csúcsot és 814 élt tartalmaz. Ekkor a *k-core*-elemzés eredménye: *k*-érték 24, csúcsok száma 27, él szám 346. Mindegyik elemzés esetén a gráfnak egy összefüggő magja maradt meg, nem vált szét különálló csoportokra.

Megbeszélés

A ClinicalTrials.gov weboldal fontos szerepet tölt be a klinikai vizsgálatokkal kapcsolatos információk összegyűjtésében és publikálásában. Bár eredetileg az Amerikai Egyesült Államok adatbázisának indult, ma már az egész világról töltnek fel bele adatokat. Nemcsak olyan vizsgálatokat tartalmaz, amelyek amerikai irányításúak és más országokban is rendelkeznek vizsgálóhellyel, hanem szerepelnek benne például kizárólag Magyarországon végzett kutatások is. Nagyobb számban a 2000-es évektől kezdve kerültek feltöltésre az adatok, így ha nem is teljes körű, de jó áttekintő képet ad az elmúlt 10–15 év magyarországi klinikai kutatásairól. A kutatásunk során elsősorban azt vizsgáltuk, hogy az egyedi vizsgálatokra vonatkozó adatokat hogyan lehet összevetni, milyen módon lehet belőlük összefoglaló leíró statisztikai kiértéke-

léseket végezni, és ezek mit mutatnak a magyarországi klinikai kutatás helyzetéről, fókuszpontjairól.

Az adatok feltöltéséért a vizsgálatot végző szervezet a felelős, és egy online űrlap kitöltésével végezhető el. Az adatmezők jelentős része szabad szöveges, emiatt az adatok minősége nem mindig megfelelő. Különösen szembevetű ez például a városneveknél, ahol (feltehetően a nem mindig anyanyelvi feltöltők miatt is) gyakran fordulnak elő elütések. Ezért az adatok elemzése előtt jelentős munkát kellett fektetni az adattisztítás elvégzésébe. A vizsgálatokat végző intézmények nevével még rosszabb a helyzet, mivel számtalan variációban szerepelnek az elnevezések: magyarul, részben vagy egészben angolra fordítva, osztály/klinika nevének megadásával vagy a nélkül, teljes hivatalos vagy rövid névvel stb. Nem segít a helyzeten az sem, hogy (elsősorban Budapesten) a folyamatos átszervezések miatt többször is megváltozott egyes intézmények neve. Szintén jellemző, hogy – talán mert a nagy gyógyszergyártók üzleti titokként kezelik az információt – nem adják meg a vizsgálati hely nevét, hanem csak egy belső azonosítót, így anonimizálva a publikusnak szánt adatot. Emiatt intézményszintű kiértékelést nem tudunk végezni, legfeljebb településszintűt.

A fent említett hiányosságok ellenére számos információt sikerült kinyernünk az adatbázisból például a magyarországi klinikai kutatások földrajzi eloszlásáról vagy a jellemző vizsgálati témákról. Megállapítható, hogy az összes vizsgálóhely közel egyharmada Budapesten található, és nem meglepő az sem, hogy a rangsorban a vidéki orvosi egyetemek városai, valamint a megyeszékhelyek következnek. (Elsősorban a megyei kórházakban folyó kutatásoknak köszönhetően.) Az országban számos olyan kisebb település is van, ahol vagy a helyi egészségügyi intézmény, vagy akár egy-egy praxis kapcsolódik be a tudományos kutatásba. A hálózatelemzés eredménye azt mutatja, hogy szoros az együttműködés a magyarországi kutatóhelyek között, sokszor vesznek részt közös projektekben. A kutatási helyek egyetlen sűrű csoportot alkotnak, nincs szegregáció, azaz a közösen részt vevő városok a geográfiai elhelyezkedés szerint nem válnak szét egymástól független, kooperációt kerülő csoportokra.

Következtetés

A leggyakrabban kutatott betegségek közé az arthritis, a diabetes, többféle daganatos betegség, valamint légzőszervi megbetegedések tartoznak. A klinikai vizsgálatok fázisait tekintve Magyarországon legnagyobb arányban a 3-as, kisebb részben a 2-es fázisú vizsgálatok történnek. A 4-es fázisú vizsgálatok aránya 10% alatti, míg az 1-es fázisú vizsgálatok száma meglehetősen alacsony.

Az interneten elérhető, nyilvános adatbázisok egyre növekvő jelentőséggel bírnak, és fontos információforrássul szolgálhatnak. Kutatásunkban az elsősorban egyedi klinikai vizsgálatok bemutatására szolgáló *ClinicalTrials.gov* adatbázisából sikerült aggregált információkat kinyernünk, ám ehhez némi programozásra is szükség

volt. Mivel az adatok számos, különböző országokban élő adatfeltöltőtől érkeznek, problémát jelent az adatok nem mindig megfelelő minősége, az egységesség hiánya. A nyílt adatbázisokat sokan használják és elemzik, ezért könnyebb felfedezni az esetleges hibákat, hiányosságokat. De ez akkor válik igazán hasznossá, ha van olyan csatorna, amelyen keresztül jelezni lehet ezeket akár az adatfeltöltő, akár az oldal üzemeltetője felé. Az adatminőség javításához hozzájárulhatna többek között az adatbevitel további strukturálása, a szabad szöveges mezők számának csökkentése, vagy pedig a feltöltők munkáját segítő irányelvek, útmutatások megfogalmazása, amelyek például egységesítik az intézmények megnevezésének megadását (angol vagy eredeti nyelvű név használata, ékezetes nevek vonatkozó ajánlások stb.). Az elemzésünk során feltárt hiányosságok, tapasztalatok hasznosak lehetnek a hazai adatgyűjtő archívumok, regiszterek fenntartóinak és tervezőinek is, hiszen jelzik, hogy mire kell odafigyelni, és megmutatják, milyen hibák csúszhatnak egy jó kezdeményezésbe.

Anyagi támogatás: A közlemény megírása, illetve a kapcsolódó kutatómunka anyagi támogatásban nem részesült.

Szerzői munkamegosztás: T. T.: A szoftver programozása, adatelőkészítés, adattisztítás, statisztikai elemzések elvégzése, a kézirat megszüvegezése, az ábrák elkészítése. D. E.: Az eredmények ellenőrzése, a kézirat végső változatának megszüvegezése. P. P., P. G.: A vizsgálati célok megfogalmazása, hálózatelemzési módszertan kialakítá-

sa. A cikk végleges változatát valamennyi szerző elolvasta és jóváhagyta.

Érdeklőségek: A szerzőknek nincsenek érdeklőségeik.

Irodalom

- [1] Food and Drug Administration Modernization Act of 1997, Public Law No. 105–115, 1997, 111 Stat. 2310
- [2] *National Library of Medicine*: Press Release: National Institutes of Health Launches “ClinicalTrials.gov”, 2000. Available from: https://www.nlm.nih.gov/archive/20040831/news/press_releases/clntrlpr00.html
- [3] *National Institutes of Health, Department of Health and Human Services*: Clinical Trials Registration and Results Information Submission. 2016. Available from: <https://www.federalregister.gov/d/2016-22129>
- [4] Hudson, K. L., Lauer, M. S., Collins, F. S.: Toward a new era of trust and transparency in clinical trials. *J. Am. Med. Assoc.*, 2016, 316(13), 1353–1354.
- [5] Zarin, D. A., Tse, T., Williams, R. J., et al.: Trial reporting in ClinicalTrials.gov – the final rule. *N. Engl. J. Med.*, 2016, 375(20), 1998–2004.
- [6] *National Library of Medicine*: ClinicalTrials.gov Background, 2016. Available from: <https://ClinicalTrials.gov/ct2/about-site/background>
- [7] OpenRefine [Computer software], 2015. Available from: <http://openrefine.org/>
- [8] Seidman, S. B.: Network structure and minimum degree. *Soc. Networks*, 1983, 5(3), 269–287.

(Tóth Tamás,
Pf. 2, 1428 Budapest,
e-mail: toth.tamas@public.semmelweis-univ.hu)

Eladó praxis!

Budapest XX. kerületében **500 fős gyermekorvosi praxis** betegség miatt eladó.

Éjszakai ügyelet van.

Érdeklődni az alábbi telefonszámon lehet:

+36 30 999 59 72