

Önidézetek: a tudománymetria mostohagyermekai?

Schubert András

MTA Könyvtár és Információs Központ, Tudománypolitikai és Tudományelemzési Osztály, Budapest

Az önhivatkozásokat, önidézeteket sokan a hiúság jelének tartják, és mellőznék őket a tudománymetriai elemzések-ből. Valójában az önidézetek más, de nem kevésbé értékes információkkal szolgálnak a tudományos kommunikáció vizsgálatában, mint a másoktól kapott idézetek. Az önidézés gyakorlata súlyos etikai kérdéseket is felvet, amelyekre a választ a szerzők, szerkesztők, kiadók együttműködése adhatja meg. *Orv. Hetil.*, 2016, 157(32), 1289–1293.

Kulcsszavak: tudománymetria, önhivatkozások, önidézetek

Self-citations: the stepchildren of scientometrics?

Self-references, self-citations are considered by some as the sign of vanity, and deemed to be omitted from scientometrics analyses. In fact, self-citations reveal information in the study of scientific communications that is different but not less valuable than citations received from others. In the practice of self-citation severe ethical issues may emerge that can be handled by the co-operation of the authors, editors and publishers.

Keywords: scientometrics, self-references, self-citations

Schubert, A. [Self-citations: the stepchildren of scientometrics?]. *Orv. Hetil.*, 2016, 157(32), 1289–1293.

(Beérkezett: 2016. február 25.; elfogadva: 2016. március 17.)

Ennek a cikknek a címét először harmincegynéhány évvel ezelőtt írtam le egy papírra, igaz, akkor még kérdőjel nélkül. Valóban, az 1980-as évek közepéig a szakirodalomban összesen két cikk foglalkozott az önidézetekkel: *Tagliacozzo* [1] a növényélettan és a neurobiológia területén vizsgálta az önidézetek gyakoriság- és koreloszlását, *Lawani* [2] pedig az önidézet és az önhivatkozás mérése közötti alapvető különbségre hívta fel a figyelmet (lásd alább).

A téma szakirodalma az 1990-es években indult folyamatos növekedésnek, és mára a Thomson-Reuters Web of Science adatbázisa közel 500 cikket tart nyilván az önidézetek témájában. Az önidézetek kérdése nemcsak a tudománymetriát foglalkoztatja, a cikkek között etikai, ismeretelméleti, pszichológiai megközelítéseket is találhatunk. A cikkek döntő többsége azonban kétségtelenül tudománymetriai, tudománypolitikai szempontból vizsgálja az önidézeteket; több mint 100 cikk a *Scientometrics* folyó-

iratban jelent meg. Az ilyen irányú vizsgálatokat összefoglaló tanulmányunkat tíz évvel ezelőtt közzétük [3].

Az önidézetek rendszeres vizsgálatához mindenképp az alapfogalmakat kell tisztázni. Önidézetten a leggyakrabban a szerzői önidézeteket értjük, tehát azokat az eseteket, amikor a szerző valamely saját korábbi munkáját tünteti fel az irodalomjegyzékben.

Az önidézet fogalma azonban ennél lényegesen tágabb. Beszélhetünk önidézetekről folyóiratok szintjén (egy folyóirat cikkeiben ugyanazon folyóirat más cikkeire vonatkozó hivatkozások esetén, szerzőktől függetlenül), intézmények szintjén (ezt esetenként „házon belüli” – in-house – hivatkozásnak is nevezik) vagy országos szinten is (hazai hivatkozások).

Önhivatkozások és önidézetek

A tudománymetriai irodalomban fontos különbséget tenni egy cikk *hivatkozásai* (az irodalomjegyzékben vagy

lábjegyzetekben szereplő tételek) és *idézetei* (más cikkeknek rá vonatkozó hivatkozásai) között. A hivatkozásokat adjuk, az idézeteket kapjuk. Az angol nyelvű irodalomban a hivatkozásoknak a „reference”, az idézeteknek a „citation” szó felel meg. A megkülönböztetés természetesen cikkhalmazokra (egy szerző művei, egy folyóirat cikkei) is érvényes.

Az önhivatkozások, önidézetek esetében még kényesebb a helyzet. Egy cikkhalmaz önhivatkozásainak és önidézeteinek a száma megegyezik; a cikkek által kapott önidézetek pontosan az ugyanezen cikkek irodalomjegyzékében szereplő önhivatkozások. Az önhivatkozások és az önidézetek százalékos aránya azonban két teljesen különböző érték. A számláló mindkét esetben azonos (az önhivatkozások, illetve önidézetek közös száma), a nevező viszont az egyik esetben a cikkek irodalomjegyzékében hivatkozott cikkek összes száma, a másikkban viszont a cikkek által kapott összes idézet.

Egy szerző a cikkeinek az irodalomjegyzékében elenyésző mértékben hivatkozhat saját cikkeire, de ha más egyáltalán nem hivatkozik rájuk, az önidézetsége akár 100%-os is lehet. Másrészt, ha egy szerző kizárólag saját munkáira hivatkozik (100%-os önhivatkozás), de a másoktól kapott idézetek száma ennek sokszorosa, akkor az önidézetsége igen alacsony lehet. A két mutatószám tehát lényegesen különbözik, és egészen mást mutat.

Sajnos a szakirodalomban is sokszor keveredik a két kifejezés, ezért rendkívül körültekintően lehet csak a különböző forrásokból származó adatokat összevetni.

Szerzői önhivatkozások és önidézetek

A szerzői önhivatkozás a tudományos kutatás természetes, mondhatni, szükséges eleme. Azt dokumentálja, hogy a szerző szóban forgó munkája egy kutatási folyamat része, a közvetlen előzményeknek nyoma van az irodalomban. Ha egy szerző sohasem hivatkozna önmagára, az azt jelentené, hogy minden egyes munkája független lenne az előzőktől. Ez egy meglehetősen abszurd és semmiképpen sem követendő gyakorlat lenne.

A szerzői önidézetek a tudományometriában nem csökkent értékű, netán értéktelen idézeteknek tekintendők; egyszerűen *mást* mutatnak, mint a másoktól kapott idézetek.

Ahogy az korábban említettük, az önhivatkozások és önidézetek abszolút száma természetesen megegyezik (minden önhivatkozás egyben önidézet is), de százalékos arányuk szélsőségesen különbözhet. Alapvető különbség, hogy míg az önhivatkozás mértéke csakis a szerző saját döntésének eredménye (milyen arányban hivatkozik magára és másokra a cikkeiben), az önidézetséget döntően a másoktól kapott idézetek száma befolyásolja.

Az önhivatkozások (önidézetek) abszolút száma a szerzők termelékenységétől is függ. Ha egy szerzőnek összesen n cikke van, akkor abban a szélsőséges esetben, ha minden egyes cikkében minden korábbi cikkére hivat-

kozik, $n \times (n-1)/2$ önidézete lehetséges, vagyis az önidézetek száma nagyságrendben a cikkszám négyzetével arányosan növekszik. Egy 10 cikk szerző esetében a szám 45, tehát a cikkenkénti átlagos lehetséges önidézetség 4,5 idézet/cikk. 100 cikk esetében ez az érték közel 50 idézet/cikkre nő. A példa természetesen extrém, de a cikkek száma és az önidézetek aránya között a valóságban is erős a pozitív korreláció. *Van Raan* vizsgálatai szerint [4] az extrém példa négyzetes függése helyett az erősen idézett cikkek esetében lineáris függvény, a gyengén idézett cikkek esetében 1,5 kitevőjű hatványfüggvény írja le az önidézetség és a publikációs szám közötti összefüggést.

Az önidézetek aránya a kapott idézetek számával csökken: *Aksnes* közel 50 000 cikkre kiterjedő vizsgálata [5] szerint 1–50 kapott idézet között 30%-ról 20%-ra; 100 körüli idézetszám esetén az arány már csak 10% [4].

Az önidézetség átlagos mértéke szakterületenként különböző. Több tanulmány [5–7] egybehangzó eredményei szerint a tudományágak közül a fizikában (különösképpen az asztrofizikában) és a matematikában a legmagasabb az önidézetek aránya (30% körüli), az orvostudományokban 15–20% között van, a társadalom- és bölcsészettudományokban 10% alatt marad.

Eléggyé nyilvánvaló, hogy egy adott cikk önidézetsége a cikk megjelenése után az eltelt évek során rohamosan csökken. A megjelenéssel azonos évben kapott idézetek fele önidézet; a cikkek 10 éves korára ez az érték 10% alá csökken [8].

Nemzetközi összehasonlításban a legalacsonyabb (25% vagy az alatti) önidézetség-értékeket az Egyesült Államok, az Egyesült Királyság, Írország és Kanada esetében találták, a legmagasabbak (50% fölött) Ukrajna, Belorusszia és Irán esetében voltak (1999. évi adatok [8]). Ebben az összehasonlításban az önhivatkozottság-értékek valószínűleg realiztikusabb képet mutatnának (sajnos ilyeneket nem találtam az irodalomban), mert nem terhelné őket az angolszász szakirodalom idézetségének jól ismert túlértékeltége.

A társszerzőség és az önidézetség kapcsolata bonyolult. A naiv spekuláció azt sugallja, hogy a nagyobb társszerzőszám nagyobb önidézetséggel jár [9]. A valóság azonban az, hogy a társszerzők számának növekedése a cikkek összes idézetségét is növeli, amit az önidézetek számának növekedésével csak részben lehet magyarázni [10]. *Aksnes* fentebb említett cikkéből [5] kitűnik, hogy a százalékos önidézetség 4–6 társszerző esetén maximumot ér (körülbelül 30%), több társszerző esetén visszatér a mintegy 20%-os átlagértékhez.

A többszerzős cikkek esetében felülvizsgálatra szorul az önidézetek értelmezése is. Ezt a témát egy korábbi cikkünkben már érintettük [11]. Onnan idézzük:

„Az önidézés szigorú definíciója szerint önidézetről beszélünk, ha az idéző és az idézett cikknek van legalább egy közös szerzője. Nyilvánvalóan azonban más jellegű az önidézés akkor, ha egy egyszerűs cikk szerzője idézi egy másik egyszerűs cikket (sietek hozzátenni, hogy természetesen ön-

magában ez sem elítélendő cselekedet), mint ha két száz-szerzős cikknek van egy közös szerzője.

A fenti szigorú definíción enyhíthetünk például azáltal, hogy az »igen-nem« megkülönböztetés helyett az önidézett-ség 0 és 1 közötti mértékét definiáljuk az idéző és idézett cikk szerzői halmazának átfedése szerint [12]. A mutatószám értéke 0, ha az idéző és idézett cikknek nincs közös szerzője, 1, ha a két szerzői lista megegyezik.

Az így definiált mutatószám (jelöljük x -szel) használatával természetesen körültekintően kell bánni. Itt is (mint minden tudománymetriai elemzésnél) mindenképp az elemzés céljára kell tekintettel lennünk. Ha cikkek idézettségét vizsgáljuk, és figyelembe akarjuk venni az önidézett-séget, akkor minden egyes kapott idézetet $1-x$ értékűnek számolhatunk. Ugyanígy járhatunk el például a folyóiratok önidézett-séggel korrigált impaktfaktorának kiszámításánál.

Ha azonban egyes kutatók vagy kutatócsoportok idézettsége az elemzés tárgya, akkor indokoltnak látszik az a megoldás, hogy azokat az idéző cikkeket, amelynek szerzői között szerepel a vizsgált személy vagy a vizsgált csoport valamely tagja, 0 súllyal vegyük figyelembe (x értékétől függetlenül), a többire használhatjuk az $1-x$ értéket.”

Az önidézett-ség mértékét befolyásoló egyéb tényezők közül egy meglepőre hívja fel a figyelmet egy friss kutatás [13]: a férfi kutatók szignifikánsan hajlamosabbak az önidézésre, mint a nők.

Folyóirat-önhivatkozások és -önidézetek

A folyóiratoknál az önhivatkozás, önidézet megállapítása többnyire egyértelmű. Itt is lehetnek természetesen problémás esetek: például a folyóiratok egyesülése, illetve szétválása esetében. Az önhivatkozás és az önidézet viszonya itt is ugyanolyan, mint a szerzők esetében: az abszolút értékek egyenlők, de a százalékos arányok egészen különbözők lehetnek.

A lokális jelentőségű folyóiratokra jellemző, hogy hivatkozásai folyóiratok széles spektrumára vonatkoznak, ezért az önhivatkozási arány kicsi, de idézetei jelentős részét saját többi cikkeitől kapja, vagyis az önidézett-ség magas. Sajnos az *Orvosi Hetilap* erősen ebbe a kategóriába esik: az önhivatkozások aránya alig több mint 3%, az önidézett-sége viszont közel 25%-os.

A nagy presztízsű folyóiratoknál (például *Lancet*, *New England Journal of Medicine*) a helyzet fordított: az önhivatkozás magas (hiszen a megkerülhetetlen források jelentős része ott jelent meg), az önidézett-ség viszont alacsony (az idézetek döntő többsége más folyóiratoktól származik).

Külön kategóriát alkotnak az összefoglaló cikkeket közlő („review”) folyóiratok, amelyeknek mind az önhivatkozási, mind az önidézett-ségi aránya rendkívül alacsony (gyakran 1% alatti), hiszen természetüknél fogva a legszélesebb körből gyűjtik össze a hivatkozásait, és kapják idézeteiket is.

A Thomson-Reuters által közzétett tanulmány [14] tanúsága szerint a folyóiratok önidézett-ségi aránya az impaktfaktor növekedésével meredeken csökken.

Az 1. ábrán bemutatjuk egy nagyobb folyóiratmintán az önhivatkozások és önidézetek százalékos arányának összefüggését. A mintában a 2008. évi Science Citation Index Journal Citation Reports (SCI JCR) adatbázisból mindazok a folyóiratok szerepelnek (körülbelül 3000), amelyekben az önhivatkozások (önidézetek) száma legalább 100 volt. Bár az ábra sajátos mintázatot mutat, a változók között lineáris korreláció nem található. Az önidézetek zöme a 0–30%, az önhivatkozásoké a 0–20% tartományban található. Ahogyan jeleztük, a legnagyobb önidézett-ségű és alacsony önhivatkozottságú folyóiratok a lokális kiadványok közül kerülnek ki; az ábra jobb alsó pontjai a *Česká a Slovenská neurologie a neurochirurgie*, a *Cuban Journal of Agricultural Science* és a *Pakistan Journal of Botany* folyóiratokat reprezentálják. Az egyetlen 40% önhivatkozást meghaladó folyóirat a szerzői önidézett-ségben is kiemelkedő értékeket mutató szakterületet, az asztrofizikát képviselő *Astrophysics Journal*. (Ennek a folyóiratnak és az egész szakterületnek a némi-képpen anomális viselkedésére a szakterület művelői is régen felfigyeltek [15].)

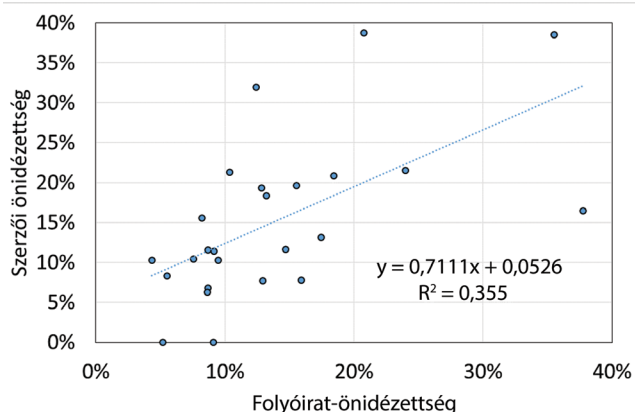


1. ábra Folyóiratok önidézett-ségi és önhivatkozási arányai (Az SCI JCR 2008. évi adatai alapján)

A szerzői önidézetek és a folyóirat-önidézetek kapcsolata

Ferrara és Romero [16] a Scopus adatbázis 2011. évi adatai alapján meghatározták a „számítógép-tudományi alkalmazások” szakterület 25 vezető folyóiratának cikkeire kapott idézetek között a szerzői önidézetek arányát. Ezeket az értékeket vetettük össze ugyanezen folyóiratoknak az SCI JCR 2013. évi adatai alapján meghatározott folyóiratönidézet-adataival. Amint azt a 2. ábrán láthatjuk, a két mutatószám között mérsékelt pozitív korreláció tapasztalható.

Igen hasonló képet kapunk, ha a szerzői és a folyóirat-önidézettségeket a természettudományok főbb szakterületeire összegezve ábrázoljuk (3. ábra). A korreláció láthatóan rontja a kémia meglepően magas szerzői ön-



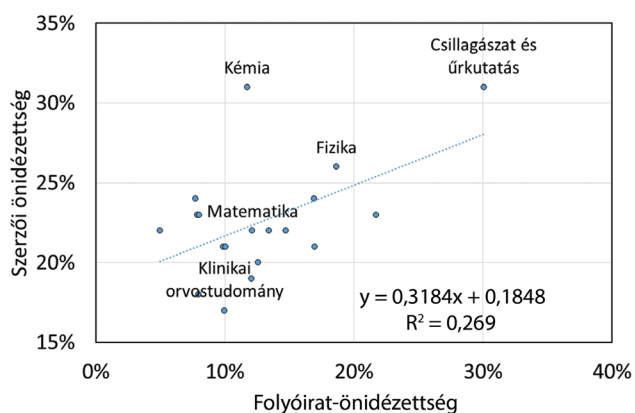
2. ábra

A szerzői és a folyóirat-önidézettek aránya a számítógép-tudományi alkalmazások szakterület 25 vezető folyóiratának cikkeire kapott idézetek között (A Scopus 2011. évi [15] és az SCI JCR 2013. évi adatai alapján)

idézettesség-értéke. Minthogy ezt semmilyen más vizsgálat nem támasztja alá, elképzelhető, hogy az *Aksnes* [5] által közölt adat hibás. A csillagászat és űrkutatás (benne az asztrofizika) kiemelkedő értéke többszörösen igazolt.

Az önidézés etikai kérdései

„Eigenlob stinkt!” – „Az öndicséret bűdös!” – tartja a mondás. Az önidézeteiket pedig sokan az öndicséret egy változatának és így a hiúság, szerénytelenség jelének tekintik. Ez általánosan véve természetesen nem igaz [17]. Amióta azonban az idézetek a kutatás értékelésében is szerepet kaptak, már nemcsak a hiúság, hanem anyagi, egzisztenciális érdekek is motiválhatják a szerzőket az idézetek számának bármilyen módon történő növelésére. A publikációk bírálóinak, a kiadványok szerkesztőinek a felelőssége, hogy a kéziratok irodalomjegyzékét is épp olyan kritikusán ellenőrizzék, mint az elméleti vagy kísérleti módszereket, és felhívják a figyelmet a hiányzó



3. ábra

A szerzői és a folyóirat-önidézettek aránya a természettudományok főbb szakterületein (*Aksnes* cikke [5] és az SCI JCR 2013. évi adatai alapján)

vagy feleslegesen idézett tételekre. (Sajnos, ez néha csak arra korlátozódik, hogy a bíráló a saját nem idézett munkáira hívja fel a szerző figyelmét.) A követendő alapelv egyszerű: a saját cikkeket is pontosan ugyanolyan mércével kell mérni, mint másokéit. Ezt persze kimondani, leírni könnyű, megvalósítani annál nehezebb.

Még komolyabb etikai problémákat vetnek fel a folyóirat-önidézettek. Ezek túlzott elterjedése ellen ugyanis pontosan azoknak nem érdekük fellépni, akik megtehetnék: a folyóiratok szerkesztőinek. Ami ennél sokkal rosszabb, egyre több esetben bizonyosodik be, hogy a szerkesztők bátorítják a kéziratokat beküldő szerzőket a folyóirat korábbi cikkeinek idézésére, sőt ennek érdekében bizonyos nyomást is gyakorolnak. Az impaktfaktor (vagy bármilyen más idézettségen alapuló mutatószám) manipulálásának ez az eszköze súlyosan elítélendő, és méltó következményekkel (például a Web of Knowledge adatbázisból való kizárással) jár. A nyilvánosságra került esetek [18–20] alighanem csak a jéghegy csúcsát jelentik, de az internet nyilvánossága ma már nem nagyon engedi, hogy a hasonló esetek rejtve maradjanak.

Zárszó

Az önhivatkozásoknak, önidézetteknek helyük van a tudományos kommunikációban, és ennek megfelelően a tudományos kommunikáció elemzésében, a tudomány-metriában is. Az önidézettek természetüknél fogva eltérnek a másoktól kapott („független”) idézetektől. Az eltérés minőségi, de nem értékbeli, az önhivatkozások, önidézettek elemzése más, de nem kevésbé értékes információkat tár fel a tudományos kommunikáció világáról, mint az idéztelemzés más területei. Elkülönítve kell őket vizsgálni, de a megoldás semmiképpen nem a teljes „kiközösítésük”, mellőzésük, hanem integrálásuk a tudománymetria eszköztárába.

Anyagi támogatás: A szerző köszöni az FP7-SSH-2013-2 #613202 (IMPACT-EV) projekt támogatását.

A cikk végleges változatát a szerző elolvasta és jóvá hagyta.

Érdekeltségek: A szerzőnek nincsenek érdekeltségei.

Irodalom

- [1] *Tagliacozzo, R.*: Self-citations in scientific literature. *J. Document.*, 1977, 33(4), 251–265.
- [2] *Lawani, S. M.*: On the heterogeneity and classification of author self-citations. *J. Assoc. Am. Soc. Inform. Sci.*, 1982, 33(5), 281–284.
- [3] *Glänzel, W., Debackere, K., Thijs, B., et al.*: A concise review on the role of author self-citations in information science, bibliometrics and science policy. *Scientometrics*, 2006, 67(2), 263–277.

- [4] *Van Raan, A. F.*: Self-citation as an impact-reinforcing mechanism in the science system. *J. Am. Soc. Inform. Sci. Technol.*, 2008, 59(10), 1631–1643.
- [5] *Aksnes, D. W.*: A macro study of self-citation. *Scientometrics*, 2003, 56(2), 235–246.
- [6] *Snyder, H., Bonzi, S.*: Patterns of self-citation across disciplines (1980–1989). *J. Inform. Sci.*, 1998, 24(6), 431–435.
- [7] *Glänzel, W., Thijs, B.*: The influence of author self-citations on bibliometric macro indicators. *Scientometrics*, 2004, 59(3), 281–310.
- [8] *Glänzel, W., Thijs, B., Schlemmer, B.*: A bibliometric approach to the role of author self-citations in scientific communication. *Scientometrics*, 2004, 59(1), 63–77.
- [9] *Rousseau, R.*: Why am I not cited, or why are multi-authored papers more cited than others? *J. Document.*, 1992, 48(1), 79–80.
- [10] *Van Raan, A. F.*: The influence of international collaboration on the impact of research results: Some simple mathematical considerations concerning the role of self-citations. *Scientometrics*, 1998, 42(3), 423–428.
- [11] *Schubert, A.*: Equally or equitably? Author roles and co-author shares in scientific publication. [Igazságosan vagy testvériesen? Szerzői szerepek és társszerzői részesedések a tudományos publikálásban.] *Orv. Hetil.*, 2016, 157(13), 512–516. [Hungarian]
- [12] *Schubert, A., Glänzel, W., Thijs, B.*: The weight of author self-citations. A fractional approach to self-citation counting. *Scientometrics*, 2006, 67(3), 503–514.
- [13] *King, M. M., Correll, S. J., Jacquet, J., et al.*: Men set their own cites high: Gender and self-citation across fields and over time. Presented by M.M.K. at the American Sociological Association Annual Meeting. Sociology of Science Session. Chicago, IL. 2015. <http://www.eigenfactor.org/gender/self-citation/Self-Citation.pdf>
- [14] Journal self-citation in the Journal Citation Reports – Science Edition (2002). <http://wokinfo.com/essays/journal-self-citation-jcr/>
- [15] *Trimble, V.*: A note on self-citation rates in astronomical papers. *Publ. Astronomic Soc. Pacific*, 1986, 98(610), 1347–1348.
- [16] *Ferrara, E., Romero, A. E.*: Scientific impact evaluation and the effect of self-citations: Mitigating the bias by discounting the h-index. *J. Am. Soc. Inform. Sci. Technol.*, 2013, 64(11), 2332–2339.
- [17] *Falagas, M. E., Kavvadia, P.*: “Eigenlob”: Self-citation in biomedical journals. *FASEB J.*, 2006, 20(8), 1039–1042.
- [18] *Smith, R.*: Journal accused of manipulating impact factor. *BMJ*, 1997, 314, 461.
- [19] *Wilbite, A. W., Fong, E. A.*: Coercive citation in academic publishing. *Science*, 2012, 335(6068), 542–543.
- [20] Gaming the impact factor puts journal in time-out. <http://scholarlykitchen.sspnet.org/2011/10/17/gaming-the-impact-factor-puts-journal-in-time-out/>

(Schubert András,
e-mail: schuba@iif.hu)

A rendezvények és kongresszusok híryanagának leadása

a lap megjelenése előtt legalább 40 nappal lehetséges, a 6 hetes nyomdai átfutás miatt.
Kérjük megrendelőink szíves megértését.

A híryanagokat a következő címre kérjük:
Orvosi Hetilap titkársága: Budai.Edit@akkrt.hu
Akadémiai Kiadó Zrt.