

## Válaszadó-vezérelt mintavétel: ritka és rejtett csoportok kvantitatív vizsgálata\*

---

**Simon Dávid,**  
az Eötvös Loránd Tudomány-  
egyetem oktatója és kutatója  
E-mail: simdav@caesar.elte.hu

A ritka és rejtett sokaságok esetén alkalmazott mintavételi és becslési technikát, a válaszadó-vezérelt mintavételt az angol nyelvterületen már kiterjedten használják, ám Magyarországon kevésbé ismert ez az eljárás. A szerző tanulmányában a módszer bemutatásán túl, a legfrissebb angol nyelvű szakirodalom ismertetésén keresztül kitér az alkalmazás során felmerülő kihívásokra, az eddigi kutatási tapasztalatokra és az etikai problémákra. A hazai és nemzetközi tapasztalatokra támaszkodva ajánlásokat tesz a válaszadó-vezérelt mintavétel felhasználási területeire, illetve a módszerrel kapcsolatos további kutatások irányára.

TÁRGYSZÓ:  
Válaszadó-vezérelt mintavétel.  
Ritka és rejtett sokaság.

\* A tanulmány a „Európai Léptékkel a Tudásért, ELTE” projekt keretében, a TÁMOP-4.2.1/B-09/1/KMR pályázat támogatásával készült. A szerző ezúton mond köszönetet *Rudas Tamásnak* a támogatásért és *Kmetty Zoltánnak* a részvételével zajlott kutatások anyagainak átadásáért.

A válaszadó-vezérelt mintavételi eljárás (respondent driven sampling – RDS) célja a nehezen körülhatárolható társadalmi csoportok kvantitatív vizsgálatának lehetővé tétele. Az RDS-t a nemzetközi gyakorlatban kiterjedten, Magyarországon csak elvétve alkalmazzák. Tanulmányomban a módszert leíró nemzetközi irodalom ismertetésén túl a legfrissebb módszertani cikkek összefoglalására, a magyar kutatások ismertetésének teljességre töreksem (természetesen ennek ellenére előfordulhat, hogy néhány erre vonatkozó kutatást nem találtam meg a szisztematikus keresés ellenére sem).

Mindenekelőtt érdemes néhány szóban definiálni, hogy mit tekintünk nehezen körülhatárolható társadalmi csoportnak, illetve mely csoportokat sorolhatjuk ide. Az angol nyelvű szakirodalomban a nehezen körülhatárolható sokaság (hard to reach population) szinonimájaként szerepel a ritka és rejtett sokaság (rare and hidden, rare and elusive population) kifejezés. A definíció szempontjából talán Spreen [1992] megközelítése a legvilágosabb, aki két jellemző alapján különíti el a nehezen körülhatárolható csoportokat:

- a kérdéses részsokaság teljes sokasághoz viszonyított aránya, valamint
- az adott részsokaság rejtőzködő, illetve nem rejtőzködő jellege szerint.

1. táblázat

*A nehezen körülhatárolható társadalmi csoportok definiálása*

Arány	Könnyen elérhető (nem rejtőzködő)	Nehezen elérhető (rejtőzködő)
Megfelelő arány	Megfelelő részsokaság	Nagy, rejtett részsokaság
Alacsony arány	Ritka részsokaság	<b>Ritka, rejtett részsokaság</b>

*Forrás: Spreen [1992].*

Rejtőzködőnek tekinthetjük azokat a részsokaságokat, amelyek tagjai az adott részsokasághoz tartozásukat igyekeznek elrejteni. Az alacsony arány meghatározása problematikusabb, azt nem elméleti, sokkal inkább praktikus szempontok határozzák meg: alapvetően azokat a részsokaságokat tekinthetjük ritkának, amelynek tagjai esetén a (rétegzett, több lépcsős) véletlen mintavétel nem finanszírozható. Az itt ismer-

tetett RDS-módszer a ritka és rejtett sokaság esetén – a megfelelő megkötések mellett – helyettesítheti a véletlen mintavételt.

## 1. A válaszadó-vezérelt mintavétel módszertani alapjai

A legfrissebb magyar szakirodalomban szerepel néhány rövid összefoglalás (*Kapitány* [2010], *Domokos et al.* [2010]) a válaszadó-vezérelt mintavételről, azonban azok – feltehetően a terjedelmi korlátok miatt – nem térnek ki a módszer részleteire, a nemzetközi irodalomban fellelhető vitás pontokra, esetenként sommás állításokat tartalmaznak. Tanulmányomban átfogó képet kívánok nyújtani a módszerről és annak legfrissebb kritikájáról.

Az RDS a hálózati mintavételek csoportjába tartozik,<sup>1</sup> amelyek közös tulajdonsága, hogy az adatfelvétel során a vizsgált sokaság tagjai úgy kerüljenek kiválasztásra, hogy őket a minta egy korábbi tagja ajánlja.

A válaszadó-vezérelt mintavétel első leírása 1997-ben jelent meg *Douglas D. Heckathorn* [1997] cikkében, amelyben a modell ugyan bizonyítja, hogy a mintavételi módszer – adott feltételek mellett – torzítatlan becslésre alkalmas, az ott szereplő feltételek közül egy azonban biztosan ellenőrizhetetlen. Ezért és mivel a későbbi cikkek olyan becslési modellt tartalmaznak, amelynek feltételei a valós mintavétellel jobban összeegyeztethetők, Heckathorn modelljét nem ismertetem. Írása azonban olyan módszertani ajánlásokat is tartalmaz, amelyek érvényessége nem változott. A válaszadó-vezérelt mintavétel szerző által javasolt lépései a következők:

- a kezdőpontok (a minta kiindulópontjait jelentő első válaszadók – seeds) kiválasztása (az erre vonatkozó statisztikai feltételekről később írok), a válaszadók jutalmat kapnak az interjúért;
- a kezdő adatfelvételek elkészítését követően a kezdő válaszadónak meghatározott összegű jutalmat ígérnek (kupon formájában – amely az anonim azonosítást is biztosítja) annak érdekében, hogy további, a célcsoporthoz tartozó válaszadókat vonjanak be a kutatásba (a jutalmat csak a bevont személlyel készült sikeres interjú után kapták meg az ajánlók);
- a folyamat következő lépéseiben a válaszadók továbbra is kettős jutalmat kapnak (az elkészült interjúért és további válaszadók bevonásáért, egészen a szükséges mintaméret kialakulásáig;

<sup>1</sup> A hálózati mintavételek történetéről lásd *Sirken* [1998].

- a bevonható további válaszadók száma válaszadónként korlátozott, annak érdekében, hogy ne legyenek félprofi bevonók;
- a célcsoporthoz tartozást a kutatók ellenőrzik, hogy elkerüljék a nem célcsoporthoz tartozók jutalom miatti válaszadását;
- az interjúkat egy adott helyen készítették, ahova a válaszadóknak el kellett menniük.

Három szempontot érdemes kiemelni a folyamatleírásból: egyrészt a válaszadókat elsődleges (válaszadásért) és másodlagos (további személyek bevonásáért) ösztönzőkkel motiválják, másrészt az interjúkat a válaszadók szempontjából semleges helyen készítik, harmadrészt a folyamat a válaszadóknak teljes anonimitást biztosít. Mindhárom szempont a rejtőzködő csoportok válaszadási hajlandóságát növeli. A másodlagos ösztönzők alkalmazása csoportközvetített (group mediated) motivációt biztosít, amelyekkel kapcsolatban Heckathorn (korábbi kutatásai alapján) nagyobb hatékonyságot feltételez, mint az elsődleges ösztönzők esetén.

Az RDS-bebecslések statisztikai modellje azon alapul, hogy az itt leírt mintavételi folyamat egy gráfon történő véletlen bolyongásnak feleltethető meg, amely leírható Markov-lánc Monte-Carlo- (Markov Chain Monte Carlo – MCMC-) mintavételként. A mintavétel során feltételezzük, hogy a következő mintatag véletlenszerűen kerül kiválasztásra a válaszadóval kapcsolatban levő célcsoporttagok közül – a hálózat ismeretében –, és meghatározható minden célcsoporttag esetén annak a valószínűsége, hogy a következő lépésben milyen valószínűséggel, mely célcsoporttag kerül a mintába. Így felírható egy átmenet-valószínűségeket tartalmazó mátrix. Ezen mátrix hatványai megadják, hogy a hatványkitevőnek megfelelő lépésszám (mintavételi hullám) után bármely kiindulópontból milyen valószínűséggel lehet eljutni a hálózat bármely más pontjába. A hálózatról feltételezzük továbbá, hogy egyetlen komponensből<sup>2</sup> áll, ezért véges számú lépés után a hálózat bármely pontjából nem nulla valószínűséggel lehet eljutni a hálózat bármely másik pontjába – ez az átmenet-valószínűségeket megadó mátrix egy szabályos Markov-láncot ír le.<sup>3</sup> A következő bebecslések a mintavételi folyamat Markov-tulajdonságán alapulnak (részletes bizonyításokat terjedelmi okokból itt nem közlök, azok elérhetők az eredeti cikkekben).

*Salganik–Heckathorn* [2004] az RDS-bebecslés továbbfejlesztett változatát publikálta, melynek statisztikai modellje a későbbiek során elterjedt. A tanulmány elnyer-

<sup>2</sup> A hálózati mintavételek módszertani leírásánál általában a hálózatelemzés szakkifejezéseit használják, jelen cikkemben is ezt fogom követni, az egyes szakkifejezések lábjegyzetes magyarázata mellett. Komponensnek a hálózat azon részét nevezik, amely esetén igaz az, hogy a hálózat bármely két tagja között létezik közvetett (több kapcsolaton keresztül) vagy közvetlen kapcsolat, és nincs olyan tagja a hálózatnak, amelyre a komponens tagjain kívül ez igaz lenne.

<sup>3</sup> Természetesen adott mintavétel esetén ezek az átmenet-valószínűségek csak becslhetők, mivel az esetek többségében nem ismerjük a teljes hálózatot.

te az Amerikai Szociológiai Társaság legjobb publikációnak járó díját a matematikai szociológia szekcióban. A szerzők által javasolt modell, a minta alapján, előbb egyes hálózati tulajdonságokra majd azok alapján a lekérdezett változók sokasági paramétereire ad becslést. A becslés feltételei a következők:

1. A válaszadók kölcsönös kapcsolatban állnak az általuk ajánlott potenciális válaszadókkal, akik a célcsoport tagjai (reciprok kapcsolatok).
2. A hálózat egy komponensből áll.
3. A mintavétel visszatevéssel történik.<sup>4</sup>
4. A válaszadó teljes pontossággal határozza meg személyes kapcsolatainak számát azaz fokszámát<sup>5</sup> (a becslések valójában a fokszámok egymáshoz viszonyított arányán alapulnak).
5. A válaszadó véletlenszerűen választ az ismerősei közül az ajánlás alkalmával.
6. A vizsgált változó diszkrét.

Ezen feltételeken túl a megfelelő becslések elvégzéséhez arra is szükség van, hogy a minta érje el (közelítse meg) az egyensúlyi állapotot. A mintavétel Markov-tulajdonságából következik, hogy a minta összetétele adott változóra nézve egyensúlyi állapothoz közelít. Az egyensúlyi állapot megközelítésének mértékét úgy határozhatjuk meg, ha az egymást követő hullámok során megvizsgáljuk, hogy milyen mértékben változik a célváltozó kiválasztott értékének aránya a mintában. Akkor tekintjük az adott célváltozó szempontjából egyensúlyban levőnek a mintát, ha a változás mértéke egy előre meghatározott értéknél kisebb.<sup>6</sup>

A felsorolt feltételek teljesülése esetén a mért változóra a következő aránybecslés tehető (itt a dichotóm változóra mutatom be a becslési eljárást, amelyet a továbbiakban RDS I becslésnek nevezek):

$$N_A D_A C_{A,B} = N_B D_B C_{B,A}, \quad /1/$$

$$P_A D_A C_{A,B} = P_B D_B C_{B,A}, \quad /2/$$

$$P_A + P_B = 1, \quad /3/$$

<sup>4</sup> Szimuláció alapján a visszatevés nélküli mintavétel nem befolyásolja jelentősen a becsléseket (*Salganik–Heckathorn* [2008]).

<sup>5</sup> A hálózat kutatásban fokszám alatt a kapcsolatok számát értik.

<sup>6</sup> A kovariánciakritérium és a becslések szórása között pontosan meg nem határozott pozitív irányú összefüggés van.

$$\hat{P}_A = \frac{\hat{D}_B \hat{C}_{B,A}}{\hat{D}_A \hat{C}_{A,B} + \hat{D}_B \hat{C}_{B,A}}, \quad /4/$$

$$\hat{D}_A = \frac{n_A}{\sum_{j=1}^{n_A} d_j^A}; \quad \hat{D}_B = \frac{n_B}{\sum_{j=1}^{n_B} d_j^B}, \quad /5/$$

$$\hat{C}_{A,B} = \frac{f_{A,B}}{n_A}; \quad \hat{C}_{B,A} = \frac{f_{B,A}}{n_B}, \quad /6/$$

ahol

$N_A; N_B$  – a változó első, illetve második értékét választó személyek száma a részsokaságban;

$n_A; n_B$  – a változó első, illetve második értékét választó személyek száma a mintában;

$D_A; D_B$  – a változó első, illetve második értékét választó személyek átlagos fokszáma a részsokaságban (illetve annak becslése a mintából);

$C_{A,B}; C_{B,A}$  – az első és második, illetve második és első értéket választó személyek közötti kapcsolat valószínűsége a részsokaságban (illetve annak becslése a mintából);

$P_A; P_B$  – az első, illetve második értéket választó személyek aránya (illetve annak becslése a mintából);

$d_j^A; d_j^B$  – az első illetve második értéket választók közül a  $j$ . válaszadó fokszáma;

$f_{AB}; f_{BA}$  – azon válaszadók száma, akik az első, illetve második értéket választották és az általuk ajánlott személy a második, illetve első értéket választotta.

A becslési eljárás során a hálózati jellemzők ( $\hat{C}$ , illetve  $\hat{D}$ ) becslésén keresztül végzünk aránybecslést ( $\hat{P}$ ) a kérdéses változó esetén. Az /1/ és /2/ egyenletek a kapcsolatok feltételezett kölcsönösségének következményei, a /3/ egyenlet triviális. A /4/ a /2/ és /3/ egyenletek következménye. A fokszám becslése Hansen–Hurwitz-becslésen alapul (az egyszerű levezetést lásd *Salganik–Heckathorn* [2004]). A /6/ annak a feltételnek a következménye, hogy a válaszadó véletlenszerűen választ a kapcsolatai közül, amikor ajánl (Heckathorn erre vonatkozó korábbi kutatása meg-

erősíti, hogy az ajánlott személyek választása valóban közel véletlenszerű). Meg kell jegyezni, hogy 3 vagy többértékű változók esetén a becslés túldeterminálttá válik, amire Heckathorn [2007] a lineáris legkisebb négyzetes közelítést vagy a saját, adat-simításos módszerét ajánlja. A cikk bizonyítása alapján belátható, hogy a becslés aszimptotikusan torzítatlan diszkrét változókra a feltételek teljesülése esetén.

Folytonos változó esetén Heckathorn kétkomponensű súly kialakítását javasolja (lásd a /7/ egyenletet), ennek kiszámítása azonban lényegében azonos azzal, mintha a folytonos változó értékeit particionálnánk és ennek megfelelően végeznénk el az ismertett becslést, majd az így kapott és a mintában szereplő eloszlás különbségét használnánk az adott változó súlyozására.

$$DW_j = DC_j RC_j = K \frac{1}{d_j} \frac{E_x}{p_x}, \quad /7/$$

ahol

- $DW_j$  – kétkomponensű súly a  $j$ . mintatag esetén adott változóra,
- ha  $j$ . mintatagra a változó értéke  $x$ ;
- $DC_j$  – fokszámfüggő súlykomponens;
- $RC_j$  – kiválasztás függő súlykomponens;
- $K$  – állandó, amely biztosítja az elemszám megőrzését;
- $E_x$  – a változó  $x$  értékéhez (értéktartományához) tartozó becsült érték, ha figyelmen kívül hagyjuk a fokszám hatását (kiszámítását lásd Heckathorn [2007]);
- $p_x$  – a változó  $x$  értékéhez (értéktartományához) tartozó mintatagok aránya.

Heckathorn [2007] a folytonos változók olyan particionálását javasolja, korábbi RDS-adatok elemzésével, ahol az értékenkénti átlagos esetszám  $12 \pm 4$ . Alacsonyabb átlagos esetszám a becslések instabilitásához, magasabb átlagos esetszám túl pontatlan súlyozáshoz vezet.

Fontos kiemelni (és ennek érzékeltetésére az említett kétkomponensű súly alkalmas leginkább), hogy a változókra egyenként kell becslést adni, a módszerrel nyert adatbázisban nem megfelelő az olyan megfigyelési egységre alkalmazott súlyozás, amely minden változó esetén biztosítaná a várható érték torzítatlan becslésének lehetőségét egyszerű átlagszámítással.

Voltz és Heckathorn [2008] egy, az előbbiektől eltérő becslésre is javaslatot tesz, amelynek feltétele az, hogy a minta egyensúlyi állapotba kerüljön (lásd korábban). Az egyensúlyi állapot elérését az biztosítja, ha az említetteken túl azt is feltételezzük,

hogy a mintavétel egy láncon elsőfajú Markov-folyamat, vagyis a láncokban bármely egymást követő három mintatagra igaz, hogy a becsleni kívánt változó értéke az első és a harmadik mintatag esetén feltételesen független a változó által a második mintatagnál felvett értékre, mint feltételre nézve (másképpen fogalmazva a mintavétel memória mentes). Ebben az esetben Voltz és Heckathorn szerint a következő becslés tehető (amelyet a továbbiakban RDS II becslésnek nevezek):

$$\hat{P}_A = \frac{n_A}{n} \frac{\hat{D}}{\hat{D}_A}, \quad /8/$$

$$\hat{D} = \frac{n}{\sum_{i=1}^n \frac{1}{d_i}}, \quad /9/$$

ahol

$\hat{D}$  – a célcsoport átlagos fokszámának a becslése a mintából (Hansen–Hurwitz-becslés).

Az adott változók eloszlására vonatkozó becsléseken túl szintén fontos kérdés a becslések szórása, ami meghatározza egyúttal az adott pontosságú becslés eléréséhez szükséges mintaméretet is. A jelenlegi irodalom alapján egzakt, torzítatlan becslés nem áll rendelkezésre a becslések szórására, *Salganik* [2006] bootstrap módszert ajánl ennek meghatározására. Korábbi adatok másodelemzése és modellezése alapján ő arra a következtetésre jutott, hogy a designhatás (def) értéke az RDS I esetében 2 körüli (azaz az RDS I becslés varianciája az egyszerű véletlen mintavételhez képest kétszeres). Ennek alapján azonos becslési pontosság elérése érdekében az egyszerű véletlen mintához képest kétszeres mintaméret alkalmazását ajánlja. *Volz* és *Heckathorn* [2008] ugyanakkor kidolgozott egy egzakt módszert diszkrét változók RDS II becslésének varianciabecslésére, amelyet az alábbiakban levezetés nélkül adok közre:

$$\hat{V}_{P_A} = \hat{V}_1 + \frac{P_A^2}{n} \left( (1-n) + \frac{2}{n_A} \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{C}^{i-j})_{A,A} \right), \quad /10/$$

$$\hat{V}_1 = \frac{\hat{V}(Z_i)}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (Z_i - P_A)^2, \quad /11/$$

$$Z_i = \frac{\hat{D}}{d_i} I_A(i), \quad /12/$$



ahol

$\hat{V}_{P_A}$  – a változó  $A$  értékét választó személyek arányára vonatkozó becslés variációjának becslése;

$(\hat{C}^{i-j})_{A,A}$  – becslés annak a valószínűségére, hogy az  $A$  értéket választó személyek  $i$ - $j$  lánchosszúság után  $A$  értéket választó személyt ajánlanak (a becslés az átmenetvalószínűség-mátrix megfelelő hatványra emelése után kapott eredménymátrix első sor, első oszlopában található valószínűség);

$I_A(i)$  – indikátorváltozó, melynek értéke 1, ha a minta  $i$ -edik tagja a változó  $A$  értékét választotta és 0, ha a minta  $i$ -edik tagja nem  $A$ -t választotta.

A leírt varianciabecslés a Hansen–Hurwitz-becslés analógiájának tekinthető, azzal a – jelentős – különbséggel, hogy figyelembe veszi, a mintatagok nem függetlenek, és felhasználja a mintavétel korábbiakban leírt Markov-megközelítését. A becslés ugyanakkor abban az esetben teljesül, ha az adott változóra nézve a mintavétel, mint Markov-folyamat, memóriamentes (ez a feltétel azokban a hálózatban, ahol olyan közel zárt csoportok jelennek meg, amelyekben a vizsgált változó értéke homogénebb lehet, mint a hálózat többi részén, nem feltétlenül teljesül). (Erről lásd később.)

Megjegyzendő, hogy a közölt becslés nem torzítatlan, mivel  $C_{A,A}$  nem ismert,  $D$ -hez hasonlóan csupán becslhető.

Az itt leírt becslési eljárások közül a folytonos és diszkrét változók RDS I becslése, a becslések bootstrap varianciabecslése elvégezhető *Volz*, *Degani* és *Heckathorn* által közreadott RDSAT szoftver segítségével (<http://www.respondentdrivensampling.org>); az RDS I és RDS II becslés elérhető R-csomagként (package RDS). A becslési eljárásokra az interneten elérhető Melléklet tartalmaz konkrét számítási példát ([www.ksh.hu/statszemle](http://www.ksh.hu/statszemle)).

## 2. Válaszadó-vezérelt mintavétel a gyakorlatban

A válaszadó-vezérelt mintavételt az angolszász irodalom tanúsága szerint már több mint 30 országban, számos esetben használták,<sup>7</sup> ennek ellenére csak a legutóbbi

<sup>7</sup> *Malekinejad és szerzőtársai* [2008] csak az AIDS- és HIV-fertőzések kapcsán, az Egyesült Államokon kívül 123 RDS-módszert használó kutatást azonosítottak, összesen 28 országban.

években jelentek meg olyan kutatások, amelyek a módszert valós adatokon tesztelik. A következőkben a *Salganik* és *Heckathorn* által 2004-ben írt cikkekre hivatkozó 466 publikáció<sup>8</sup> közül azokat ismertetem, amelyek a módszer értékelésére törekedtek, illetve az egyes felmerülő problémákra javasolnak megoldásokat, valamint azokat, amelyek egyes kutatási területeken foglalják össze a módszer tapasztalatait.

## 2.1. Módszertani megfontolások

Hét, kifejezetten módszertani tanulmány ismertetek, amelyek eltérő adatokon vizsgálják a válaszadó-vezérelt mintavétel tulajdonságait. *Kendall* és *szerzőtársai* [2008] fortalezai (Brazília) homoszexuális férfiakkal kapcsolatos kutatásuk során hasonlították össze a hólabda (2002. év,  $N = 127$ ), a hely-idő (2002. év,  $N = 274$ ) és a válaszadó-vezérelt mintavétel (2005. év,  $N = 406$ ) eredményeit. *Wejnert* [2009] valós sokaságon, egy egyetem hallgatóinak körében 2004-ben és 2008-ban végzett webalapú válaszadó-vezérelt mintavétel becsléseit hasonlítja össze a sokaság ismert paramétereivel. *Gile* és *Handcock* [2009] 1000 modellezett hálózatot vizsgált, amelyek egyenként 1000 fős sokaságot szimuláltak (a nem vizsgált hálózati paramétereket egy korábbi kutatás valós hálózati adatai alapján állították be). Az 1000 darab 1000 elemű hálózatból 500 elemű mintákat vettek. *Goela* és *Salganik* [2010] egy valós nagymintás kutatás hálózatának 4430 fős részhálóját használták (ezen belül 85 különálló komponens), amelyen az RDS I becslés feltételeinek megfelelő 500 fős mintákat szimuláltak, a mintavételt 10 ezer alkalommal ismételték meg. *Tomas* és *Gile* [2010] 1000 fős szimulált hálózattal modellezték a sokaságot (a hálózat tulajdonságait a szerzők nem részletezik, azt azonban állítják, hogy a hálózatot úgy paraméterezték, hogy az egy korábbi kutatás valós paramétereinek, például átlagos fokszámának feleljen meg). *Kral* és *szerzőtársai* [2010] San Franciscó-i intravénás droghasználók kapcsán végzett kutatásuk során hasonlították össze egy-egy célzott mintavétel (targetted sample,  $N = 651$ ), illetve válaszadó-vezérelt mintavétel ( $N = 534$ ) segítségével vett mintát. A válaszadó-vezérelt mintavétel modellezése során 200 fős mintákat vettek. *Verdery* és *Mouw* [2011] a Facebook 100, egyetemalapú ismeretségi hálóján modellezett mintavétellel vizsgálták a válaszadó-vezérelt mintavétel tulajdonságait. A kutatás során 100 ezer mintavételt modellezték.

A 2. táblázat összefoglalja a cikkekben használt mintákat, valamint azt, hogy az hét tanulmány közül melyik milyen tényezők hatásait vizsgálta.

<sup>8</sup> Salganik és Heckathorn írására hivatkozó cikkek áttekintése a Google Scholar segítségével történt, és a jelen tanulmány írásakor aktuális állapotot tükrözi.

2. táblázat

Az idézett cikkekben használt minták sajátosságai és a vizsgált tényezők összefoglalása

Sokaság, tényező	Kendall <i>et al.</i> [2008]	Wejnert [2009]	Gile– Handcock [2009]	Goela– Salganik [2010]	Tomas– Gile [2010]	Kral <i>et al.</i> [2010]	Verdery– Mouw [2011]
	A minták sajátosságai						
Modellezett sokaság, modellezett mintavétel			+		+		
Valós sokaság (hálózat), modellezett mintavétel				+			+
Valós sokaság, valós mintavétel		+				+	
	Vizsgált tényezők						
1. Homofília (hasonlók ajánlásának tendenciája)			+		+		
2. Hálózati sajátosságok			+	+			+
3. Kiindulópont választásából származó torzítás			+				
4. Ajánlás sajátosságai			+		+		
5. Fokszámmérés hatása		+					
6. Minta paraméterei (méret, visszatevés, egyensúlyon kívüli minta)		+	+				
7. Válaszmehtagadás					+		
8. Eltérő becslések összevetése		+	+		+		
9. Eltérő módszertanú mintavételek összevetése	+					+	

A következőkben tematikusan tekintem át az egyes cikkek által vizsgált területeket és témánként közlöm az eredményeket.

**1. A homofília hatásának vizsgálata.** Homofília alatt azt értjük, ha az ajánlások során a hasonló tulajdonságú, azaz a kutatásban vizsgált változókon azonos értékkel szereplő személyek preferenciája jellemző az egyszerű véletlen választáshoz viszonyítva. *Tomas* és *Gile* [2010] vizsgálata szerint *a*) a célcsoporthoz viszonyítva kis minta és jelentős mértékű homofília esetén az RDS I becslés kisebb torzítást eredményez (bár a becslések szórása nagyobb), míg kis homofília mellett, a becslések kisebb szórása miatt, az RDS II becslés alkalmazandó. *Gile* és *Handcock* [2009] szimulációjuk során azt találták, hogy *b*) jelentős mértékű homofiliánál egyáltalán nem javasolt a válaszadó-vezérelt mintavétel használata, ezért *c*) előzetes kutatást ajánlanak a homofília meghatározása érdekében.

**2. A hálózati sajátosságok hatása.** *Gile és Handcock* [2009] azt vizsgálták, hogy milyen hatása van annak, ha a vizsgált változó egyes értékei által kijelölt csoportok átlagos fokszáma eltér egymástól (például, ha nemek szerinti megoszlásra vagyunk kíváncsiak és a nők átlagos kapcsolatszáma jelentősen eltér a férfiakétól). A szerzők azt találták, hogy *a*) az átlagos fokszámban mutatkozó különbségek a nagyobb átlagos fokszámú érték gyakoriságának alulbecsléséhez vezet. *b*) Ez a hatás számottevővé válik, ha a gyakoriságok nagymértékben eltérnek az egyenletestől. *Goela és Salganik* [2010] valós hálózaton modellezve a válaszadó-vezérelt mintavételt azt tapasztalta, hogy *c*) a becslések aszimptotikusan torzítatlanok, ugyanakkor *d*) a becslések varianciája jelentős, a számított designhatás 5 és 10 között van, ami ötszörös, tízszeres mintát tenne szükségessé az egyszerű véletlen mintavételhez képest. *Verdery és Mouw* [2011] kimutatta, hogy *e*) azon változók esetén, amelyek jelentős klaszterezettséghez vezetnek (azaz a hálózaton belül olyan alcsoportok vannak, amelyekre jellemző a nagyszámú alcsoporton belüli és kevés azon kívüli kapcsolat, valamint ezen alcsoportok esetén a vizsgált változó szóródása kisebb) a designhatás jelentőssé válhat (egyes esetekben akár 85-re is növekedhet). A szerzők arra is rámutattak, hogy *f*) a tapasztalt designhatás és a becslések ismert varianciabecslései nem korrelálnak egymással.

**3. A kiindulópont választásából származó torzítás hatása.** Bár mind az RDS I, mind az RDS II becslések aszimptotikusan torzítatlanok (nagy mintaelemszám esetén a becslések nem függenek a kiindulópontoktól), az elméleti megfontolások is hangsúlyozzák a hosszú láncok fontosságát. *Gile és Handcock* [2009] kimutatták, hogy *a*) különösen nagy homofília esetén fontos a hosszú láncok alkalmazása (a modellezés során 4-ről 6-ra emelték a láncok hosszát).

**4. Az ajánlás sajátosságainak hatása.** A válaszadó-vezérelt mintavétel feltételezése az, hogy a válaszadók az ismerőseik közül véletlenszerűen választanak. *Gile és Handcock* [2009] az ettől eltérő ajánlás hatásait vizsgálták. A szerzők azt találták, hogy *a*) adott változó mellett az ajánlás során valamely érték felé mutató torzítás azonos irányba torzítja a becsléseket (RDS II becslést vizsgálva, 20 százalékos ajánlási torzítás esetén a szimulált eredmények 2 százalékpont körüli torzítást mutattak az ismert designhatásnál).

**5. A fokszámmérés hatása.** Amint a leírt becslési eljárásokból látszik, a válaszadók fokszáma döntő jelentőségű, a leírt feltételek egyike szerint annak pontos meghatározására is szükség van. *Wejnert* [2009] vizsgálta a fokszámra vonatkozó eltérő adatfelvételi módok hatásait. A kutatás során a következőképpen vizsgálták a fokszámot (2004-ben és 2008-ban némileg eltérő módon): hány embert vett fel a válaszadó számítógépes csevegő programjába,<sup>9</sup> hány embert tudnának ajánlani a kutatáshoz (2008-ban hány emberrel beszélnek meg fontos dolgokat), az elmúlt 30 napban

<sup>9</sup> Például hány ismerőse van MSN vagy Skype stb. programban.

(2008-ban az elmúlt 14 napban) hány emberrel leveleztek e-mailen, a válaszadónak hány ismerőse, barátja és közeli barátja van (2008-ban a mobiltelefonban szereplő ismerősök száma is szerepelt). Az eredmények azt mutatták, hogy *a*) a becslések nem tértek el jelentősen a különböző fokszámadatok felhasználásával, de kijelenthető, hogy *b*) annak a fokszámadatnak a felhasználása adja a sokaságra vonatkozó legjobb becslését, amely a legjobban köthető az ajánláshoz, azaz elsősorban hány embert tudna ajánlani, másodsorban hány ismerőse van a megkérdezettnek a számítógépes csevegő programjában (általában így egyeztettek a válaszadók, mielőtt ajánlottak volna valakit).

**6. A minta paramétereinek (mintaméret, visszatevés, egyensúlyi állapoton kívüli esetek) hatásai.** *Wejnert* [2009] kimutatta, hogy *a*) kis mintákra az RDS I bootstrap szórásbecslése alulbecsül – így a becslés gyakran a konfidencia-intervallumon kívül esik –, ezzel szemben az RDS II egzakt szórásbecslése felülbecsül (ami az intervallumbecslés pontatlanságát indokolatlanul növeli, illetve szükségtelenül nagyméretű mintát igényel). *Gile* és *Handcock* [2009] azt tapasztalták, hogy *b*) amennyiben a minta mérete közelít a célcsoport méretéhez (ami nagyon kisméretű célcsoportok esetén fordulhat elő), a vizsgált változó egyes értékeihez tartozó eltérő átlagos fokszám torzító hatása jelentős mértékben megnő. Fontos tényező (különösen az RDS II becslés esetén), hogy a minta mikor éri el az egyensúlyi állapotot. Egyes kutatások az egyensúlyi állapot elérése előtti mintát elhagyták az elemzésnél. *Wejnert* [2009] szerint *c*) a közel egyensúlyi állapot elérése előtti adatok elhagyása nem javította a becslések pontosságát, viszont növelte azok szórását, ezért a szerző nem javasolja az egyensúlyi állapot elérése előtti adatok elhagyását (ezt *Gile* és *Handcock* is megerősítette). A válaszadó-vezérelt mintavétel modelljének feltételezése, hogy a mintavétel visszatevéses. *Gile* és *Handcock* [2009] az is vizsgálta, hogy milyen hatást okoz a gyakorlatban alkalmazott visszatevés nélküli mintavétel a becslések esetén. A szerzők azt találták, hogy *d*) a visszatevés nélküli mintavétel csökkenti a becslések szórását (valamint a kiinduló minta torzító hatását).

**7. A válaszmegtagadás hatása.** *Gile* és *Handcock* [2009] azt vizsgálta, hogy a változó egyes értékeihez tartozó eltérő válaszmegtagadási arányok hogyan befolyásolják a becsléseket. A szerzők azt tapasztalták, hogy *a*) a változó értékei szerint eltérő arányú válaszmegtagadás az alacsonyabb válaszmegtagadással jellemezhető érték felülbecsléséhez vezet (nincs szignifikáns különbség az RDS I és RDS II becslések között, azonban mindkettő hatékonyabb, mint a naiv becslés).

**8. Az eltérő becslések összehasonlítása.** Az eddigiekben két eltérő becslést ismerttettem, emellett *Tomas* és *Gile* [2010] három további becslést is bemutat, amelyek hatékonyságát szintén vizsgálja. Ezek közül azonban az ún. naiv becslésre ismertetem a cikk eredményeit (naiv becslés alatt azt értik a szerzők, amikor a keletkezett mintát egyszerű véletlen mintaként kezelve adunk becslést). További két, a cikkben vizsgált becslést az irodalomban nem használták kiterjedten, így azokkal itt ter-

jedelmi okok miatt nem foglalkozom. Szimulációjuk eredményei azt mutatták, hogy *a*) egyik becslési eljárás sem jobb minden vizsgált körülmény esetén a másiknál. *Wejnert* [2009] az RDS I és RDS II becsléseket hasonlítja össze, amely alapján elmondható, hogy *b*) a két módszer becslései nem térnek el jelentősen egymástól. *Gile* és *Handcock* [2009] azt mutatta ki, hogy *c*) a becslések hibája tekintetében az RDS II becslés jelentős homofília (lásd /1/) és a vizsgált változó értékeihez kapcsolódó eltérő átlagos foksám (lásd /5/) esetén pontosabb, mint az RDS I becslés. Ugyanakkor *d*) rövid láncok, a kiinduló-pontok jelentős torzítása és az ajánlás torzítása (lásd /4/) mellett az RDS I becslés átlagos hibája kisebb.

**9. Az eltérő módszerekkel vett minták összevetése az RDS-mintákkal.** A két idézett cikk (*Kendall et al.* [2008], *Kral et al.* [2010]) egyaránt megerősítik az RDS-mintavétel használhatóságát. *Kendall*ék a hely-idő, illetve hólabda mintával összevetve *a*) jelentős eltérést találtak a gazdasági, társadalmi státus eloszlása tekintetében az RDS-mintánál, ugyanakkor *b*) az RDS-minta esetén kapott eredmények jobban közelítették a vizsgált város teljes népességének eloszlását.<sup>10</sup> *Kral* és *szerzőtársai* *c*) a célzott mintavétellel összevetve az RDS-t, a legtöbb vizsgált változóra tett intervallumbecsléseket jelentős mértékben átfedőnek találták, kivétel volt ez alól a feketék eloszlása, illetve egy többségében zárt közösséget alkotó feketék által lakott posztai körzet. Utóbbi eltérés alapján *Kral* és *szerzőtársai* azt javasolják, hogy *d*) az RDS-mintavétel esetén a kezdőpontok kiválasztását egyéb mintavételi technikák segítségével tegyék heterogénebbé.

## 2.2. Gyakorlati tapasztalatok

A továbbiakban két tanulmányt foglalok össze – amelyek egy átfogó nemzetközi metaelemzést ismertetnek –, valamint kitérek azokra a területekre, amelyeken a nemzetközi irodalom alapján az RDS-mintavételt rendszeresen alkalmazzák.

A bemutatandó metaelemzés célja az RDS-módszer alkalmazásának nemzetközi áttekintő elemzése volt a HIV-fertőzés terjedésének biológiai és magatartáskutatási vizsgálatai kapcsán (*Johnston et al.* [2008], *Malekinejad et al.* [2008]). Az elemzés során 123 Egyesült Államokon kívül végzett kutatást találtak, amelyek megfeleltek a felállított kritériumoknak (megfelelő kutatási kérdést vizsgáltak és RDS-módszert alkalmaztak). A 3. táblázat a 123 kutatás főbb paramétereit foglalja össze, és többé-kevésbé körülhatárolja az RDS-mintavétel szokásos kereteit. Jól látszik, hogy sok kérdésben a vizsgált jelentős számú kutatás rendkívül kis eltérést mutatott (megelőző kutatás végzése, kezdőpontok diverzifikálása, személyes interjúk készítése, egy kérdezési helyszín

<sup>10</sup> A vizsgált célcsoport (férfiakkal szexuális kapcsolatot létesítő férfiak) valós gazdasági-társadalmi jellemzői természetesen nem voltak ismertek.

alkalmazása, pénzjutalom a részvételért, az elemzéshez az RDSAT szoftver használata). Ugyanakkor más vizsgált szempontokat tekintve jelentős eltérések voltak a kutatások között (kezdőpontok száma, rekrutációs határidő megszabása, mintaméret), vagy relatíve kis számban történt róluk említés (jutalom a rekrutációért).

3. táblázat

*RDS-mintavétel főbb paramétereit 123 nemzetközi HIV-fertőzéssel kapcsolatos kutatás metaelemzése alapján 2003–2007 között*

Paraméter, szempont	A kutatások leírása
Előzetes kutatás	101 esetben történt, 11 esetben nem történt és 11 esetben nem volt információ
Interjúkészítés módja	110 esetben személyes kérdés, 8 esetben számítógéppel segített strukturált interjú, 5 esetben önkitöltős kérdőív
A kérdés helyszíne	92 esetben egy helyszínen, a helyszínek száma maximálisan 5 volt
Kezdőpontok kiválasztása	120 esetben demográfiai vagy magatartási kockázatok alapján diverzifikáltan történt, 3 esetben nem volt ilyen diverzifikáció
Kiegészítő kezdőpontok	72 esetben az elsődleges kezdőpontokon kívül további kezdőpontokat vontak be
Kezdőpontok száma	Átlagosan 10 (2 és 32 között változott)
Rekrutációs idő	59 kutatás esetén nem szabtak határidőt a válaszadóknak a rekrutálásra, 44 esetben 7–60 nap közötti határidőt szabtak (20 esetben erről nem volt elérhető információ)
Jutalom a részvételért	89 esetben pénzjutalmat adtak, 11 esetben ajándékot vagy utalványt, 3 esetben tárgyi ajándékot, 4 esetben nem volt ilyen jutalom (16 esetben a jutalomról nem volt információ)
Jutalom a rekrutációért	58 esetben adtak pénzjutalmat, 14 esetben egyéb jutalmat, 6 esetben nem volt ilyen jutalom (45 esetben nem volt információ erről a kérdéstről)
A mintavétel időtartama	Átlagosan 9,2 hét (2–56 hét között változott)
Designhatás	1–2,5 között változott, ebből 34 esetben volt nagyobb 1,5-nél (32 esetben ilyen adatot nem közöltek)*
Mintaméret	Az átlagos mintaméret 273 volt (59 és 963 között változott), 5 esetben nem közöltek ilyen adatot; a kitűzött mintaméret átlagosan 98 százalékát érték el a kutatások (10 esetben erről nem volt adat)
Egyensúlyi állapot elérése	99 esetben a minta egyensúlyi állapotot ért el, 6 esetben nem ért el egyensúlyi állapotot (18 esetben nem állt rendelkezésre adat erről)**
Becslés módja	121 esetben az RDSAT szoftvert használták, 2 esetben más módszert

\* *Malekinejad* és *szerzőtársai* nem közlik, hogy a designhatás kiszámítása milyen módon történt az egyes kutatások során.

\*\* A szerzők nem közlik, hogy mit fogadtak el egyensúlyi állapotként (például hány változóra vonatkozóan vizsgálták az egyensúlyi állapot meglétét az egyes kutatásokban).

*Forrás: Malekinejad et al. [2008].*

A 4. táblázat azokat a kvantitatív, illetve kvalitatív tényezőket foglalja össze, amelyek a vizsgált RDS-mintavétellel készült kutatások eredményességét leginkább befolyásolták.

4. táblázat

*Az RDS-mintavételel siker- és kudarctényezői 128 nemzetközi HIV-fertőzéssel kapcsolatos kutatás metaelemzése alapján*

Sikertényezők	Kudarctényezők
Egyes csoportok esetén nagyobb valószínűséggel teljesül a kitűzött mintaméret cél (például intravénás droghasználók).	A rosszul meghatározott célcsoport a mintavétel során torzításhoz vezethet (például amennyiben a célcsoport korhoz kötött és a kapcsolatszám kérdésénél ezt nem veszik figyelembe, akkor a kapcsolatszámok torzítottak lesznek).
A rekrutálás határidőhöz kötése szignifikánsan növelte a tervezett és megvalósult minta méretének hányadosát.	A túl magas jutalmak a kuponok kereskedelméhez vezettek, míg a túl alacsony jutalmak a mintavételi időszak megnyúlását okozták.
A minta elemezhetőségéhez szükség van arra, hogy a minta egyensúlyi állapotot érjen el, amennyiben ez nem következik be, további mintát kell venni az egyensúlyi állapot eléréséig.	Az RDS-mintavétel nem használható olyan esetekben, amikor a célcsoport több zárt csoportból áll, amelyek között nincs kapcsolat (például szexmunkások, akiket akadályoznak a szabad kapcsolatfelvételen).
Pont- és intervallumbecslések készítéséhez hasznos az RDSAT szoftver alkalmazása.	

*Forrás: Johnston et al. [2008].*

Végül az áttekintett cikkek alapján néhány szó az RDS-mintavétel felhasználási területeiről. Az eddig idézett írások döntő többsége elsősorban a közegészségügyi kutatások köréből került ki (azon belül a(z) intravénás) droghasználók, szexuális kisebbségek, szexmunkások a leggyakrabban vizsgált célcsoport). Több kutatás az internethasználókat célozta meg (például a már idézett *Wejnert* [2009] tanulmány). Szórványosan egyéb szociológiai tárgyú RDS-mintavételt alkalmazó kutatások is megtalálhatók az irodalomban, mint például egy gyermekmunkával kapcsolatos elemzés (*Bjørkhauga–Hatloy* [2009]). Számos más területen is megjelennek olyan vizsgálatok, amelyek ugyan hivatkoznak az RDS-módszertanra, de valójában egyszerű hólabda mintavételen alapulnak, és nem veszik figyelembe az RDS kritériumait, és nem használják az ismertetett becsléseket (például *de Bussya* és *Wolf* [2009] kutatása az ausztrál PR professzionalizációjáról).



### 2.3. Etikai kérdések

Az áttekintett szakirodalomban az ismertetett módszertani és gyakorlati kérdéseken túl etikai dilemmák is felmerülnek (nem egy közülük gyakorlati problémát is jelent). Az itt leírt etikai kérdések elsősorban az intravénás droghasználók, illetve HIV-fertőzöttek kutatása kapcsán fogalmazódtak meg, azonban tanulságosak lehetnek bármely RDS-mintavétel tervezése és lebonyolítása esetén (természetesen a megfelelő módosításokkal).

Scott 2007-ben végzett kutatása az intravénás droghasználókat tekintve vizsgálja a válaszadó-vezérelt mintavétellel kapcsolatos problémákat, eredményei azonban más felhasználási területeken is fontosak lehetnek (Scott [2008a]). Publikációja jelentős vitát váltott ki, melyet vázlatosan ismertetek. Scott egy RDS-módszerrel végrehajtott kutatás során résztvevő-megfigyeléseket végzett és azt követően interjúkat készített. Ennek eredményei alapján elsősorban a módszerhez kapcsolódó jutalmazási rendszer alkalmazása és annak következményei vezettek olyan folyamatokhoz a vizsgált célcsoportban, amelyek etikai kérdéseket vetnek fel. Ezek a folyamatok a következők voltak:

– „Tülekedés a pénzért”: a jutalmazás elterjedő híre a célcsoportban a pénzforrásért való harchoz vezetett, ami létrehozta a kutatáshoz kapcsolódó kuponok másodgazdaságát.

– Kuponátadás, mint haszonszerzési forma: a válaszadók ajánlottaikat nem a közvetlen ismerőseik között keresték meg, hanem gyakran pénzért adták tovább a kuponokat (ezzel egyrészt további bevételre tettek szert, másrészt biztosabbak lehettek abban, hogy az adott személy ténylegesen részt vesz a kutatásban és megkapják a rekrutálásért járó összeget is).

– Félrevezető reklám: több válaszadó választotta azt a stratégiát, hogy téves információkat adott el pénzért a kutatásról, azt állítva, hogy kupon nélkül is részt lehet venni a kutatásban.

– Kockázatos hálózatok kiépülése: mivel a kuponok egy részét nem a közvetlen ismerősöknek adták tovább a megkérdezettek, új kapcsolatok jöttek létre. E kapcsolatokat vizsgálva Scott arra jutott, hogy ezekben gyakran szerepeltek valamilyen fertőzést hordozó (HIV, hepatitis) és nem fertőzött személyek.

– Életkörülményből adódó különbségek: a célcsoport kevesebb erőforrással rendelkező tagjai rosszabbul járhattak, mint a több erőforrással rendelkezők (például az autóval nem rendelkezőknek fizetniük kellett azért, hogy a kutatás helyszínére szállítsák őket).

– „Belövőhely” (shooting gallery) tulajdonosok válaszadó szervezése: akik rendszeresen használt belövőhelyeket tartottak fenn, azok

maguk szervezték, hogy kik, milyen módon vehetnek részt a kutatásban (ezzel minden résztvevő jutalmából részesedéshez jutva).

– Nem intravénás droghasználók „betanítása”: a kutatás során egyre kifinomultabb módszerek jelentek meg arra, hogy hogyan lehet megtéveszteni a kutatókat és elhíttetni velük, hogy intravénás droghasználók. Ezekért a „képzésekért” az „oktató” díjat kért.

A leírt mechanizmusok azon túl, hogy beavatkoztak a vizsgált közösség működésébe, hatalmi játszmákat erősítettek meg, és a mintavételt is károsan befolyásolhatták (például a belövőhelyek tulajdonosai által szervezett rekrutálás).

A felvetett problémákra négy cikk válaszolt, valamint egy viszontválasz született. Az egyes válaszcikkeket itt ismertetem, Scott viszontválaszát ezekhez kapcsolódóan foglalom össze.

*Broadhead* [2008] válaszcikkében elsősorban amellel érvel, hogy a felsorolt problémák nem csupán az RDS-módszert érintik, azok minden intravénás droghasználókkal kapcsolatos kutatásban előfordulhatnak. Emellett kiemeli, hogy a kutatás esetleges fertőzési kockázat növelő hatását maga Scott sem tudta bizonyítani. *Broadhead* kritikákat fogalmaz meg a kutatómódszertan kapcsán is: Scott kutatásában csupán töredéke vett részt az RDS-mintavétel alanyainak. Ezen túl *Broadhead* az RDS-kutatás vezetőivel készített interjú alapján leírja, hogy összesen 1500 kupon került kiosztásra, azonban csupán 500 válaszadó vett részt a kutatásban, vagyis a Scott által leírt fő mechanizmus, a kuponok kereskedelme valójában nem valósult meg, vagy nem ért el széles kört. *Scott* [2008b] viszontválaszában elismeri, hogy kutatásában csak a sikeres rekrutálók vettek részt (akik mindhárom kuponjukat sikeresen továbbították), azonban kiemeli, hogy kutatásának ez volt a célja.

*Lansky* és *Mastro* [2008] válaszcikkükben szintén említik, hogy a Scott által leírt problémák jelentős része más módszerrel, de hasonló célcsoportra vonatkozó kutatás esetén már korábban is felmerült. Ezen túl a szerzők kiemelik, hogy az Egyesült Államok kormányának országos közegészségügyi intézménye, a Betegségellenőrzési és Megelőzési Központok (Centers for Disease Control – CDC) finanszírozásával folyó kutatás szigorú protokollt követett, amelynek során nagy hangsúlyt fektettek a tájékoztatásra, a résztvevők beleegyezésére, valamint a vizsgált területen esetlegesen előforduló nem szándékolt események, incidensek folyamatos követésére. A szerzők (*Broadhead*hez hasonlóan) amellel is érvelnek, hogy a kutatás során alkalmazott jutalmak összege – csekély mértéke miatt –nem lehetett alkalmas arra, hogy a vizsgált közösség gazdasági és társadalmi viszonyait jelentősen befolyásolja. *Scott* [2008b] viszontválaszában kiemeli: cikkében azt állította, hogy a jutalmazási rendszer nem az egész közösségre, hanem csupán egyes résztvevőkre gyakorolt jelentős hatást, vagyis az általa ismertetett jelenségek fellelhetők a közösségben.

*Oullet* [2008], mint a chicagói droghasználatot kutató etnográfus három ponton fogalmaz meg kritikát Scott munkájával kapcsolatban: csaknem teljesen figyelmen kívül hagyja a lokális kontextust; tévesen hiszi, hogy kutatása nem befolyásolja a megfigyelt folyamatokat és interjúi során a válaszadók a teljes igazságot mondják el; valamint tévesen értelmezi a hatalmi viszonyokat, egyedi eseteket ír csupán le. *Scott* [2008b] viszontválaszában kifejti, hogy a jelenségek, amelyekről interjúalanyai beszéltek az ő megjelenése előtt történtek, így azokra nem lehetett hatással, továbbá érvel amellett, hogy az általa megfigyelt hatalmi játszmák gyors elterjedése jellemző hasonló csoportokban.

*Prachand* és *Benbow* [2008] az idézett RDS-kutatást végző hatóság (Chicago város Közegészségügyi Osztálya – Chicago Department of Public Health) részéről azt az etikai problémát vetik fel Scott publikációjával szemben, hogy olyan RDS-kutatást végzett, amelyet senki nem felügyelt és amely túlmutat kutatási megbízásán (Scott azt a feladatot kapta, hogy etnográfiai módszerekkel járuljon hozzá a mintavételi folyamat javításához). Viszontválaszában Scott kiemeli: saját egyetemének intézményi ellenőrző testülete (Institutional Review Board – IRB)<sup>11</sup> dokumentumokkal bizonyítható módon felügyelte kutatási tevékenységét, valamint azt, hogy a megfogalmazott kutatási célokat nem lépte túl.

A vitától függetlenül *Semaan* [2009] az RDS-módszerrel végzett AIDS-kutatások során felmerülő etikai és szabályozási kérdéseket összefoglaló cikkében első helyen szintén a jutalmak kérdését emeli ki. Véleménye szerint az RDS-nél alkalmazott alacsony jutalmak esetén általánosan elfogadott, hogy nem okoznak nem szándékolt hatásokat (ez alatt értve olyan cselekményeket, amelyeket a kutatás által érintett személyek a jutalom nélkül nem tennének meg). A módszerhez fűzött etikai kérdéseit az 5. táblázat foglalja össze.

5. táblázat

*AIDS-kutatások során felmerülő etikai problémák, kérdések és megoldások, válaszok*

Terület	Problémák	Megoldások, válaszok
A résztvevők tiszteltetésben tartása	A jutalmat a résztvevők kábító-szer-vásárlásra használhatják.	Egy idézett kutatás szerint a jutalmak harmadát a résztvevők saját vagy háztartási kiadásokra, negyedét közlekedésre, valamivel kevesebb, mint ötödét számlák és tartozások megfizetésére használták, a hat hónapos megfigyelési időszakban új droghasználat nem jelent meg. E mellett a kapott jutalom mértéke az átlagos havi drogadagár töredékének felelt meg.

*(A táblázat folytatása a következő oldalon.)*<sup>11</sup> Az Egyesült Államokban intézményi ellenőrző testületek felügyelik a humán kutatásokat.

(Folytatás.)

Tertület	Problémák	Megoldások, válaszok
A kölcsönösség szociológiai és pszichológiai hatása	A jutalmak olyan érzetet kelthetnek a résztvevőkben, hogy azért cserébe teljesíteniük kell valamit.	A szerző érvei szerint a felhasznált időért és fáradságért cserébe adott jutalomnak nincs ilyen hatása.
A résztvevők motivációja	A jutalmazás aláássa az önkéntes részvételi szándékot és a belső motivációt.	Egy idézett kutatás azt mutatja, hogy a válaszadók többsége (58%-a) a jutalmazás mellett több más indokot is adott a kutatásban történő részvételére.
HIV-fertőzött ember partnerének értesítése	Olyan kutatás esetén, ahol HIV-tesztet is végeznek, felmerül a szexuális partner/egészségügyi intézmények értesítésének kérdése.	Az értesítés kérdését a kutatás tervezésekor mérlegelni kell. Az értesítés akkor tehető meg, ha az érintett tud róla és beleegyezett (ennek az időpontját is meg kell tervezni: a teszteredmény ismeretében vagy a teszt végzését megelőzően). Fel kell ajánlani a segítséget a partnerekkel történő megbeszéléshez.

Forrás: Seeman [2009].

Az itt összefoglalt problémák és válaszok jelentős része bármely célcsoport kapcsán fennáll. Fontos azonban, hogy a kutatás tervezési fázisában hasonló lista készüljön a kutatásban résztvevők esetén felmerülő etikai problémákról, illetve terv készüljön a nem szándékolt hatások elkerülésére és a monitorozásra.

### 3. A válaszadó-vezérelt mintavétel a hazai irodalomban

A következőkben a hazai irodalomban előforduló (kisszámú) válaszadó-vezérelt kutatást tekintem át az eddigi módszertani fejtegetések tükrében.<sup>12</sup> Korábbiakban már említettem, hogy az RDS-mintavétel elmélete mindössze két helyen kerül említésre a magyar módszertani szakirodalomban: *Domokos és szerzőtársai* [2010] több mint 600 oldalas módszertani összefoglalójuk „Rejtett populációk becslési módszerei” című fejezetében valamivel kevesebb, mint három oldalon foglalják össze. A

<sup>12</sup> A folyóiratközléseken túl további kutatásokban is használták már a módszert, ezek egy része azonban piackutatás célú, például a TÁRKI speciális vásárlói szegmens kutatása (*Balogh Anikó* szóbeli közlése) vagy a FORSENSE Piackutató és Stratégiai Tanácsadó Kft. banki ügynök kutatása (V. HUNNET kapcsolathálózati konferencián *Bozsonyi Károly, Jelenfi Gábor és Kmetty Zoltán* prezentációja 2008-ban) vagy Bozsonyi Károly és Kmetty Zoltán 2007-ben elhangzott előadása a VI. Névtudományi konferencián a keresztnevek eloszlásának felhasználásáról a hólabda mintavétel torzításának vizsgálatához a piackutatásban.

szerzők *Heckathorn* (illetve szerzőtársai) három cikkére hivatkoznak, a mintavétel módszerét vázolják fel, annak becslési háttere nélkül (így nem alkalmasak arra, hogy kutatást tervezzenek a leírás nyomán). Domokosék egy pontatlanságát azonban érdemes korrigálni: a módszer nem valószínűségi mintát, hanem torzítatlan becslésre alkalmas mintát ad. A módszer másik magyar irodalomban megjelenő leírását *Kapitány* [2010] adja. A szerző a ritka populációk mintavételi módszereit foglalja össze, a 16 oldalas cikkben szűk egy oldalt szentel az RDS leírásának, és *Heckathorn* [1997], valamint *Salganik–Heckathorn* [2004] tanulmányára hivatkozik. Kapitány véleménye szerint „az alkalmazott súlyozási módszerek (...) nem képesek kompenzálni a mintavétel alapvető hiányosságait”, amely alatt a szerző elsősorban azt érti, hogy a célcsoportok esetén ritkán teljesülhet az eljárás azon feltétele, hogy bármely két célcsoporttagot összeköti az ismeretségi háló. Összefoglalva tehát, a magyar szakirodalom eddig nem biztosított alapos áttekintést az angolszász területen egyébként jelentős mértékben használt módszerről.

A magyar kutatási irodalmat vizsgálva összesen négy cikkben található nyoma a válaszadó-vezérelt mintavétel alkalmazásának.

6. táblázat

*Válaszadó-vezérelt kutatások a hazai irodalomban*

Szerző(k)	Vizsgált csoport	Minta mérete	Alkalmazott becslési módszer
<i>Bernát</i> [2006]	Romák	$N = 775$	n.a.
<i>Márványkövi et al.</i> [2008]	Roma intravénás droghasználók	$N = 70$	n.a.
<i>Horváth–Kmetty</i> [2008]	Külföldön élő magyar diplomások	$N = 358$	n.a.
<i>Hárs</i> [2009]	Magyar munkavállalók Ausztriában	$N = 418$	Fokszám reciprokával arányos súlyozás

A magyar szakirodalomban szereplő válaszadó-vezérelt cikkek közös jellemzője, hogy viszonylag kevés információt tartalmaznak a mintavételről, különösen annak tényleges lebonyolításáról és az alkalmazott becslési eljárásokról. A következőkben ismertetem az egyes cikkekben elérhető módszertant.

*Dávid és Snijders* [2000] tanulmánya a magyarországi válaszadó-vezérelt mintavétel előzményének tekinthető. A kutatás célja a budapesti hajléktalanok számának becslése volt, melynek során fogás-visszafogás (capture-recapture) és hólabda alapú becslési módszert használtak. Az utóbbi eljárás kapcsán került sor hólabda mintavételre, ami azonban kifejezetten csak (Snijders által kidolgozott módszer alapján) a hajléktalanok lélekszámának becslésére törekedett (az írás részletesen ismerteti az alkalmazott módszertant).

Az első hazai, valóban válaszadó-vezérelt mintavételt végrehajtó kutatást *Bernát* [2006] publikálta. A szerző célja a roma és nem roma népesség migrációs potenciáljának vizsgálata volt, különös tekintettel annak hálózati hátterére. Válaszó-vezérelt mintavételre a romákra vonatkozó minta kialakításánál került sor: 150 kistérségből vett mintát, kistérségenként 3 kezdőpontból 42 fős minta alakított ki úgy, hogy minden válaszadó 6 ajánlást adhatott, amiből egyet kerestek fel a kérdezőbiztosok (ez a design szerint 14 elemű láncokat jelentett). Így összesen 775 fős minta jött létre. A kistérségek mintáit a 2001-es népszámlálás roma válaszadóinak arányában súlyozták, így alakult ki az országos minta. A cikk nem tartalmaz információt arról, hogy milyen becslést alkalmaztak a kutatók.

*Márványkövi és szerzőtársai* [2008] kutatásának célja a budapesti roma és nem roma intravénás droghasználók ismereteinek, szerhasználói szokásainak stb. felmérése. A szerzők 3 független kezdőpontból alakították ki a 70 fős roma mintát válaszadó-vezérelt módszerrel (azaz a design szerint három 23 elemű láncban történt a mintavétel). A nem roma mintát illesztéssel alakították ki (az illesztés szempontjai: szerhasználói karrier, nem, életkor). Válaszóként három ajánlásra került sor, azonban azok közül csak egy további személyt kérdeztek meg (a nem roma minta illesztése folyamatosan történt), akik 1000 forintos ajándékutalványt kaptak (elsődleges ösztönzőként). A kérdezés négy hónapon keresztül zajlott.

*Horváth és Kmetty* [2008] kutatásának célja a Magyarországon és a külföldön élő magyar diplomások értékrendjének összehasonlítása. A hazai diplomások értékrendjének vizsgálatához az Európai társadalmak összehasonlító vizsgálatának (European Social Survey) magyarországi adatait (az 1500 fős országos minta diplomás almintáját) használták fel. A külföldön élő diplomások vizsgálatához az RDS-t alkalmazták. Az adatfelvétel interneten keresztül történt, az e-mail címet egyedi azonosítónak használták, és több különböző forrás szolgált kiindulópontként (volt, aki közvetlen személyre szóló e-mailt kapott, és volt aki levelezési listán jutott az online kérdőív címéhez). A szerzők szerint a minta „sokkal inkább tekinthető egy sokféle heterogén listáról kiinduló mintavételnek, amit kiegészített egy hólabda adatfelvételi módszer is”. Ezzel együtt megjegyzik, hogy „több vizsgált változó mentén is beállt a mintában az egyensúlyi állapot, ami a minta reprezentativitásának indikátora”. Nem található azonban információ arról, hogy hány lépésben alakították ki a mintát, illetve mekkora volt az átlagos lánc hosszúság.

Végül a legfrissebb publikált kutatás *Hárs* [2009] nevéhez fűződik, melynek célja a rendszerváltás óta Ausztriában munkát vállaló magyarok attitűdjének, valamint munkavállalásukkal kapcsolatos adatainak felmérése volt. A szerző két almintát alakított ki (a Nyugat-Magyarországi régióban, illetve az ország többi részében), valamint meggyenként adott számú kezdőpontból indult ki (a kezdőpontok száma a célcsoport korábbi kutatások alapján becsülhető arányának felelt meg). A kérdezőbiztosoknak kezdőpontként 14 elemű láncot kellett lekérdeznük úgy, hogy minden válaszadó több személyt ajánlott, akik közül meghatározott kulcs alapján véletlenszerűen választották

ki a kérdezőbiztosok a következő válaszadót. Az eredeti tervek szerint 252 fős nyugat-magyarországi és 238 fős, az ország többi részére vonatkozó mintát kellett volna létrehozni, azonban „a lekérdezés során megghiúsulások, alacsony kapcsolatszám és egyéb nehézségek miatt az országos minta 163 fős lett, míg a nyugat-magyarországi mintába 255-en kerültek”. A válaszadók mindössze 19 százaléka ajánlott kettő vagy több, 9 százaléka három vagy több személyt. A kérdezés négy hónapon keresztül zajlott. Hárs az adatokat az ajánlott személyek számával (fokszám) fordítottan arányosan, valamint (korábbi kutatás alapján) NUTS I régió, iskolai végzettség és kor alapján súlyozta, oly módon, hogy a korábbi kutatásból csak a peremeloszlásokat használta, és iteratív arányos illesztéssel alakította ki a súlyt. Az adatok érvényességét független adatok (munkavállalási statisztikák, amelyek azonban nem tartalmazzák a teljes vizsgált csoportot) segítségével ellenőrizte, amely alapján nem tekintetében jól közelített a súlyozott minta, amely hasonlóan jól becsülte a 25–44 év közötti korosztályt, ugyanakkor alulbecsülte a 45 év felettieket és a burgenlandi munkavállalókat, míg felülbecsülte a 25 év alatti és a bécsi munkavállalókat.

Összegezve a magyarországi tapasztalatokat, az irodalomban szereplő kutatások a következő jellemzőkkel írhatók le:

- lényegében nem jelennek meg a módszertanban leírt ösztönzők (a másodlagos ösztönző egyáltalán nem, az elsődleges egyetlen esetben);
- a leírt minták egy kivételével megfelelő lánchosszúságúnak tűnnek, bár a módszertani leírások általában a mintavételi tervet közlik és nem a megvalósult mintavétel leírását;
- a leírt minták nagysága változó, a becslések szórásának hiányában nem dönthető el, hogy megfelelő-e (a 70 fős minta önmagában biztosan nem megfelelő, de a kutatás célja ebben az esetben az összehasonlítás, az illesztett minta kialakítása volt);
- mindegyik ismertett kutatásra jellemző ugyanakkor, hogy nem tartalmaz információt arra vonatkozóan, vajon milyen becslést alkalmaztak a kutatók, illetve mekkora volt azok szórása; nem vizsgálták a homofiliát, sem az értékenkénti fokszám-megoszlást, illetve a hálózatok változó értékenkénti klaszterezettségét.

#### 4. Ajánlások és további kutatási lehetőségek

Az angol nyelvű irodalmat összefoglalva a következő ajánlások adhatók a hazai kutatók számára. A válaszadó-vezérelt mintavétel körültekintéssel alkalmazható, ha

- a kutatási kérdés tekintetében jelentős homofília várható;
- ettől nem függetlenül a hálózat a vizsgált változó(k) szerint klaszterezett;
- a vizsgált változó(k) értékeihez rendelhető átlagos foksám jelentős eltéréseket mutat, és a változó az egyenletestől jelentősen eltér, vagy a jelentős értékenkénti átlagos foksám-különbség mellett a minta mérete megközelíti a vizsgált sokaság méretét.

Fontos, hogy a kutatás tervezése és az elemzés esetén a következőket tartsuk szem előtt

- az RDS-mintavétel nem használható, amikor a célcsoport több zárt csoportból áll, amelyek között nincs kapcsolat;
- fontos a célcsoportot pontos meg határozása és ezt a meghatározást kell alkalmazni a kérdezés során, amikor a kapcsolatszám felmérésére kerül sor;
- a kiindulópontok megválasztásakor törekedni kell a heterogenitásra, esetleg kvóták alkalmazásával (ennek érdekében előzetes kutatást is érdemes végezni, vagy figyelembe venni a korábbi kutatások eredményeit);
- a kiindulópontok torzítását a hosszú (legalább 6 lánchosszúságú) láncok csökkentik megfelelően;
- számítani kell rá, hogy a kívánt méretű minta eléréséhez a kiegészítő kezdőpontokat kell kijelölni;
- a megkérdezettek jutalmazása tekintetében a célcsoport gazdasági státusát is érdemes figyelembe venni (túl alacsony jutalmak lassítják a mintavételt, túl magas jutalmak torzíthatják azt); a jutalmazás hatásainak etikai vonatkozásait is érdemes megfontolni az adott célcsoport esetén;
- ellentmondásos, hogy mekkora méretű mintát érdemes alkalmazni: amennyiben feltételezhető, hogy a minta nem jelentősen klaszterezett a vizsgált változók értékei szerint az egyszerű véletlen mintával összehasonlítva, elégséges a kétszeres mintaméret,<sup>13</sup> de a mintának mindenképpen egyensúlyi állapotot kell közelíteni a legfontosabb változók szempontjából (lásd korábban);

<sup>13</sup> Itt hívom fel a figyelmet arra, hogy az általánosságban használt többlépcsős, rétegzett mintavétel esetén a kutatók gyakran úgy számolnak, mintha egyszerű véletlen minta lenne, pedig ebben az esetben is van designhatás, amelynek mértéke változónként és kutatásonként eltérő módon 2–5 között alakulhat. Az RDS-mintavétel design hatásának mérlegelésénél ezt a tényt is érdemes figyelembe venni, amikor a kutatás költséghatékonyágát mérlegeljük.



- a rekrutációs idő korlátozása a korábbi kutatások tükrében javítja a mintavétel hatékonyságát;
- a foksám becsléséhez érdemes a mintavételhez kötődő, jól megfontolt kérdést feltenni;
- általában az RDS II becslést és a hozzá kapcsolódó szórásbecslést érdemes használni;
- a becslésekhez az RDSAT szoftver használata javasolt.

A hazai tapasztalatokat is figyelembe véve ezen ajánlásokra nagyobb hangsúlyt kell fektetni a további kutatások során, ugyanakkor számos csoport esetén (ahogy korábban említettem, akár piackutatási célú felhasználásra is) érdemes a válaszadó-vezérelt mintavételt alkalmazni Magyarországon is.

A nemzetközi irodalom alapján a módszerrel kapcsolatos további kutatási irány a legújabb publikációk alapján azon változók meghatározása, amelyek sem a foksámmal, sem a hálózat klaszterezettségével nem függnek össze, és így alacsony designhatás mellett a véletlen mintavételhez hasonló hatékonysággal kutathatók a válaszadó-vezérelt mintavétel módszerével.

## Irodalom

- BERNÁT, A. [2006]: A kapcsolathálózat jelentősége a migrációban – etnikai metszetben. In: *Némedi D. – Somlai P. – Szabari V. – Szikra D. (szerk.): Kötő-jelek 2005*. ELTE Társadalomtudományi Kar Szociológia Doktori Iskola. Budapest. 123–146. old.
- BJØRKHAUG, I. – HATLØY, A. [2009]: Utilization of Respondent-Driven Sampling Among a Population of Child Workers in the Diamond-Mining Sector of Sierra Leone. *Global Public Health*. Vol. 4. No 1. pp. 96–109.
- BROADHEAD, R. S. [2008]: Notes on a Cautionary (Tall) Tale About Respondent-Driven Sampling: A Critique of Scott's Ethnography. *International Journal of Drug Policy*. Vol. 19. No. 3. pp. 235–237.
- DE BUSSYA, N. M. – WOLF, K. [2009]: The State of Australian Public Relations: Professionalisation and Paradox. *Public Relations Review*. Vol. 35. No. 4. pp. 376–381.
- DÁVID B. – SNIJDERS, T. [2000]: A budapesti hajléktalanok számának becslése. *Szociológiai Szemle*. 10. évf. 3. sz. 60–75. old.
- DOMOKOS T. – FÁBIÁN R. – HORVÁTH G. C. – MÁRVÁNYKÖVI F. – MERVÓ B. – RÁCZ J. [2010]: *A szükséglet-meghatározás nemzetközi és hazai tapasztalatainak, az alkalmazott eszközök használhatóságának összefoglalása*. Nemzetközi Drogmegelőzési Intézet. Budapest. [http://elte.academia.edu/JozsefRacz/Books/1178599/A\\_szukseklet-meghatározás\\_nemzetkozi\\_es\\_hazai\\_tapasztalatainak\\_az\\_alkalmazott\\_eszkozok\\_használhatosaganak\\_összefoglalasa](http://elte.academia.edu/JozsefRacz/Books/1178599/A_szukseklet-meghatározás_nemzetkozi_es_hazai_tapasztalatainak_az_alkalmazott_eszkozok_használhatosaganak_összefoglalasa)
- GOELA, S. – SALGANIK, M. J. [2010]: Assessing Respondent-Driven Sampling. *Proceedings of the National Academy of Sciences*. Vol. 107. No. 15. pp. 6743–6747.

- GILE, K. J. – HANDCOCK, M. S. [2009]: Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methodology*. Vol. 40. No. 1. pp. 285–327.
- HÁRS, Á. [2009]: *Magyarok az osztrák munkaerőpiacon: Ingázók, bevándorlók, munkaerő-migránsok?* KOPINT Konjunktúra Kutatási Alapítvány.  
<https://skydrive.live.com/?cid=20226f10b70b2c25&id=20226F10B70B2C25%21311>
- HECKATHORN, D. D. [1997]: Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*. Vol. 42. No. 2. pp. 174–199.
- HECKATHORN, D. D. [2007]: Extensions of Respondent-Driven Sampling: Analyzing Continuous Variables and Controlling for Differential Recruitment. *Sociological Methodology*. Vol. 37. No. 1. pp. 151–207.
- HORVÁTH Á. – KMETTY Z. [2008]: Mást gondolnak a világról? A külföldön élő magyarországi diplomások értékrendje. In: *Füstös L. – Guba L. – Szalma I. (szerk.): Társadalmi regiszter*. 1. sz. 153–171. old.
- JOHNSTON, L. G. – MALEKINEJAD, M. – KENDALL, C. – IUPPA, I. M. – RUTHERFORD, G. W. [2008]: Implementation Challenges to Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance: Field Experiences in International Settings. *Aids and Behavior*. Vol. 12. No. 4. pp. 131–141.
- KAPITÁNY B. [2010]: Mintavételi módszerek ritka populációk esetén. *Statistikai Szemle*. 88. évf. 7–8. sz. 739–754. old.
- KENDALL, C. – KERR, L. R. – GONDIM, R. C. – WERNECK, G. L. – MACENA, R. H. – PONTES, M. K. – JOHNSTON, L. G. – SABIN, K. – MCFARLAND, W. [2008]: An Empirical Comparison of Respondent-Driven Sampling, Time Location Sampling, and Snowball Sampling for Behavioral Surveillance in Men Who Have Sex with Men, Fortaleza, Brazil. *AIDS and Behavior*. Vol. 12. No. 4. pp. 97–104.
- KRAL, A. H. – MALEKINEJAD, M. – VAUDREY, J. – MARTINEZ, A. N. – LORVICK, J. – MCFARLAND, W. – RAYMOND, H. F. [2010]: Comparing Respondent-Driven Sampling and Targeted Sampling Methods of Recruiting Injection Drug Users in San Francisco. *Journal Urban Health*. Vol. 87. No. 5. pp. 839–850.
- LANSKY, A. – MASTRO, T. D. [2008]: Using Respondent-Driven Sampling for Behavioural Surveillance: Response to Scott. *International Journal of Drug Policy*. Vol. 19. No. 3. pp. 241–243.
- MALEKINEJAD, M. – JOHNSTON, L. G. – KENDALL, C. – KERR, L. E. F. S. – RIFKIN, M. R. – RUTHERFORD, G. W. [2008]: Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance in International Settings: A Systematic Review. *AIDS and Behavior*. Vol. 12. No. 4. pp. 105–130.
- MÁRVÁNYKÖVI F. – MELLES K. – LÉGMÁN A. – RÁCZ J. [2008]: A kezelésbe jutás akadályai roma és nem roma, kezelésen kívüli intravénás használok körében Budapesten. *Addiktológia*. VII. évf. 3–4. sz. 205–233. old.
- OUELLET, L. J. [2008]: Cautionary Comments on an Ethnographic Tale Gone Wrong. *International Journal of Drug Policy*. Vol. 19. No. 3. pp. 238–240.
- PRACHAND, N. G. – BENBOW, N. [2008]: Clarifying the Ethnographer's Role in Chicago's HIV Behavioural Surveillance-Injecting Drug Users Cycle, 2005: Response to Scott. *International Journal of Drug Policy*. Vol. 19. No. 3. pp. 244–245.

- SALGANIK, M. J. – HECKATHORN, D. D. [2004]: Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling. *Sociological Methodology*. Vol. 34. No. 1. pp. 193–239.
- SALGANIK, M. J. [2006]: Variance Estimation, Design Effects, and Sample Size Calculations for Respondent-Driven Sampling. *Journal of Urban Health: Bulletin of the New York Academy of Medicine*. Vol. 83. No. 7. pp. 98–112.
- SCOTT, G. [2008a]: “They Got Their Program and I Got Mine”: A Cautionary Tale Concerning the Ethical Implications of Using Respondent-Drive Sampling to Study Injection Drug Users. *International Journal of Drug Policy*. Vol. 19. No. 1. pp. 42–55.
- SCOTT, G. [2008b]: Responses on a Cautionary Tale Concerning the Ethics of Using Respondent-Driven Sampling to Study Injection Drug Users. *International Journal of Drug Policy*. Vol. 19. No. 3. pp. 246–247.
- SEMAAN, S. [2009]: Ethical and Regulatory Considerations in HIV Prevention Studies Employing Respondent-Driven Sampling. *International Journal of Drug Policy*. Vol. 20. No. 1. pp. 14–27.
- SIRKEN, M. G. [1998]: *A Short History of Network Sampling*. National Center for Health Statistics. Hyattsville. [http://www.amstat.org/sections/srms/Proceedings/papers/1998\\_001.pdf](http://www.amstat.org/sections/srms/Proceedings/papers/1998_001.pdf)
- SPREEN, M. [1992]: Rare Populations, Hidden Populations, and Link-Tracing Designs: What and Why. *Bulletin de Méthodologie Sociologique*. Vol. 36. No. 1. pp. 34–58.
- TOMAS, A. – GILE, K. J. [2010]: The Effect of Differential Recruitment, Non-response and Non-recruitment on Estimators for Respondent-Driven Sampling. <http://arxiv.org/abs/1012.4122v1>
- VERDERY, A. M. – MOUW, T. [2011]: *Estimated Sampling Variance in Respondent Driven Sampling: Do You Know If You Have an Imprecise Sample?* Working Paper.
- VOLZ, E. – HECKATHORN, D. D. [2008]: Probability Based Estimation Theory for Respondent Driven Sampling. *Journal of Official Statistics*. Vol. 24. No. 1. pp. 79–97.
- VOLZ, E. – WEJNERT, C. – DEGANI, I. – HECKATHORN, D. D. [2007]: Respondent-Driven Sampling Analysis Tool (RDSAT) Version 5.6. Cornell University. Ithaca. <http://respondentdrivensampling.org/>
- WEJNERT C. [2009]: An Empirical Test of Respondent-Driven Sampling: Point Estimates, Variance, Degree Measures and Out-of-Equilibrium Data. *Sociological Methodology*. Vol. 39. No. 1. pp. 73–116.

## Summary

The sampling and estimating method, called respondent driven sampling, is widely used in English language literature, but it is not well known in Hungary. The author in his article beyond introducing the method discusses challenges, experience and ethical problems based on the review of current English literature. He gives recommendations for using the method in Hungary based on the national and international experience and suggests further ways of methodological research as well.