

Simon Dávid

A válaszadó-vezérelt mintavétel –
kísérlet egy módszer társadalomtudományi szemléletű
bemutatására

2019

EÖTVÖS LÓRÁND TUDOMÁNYEGYETEM
TÁRSADALOMTUDOMÁNYI KAR
SZOCIOLÓGIA DOKTORI ISKOLA
SZOCIOLÓGIA KÉPZÉSI PROGRAM

KONZULENS: DR. FOKASZ NIKOSZ EGYTEMI TANÁR DSC

Tartalom

I.	Bevezetés.....	4
I.1.	Egy mintavétellel szemben támasztott módszertani követelmények.....	6
I.2.	A válaszadó-vezérelt mintavétel, mint tudományos konstrukció.....	9
II.	A válaszadó-vezérelt mintavétel bemutatása.....	13
II.1.	A válaszadó-vezérelt mintavétel helye a mintavételi eljárások között.....	13
II.2.	A válaszadó-vezérelt mintavétel előzményei és módszertani alapjai.....	18
II.3.	A válaszadó-vezérelt mintavétel matematikai statisztikai alapjai.....	22
II.4.	Módszertani megfelelés és matematikai környezet.....	34
III.	A válaszadó-vezérelt mintavétel gyakorlati alkalmazása.....	36
III.1.	A válaszadó-vezérelt mintavétel alkalmazása nemzetközi kutatásokban.....	36
III.2.	A válaszadó-vezérelt mintavétel alkalmazása hazai kutatásokban.....	43
III.3.	A válaszadó-vezérelt mintavétel alkalmazói, megrendelői és társadalmi környezete.....	54
IV.	A válaszadó-vezérelt mintavétel alkalmazása kapcsán felmerülő módszertani kérdések....	56
IV.1.	A válaszadó-vezérelt mintavétel tulajdonságaira irányuló kutatások.....	56
IV.2.	A válaszadó-vezérelt mintavétel egyes feltételeinek vizsgálata és kezelésük.....	65
IV.3.	A válaszadó-vezérelt mintavétel vizsgálata szimuláció segítségével.....	69
IV.3.1.	Korábbi szimulációs kutatások összefoglalása.....	69
IV.3.2.	A szimulációs vizsgálat kutatási kérdése.....	72
IV.3.3.	A szimulációs kutatás módszertana.....	73
IV.3.4.	A szimulációs kutatás eredményei.....	74
IV.3.5.	A szimulációs kutatás konklúziója és további kutatási lehetőségek.....	79
IV.4.	Módszertani megfelelés és alkalmazói környezet.....	82
V.	A válaszadó-vezérelt mintavétel és szociológiai elméletek.....	83
V.1.	Elméleti megközelítések.....	83
V.2.	A válaszadó-vezérelt mintavétel szociológiai relevanciáját megerősítő empirikus kutatás.....	87
V.3.	A válaszadó-vezérelt mintavétel szociológiai kutatási környezete (összefoglaló).....	89
VI.	A válaszadó-vezérelt mintavétel jogi és etikai szempontjai.....	91
VI.1.	Jogi, adatvédelmi kérdések.....	91
VI.2.	Etikai kérdések.....	99
VI.3.	A válaszadó-vezérelt mintavétel jogi és etikai környezete.....	102
VII.	Következtetések.....	104
VII.1.	A válaszadó-vezérelt mintavétel módszertani szempontból.....	104
VII.2.	A válaszadó-vezérelt mintavétel szociológiai kontextusa.....	105
VII.3.	Záró megjegyzések és további kutatási lehetőségek.....	107
VIII.	Irodalom.....	109

I.	Melléklet.....	120
II.	Melléklet.....	121
III.	Melléklet.....	124
IV.	Melléklet.....	126
V.	Melléklet.....	127

I. Bevezetés

Dolgozatomban egy mintavételi eljárás, az úgynevezett válaszadó-vezérelt mintavétel (respondent driven sampling, RDS) alkalmazási kérdéseit járom körül. Dolgozatom célja azonban nem egyszerűen az eljárás bemutatása, a felmerülő szempontok (statisztikai, módszertani, etikai, jogi) áttekintése. A szűkebb módszertani megközelítésen túl dolgozatomban arra törekszem, hogy egy módszertani eszközt, mint a társadalomtudományi kutatás tárgyát, mint tudományos konstrukciót mutassam be, amely különféle társadalmi szereplőkkel és társadalmi jelenségekkel interakcióban működik, használható, értelmezhető.

Nem titkolt szándékom ezzel a megközelítéssel egyfajta új nézőpontot teremteni a társadalomkutatási módszertani kérdéseknek és ezen keresztül új értelmezést adni a társadalomkutatás módszerének. Céлом egyrészt, hogy példát mutassak egy módszertani kérdés társadalomtudományi érzékenységű kezelésére, másrészt felhívjam a figyelmet arra, hogy a társadalomkutatás módszertana a szociológia integráns, bizonyos értelemben meghatározó része. Ez a gondolat természetesen nem tekinthető újnak. A módszer kiemelt szerepére a szociológiában már Simmel [1909] felhívta a figyelmet, amikor a szociológia tudományokon belül elfoglalt helyét nem kutatása tárgyával, hanem annak módszerével specifikálta. Természetesen a módszer fogalma Simmelnél nem tekinthető azonosnak a módszertan mai általános értelmezésével. Míg Simmel a módszert, mint megközelítést, azaz a kutató és a kutatás tárgya közötti viszonyt értelmezte, addig mára a módszertan lényegében az eszközök megfelelő használatként értelmeződik át. Ugyanakkor a módszertani eszközök befolyásolják a szociológiai elméletek verifikálhatóságát (vagy falszifikálhatóságát) és fordítva, a szociológiai elméletek új módszertani eszközök kifejlesztését indukálhatják vagy legalább is tehetik lehetővé.

Van példa arra is, hogy a mai értelemben vett módszertan, illetve bizonyos módszertan alkalmazása paradigmászerű képződmények létrejöttéhez vezet. Ilyen a kapcsolathálózati megközelítés szociológiai alkalmazása, amely egyrésztől merőben módszertani alkalmazás (hálózatelemzés), másrésztől felveti a szociológián belül egy, a korábbiaktól eltérő elmélet lehetőségét, bár korántsem mondható ki, hogy valóban paradigmáról van szó: ahogy a kérdést alaposan körüljáró Tardos Róbert [1995] cikkében fogalmaz „inkább kezdeményekről, egyedi építőkövekről beszélhetünk, mint egységes építményről”.

Természetesen a leírttal ellentétes irányú hatások is érvényesülnek: a szociológia elméleti változásai is befolyásolják a módszertan alakulását. Nyilvánvalóan más eszköztár alkalmazását

teszik szükségessé a társadalmi jelenségeket külsődleges társadalmi tényekkel magyarázó elméletek, amint azok, amelyek elsősorban az intenciókra, attitűdökre koncentrálnak, vagy azok, amelyek a társadalom működését éppen a különféle társas interakciókból vezetik le. Ebben az értelemben a szociológia elméleti változásai gyakorolnak nyomást új mérőeszközök kifejlesztésének irányába.

Dolgozatomnak a leírtaknak megfelelően két olvasata van: egyrészt egy, a módszertan keretein belül értelmezhető elemzés egy konkrét módszer alkalmazhatósága kapcsán, másrészt egy módszer elhelyezése, kontextualizálása a társadalomtudományos, illetve szűkebben véve szociológiai fogalomtérben. Ennek megfelelően két kérdésre keresem a választ: mennyiben alkalmazható a válaszadó-vezérelt mintavétel a szociológiai kutatásokban a szociológiai módszertan elfogadott követelményei alapján és milyen módon áll kölcsönhatásban e módszer egyes társadalmi jelenségekkel és a szociológiai elméletekkel.

Előre kell bocsátanom, hogy egyik kérdésre sem adhatok sem végleges, sem egyértelmű választ. A válaszadó-vezérelt mintavétel mint módszer, amint a dolgozat végére nyilvánvalóvá válik, sok problémával küzd, ugyanakkor számos lehetőséget is rejt. A problémák bármelyike válhat olyan súlyossá, hogy későbbi kutatások alapján ezt a módszert végképp száműzni kell a társadalomkutatás módszertani tárházából. Ugyanakkor a lehetőségek közül több is kecsegtet azzal, hogy esetleg új utakat nyithat a társadalom, illetve annak folyamatai jobb megismeréséhez, akár paradigma szintű új elméletek megalkotásához is. A társadalomkutatások módszertanának elhelyezése a tudomány egészén belül pedig mindig nézőpont, tudományfelfogás kérdése marad. Bár igyekezni fogok érvekkel alátámasztani állásponantomat, az a dolgozat végére is megmarad álláspontnak, tekintettel arra, hogy a tudomány egésze, annak felosztása, az egyes területek besorolása társadalmi konstrukció, abban az értelemben, hogy a társadalom és azon belül a tudománnyal kapcsolatba hozható csoportok (tudósok, tudománypolitikusok, az eredményeket felhasználó vállalkozók, az eredményeket befogadó társadalom) interakcióján és – többé kevésbé létrejövő, ugyanakkor dinamikus – konszenzusán alapul.

Dolgozatomban a felvetett kérdéseket hét fejezetben tárgyalom. Minden egyes fejezetben egyszerre próbálom megválaszolni mindkét fő kérdést, de visszautalásokkal, értelmezéssel igyekszem világossá tenni, hogy mely állításaim melyik kérdés szempontjából milyen jelentőséggel bírnak. Természetesen az egyes áttekintett témák különböző mértékben érintik a két feltett kérdést.

A bevezető fejezet további részében bővebben kifejtem dolgozatom fent leírt két kutatási kérdését. A második fejezetben a módszerrel kapcsolatos alapvető irodalomra támaszkodva magát az eljárást mutatom be. A harmadik fejezetben a módszer gyakorlati alkalmazását mutatom be a nemzetközi és hazai kutatások gyakorlatának összefoglalásán keresztül. A negyedik fejezet a válaszadó-vezérelt mintavétellel kapcsolatos szorosan vett módszertani kérdéseket tisztázza a vonatkozó irodalom összefoglaló, kritikai értelmezésén keresztül. E fejezet részeként egy saját szimulációs kutatás eredményeit is bemutatom a vizsgált módszer alkalmazhatóságához kapcsolódóan. Az ötödik fejezetben a vizsgált módszer és a társadalomtudományi elméletek között keresek kapcsolatot. A hatodik fejezetben a válaszadó-vezérelt mintavétel tágabb társadalmi kontextusaként a kapcsoló jogi és etikai kérdésekkel foglalkozom a vonatkozó irodalom, jogszabályok, illetve korábbi kutatásaim ilyen irányú „melléktermékei” kapcsán. Végül a hetedik fejezetben összefoglalom a feltett két kérdésre adott válaszaimat.

I.1. Egy mintavétellel szemben támasztott módszertani követelmények

Ahhoz, hogy a válaszadó-vezérelt mintavétel tudományosan elfogadott módszertani követelményeknek való megfelelését megvizsgáljam, tisztázni kell röviden, hogy melyek ezek a követelmények. A survey alapú kutatások módszertani szakirodalma a XX. század közepére alaposan körülhatárolt elvárásokat fogalmaz meg a mintavételi eljárásokkal szemben. Nehézséget elsősorban az okoz, hogy a kidolgozott követelményrendszer a szinte egyeduralkodó véletlen mintavételt veszi alapul. Ennek ellenére a leírtak használhatók a hagyományostól eltérő mintavétel esetén is. Mára már minden survey kutatásokkal foglalkozó módszertani kézikönyv implicit vagy explicit módon rögzíti ezeket a követelményeket (pl.: Pfeffermann, Rao [2009], Babbie [2008] vagy Freedman és mtsai. [2005]). A szociológiai kutatás céljából – azaz a társadalmi jelenségek megismeréséből – kiindulva a kvantitatív empirikus kutatásokkal szemben egyetlen rövid állítással megfogalmazható az elvárt követelmény: az adatok írják le pontosan a vizsgált jelenséget. Közelebről megvizsgálva a mintavétel szemszögéből ez azt jelenti, hogy a szociológiai kutatási kérdés (vizsgált jelenség) alapján definiált populációra (továbbiakban a célcsoportra) vonatkozóan képesek legyünk levonni pontos következtetéseket¹. A következtetéseket mérések alapján vonjuk le. Ez

¹ Egy vizsgált társadalmi jelenségre vonatkozó következtetés levonása kvantitatív empirikus alapokon meglehetősen bonyolult, soktényezős probléma (magában foglalja a kutatási kérdések alapján hipotézisek megalkotását, konceptualizálását, operacionalizálását, a mintavételt, a lekérdezést és a statisztikai elemzőeszközök alkalmazását), amelyek nézetem szerint együttesen alkotják a kvantitatív empirikus szociológiai kutatás módszertanát. Mindezekből jelen dolgozat szempontjából egy elemet, a mintavételt emelek ki érintve

lehetséges elvileg olyan módon, hogy a vizsgálatba bevonjuk a teljes célcsoportot. Az adatgyűjtés szempontjából – más tényezők azonossága esetén – ez vezetne a legpontosabb eredményekhez. Az, hogy ez a leggyakrabban – kisméretű célcsoportok kivételével – nem következik be egyszerűen az erőforrások (pénz, idő, információ) szűkösségének a következménye. Marad tehát az a módszer, hogy a célcsoport egy része körében gyűjtünk adatokat. A részpopuláció kiválasztása a célcsoportból annak érdekében, hogy a teljes célcsoportra következtetéseket vonhassunk le az, amit mintavételként definiálhatunk. Az idézett kézikönyvek két fogalommal ragadják meg az ilyen részpopuláció képessé kapcsolatos követelményeket: az érvényességgel és a megbízhatósággal. Érvényesség alatt azt értjük, hogy azt mérjük, amit elméleti szempontból mérni akarunk. Ha feltesszük, hogy az egyedi mérések esetén a mérőeszközünk érvényes, azaz az egyedi mérések pontosan azt mérik, amit a kutatási kérdésből levezetve mérniük kell, akkor a mérés érvényessége a mintavételen és annak megvalósulásán, valamint az alkalmazott statisztikai elemzési eljárásokon múlik. Statisztikai fogalmakat használva érvényes az a mintavétel, amely alapján torzítatlan becslés adható a célcsoport valamely tulajdonságára. A két utóbbi kijelentés egymással vizsgálatom szempontjából ekvivalens. A másik szempont, a megbízhatóság azt mutatja meg, hogy amennyiben a vizsgálatot többször megismételjük, az eredmények milyen mértékben térnének el egymástól. Statisztikai értelemben a megbízhatóság az alkalmazott mintavétel során létrehozott minta alapján készített célcsoportra vonatkozó becslési eljárások szórásaként fogható fel. Fontos megérteni a két fogalom különbözőségét a szociológiai kutatás egésze szempontjából. Ha feltesszük, hogy egy olyan kutatómódszertant alkalmaznak, amely nem érvényes, az folyamatosan téves következtetésekhez vezet az azonos tartalmú kutatások esetén, míg a kevésbé megbízható, de egyébként érvényes kutatómódszertan eredményei esetleg ellentmondásosak, de többségükben tükrözik a mérhető valóságot (ismét csak nem foglalkozva a valóság mérhetőségének egyéb kérdéseivel). Ennek folytán egy módszer, és így egy mintavétel szempontjából is elsődleges, hogy az minél inkább érvényes legyen és másodlagos, hogy minél inkább megbízható legyen. Továbbá következik ebből az is, hogy míg az érvényességnek létezik elvi maximuma, a megbízhatóság maximuma csak a teljes körű adatfelvételek esetén érhető el. Ugyanakkor fontos szempont, hogy egy eljárás megbízhatósága az elérendő cél szempontjából megfelelő legyen, azaz legyen lehetőség a rendelkezésre álló forrásokból megfelelő pontosságú becslések kiszámítására.

esetenként a minta lekérdezésének praktikus kérdéseit, illetve a becslés kérdését is tekintettel arra, hogy az itt vizsgált módszer esetén a két tényező nehezen választható szét.

Az érvényesség kérdését a mintavétel területén egy további tényező alapvetően befolyásolja, így ezzel külön is érdemes foglalkozni. Egy mintavétel érvényességének szükséges – de nem elégséges – feltétele, hogy a célcsoport minden tagja elvileg bekerülhessen a mintába, amennyiben nem rendelkezünk valamilyen előzetes információval a populáció tagjainak hasonlóságáról a vizsgált jelenségek szempontjából. A mintavétel területén ezt lefedettségként (coverage) definiáljuk. Amennyiben a célcsoport valamely szegmense nem kerülhet a mintába, akkor az lefedettségi hibához (coverage error) vezethet, amely pontosan meg nem határozható² torzítást, érvényességi problémát (torzítást) okoz. Ezért az érvényesség értékelésénél fontos számba venni a lefedettség kérdését is.

Az adatfelvétel lebonyolítása további érvényességet is érintő problémát vet föl: a meghiusulás (unit nonresponse) problémáját. A meghiusulás problémája annyiban hasonló a lefedettség problémájához, hogy ennek következménye is az, hogy a célcsoport mintába bekerült alcsoportjának egy része esetén nem fog megtörténni a mérés. Fontos különbség azonban, hogy a meghiusulás kétféle módon következhet be: véletlenszerűen vagy szisztematikusan. Míg az első esetben többször megismételt azonos kutatási módszertan mellett az adott populáció tag bekerülhet a mintába, addig az utóbbi esetben az adott populáció tag soha nem kerülhet be a mintába. Ebből könnyen levezethető, hogy a szisztematikus meghiusulás lényegében megfeleltethető egy rejtett lefedettségi problémának, így minden igaz rá, ami a lefedettségre, tehát érvényességi hibát, azaz torzítást okozhat. Ezzel szemben a véletlen meghiusulás az érvényességet nem érinti, ha ismerjük annak bekövetkezési valószínűségét minden célcsoporttagnál³.

Végül, de nem utolsó sorban egy mintavételi eljárás kapcsán érdemes vizsgálni, hogy a fenti paraméterekre adható-e előzetes becslés. Ennek a szempontnak a jelentősége kettős: részben befolyásolja a kutatás tervezhetőségét, részben – és ez a fontosabb szempont – nagyon jelentősen meghatározza az eredmények interpretációját. Amennyiben az érvényesség és megbízhatóság a kutatás tervezése során jó közelítéssel becsülhető és egyes tervezési paraméterekkel befolyásolható, akkor a tervezett minőségű (pontosságú) kutatás hozható létre.

² A célcsoport le nem fedett részére vonatkozóan bizonyos feltételezések mellett élhetünk következtetésekkel, azonban ezek a feltételezések csak akkor ellenőrizhetőek, – azaz akkor fogadhatók el tudományos értelemben igaznak, – ha egy másik kutatás, adatgyűjtés lefedti ezt az alcsoportot is.

³ Könnyen belátható egy egyszerű véletlen minta esetén, ahol minden mintatagnak azonos a mintába kerülési valószínűsége, hogy a célcsoport azon alpopulációja, akiknél a véletlen meghiusulás előfordul, összességében minden mintába kisebb valószínűséggel kerülnek be. Így, amikor olyan becslési eljárást használunk, amelyik egyetlen mintába kerülési valószínűséget feltételez torzított, nem érvényes eredményt kapunk (kivéve, abban a ritka esetben, amikor az adott alpopuláció pontosan azonos a vizsgált jelenség szempontjából a célcsoport többi részével).

Másrészt az érvényesség és megbízhatóság ismeretében megadható, hogy milyen mértékű eltérések mutathatók ki az adott kutatás segítségével, ami megóvja a kutatót téves következtetések levonásától

Az alfejezet összefoglalásaként a leírt kutatási kérdés kapcsán megfogalmazhatók módszertani fő- és alhipotézisem is:

H_I: Főhipotézisem szerint a válaszadó-vezérelt mintavétel a társadalomtudományi kutatás módszertani követelményeinek megfelel.

H_{Ia}: A válaszadó-vezérelt mintavétel alkalmas a vizsgált célcsoportra vonatkozó érvényes következtetések levonására.

H_{Ib}: A válaszadó-vezérelt mintavétel megbízhatósága megbecsülhető és csak annyival kisebb, mint a hagyományos mintavételeké, hogy az még az alkalmazhatóságának nem gátja.

I.2.A válaszadó-vezérelt mintavétel, mint tudományos konstrukció

Ebben a fejezetben – a korábban leírt kutatási kérdések kifejtéseként – felvázolom a vizsgált jelenség, a válaszadó-vezérelt mintavétel mint tudományos konstrukció modelljét (1. ábra). A tudományos konstrukció fogalmát azért tartom célszerűnek bevezetni, hogy elemzésem tárgyát entitásként értelmezhessem, vizsgálhassam annak környezetével együtt, mintegy interakciós helyzetben. Ezt azért tartom fontosnak, mert így lehetőség van arra, hogy ne egyszerűen egy eszközt és a vele szemben támasztott követelményeket vizsgáljam, hanem – bizonyos távolságot tartva – az eszköz keletkezésének, használatának, működésének körülményeit is. Ennek megfelelően a tudományos konstrukció fogalmának bevezetése csupán analitikus eszköznek tekinthető. Tudományos konstrukció alatt egy olyan entitást értek, amely tudományos eredményeken alapul és interakcióban áll a későbbi kutatásokkal, illetve tudományos és tágabb környezetével, ebben az értelemben tehát egy konstrukció, amelyet a tudományos közösség hoz létre, és a tudományos közösség – legalább és általában – egy része elfogad.

1. ábra A válaszadó-vezérelt mintavétel, mint tudományos konstrukció modellje



Forrás: saját ábra

A válaszadó-vezérelt mintavétel, mint a szociológiai kutatás egy lehetséges módszere, tudományos konstrukció – más szociológiában alkalmazott módszerekhez hasonlóan – komplex környezetben helyezkedik el. Az ábrán az egyes tényezők a vizsgált tudományos konstrukció (módszer) szempontjából környezetként jelennek meg. Környezet alatt ebben az esetben a módszer működésére, fejlődésére, használatára és használhatóságára hatást gyakorló szereplőket, tényezőket és tudományos konstrukciókat értek. A környezet egyes elemei különféle hatások csoportjait jelentik, amelyek részben saját önkényes alkotásaim, részben a későbbiekben idézett irodalmak alapján hoztam létre. A környezet egyes elemei között az ábrázolt kapcsolatok semmiképpen sem tekinthetők teljesnek, inkább azt jelzik, hogy mely kapcsolatokat tekintek meghatározónak, illetve mely kapcsolatokat kívánok vizsgálni elemzésem során. Így például a társadalom, mint a környezet egy eleme biztosan kapcsolatban van az összes többi környezeti elemmel, azonban elemzésem szempontjából a vizsgált tudományos konstrukció és a környezet egyes itt definiált elemei közötti kapcsolatok tekinthetők elsődlegesnek és vizsgálandónak. Ezt az érvelést alapul véve viszont külön indoklásra szorul minden olyan jelölt – és vizsgálandó – kapcsolat, amely nem a vizsgált tudományos konstrukció és a környezet egy eleme között húzódik. Az etikai, illetve jogi környezet esetén a kérdések megvitatása során a hivatkozott szerzők a vizsgált tudományos konstrukció és a társadalom közötti kölcsönhatást elemzik elsősorban, ezért szükséges jelezni, hogy itt ezt a kapcsolatot érinti elemzésem. A módszertani és megrendelői környezet esetén

pedig fontos megkülönböztetni az adott környezeti elemen belül elkülöníthető két elemet, amelyek hasonló irányból hatnak, ugyanakkor indíttatásukban eltérnek. A „megrendelői” környezet elemei a vizsgált tudományos konstrukciót, mint az általuk megfogalmazott kérdések megválaszolásának eszközt alkalmazzák, de míg a szociológiai kutatási környezet célja a társadalmi jelenségek összefüggéseinek feltárása, a piackutatási környezet célja a profit maximalizálása⁴. Ez a különbségtétel a vizsgált tudományos konstrukcióval szemben támasztott elvárások különbözőségében és végső soron a gyakorolt hatások különbözőségében is jelentkezik. A módszertani környezet esetén indokolt elválasztani egymástól az elméleti háttérrel jelentő matematikai környezetet és az alkalmazói környezetet, amely a módszer praktikus kivitelezőit, illetve az általuk meghonosított gyakorlatot foglalja magában. Ez utóbbi két környezeti elem többé-kevésbé elkülöníthető lesz a későbbiek során, amikor a válaszadó-vezérelt mintavétellel kapcsolatos módszertani kutatásokat vizsgálom.

A modell alkalmazásával az a céltom tehát, hogy világosabban fel tudjam vázolni második kutatási kérdéseimre adandó válaszaimat, azaz azt, hogy a válaszadó-vezérelt mintavétel milyen módon áll kölcsönhatásban különböző társadalmi jelenségekkel és a szociológiai kutatásokkal.

Érdekes azt is itt tisztázni, hogy milyen kérdések vethetők fel a fent felvázolt modellben a hatások szempontjából. Az alapvető kérdés arra vonatkozik, hogy az adott tudományos konstrukció, mint tudományos konstrukció felhasználható-e. Esetünkben ez azt jelenti, hogy a válaszadó-vezérelt mintavétel, mint módszer alkalmazható-e. Ez a kérdés a módszertani környezet esetén megegyezik kutatásom első kérdésével, azonban a többi környezet szempontjából merőben eltér attól. Nem alkalmaznám azonban ezt a modellt, ha annak egyetlen előnye az lenne, hogy a módszertani környezetben (a módszertani normák alapján) megfogalmazott alkalmazhatósági kritérium kiterjeszhetővé válik más környezeti elemekre. Az adott környezeti elem szempontjából értelmezhető alkalmazhatóságon túl ugyanis fontos, tanulságos és megértendő szempontnak látom azt, hogy az egyes környezeti elemek milyen összefüggésben állnak a tudományos konstrukcióval, annak alkalmazásával, illetve alakulásával. Utóbbi összefüggések megértése – az tudományos megismerési vágy kielégítésén túl – azzal a praktikus haszonnal is bírhat, hogy ráirányítja a figyelmet a vizsgált tudományos

⁴ Bár a piackutatási környezetre nem térek ki dolgozatomban részletesen, könnyű belátni, hogy a profitmaximalizálás a kutatási eszközök terén a minőségi paraméterek és a költségek, illetve idő tényező közül az utóbbiak optimalizálását eredményezi, így egy módszer alkalmazására teljesen már hatást gyakorol, mint a kutatási környezet.

konstrukcióról megfogalmazott megállapítások kontextusára. Ez a kontextus adott esetben módosíthatja a megállapításokat, azok jelentőségét, illetve azok értelmezését.

Az alfejezet összefoglalásaként megfogalmazhatók fő- és alhipotéziseim a válaszadó-vezérelt mintavétel és egyes társadalmi jelenségek, valamint szociológiai elméletek kapcsolatáról:

H_{II}: Főhipotézisem szerint a válaszadó-vezérelt mintavétel alkalmazását és alkalmazhatóságát befolyásolja a jogi, etikai, társadalmi, megrendelői és alkalmazói környezete.

H_{IIa}: A válaszadó-vezérelt mintavétel az alkalmazói környezet matematikai ága szempontjából kevés kihívást jelent.

H_{IIb}: A válaszadó-vezérelt mintavétel az alkalmazói környezet módszertani ága szempontjából jelentős kihívást jelent, ami az alkalmazási terület, célcsoport társadalmi sajátosságai miatt alakul ki.

H_{IIc}: A válaszadó-vezérelt mintavétel szociológiai kutatási környezete jelenlegi tendenciái alapján megrendelőként lép(het) fel a módszer kapcsán az elméletekben megfogalmazott és empirikusan igazolt társadalmi változások miatt.

H_{IIId}: A jogi és etikai környezet a válaszadó-vezérelt mintavétel alkalmazásának komoly korlátjává válhat, akár el is lehetetleníti a módszer alkalmazását.

II. A válaszadó-vezérelt mintavétel bemutatása⁵

A hazai módszertani szakirodalomban – a jelen fejezet alapját jelentő cikken kívül is – szerepel néhány rövid összefoglalás [Kapitány, 2010; Domokos és szerzőtársai, 2010] az itt ismertetésre kerülő módszerről, azonban azok – feltehetően a terjedelmi korlátok miatt – nem térnek ki a módszer részleteire, a nemzetközi irodalomban fellelhető vitás pontokra, esetenként sommás állításokat tartalmaznak. Ebben a fejezetben bemutatom a vizsgált módszer előzményeit, alapjait, annak módszertani és matematikai kapcsolódásait, azaz a vizsgált jelenség elsősorban matematikai másodsorban módszertani környezetét.

II.1. A válaszadó-vezérelt mintavétel helye a mintavételi eljárások között

A szociológia kvantitatív empirikus megalapozásának törekvése a tudomány születése óta érvényesül (gondoljunk akár Durkheim Öngyilkosságára vagy Spencer munkásságára). Ez a XX. század elejéig elsősorban a teljes körű (és általában intézményi célú) adatgyűjtések⁶, esetenként bizonytalan adatgyűjtések adatainak felhasználását jelentette. Már az 1780-as évek közepén Laplace megkísérelt a teljes populáció egy részére vonatkozó adatok alapján becslést adni a populáció egészére, azonban módszere nem terjedt el és komoly hiányosságokkal küzdött (Brewer és Gregoire [2009]). A XIX. század végén Anders Nicolaj Kiaer a Norvég Központi Statisztikai Hivatal igazgatója több évtizeden keresztül végzett olyan adatfelvételeket, amelyek az akkoriban szokásos teljeskörű adatfelvétel helyett csak a populáció egy részének megkeresésével történtek. Módszerére a részleges vizsgálat (partial investigation) kifejezés mellett először alkalmazta a reprezentatív módszer kifejezést. Az alkalmazott kiválasztási módszer azonban matematikai szempontból nem volt kellően megalapozott (Brewer és Gregoire [2009]). Az első, a mai módszertani kritériumoknak is megfelelő mintavétel alapú survey kutatást 1912-ben Bowley végezte 5 angliai város munkásságának helyzetét vizsgálva. Ezt a kutatását 1909-ben egy kísérlet előzte meg, melynek során 3878 cég adataiból 40 darab 400 elemű véletlen mintát kiválasztva demonstrálta, hogy a kapott átlagok normális eloszlást követnek, amelynek a várható értéke éppen a teljes sokaságban megfigyelhető átlag (Aldrich [2008]). A mintavétel módszere gyors fejlődésnek indult a következő évtizedekben.

⁵ A fejezet a szerző Statisztikai Szemlében megjelent cikke alapján készült (Simon [2012])

⁶ A teljeskörű adatgyűjtések története sokkal korábbra nyúlik vissza: ilyen jellegű adatgyűjtéseket a nagy ókori birodalmak is több kevesebb rendszerességgel végeztek, a Római birodalomban például öt éves rendszerességgel 77 népszámlálást hajtottak végre. Ez a fajta rendszeres adatgyűjtés Európában nem maradt fenn, de különféle összeírások a középkorban és a korajukban is végeztek, amelyek közül dokumentáltsága, szervezettsége és teljessége miatt kiemelkedik az 1085 és 1086 között végrehajtott angliai összeírás, amely alapján az úgynevezett Doomsday Book íródott.

Ebben egyaránt szerepet játszottak a létrejövő közvélemény kutató cégek (Gallup és a Roper intézet) és a statisztikai hivatalok, elsősorban az Egyesült Államok Census Bureau-ja olyan munkatársak részvételével, mint Morris H. Hansen és William N. Hurwitz, akikről egy ma általánosan alkalmazott becslési eljárást is elneveztek (Brewer és Gregoire [2009]). A mintavételi eljárások sorában kiterőnek bizonyult (legalább is a tudományos alkalmazás szempontjából) a Gallup által kidolgozott kvótás mintavételi eljárás. Az eljárás azon alapult, hogy az akkorra már a rendszeres censusok révén megbízhatóan rendelkezésre álló populációs adatok alapján előre meghatározták, hogy a mintában különböző tulajdonságokkal rendelkező személyeknek milyen számba kell bekerülniük (kvóta), úgy, hogy a minta eloszlása megfeleljen a populáció eloszlásának. A mintavételi eljárás ezt követően a kialakított kvóták feltöltésével zajlott. A módszer problémája, hogy míg a kvótát meghatározó ismérvek alapján a minta eloszlása biztosan megfelel a populációs eloszlásoknak, addig a mért egyéb ismérvekről ez nem feltétlenül mondható el. Utóbbi probléma miatt a módszer ugyan használatos maradt a piackutatás területén, de szociológiai kutatásokban a véletlen mintavételen alapuló eljárások terjedtek el, amely alapjaiban Bowley-nál jelent meg, de végleges elfogadottságát (az úgynevezett randomizációs ortodoxiát) Neymannak köszönheti (Brewer és Gregoire [2009]). A módszer paradigmászerű elterjedésére jellemző, hogy a vezető angolszász szociológiai folyóiratokban megjelenő cikkek közül egyre nagyobb számban voltak a survey-n alapuló kutatások: arányuk 1949/50-ben 24%, 1964/65-re már 55% (az arány 1979/80-ban is hasonló volt 56%) (Heer és mtsai. [1999]). A 70-es években azonban a módszer elérte határait, a társadalom bizalmatlanabbá vált, jelentős problémává vált a válaszmegtagadás. Az 80-as és 90-es években az új technikai eszközök megjelenése lökést adott a survey alapú kutatásnak és módosította annak mintavételi technikáit is: megjelent és elterjedt a telefonos interjú, majd a CATI, a számítógéppel segített telefonos interjú (Heer és mtsai. [1999]). A 90-es években jellemzően a minőségirányítás kérdésköre került előtérbe, miközben a válaszmegtagadás mértéke továbbra is komoly problémát jelentett (Heer és mtsai. [1999]). A 1990-es évek közepe óta elsősorban a piackutatás területén egyre nagyobb szerepet játszik az online survey, ami mintavételi szempontból teljesen új kihívásokat jelent, mivel ebben az esetben nem beszélhetünk ismert mintavételi keretről (az internethasználókról nem létezik lista, nincs olyan ismert eljárás, amellyel ismert valószínűséggel lehetne választani közülük). Erre a problémára több megoldás is létezik, ezek közül azonban kevés az, amelyik a megbízhatóság és érvényesség kritériumainak is eleget képes tenni. Ilyen technika a véletlen mintán alapuló kiválasztás, majd az internettel nem rendelkező mintatagok számára internet biztosítása (elsősorban panel minta esetén), illetve a véletlen tárcsázáson alapuló telefonos mintából az

internettel rendelkezők szűrése. A leggyakrabban alkalmazott úgynevezett panel alapú online kutatás alapján csak akkor lehet érvényes következtéseket levonni és megállapítani a megbízhatóság mértékét, ha a paneltagság véletlen kiválasztáson alapul⁷. Úgy tűnik, hogy bár a válaszadók elérése ilyen módon alacsonyabb költségű, a válaszadási arány nagyon alacsony, ami további torzításokhoz vezethet (Velu – Naidu [2009]). A 2008-as évek válságának hatására a hagyományos survey kutatások iránti piaci megrendelés jelentősen csökkent, miközben a tudományos kutatások területén a hagyományos survey kutatás továbbra is megkerülhetetlennek tűnik (Smith [2013]).

A mintavétel alapú survey módszertan fent vázolt rövid története a főáramra koncentrál és csak kis mértékben (az online survey esetén) érinti azt a kérdéskört, hogy mi történik akkor, ha az általánosan elfogadott véletlen mintavétel szükséges előfeltételei nem állnak rendelkezésre. A véletlen mintavétel megkerülhetetlen előfeltétele, hogy a kutató rendelkezzen egy teljes listával a vizsgálni kívánt célcsoportról, amely tartalmazza a célcsoporttagok elérhetőségét is. Ez a feltétel azonban gyakran nem teljesül és így a véletlen mintavétel elve nem alkalmazható. Ezeket a célcsoportokat (mintavételi értelemben populációkat) összefoglaló néven nehezen elérhető (*hard to reach*) csoportoknak nevezik.

Az angol nyelvű szakirodalomban a nehezen elérhető populáció szinonimájaként szerepel a ritka és rejtett csoportok (*rare and hidden, rare and elusive population*) kifejezés. A definíció szempontjából talán Spreen [1992] megközelítése a legvilágosabb, aki két dimenzió mentén különíti el a nehezen elérhető csoportokat: a kérdéses csoport teljes sokasághoz viszonyított aránya és az adott csoport rejtőzködő, illetve nem rejtőzködő jellege mentén, amelyet az alábbi táblázattal szemléltet (1. táblázat).

1. táblázat A nehezen elérhető társadalmi csoportok definiálása

	Könnyen elérhető (nem rejtőzködő)	Nehezen elérhető (rejtőzködő)
Megfelelő arány a populációban	Megfelelő részpopuláció	Nagy, rejtett részpopuláció
Alacsony arány a populációban	Ritka részpopuláció	Ritka, rejtett részpopuláció

Forrás: Spreen [1992] nyomán a szerző módosításaival

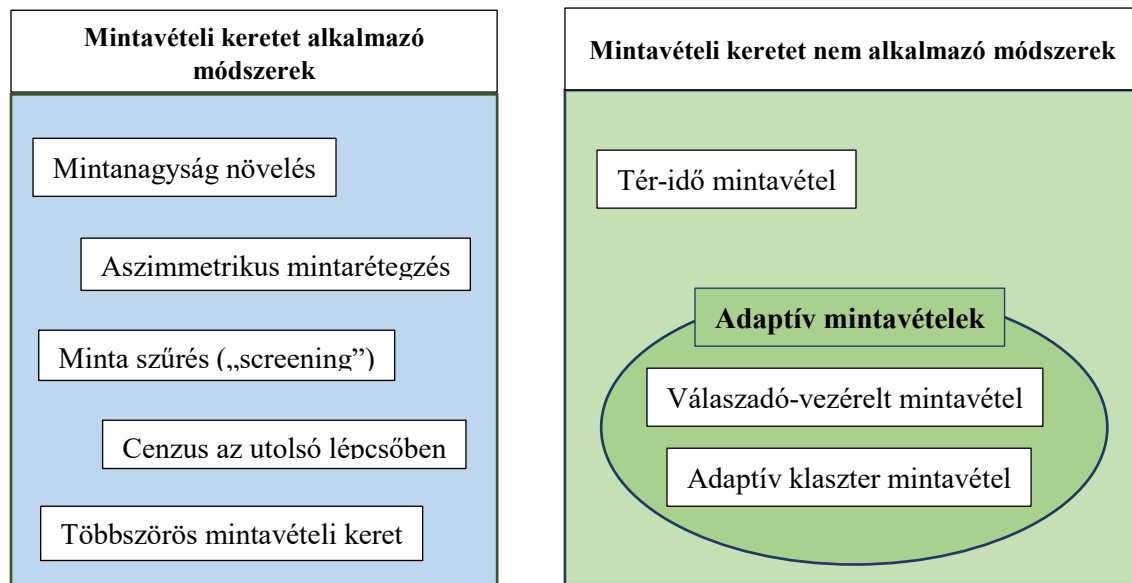
Rejtettnek tekinthetjük azokat a csoportokat, amelyek tagjai az adott csoporthoz tartozásukat igyekeznek elrejtteni, vagy az nem ismert. Az alacsony arány meghatározása problematikusabb, azt nem elméleti, sokkal inkább praktikus szempontok határozzák meg: alapvetően azokat a

⁷ Bolla Zsófia [2011] kutatása alapján a megkeresett magyarországi online kutatással foglalkozó 5 cég közül mindössze egy panel rekrutációs eljárása tekinthető ebben az értelemben megalapozottnak.

csoportokat tekinthetjük ritkának, amelynek tagjai esetén a (rétegzett, több lépcsős) véletlen mintavétel nem finanszírozható⁸.

A nehezen elérhető populációkból történő mintavételre több eljárás létezik, amelyek egy lehetséges csoportosítását a **Hiba! A hivatkozási forrás nem található.** mutatja be (Balogh-Simon [2011], illetve Kapitány [2010]) alapján, de saját csoportosításomban).

2. ábra Mintavételi módszerek csoportosítása nehezen elérhető populációk vizsgálata esetén



Forrás: saját ábra

Az egyes módszerek egymástól jelentősen eltérő megközelítést alkalmaznak, eltérő feltételezések mellett.

A módszerek egy része valamilyen mintavételi keretből indul ki. A *mintanagyság növelése* akkor jöhet szóba, ha a célcsoport a teljes populáció⁹ megfelelően nagy arányát teszi ki. Így a szokásosnál nagyobb, de még költséghatékonyan kivitelezhető mintavétel mellett várhatóan elég magas arányban kerülnek a célcsoport tagjai a mintába. Ezt a módszert módosíthatjuk, ha van olyan rétegzési szempont, amely mentén egyes rétegekben a célcsoport a teljes populáción belül jelentősen nagyobb arányban képviselteti magát (sűrű réteg). Ebben az esetben a

⁸ Meg kell jegyeznem ugyanakkor, hogy például mintavételi szempontból nagy rejtett populációnak tekinthetjük az internethasználókat és körükben a mintavétel hasonló problémákat vethet fel, amennyiben az internetet, mint eszközt kívánjuk használni.

⁹ Teljes populáció helyett alkalmazható olyan populáció is, amellyel kapcsolatban rendelkezünk mintavételi kerettel és teljes egészében magában foglalja a vizsgálni kívánt populációt (például, ha a vizsgálandó populáció a roma fiatalok, akkor használhatjuk teljes populációként a fiatalokat, mivel korcsoport alapján létezik mintavételi keret).

célcsoport szempontjából sűrűbb rétegeket felül, más rétegeket alulreprezentálhatunk (*aszimmetrikus mintarétegzés*), ezzel ugyanakkor kevésbé hatékony lesz a megfigyelésünk (kevesbé megbízhatóan, nagyobb mintavételi hibával mérünk) a célcsoport azon részében, amelyik a ritka rétegekbe tartozik (Kapitány [2010]). Ez utóbbi eljárást a piac- és közvéleménykutatási gyakorlatban „boost” mintának is nevezik (Pintér és Kátay [2010]). A hagyományos véletlen mintavételi eljárás egy másik módosításának tekinthető, amikor nagy teljes populációra vonatkozó, hagyományos véletlen mintán csak azt vizsgáljuk, hogy a célcsoporthoz tartozik-e a mintatag és a kiválasztottak képezik a tulajdonképpeni kutatás mintáját (*mintaszűrés*). Ebben az esetben lehetőség van a szűrést költséghatékonyabb eljárással (például telefonon) végezni. Fontos azonban ebben az esetben figyelembe venni, hogy az így kapott minta megbízhatósága (becslésének szórása) az azonos méretű hagyományos mintánál nagyobbak (lásd Balogh-Simon[2011]). Végül lehetséges a nem létező mintavételi keret pótlása olyan módon, hogy többlépcsős véletlen mintavétel során a megfelelő szintű mintavételi egységben teljeskörűen összeírják a célcsoport tagjait, majd közülük választják véletlen mintavétellel a mintatagokat (*cenzus az utolsó lépcsőben*) (pl.: Kertesi-Kézdi [1998]). A módszer mind etikai (teljes összeírás sértheti a célcsoport érdekeit), mind megvalósítási (ki rendelkezik a teljes összeíráshoz megfelelő ismeretekkel) szempontból problémákat vethet föl és meglehetősen költséges. Problémát okozhat az is, ha több, de nem teljes mintavételi keret áll rendelkezésre, vagyis léteznek listák a célcsoportról, de azok egyike sem fedi le a teljes populációt. Ebben az esetben alkalmazható a *többszörös mintavételi keret* módszere. A módszer lényege, hogy figyelembe véve a keretek közötti átfedést és a keretek eltérő megkeresési költségét optimális minta állítható fel, amelyben az alacsony költségű mintavételi keretek alapján állítjuk elő a minta jelentős részét¹⁰ (Lohr [2009]). Ez utóbbi módszer speciális esete, amikor az eltérő mintavételi keretekhez eltérő adatfelvételi módok (pl.: online, telefonos vagy személyes) társulnak, így csökkentve az eltérő adatfelvételi módszerek hibáit. Ezt az eljárást a piac- és közvéleménykutatási gyakorlatban hibrid vagy multimodális adatfelvételnek is nevezik (Pintér és Kátay, [2010]).

A további módszerek azok, amelyek mintavételi keret nélkül alkalmazhatók. A *tér-idő mintavétel* olyan esetekben alkalmazható, amikor a célcsoportról nem áll rendelkezésre

¹⁰ Lohr példája: ha az USA-ban élő statisztikusokból kívánunk mintát venni, akkor felhasználható az Amerikai Statisztikai Szövetség (ASA) taglistája, a Matematikai Statisztikusok Intézetének (IMS) taglistája, azonban a kettő együtt sem fedi le a teljes célcsoportot, ezért szükség van harmadik mintavételi keretként az Egyesült Államok teljes népességére. Míg az első kettő alapján alacsony költséggel végezhető el a mintavétel, az utóbbi alapján igen magas költséggel (alacsony találati valószínűség), azonban az utóbbi keret csak a célcsoport első két keretben nem szereplő részére kell mintát venni, ami csökkenti a költségeket.

mintavételi keret, de a populáció tagjai rendszeresen megjelennek térben jól körülhatárolható helyeken (ilyen például az idénymunkások populációja, a nyilvános helyen sakkozók populációja, vagy a regisztrált munkanélküliek). A módszer lényege, hogy a helyek és a megjelenés ciklusainak előzetes felmérését követően a tér-idő egységekből véletlen mintát véve a kiválasztott egységekben meghatározott frekvenciával kerülnek megkérdezésre a populáció tagjai (például minden második ott megjelenő). Fontos kezelni azt a problémát, hogy a populáció egyes tagjai több mintavételi helyszínen is megjelenhetnek eltérő gyakorisággal így a mintába kerülés valószínűsége eltérő lehet. A módszer hiányossága, hogy a célcsoport minden tagja szinte soha nem érhető el, illetve a bekerülés valószínűsége csak hozzávetőlegesen becsülhető meg (az egyes helyeken történő megjelenés gyakoriságára történő rákérdezéssel). A módszer alkalmazásáról bővebben ír Balogh-Simon [2011], illetve konkrét példával Muhib és mtsai. [2001]. Végül, de nem utolsó sorban a csoporton belül külön alcsoportot alkotnak az adaptív mintavételek, amelyek a populáció tagjai közötti ismeretségi kapcsolatokat használják fel. Ide tartozik vizsgálatom tárgya a válaszadó-vezérelt mintavétel. E mellett ennek a csoportnak a tagja az úgynevezett *adaptív klaszter mintavétel*. A módszer lényege, hogy több lépcsős véletlen mintavételt alkalmazva az utolsó előtti mintavételi lépcsőben kiválasztott mintavételi egységben, amelyben az utolsó mintavételi egységek között a célcsoport tagja megtalálható nagyobb arányú mintavételt végeznek. A lebonyolítás lépcsői tehát: egy hagyományos többlépcsős mintavétel alapján adatfelvétel lebonyolítása, majd azokban az utolsó előtti mintavételi egységekben, ahol a célcsoport tagjai felbukkantak újabb, nagyobb kiválasztási arányú mintavétel és adatfelvétel. A módszer problémája, hogy előre nem mérhető fel, hogy mekkora arányú mintát kell venni a mintavétel második fázisában. Figyelembe kell venni továbbá, hogy az így kiválasztott minta alapján készülő becslések szórása nagyobb, mint egy hasonló méretű többlépcsős mintavételé.

II.2. A válaszadó-vezérelt mintavétel előzményei és módszertani alapjai

A válaszadó-vezérelt mintavétel előzményének kell tekintenünk az úgynevezett hólabda mintavételt. Ahogy Biernacki és Waldorf [1981] fogalmaz, ezt a módszert a kvalitatív kutatások területén elterjedten alkalmazták elsősorban szenzitív kérdések vizsgálatánál, ahol feltételezhető, hogy a célcsoport tagjai ismerik egymást. Szociológiaelméleti szempontból ugyanakkor – hivatkoznak a szerzők – már Coleman kardoskodott amellett, hogy az ilyen típusú (hálózati) mintavétel kifejezetten alkalmas szociológiai kutatások lebonyolítására, mivel lehetővé teszi a természetes interakciós egységekből történő mintavételt (és itt visszautalhatunk Simmelre, aki a szociológiai módszer szempontjából ezt szintén alapvetőnek tartotta).

Goodman [1961] javasolta először a hólabda mintavétel alkalmazását populációs adatok – kapcsolathálózati, strukturális információk megbecslése érdekében. Módszerének lényege a véletlen kiindulópont volt, célja pedig kifejezetten a strukturális változók mérése, becslése. Frank [1977] hívja fel a figyelmet arra, hogy a hólabda mintavétel alkalmas a strukturális paraméterek mellett más populációs paraméterek becslésére is. Szintén véletlen kiindulópontokat feltételezve a véletlen mintavételeknél is általánosan alkalmazott Horovitz-Thompson becslés alkalmazhatóságát bizonyítja. Módszerét Snijders [1992] tovább finomította és több problémáját kiemelte. Amellett érvelt, hogy a hólabda mintavétel alkalmasabb a hálózati, strukturális paraméterek becslésére, mint más populációs paraméterek becslésére. Felhívta a figyelmet emellett a módszer véletlen kiinduló mintára való érzékenységére is. Összefoglalva ki kell tehát emelni, hogy a hálózati mintavételt a válaszadó-vezérelt mintavétel megjelenéséig csak véletlen kiindulópontok mellett és főként a hálózati struktúrára vonatkozó becslésekre tartották alkalmazhatónak, illetve erre álltak rendelkezésre bizonyítások.

A válaszadó-vezérelt mintavétel első leírása 1997-ben jelent meg Douglas D Heckathorn [1997] cikkében. Matematikai szempontból a kiválasztás kívánatos folyamatát a szerző Markov folyamatként értelmezi. Ennek az a feltétele, hogy a kiválasztót (legyen B) kiválasztó (legyen A) és a kiválasztó (B) által kiválasztott (C) kategoriális tulajdonságai csak B-n keresztül függenek össze. Ez azt jelenti, hogy adott változó szerint A és C legyen feltételesen független egymástól B-re, mint feltételre nézve. Másképpen az ilyen folyamatokat nem emlékező folyamatoknak (memoryless process) is hívják. Egy másik fontos szempont az, hogy az adott hálózat ergodikusan, azaz minden pontjából elérhető-e az összes többi. Ha az utóbbi feltétel nem teljesül, akkor két külön mintavételi problémaként kell kezelni a két populációt. Amennyiben mindkét feltétel teljesül, Heckathorn állítása szerint a minta a mintavételi lépcsők (egyre újabb válaszolók bevonása) során egy egyensúlyi állapothoz közelít. Az egyensúlyi állapot becsülhető az alábbi egyenletrendszer megoldásával¹¹:

$\sum_{n=1}^N P_n = 1$ $\sum_{i=1}^N \frac{P_{ij}}{P_{i+}} P_i = P_j$	<p><i>Ahol:</i></p> <p>P_n <i>adott változó n-ik attribútumának arányára adott becslés a népességben</i></p> <p>p_{ij} <i>annak a valószínűsége a mintában, hogy adott változón i-ik kategóriába eső kiválasztó j-ik kategóriába eső kiválasztottat válasszon</i></p> <p>p_{i+} <i>annak a valószínűsége, hogy adott változón a mintában a kiválasztó éppen az i-ik kategóriába essen</i></p>
---	--

¹¹ $i=2$ esetén az egyenletrendszer egyértelmű megoldást ad, többértékű változó esetén azonban közelítő megoldásra van szükség, mivel az egyenletrendszer túldeterminálttá válik.

Akkor tekinthető elegendően nagyoknak a minta, amikor az így kiszámított egyensúlyi állapotot kellően megközelíti a mintabeli eloszlás (a cikk szerzője 1%-ot javasol), vagyis az egyenletrendszer megoldásából származó becslés kis mértékben tér el a minta tényleges eloszlásától.

Heckathorn azt állítja, hogy az egyensúlyt a minta gyorsan, azaz geometrikusan közelíti. Az egyensúly gyors kialakulását segíti értelemszerűen, ha minél inkább szórt a kiinduló minta. Lelassítja a konvergenciát, ha a vizsgált közösség kapcsolathálózata közelít a nem ergodikushoz¹².

Heckathorn talán legfontosabb állítása az, hogy az egyensúlyi minta reprezentatív a véletlen mintavétel esetében vett értelemben is, ha a vizsgált alapsokaság alcsoportjai között nincs különbség a csoporton belülrre, illetve kívülrre mutató kiválasztások arányában. Ez azt jelenti, hogy minden alcsoportban azonos kellene legyen a belül-választási tendencia (*inbreeding*)¹³. Amennyiben ez a feltétel sérül az egyensúlyi minta torzul az alapsokasághoz képest. Ezt az alábbi levezetés illusztrálja (két értékű változó esetén):

(1) $E_1 + E_2 = 1$	<i>Ahol</i>	
(2) $E_1 = S_{11} E_1 + S_{21} E_2$	E_i	Becslés a vizsgált változó <i>i</i> -ik érték populációs arányára
(3) $S_{11} = I_1 + (1 - I_1) P_1$	S_{ij}	<i>i</i> értékű mintatag <i>j</i> értékű tagválasztásának valószínűsége
(4) $S_{21} = (1 - I_2) P_2$	P_i	<i>i</i> érték populációs aránya
(5) $E_1 = \frac{P_1(1 - I_2)}{P_1 I_2 - P_1 I_1 + I_1 - 1}$	I_i	Belülválasztási arány (annak a többletvalószínűsége, hogy valaki magához hasonló értékkel rendelkező személyt választ)

Az (5) alapján belátható, hogy a vizsgált változóra vonatkozó becslés (E_i) csak akkor lehet torzítatlan (vagyis egyenlő a populációs aránnyal (P_i)), ha a belül-választás mindkét értékre nézve egyenlő ($I_1 = I_2$).

¹² A nem ergodikushoz közelítő kapcsolathálózat a vizsgált célcsoport kapcsolathálózatában a szereplők közötti hálózati értelemben vett utak száma kisebb és hossza nagyobb (ez a hálózat klaszterezettségéhez is köthető, amely IV.1 fejezetben bővebben kifejtésre kerül).

¹³ A belül-választási tendencia a hálózatelemzésben ismert homofília jelenségéhez kapcsolódik, amely adott jellemző alapján hasonló tulajdonságú szereplők közötti kapcsolat véletlenszerű kapcsolódáshoz képest mutatott magasabb valószínűségeként értelmezhető.

Ez a feltétel szigorúnak tűnik, de a szerző Simmelre hivatkozva azt állítja, hogy ha egy csoport megerősíti a határait, akkor a többi környező csoport is hasonlóképpen alakul át. Ez hat a belül-választási tendenciára, így, ha egyenlőség nem is, de pozitív korreláció fennáll e tekintetben.

A szerző cikkében megvizsgálja a módszer érzékenységét a belül-választási tendencia egyenlőtlenségeire is. A modellben a belül-választási tendencia terjedelmének és terjedelmközepének hatását tesztelte. Az alapsokaságot négy alcsoport alkotta, amelyek méretét véletlenszerűen határozta meg. Minden paraméter beállításnál 1000 kísérletet végzett egyenként új csoportméretekkel és 10 hullámon keresztül végzett mintavétellel. A keletkezett minták esetén meghatározta a csoportok mintabeli és alapsokaságbeli arányainak átlagos eltérését. Ezekből az eredményekből eloszlást hozott létre, ami alapján pont és intervallum becslést tett adott paraméterek mellett a várható átlagos eltérésre.

Eredményei azt mutatják, hogy a belül-választási tendencia terjedelmének változtatása (véletlen terjedelmközep mellett) viszonylag kevésbé befolyásolja a mintabeli arányokat. A legszélsőségesebb eltéréseket megengedő 1-es terjedelem esetén a 0,1-et alig meghaladó a várható átlagos eltérés (konfidencia intervallum közelítőleg 0,05-0,2). 0,5-es terjedelemig a konfidencia intervallum felső határa 0,1 alatt maradt.

A belül-választási tendencia terjedelmközepének hatása hasonló: 0,5-es rögzített terjedelem nagyság mellett 0,6-es terjedelmközepig az átlagos eltérés konfidencia intervallumának felső határa 0,1 alatti, a maximális 0,75-ös terjedelmközep mellett pedig 0,2 körüli (várhatóérték 0,1 körül, alsó határ 0,05 körül).

A leírt statisztikai becslésre vonatkozó modell ugyan bizonyítja, hogy a mintavételi módszer – adott feltételek mellett – torzítatlan becslésre alkalmas, az ott szereplő feltételek közül a belül-választás azonban biztosan ellenőrizhetetlen, így ezt a modellt már nem alkalmazzák, köszönhetően annak, hogy a későbbi kutatások olyan becslési modelleket fogalmaztak meg, amelyek a valós mintavétellel jobban összeegyeztethető feltételeken alapultak. A becslő modellen túl azonban a cikk módszertani ajánlásokat is tartalmaz, amelyek érvényessége nem változott. A válaszadó-vezérelt mintavétel szerzők által javasolt lépései a következők:

- kezdőpontok (a minta kiindulópontjait jelentő első válaszadók, *seeds*) kiválasztása (erre vonatkozó statisztikai feltételekről és azok vizsgálatáról a IV.1 és IV.2 alfejezetekben írok részletesen), a válaszadóknak jutalmat ígérnek az interjúért
- a kezdő adatfelvételek elkészítését követően a kezdő válaszadóknak meghatározott összegű jutalmat ígérnek (kupon formájában) 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 interjúért és további válaszadók bevonásáért, egészen a kellő mintaméret kialakulásáig (lásd korábban egyensúlyi eloszlás kialakulása)
- a bevonható további válaszadók száma válaszadónként korlátozott, annak érdekében, hogy ne legyenek fél-profi bevonók
- a célcsoporthoz tartozást (a cikk esetében intravénás droghasználók) a kutatók ellenőrzik, hogy elkerüljék a nem célcsoporthoz tartozók jutalom miatti válaszadását
- az interjúkat egy helyen készítették, ahova a válaszadóknak kellett elmennie.

Két szempontot érdemes kiemelni a fenti folyamatleírásból: egyrészt a válaszadókat elsődleges (válaszadásért) és másodlagos (további személyek bevonásáért) ösztönzőkkel motiválják, másrészt az interjúkat a válaszadók szempontjából semleges helyen készítik. Mindkét szempont a rejtőzködő csoportok válaszadási hajlandóságát növeli. A másodlagos ösztönzők alkalmazása csoport közvetített (*group mediated*) motivációt biztosít, amelyekkel kapcsolatban Heckathorn (korábbi kutatásai alapján) nagyobb hatékonyságot feltételez, mint az elsődleges ösztönzőkkel kapcsolatban. Hasonlóan fontos, hogy a semleges helyen készített interjú az anonimitás megőrzésével hozzájárult a válaszadási hajlandóság növeléséhez.

II.3. A válaszadó-vezérelt mintavétel matematikai statisztikai alapjai

A válaszadó-vezérelt mintavétel további becsléseinek statisztikai modellje azon alapul, hogy a fent leírt mintavételi folyamat egy gráfon történő véletlen bolyongásnak feleltethető meg, amely leírható Markov-lánc Monte Carlo mintavételként (MCMC mintavétel) (Goel és Salganik [2009]). Az MCMC módszer azt feltételezi, hogy egy V állapottéren értelmezhető egy Markov-lánc. A válaszadó-vezérelt mintavétel esetén az állapotter elemei célcsoport tagjait tartalmazza. Létezik ezen felül egy $K(v_i, v_j)$ kernel mátrix, amelynek elemei megadják a v_i állapotból a v_j állapotba történő átmenet valószínűségét, ami a válaszadó-vezérelt mintavétel esetén annak a valószínűségét jelenti, hogy v_i személy v_j személyt ajánlja. A K mátrix segítségével a V állapottéren definiálható a Markov-lánc. Amennyiben igaz az, hogy v_i -ből (bárhány lépés alatt) elérhető v_j tetszőleges i és j esetén, akkor létezik egyetlen π eloszlás, amely a Markov-lánc egyensúlyi eloszlása és amely eloszlás felé tart a Markov folyamat. Ha egy függvénnyel minden v_i -hez hozzárendelünk egy tetszőleges értéket (a személyek attribútumát), akkor az egyensúlyi állapot elérése esetén a π eloszlás ismeretében megadható a függvény (vagyis az attribútum) várható értékének torzítatlan becslése. Amennyiben igaz a gráfon történő ily módon leírt véletlen sétára az, hogy minden v_i -ből csak egy v_j felé haladunk és véletlenszerűen, azonos

valószínűség mellett kerül kiválasztásra v_j , akkor a π eloszlás éppen a fokszámmal lesz arányos. Mindezek alapján lehetségessé válik a torzítatlan becslés, amennyiben a minta eléri az egyensúlyi állapotot. A szerzők ugyanakkor kiemelik, hogy akkor és csak akkor tekinthető a válaszadó-vezérelt mintavétel MCMC folyamatnak, és akkor érvényes rá a fenti levezetés, ha minden válaszadó (a) egyetlen következő válaszadót választ ki saját ismeretségi hálójából, (b) azonos valószínűséggel, véletlenszerűen és (c) a mintavétel visszatevéses. Az egyensúlyi állapot ebben az esetben is számos lépést követően alakul csak ki.

Salganik és Heckathorn [2004] a válaszadó-vezérelt mintavételen alapuló becslés továbbfejlesztett változatát publikálta, melynek statisztikai modellje a későbbiek során elterjedt. A cikk elnyerte az Amerikai Szociológiai Társaság legjobb publikációnak járó díját a matematikai szociológia szekcióban¹⁴. A szerzők által javasolt modell a minta alapján egyes hálózati tulajdonságokra vonatkozó becslés felhasználásával ad becslést az egyes, lekérdezett változók populációs paraméterére. A becslés feltételei a következők:

- i. A válaszadók kölcsönös kapcsolatban állnak az általuk ajánlott potenciális válaszadókkal, akik a célcsoport tagjai (reciprok kapcsolatok).
- ii. A hálózat egy komponensből áll össze.
- iii. A mintavétel visszatevéssel történik¹⁵.
- iv. A válaszadó teljes pontossággal határozza meg a fokszámát¹⁶, a hálózatának a méretét (a becslések valójában a fokszámok egymáshoz viszonyított arányán alapulnak)
- v. A válaszadó véletlenszerűen választ az ismerősei közül az ajánlás alkalmával (egyetlen személyt az alább ismertetendő becslés esetén)
- vi. A vizsgált változó diszkrét.

A fenti 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 egy egyensúlyi állapothoz közelít. A közelítés mértékét úgy határozhatjuk meg, ha az egymást követő hullámok során megvizsgáljuk, hogy a célváltozó kiválasztott értékének aránya a mintában milyen mértékben

¹⁴ Ez az önmagában nem feltétlenül említésre méltó tény elemzésem szempontjából azért fontos, mert önmagában jelzi a válaszadó-vezérelt mintavétel, mint tudományos konstrukció elfogadottságát – természetesen a kérdés további vizsgálatot érdemel, amire a III.1 fejezetben a módszer alkalmazásának elterjedtségének vizsgálatánál még visszatérek.

¹⁵ Szimuláció alapján a visszatevés nélküli mintavétel nem befolyásolja jelentősen a becsléseket (Salganik-Heckathorn [2004])

¹⁶ Fokszám alatt a hálózat kutatásban a kapcsolatok számát értik.

változik. Akkor tekintjük az adott célváltozó szempontjából egyensúlyban lévőnek a mintát, ha a változás mértéke egy előre meghatározott értéknél kisebb¹⁷.

A fenti feltételek teljesülése esetén egy mért változóra az alábbi aránybecslés tehető (az alábbiakban dichotóm változóra mutatom be a becslési eljárást, amelyet a továbbiakban RDS I becslésnek nevezek):

$$N_A D_A C_{A,B} = N_B D_B C_{B,A} \quad (6)$$

$$P_A D_A C_{A,B} = P_B D_B C_{B,A} \quad (7)$$

$$P_A + P_B = 1 \quad (8)$$

$$\hat{P}_A = \frac{\hat{D}_B \hat{C}_{B,A}}{\hat{D}_A \hat{C}_{A,B} + \hat{D}_B \hat{C}_{B,A}} \quad (9)$$

$$\hat{D}_A = \frac{n_A}{\sum_{j=1}^{n_A} \frac{1}{d_j^A}}; \quad \hat{D}_B = \frac{n_B}{\sum_{j=1}^{n_B} \frac{1}{d_j^B}} \quad (10)$$

$$\hat{C}_{A,B} = \frac{f_{A,B}}{n_A}; \quad \hat{C}_{B,A} = \frac{f_{B,A}}{n_B} \quad (11)$$

Ahol:

$N_A; N_B$ a változó első, illetve második értéket választó személyek száma a célcsoportban

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

$D_A; D_B$ a változó első, illetve második értéket választó személyek átlagos fokszáma a célcsoportban (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 célcsoportban (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 1. illetve 2. értéket választók közül a j . válaszadó fokszáma

$f_{AB}; f_{BA}$ azon válaszadók száma, akik az 1., illetve 2. értéket választották és az általuk ajánlott személy a 2., illetve 1. 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 (6) és (7) egyenletek a kapcsolatok feltételezett kölcsönösségének következményei, a (8) egyenlet triviális. A (9) egyenlet a (7) (8) egyenletek következménye. A fokszám becslése (10) Hansen-Hurwitz becslésen alapul (az egyszerű levezetést lásd Salganik és Heckathorn [2004]). A (11) annak a feltételnek a következménye, hogy a válaszadó véletlenszerűen választ a kapcsolatai közül, amikor ajánl (Heckathorn erre vonatkozó korábbi kutatása megerősíti, hogy az ajánlott személyek választása valóban közel véletlenszerű). Meg kell jegyezni, hogy 3 vagy több értékű változók esetén a becslés túldeterminálttá válik, amire Heckathorn [2007] a lineáris legkisebb négyzetes közelítést vagy saját adatsimításos módszerét ajánlja. A cikk bizonyítása alapján belátható, hogy a becslés aszimptotikusan torzítatlan a feltételek teljesülése esetén diszkrét változókra.

¹⁷ A konvergencia kritérium és a becslések szórása között pontosan meg nem határozott pozitív irányú összefüggés van.

Folytonos változó esetén Heckathorn kétkomponensű súly kialakítását javasolja (lásd (12) egyenlet), ennek kiszámítása azonban lényegében azonos azzal, mintha a folytonos változó értékeit partícionálnánk és így végeznénk el a fent ismertetett becslést, majd az így kapott és a mintában szereplő eloszlás különbségét használnánk az adott változó súlyozására.

$$DW_j = DC_j RC_j = K \frac{1}{d_j} \frac{E_x}{p_x} \quad (12)$$

Ahol:

DW_j Kétkomponensű súly a j . mintatag esetén adott változóra, ha j . mintatagra a változó értéke X

DC_j Fokszám-függő 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 partícionálását javasolja korábbi RDS adatok elemzésével, ahol az értékenkénti átlagos esetszám 12 ± 4 . Alacsonyabb átlagos esetszám a becslések instabilitásához, magasabb átlagos esetszám túl pontatlan súlyozáshoz vezet.

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

Voltz és Heckathorn [2008] egy, a fentiekől eltérő becslés alkalmazására is javaslatot tesz, amelynek a szintén feltétele, hogy a minta egyensúlyi állapotba kerüljön, azaz elérje azt az állapotot, amikor a két egymás követő mintavételi hullám (egymás követő lépések a mintavételben) már nem eredményeznek számottevő eltérést a becslésben (a számottevő eltérés egy előre meghatározott érték). Amennyiben a mintavétel a korábban leírtak szerint Markov folyamatnak tekinthető, a szerzők szerint az alábbi becslés tehető (amelyet a továbbiakban RDS II becslésnek nevezek):

$$\hat{P}_A = \frac{n_A}{n} \frac{\hat{D}}{\hat{D}_A} \quad (13) \quad \begin{array}{l} \text{Ahol} \\ \hat{D} \end{array} \quad \begin{array}{l} A \text{ célcsoport átlagos fokszámának a becslése a} \\ \text{mintából (Hansen-Hurwitz becslés)} \end{array}$$

$$\hat{D} = \frac{n}{\sum_{i=1}^n \frac{1}{d_i}} \quad (14) \quad \hat{D}_A \quad \begin{array}{l} A \text{ vizsgált változón } A \text{ értékkel jellemezhető} \\ \text{célcsoporttagok átlagos fokszámának a becslése a} \\ \text{mintából (Hansen-Hurwitz becslés)} \end{array}$$

$$\hat{D}_A = \frac{n_A}{\sum_{i=1}^{n_A} \frac{1}{d_{Ai}}} \quad (15)$$

Az adott változók eloszlására vonatkozó becsléseken túl fontos kérdés a becslések szórása (megbízhatósága) is – 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 a becslések szórásának meghatározására. Salganik korábbi adatok másodelemzése és modellezés alapján arra a következtetésre jutott, hogy a design hatás az RDS I esetében 2 körüli (azaz az RDS I becslés varianciája az egyszerű véletlen mintavételhez képest kétszeres). Ez 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 szórásbecslésére, amelyet az alábbiakban levezetés nélkül adok közre:

$$\hat{V}_{P_A} = \hat{V}_1 + \frac{P_A^2}{n} \left((1-n) + \frac{2}{n_A} \sum_{i=1}^n \sum_{j=1}^{i-1} (\hat{C}^{i-j})_{A,A} \right) \quad (16) \quad \begin{array}{l} \text{Ahol} \\ \hat{V}_{P_A} \end{array} \quad \begin{array}{l} A \text{ változó } A \text{ értékét választó} \\ \text{személyek arányára vonatkozó} \\ \text{becslés varianciájának becslése} \end{array}$$

$$\hat{V}_1 = \frac{\hat{V}(Z_i)}{n} = \frac{1}{n(n-1)} \sum_{i=1}^n (Z_i - P_A)^2 \quad (17) \quad (\hat{C}^{i-j})_{A,A} \quad \begin{array}{l} \text{Becslés annak a valószínűségére,} \\ \text{hogy az } A \text{ értéket választó} \\ \text{személyek } i\text{-j lánchosszúság után } A \\ \text{értéket választó személyt ajánlanak} \\ \text{(a becslés az átmenet valószínűség} \\ \text{mátrix megfelelő hatványra} \\ \text{emelése után kapott} \\ \text{eredménymátrix első sor, első} \\ \text{oszlopában található valószínűség)} \end{array}$$

$$Z_i = \frac{\hat{D}_A}{d_i} I_A(i) \quad (18) \quad I_A(i) \quad \begin{array}{l} \text{Indikátorváltozó, melynek értéke } 1, \\ \text{ha a minta } i\text{-ik tagja a változó } A\text{-} \\ \text{értékét választotta és } 0, \text{ ha a minta} \\ i\text{-ik tagja nem } A\text{-t választotta} \end{array}$$

A leírt szórás becslés a Hansen-Hurwitz becslés analógiájának tekinthető, azzal a – jelentős – különbséggel, hogy figyelembe veszi azt, hogy a mintatagok nem függetlenek és felhasználja a

mintavétel fent leírt Markov megközelítését. A becslés ugyanakkor abban az esetben teljesül, ha az adott változóra nézve a mintavétel, mint Markov folyamat memória mentes, azaz az adott változó értékei bármely megkérdéztet esetén csak az ajánlón keresztül függenek össze a változó lánc korábbi tagjainál mért értékeivel. Ez a feltétel olyan hálózatban, ahol közel zárt csoportok jelennek meg, amelyekben a vizsgált változó értéke homogénebb lehet, mint a hálózat többi részén nem feltétlenül teljesül (erről részletesebben lásd IV.1 fejezet vonatkozó részét).

Megjegyzendő, hogy a közölt becslés nem torzítatlan, mivel $C_{A,A}$ nem ismert, csupán becsülhető hasonlóan D_A -hoz.

Az itt leírt becslési eljárások közül a folytonos és diszkrét változók RDS I eloszlás becslése, a becslések bootstrap szórás becslése elvégezhető Volz, Degani és Heckathorn által közreadott szoftver segítségével [RDSAT, 2007]. Az RDS I és RDS II eloszlásbecslés elérhető R package-ként is (package 'RDS', W. Whipple Neely). 2012 óta STATA parancs is létezik RDS becslések kiszámítására (Schonlau és Liebau [2012]).

Lunagómez és Airoidi [2012] tanulmányukban azt a problémát vetik föl, hogy a használt becslések nem veszik figyelembe az eredmények függését a hálózat felépítésétől. Tanulmányukban egy Bayes-i becslést javasolnak, amely apriori feltételként rögzít egy hálózati modellt, és ezt figyelembe véve adja meg a vizsgált paraméter becslését. A szerzők tanulmányukban szimulációs eredményeket és valódi adatokon végzett számításokat is közölnek. A tanulmány egy érdekes felvetést ad, amely nem nélkülözi a valós alapot¹⁸, azonban a szimulációs példa meglehetősen távol áll a társas hálózatok tényleges struktúrájától: a kutatók apriori modellként az Erdős-Rényi random gráfot¹⁹ használták, míg a szimulált hálózat kis-világ volt – melyek a legtöbb esetben nem képezik le az emberi kapcsolathálózatok foksám megoszlását (Barabási és munkatársai, 1999), így az eredmények nem feltétlenül megbízhatóak. Mindezekon túl a szimuláció eredményeként vizsgált torzítás legtöbb esetben nagyobb volt, mint az RDS II becslés esetén. A szerzők becslési módszere csak a feltételként alkalmazott véletlen gráftól kevéssé eltérő (magas újrakapcsolódási valószínűségű kis-világ modell) gráf esetén adott kis mértékben jobb eredményt az RDS II becsléshez viszonyítva. Összességében a tanulmány problémafelvetése fontos, azonban nem ad meggyőző bizonyítékot

¹⁸ saját szimulációs eredményeim is megerősítették a hálózati struktúra hatását a becslésekre, bár ez meglehetősen gyengének bizonyult (lásd IV.3 alfejezet)

¹⁹ Az egyik első, legegyszerűbb hálózati modell, amely a hálózati kapcsolatokat egy egyenletes eloszlású p valószínűségű valószínűségi változóval modellezi (Erdős P - Rényi A. [1959]).

a javasolt becslési módszer használatához, életszerűbb hálózati modell alkalmazása esetén viszont a számítások jóval bonyolultabbak lehetnek.

Gile [2011] kis populációk melletti relatív nagy minta problémájára dolgozott ki becslési eljárást. Tekintettel arra, hogy a válaszadó-vezérelt mintavétel kisméretű célcsoportokban történő alkalmazása meglehetősen elterjedt a kérdést érdemes érinteni²⁰. A szerző az RDS II becslést korrigálja és ismert populáció méret mellett iterációs eljárással becsüli meg az egyes mintatagok mintába kerülési valószínűségét. A módszer alapja két könnyen belátható összefüggésen alapul. Egyrészt ismert populációs foksám eloszlás és populációméret alapján, amennyiben feltételezzük, hogy a kiindulópontot követően a mintába kerülés valószínűsége arányos a foksámmal, tetszőleges minta esetén (egyensúlyi állapot elérése nélkül is) megbecsülhető a mintatagok mintába kerülési valószínűsége. Másrészt ismert bekerülési valószínűségek mellett a minta alapján megbecsülhető a populációs foksám eloszlás. A válaszadó-vezérelt mintavétel esetén azonban sem a populációs foksám eloszlás, sem a bekerülési valószínűségek nem ismertek. A szerző bizonyítás nélkül, de szimulációval alátámasztva azt javasolja, hogy a bekerülési valószínűségeket a foksámok mintabeli eloszlása alapján becsüljük, majd ezt követően tegyünk becslést a populációs foksám eloszlásra a becsült bekerülési valószínűségek alapján. A becsült foksám eloszlással szimuláljunk M foksám arányos mintavételt a mintával azonos mérettel és a bekerülési valószínűségeket számítsuk ki ismételten a szimulációs eredmények alapján. Az utóbbi két lépés r számú iterációjával érhető el a becsült bekerülési valószínűség minden mintatagra (a szerzők a számítási igényekre hivatkozva 2000 mintavétel szimulációt és 3 iterációt javasolnak). A szerző szimulációval veti össze saját becslését az RDS II becsléssel. A szimulációk során saját becslése a populációs méret becslésének hibájára, a kiindulópontok torzítására, a vizsgált változó és a foksám közötti interakcióra és a homofíliára nézve is kisebb torzítást és kisebb becslési hibát biztosított, mint az RDS II becslés (a populáció 50%-át meghaladó mintaméret esetén). A becslés értékelésekor meg kell jegyezni, hogy egyrészt a becslést lényegében semmilyen levezetés nem támasztja alá, másrészt a szerző nem vizsgálja azokat a szituációkat, amikor a mintaméret a populációs méret kisebb arányát teszi ki és a homofília jelentős (ez azért fontos, mert a téves populáció méret becslés összességében akár ilyen eredményre is vezethet). Ez a becslési eljárás ugyanis kizárólag bekerülési valószínűségek és a foksám viszonyával számol és nem veszi figyelembe sem a hálózat klaszterezettségét sem a homofíliát. A jó eredmények

²⁰ A IV.1 fejezetben a minta paramétereinek hatását vizsgáló kutatások segítségével bemutatom, hogy az itt ismertetett és legelterjedtebben használt becslési eljárással a populáció méretéhez közelítő minta komoly torzítást okoz.

feltehetően annak köszönhető, hogy a homofília hatása vagy a klaszterezettség a nagy relatív mintaméret mellett már nem okoz komoly torzítást a minta összetételében. Összességében a Gile által ajánlott becslési eljárás további kutatásra érdemes, de használata jelenleg csak a biztosan kis célcsoportok esetén javasolható.

Lu és munkatársai [2012a] irányított hálózatok esetére ad javaslatot az irányítottságból fakadó torzítás és becslés szórás növekedés kezelésére. A szerzők a két leginkább használt becslés (az általam RDS I és RDS II becslésként hivatkozott becslések) adaptációját mutatja be. Az RDS I becslés módosítására a szerzők az alábbi javaslatot teszik:

$$N_A \bar{D}_A^{ki} C_{AA} + N_B \bar{D}_B^{ki} C_{BA} = N_A \bar{D}_A^{be} \quad (19)$$

$$N_A \bar{D}_A^{ki} C_{AB} + N_B \bar{D}_B^{ki} C_{BB} = N_B \bar{D}_B^{be} \quad (20)$$

$$m = \frac{\bar{D}_A^{be}}{\bar{D}_B^{be}}; w = \frac{\bar{D}_A^{ki}}{\bar{D}_B^{ki}} \quad (21)$$

$$\frac{N_A}{N_B} = \frac{wC_{AA} - mC_{BB}}{2mwC_{AB}} + \sqrt{\frac{C_{BA}}{mwC_{AB}} + \left(\frac{mC_{BB} - wC_{AA}}{2mwC_{AB}} \right)^2} \quad (22)$$

Ahol

N_X Az X értékkel jellemezhető célcsoporttagok száma (arányuk a becslendő paraméter)

\bar{D}_X^{ki} Az X értékkel jellemezhető célcsoporttagok átlagos ki- és befokszáma

C_{XY} A X értékkel jellemezhető csoporttagoktól az Y értékkel jellemezhető csoporttagok felé mutató kapcsolatok aránya az X értékkel jellemezhető csoporttagok ki-fokszámához viszonyítva

m Be-fokszám hányados

w Ki-fokszám hányados

A módosított RDS I becslés logikája – az eredeti becsléshez hasonlóan, azonban irányított hálózatot feltételezve²¹ – kétértékű változóra (értékei A és B) a következőn alapul: a populáció A értékkel jellemezhető alcsoportját nevezzük A csoportnak, a B értékkel jellemezhető alcsoportját nevezzük B csoportnak, ekkor az A alcsoport befokszáma²² (*indegree*, a (19) egyenlet jobb oldala) megegyezik az A csoporton belülré mutató kifokszám²³ (*outdegree*) és a B csoportból az A csoport felé mutató kifokszáma összegével (a (19) egyenlet jobb oldala, melynek tagjai egyenként megadhatók a csoportonkénti átlagos fokszám, illetve a B csoportból A csoportba mutató kapcsolatok aránya, valamint a csoportok mérete alapján). Az egyenletet mindkét értékre felírva (19)-(20) megoldható egyenletrendszer keletkezik, amely az egyes értékekkel jellemezhető csoportok be- illetve ki-fokszám hányadosai (m és w) (21) valamint a

²¹ Az RDS I becslés kölcsönös kapcsolatokat feltételez. Lásd a fejezet elején az RDS I becslés leírását.

²² Befokszám (*indegree*): egy hálózat adott szereplője felé irányuló kapcsolatok száma.

²³ Kifokszám (*outdegree*): egy hálózat adott szereplőjétől más szereplők felé irányuló kapcsolatok száma.

csoporthoz közti kapcsolatok arányai ismeretében (azokra adott becslések segítségével) megbecsülhetővé teszik a vizsgált változó értékeinek arányát (22).

Az RDS II becslés (amely a Hansen-Hurwitz becslés adaptált változata) irányított hálózatokra átalakított változata a (23-25) képletekkel írható fel.

$$\hat{P}_A = \frac{n_A}{n} \frac{\hat{D}^{be}}{\hat{D}_A^{be}} \quad (23)$$

Ahol

\hat{D}^{be} *A célcsoport átlagos be-fokszámának a becslése a mintából (Hansen-Hurwitz becslés)*

$$\hat{D}^{be} = \frac{n}{\sum_{i=1}^n \frac{1}{d_i^{be}}} \quad (24)$$

\hat{D}_A^{be} *A vizsgált változón A értékkel jellemezhető célcsoporttagok átlagos be-fokszámának a becslése a mintából (Hansen-Hurwitz becslés)*

$$\hat{D}_A^{be} = \frac{n_A}{\sum_{i=1}^{n_A} \frac{1}{d_{Ai}^{be}}} \quad (25)$$

Az RDS II becslés korrekciója lényegében arra az eltérésre reflektál, hogy irányított hálózaton történő véletlen bolyongás esetén nem a fokszámmal, hanem a be-fokszámmal arányos a bekerülés valószínűsége.

A két becslést a szerzők valós hálózaton történő szimulációval is vizsgálták és mindkettő jobb eredményeket adott (kisebb torzítás és kisebb becslés szórás), mint az eredeti becslések. Ugyanakkor a szerzők is elismerik – és ez lehet a legkomolyabb kritika a becsléssel kapcsolatban –, hogy a be-fokszámra vonatkozóan RDS adatfelvételben nem kaphatunk érvényes információt, márpedig ez mindkét becslés esetén esszenciális volna. Meg kell továbbá említeni, hogy nem egyértelmű mennyire tekinthető valószínűnek, hogy az RDS mintavétel alapját jelentő hálózat milyen mértékben rendelkezik aszimmetrikus kapcsolatokkal, azaz mennyire szükséges az irányított hálózat feltételezése. Egy irányított hálózat két típusú hibához vezethet a mintavétel során: egyirányú kapcsolattal ajánl az ajánló következő szereplőt a kutatásba – ebben az esetben megnő a valószínűsége a válaszmegtagadásnak, illetve a saját kapcsolatszámának becslése esetén az egyirányú kapcsolatok miatt felfelé torzít az átlagos kapcsolatszámra adott becslés (ami azonban megfelelő kérdezéstechnikával csökkenthető²⁴).

Lu egy másik cikkében (Lu [2012]) ugyanakkor egy gyakorlatban jobban használható javaslatot is tesz a becslés javítására. Felveti, hogy a változók egy részére vonatkozóan érdemes úgynevezett egonetwork²⁵ kérdéseket is feltenni, amit a becslés során lényegében az adatok

²⁴ Egy saját kutatásomban például arra kértem a válaszadót, hogy becsülje meg, hányan ajánlanák őt részvételre az adott kutatásban.

²⁵ Hálózati szakkifejezés a személyes kapcsolathálóra, azaz a személy és a vele közvetlen kapcsolatban állók hálózatára.

bővítéseként lehet használni. Lu azt javasolja, hogy a válaszadókat kérjük arra, becsüljék meg egyes vizsgált változók eloszlását ismerőseik esetén (triviális lehet a korcsoport, a nem, az iskolai végzettség). A hasonló típusú kérdések megszokottnak tekinthetők a jelenleg használatos hálózatkutató kérdőívekben. Az így kapott eredmények az adott személy kapcsolatszámával súlyozva az RDS I becslés pontosítását teszik lehetővé. A szerzők javaslata a (6) képlet C_{XY} paraméterbecslésének módosítására vonatkozik.

$$\hat{C}_{XY}^{ego} = \frac{1}{n_X} \sum_{i=1}^{n_X} \frac{n_{Xi}^Y}{d_{Xi}} \quad (26)$$

Ahol

\hat{C}_{XY}^{ego} *Az X értékkel jellemezhető csoporttagoktól az Y értékkel jellemezhető csoporttagok felé mutató kapcsolatok arányára vonatkozó becslés az X értékkel jellemezhető csoporttagok ki-fokszámához viszonyítva*

n_X *Az X értékkel jellemezhető mintatagok száma*

n_{Xi}^Y *Az X értékkel jellemezhető mintatagok közül az i-ik Y értékkel rendelkező célcsoporttag felé mutató kapcsolatainak száma (egonetwork adat, beleértve az ajánlót és ajánlottat is)*

d_{Xi} *Az X értékkel jellemezhető mintatagok közül az i-ik fokszáma*

A (26) képlet tulajdonképpen a (fokszám arányos) bekerülési valószínűség figyelembevételével végzett Hansen-Hurwitz becslés. A szerzők valós és szimulált hálózaton szimulált mintavételekkel hasonlították össze a javasolt új becslést a korábbi RDS I becsléssel és azt találták, hogy az új becslés, amennyiben az RDS mintavétel feltételei teljesülnek, kisebb szórású. Ezen túl szimulálták azt a helyzetet is, amikor a válaszadó nem véletlenszerűen ajánl következő megkérdezendő személyt. Ebben az esetben az egonetwork alapú becslés nemcsak kisebb szórású, de a preferenciális ajánlásból fakadó torzítást is jelentősen csökkentette. A becslés előremutató abban az értelemben, hogy úgy vezethet kisebb szórású és bizonyos esetekben pontosabb becslésekhez, hogy teljesíthető feltételeket támaszt a kutatóval szemben. Ugyanakkor ki kell emelnem és ki kell egészítenem Lu azon megjegyzését, miszerint csak nem szenzitív információk kapcsán képzelhető el egonetwork adatgyűjtés: ezen túl a kérdőív és így az interjú terjedelmét is jelentős mértékben növelhetik az ilyen típusú kérdések, ráadásul a másokra vonatkozó információk bizonyos kérdések esetén ki nem számítható módon növelhetik a bizonytalanságot. A problémák ellenére az egonetwork alapú adatok felhasználásával a becslés során komoly eredmény és jelentős design hatás csökkenés érhető el segítségével, különösen olyan változó esetén, amelyek összefüggnek a vizsgált célcsoport hálózati klaszterezettségével.

Ott és munkatársai [2017] megjelenés előtt álló cikkükben figyelemre méltó becslési eljárást javasolnak, amely visszatevés nélküli mintavétel esetén kezeli a szereplők mellett az élek eltérő mintába kerülési valószínűségét is. A módszer Gile [2011] iteratív eljárásán alapul, azonban a szereplők mellett az élek mintába kerülésének egyenlőtől eltérő valószínűségével is számol. A szerző által javasolt becslés két lépésben történik: az első lépésben a Gile korábbi becslése alapján az adott fokszámú szereplők mintába kerülési valószínűségét becsüli meg, majd az adott fokszámú szereplők közötti kapcsolatok valószínűségére ad iteratív becslést. A két becsléssel korrigálja az RDS I becslés két komponensét (változó értékei szerint képzett csoportok). Ott és munkatársai szimulációs eredményei alapján a célcsoporthoz képest nagy minták esetén a javasolt becslés Gile korábbi becsléséhez hasonlóan teljesített mind a becslés torzítása, mind annak szóra szempontjából. A becslés jelentősebb homofília esetén, kis mintaméret mellett is elenyésző torzítást és jelentősen kisebb szórást mutat szemben az RDS II becsléssel vagy Gile fent leírt becslésével.

Heckathorn [2011] cikkében összefoglalást ad az alkalmazott becslési eljárásokról, amelyet a fejezetben ismertetett további eljárásokkal kiegészíték (**Hiba! A hivatkozási forrás nem található.**).

2. táblázat Az RDS becslések evolúciója (Heckathorn [2011]) alapján kibővítve a fejezetben leírt becslésekkel)

RDS becslés	Felhasznált információ	Becslés típusa	Korlátok	A becslés szórása	Összevetés
(B#01) Heckathorn [1997]	Ajánlási mátrix	Markov egyensúly	Csak nominális változókra; nem kontrollált a homofília különbségekre	Nincs becslés	Kimutatja, hogy a minta önsúlyozó, ha a homofília azonos a csoportok között
(B#02) Heckathorn [2002]	Ajánlási mátrix, önbevallásos fokszám	Kölcsönösségi modellen alapuló becslés; lineáris legkisebb négyzet becslés többértékű változókra	Csak nominális változókra	Bootstrap becslés	Kontrollált a változó által képzett csoportok fokszám és homofília eltéréseire, becslés a becslés szórására
(B#03) Heckathorn [2002]	Ajánlási mátrix, önbevallásos fokszám	Kölcsönösségi modellen alapuló becslés; adatsimításos eljárás többértékű változókra	Csak nominális változókra	Bootstrap becslés	Adatsimítással szűkebb konfidencia intervallumok
(B#04=RDS I) Salganik, Heckathorn [2004]	Ajánlási mátrix, önbevallásos fokszám	Kölcsönösségi modellen alapuló becslés	Csak nominális változókra	Bootstrap becslés	Bizonyítás az asszimptotikus torzítatlanságra és jobb hálózatbecslés

RDS becslés	Felhasznált információ	Becslés típusa	Korlátok	A becslés szórása	Összevetés
(B#05=RDS II) Volz, Heckathorn, [2008]	Ajánlási mátrix, önbevallásos foks szám	Hálózat alapú RDS becslés	Nem kontrollált az ajánlási különbségekre	Analitikus becslés	Analitikus becslés a becslés szórására; lehetővé teszi folytonos változók elemzését
(B#06) Volz and Heckathorn, [2008]	Ajánlási mátrix, önbevallásos foks szám	Hálózat alapú RDS becslés adatsimítással	Csak nominális változókra	Analitikus becslés	Kimutatja a kölcsönösségi alapú becslések és a hálózati becslések konvergenciáját
(B#07) Heckathorn [2007]	Ajánlási mátrix, önbevallásos foks szám	Kétkomponensű RDS becslés	RDS módszer általános korlátai	Bootstrap becslés	Lehetővé teszi folytonos változók elemzését; kontrollált az ajánlási különbségekre
[B#08] Lunagómez és Airoidi [2012]	Hálózati modell, ajánlási mátrix, foks szám	Bayes-i becslés	RDS módszer korlátain túl az eredmény modelfüggő	nincs	Megfelelő hálózati modell mellett pontosabb becslés
[B#09] Gile[2011]	Ajánlási mátrix, önbevallásos foks szám	Iteratív foks szám becslés (a célcsoport méretét közelítő mintáknál)	RDS módszer általános korlátai, jelentős klaszterezettség	nincs	RDS II becslésnél megbízhatóbb becslés a célcsoport méretét közelítő mintáknál.
[B#10-11] Lu és mtsai. [2011a]	Ajánlási mátrix, önbevalláson alapuló ki- és befoks szám	RDS I és II becslés irányított hálózaton (a kapcsolatok kölcsönösségének sérülése esetén is alkalmazható)	RDS módszer általános korlátai, befoks szám bizonytalansága	nincs	Érvényes és megbízhatóbb, mint az RDS I és II becslés, de függ a befoks szám ismeretétől.
[B#12] Lu és mtsai. [2012]	Ajánlási mátrix, egonetwork adatok	Egonetwork becsült megoszlását felhasználó becslés	RDS módszer általános korlátai, az egonetwork adatok bizonytalansága	nincs	RDS I becslésnél megbízhatóbb becslés, amennyiben az egonetwork adatok pontosak
[B#13] Ott és mtsai. [2017]	Ajánlási mátrix, önbevallásos foks szám	Visszatevés nélküli mintavétel miatt eltérő él mintába kerülési valószínűségek kompenzálása		Bootstrap becslés	

A táblázat a vizsgált módszert az alkalmazott becslések szempontjából veti össze. Az összevetés szempontjai: a felhasznált információ, a becslés típusa, a becslés korlátai valamint a becslés szórása.

A jelen alfejezetben összefoglaltam, hogy milyen módon adható becslés az RDS mintavételből származó adatok alapján egyes változókra, azonban nem foglalkoztam azzal a kérdéssel, hogy milyen módon végezhető összefüggés vizsgálatok az ilyen módon keletkezett adatok alapján. A véletlen mintavételen alapuló minták elemzésére kidolgozott összefüggés-vizsgálatok többsége a mintavételi egységek függetlenségét feltételezi. Ez a feltétel a válaszadó-vezérelt mintavétel esetén nem áll fenn²⁶, ezért alternatív elemzési eljárások alkalmazására van szükség. Bastos és munkatársai [2012] egy más területen régóta ismert és alkalmazott módszer bevezetését javasolja az RDS minták dichotóm függőváltozójú magyarázómodelljeinek elemzésére. Elemzési módszerüket látens hálózati hatással bővített logisztikus regresszióknak (*logistic regression with latent network effect*) nevezik. Az eljárás alapjai a térstatisztikához vezethetők vissza, ahol a térbeli autokorreláció kezelésére használatosak. A szerzők nem ismertetik részletesen az elemzés alapját, de munkájuk általánosítása komoly előrelépést jelenthet az RDS alapú minták elemzése terén, mivel eddig figyelembe nem vett információt is felhasznál a becslések során.

II.4. Módszertani megfelelés és matematikai környezet

A második fejezet zárásaként érdemes megvizsgálni, hogy a fejezet fényében hogyan értelmezhető a két kutatási kérdés, azaz a válaszadó-vezérelt mintavétel módszertani megfelelése, illetve az, hogy milyen kapcsolatok, hatások merültek fel a vizsgált módszerrel tudományos konstrukcióként.

A I.1 alfejezetben megfogalmazott szempontok közül ebben a fejezetben közvetlenül érintettem az érvényesség kérdését. Az idézett tanulmányok alapján kijelenthető, hogy ideális körülmények között a válaszadó-vezérelt mintavétel alapján adható torzítatlan becslés, ugyanakkor a becslő eljárások nagy száma és eltérő – javasolt – alkalmazási feltételeik arra utalnak, hogy a valós körülmények kihívások elé állítják a vizsgált mintavételi módszert. Ezek a kihívások elsősorban azonban a megbízhatóság terén jelentkeznek.

A I.2 alfejezetben megfogalmazott tudományos konstrukció modell számos környezeti eleme közül a jelen fejezetben a matematikai környezetet érintettem. Matematikai szempontból a válaszadó-vezérelt mintavétel egy gráfon végrehajtott véletlen bolyongás probléma, amely Markov-lánc Monte Carlo mintavételként jellemezhető. Matematikai kihívást elsősorban az

²⁶ Könnyen belátható, hogy a mintavétel során egymást ajánlók semmiképpen sem tekinthetők egymástól függetlennek (pontosabban bekerülésük valószínűsége összefügg egymással, illetve az esetükben mérhető változók is nagy valószínűséggel összefüggnek egymással).

eljáráshoz kapcsolódó becslő módszerek megbízhatóságának becslése jelent, amit sok esetben általános megoldásokkal (bootstrap tesztekkel) pótolnak.

III. A válaszadó-vezérelt mintavétel gyakorlati alkalmazása

A következő fejezetben áttekintem a válaszadó-vezérelt mintavétel nemzetközi és hazai alkalmazását. A fejezet célja egyrészt a módszer széleskörű és bővülő alkalmazási területének demonstrálása a nemzetközi kutatásokról elérhető tanulmányok alapján. Ezen túl egy metaelemzés adatain keresztül bemutatom azokat a szokásos feltételeket, amelyek mellett a nemzetközi gyakorlatban alkalmazzák a válaszadó-vezérelt mintavételt. A hazai alkalmazások részletes bemutatásával elsősorban a módszer alkalmazása során felmerülő gyakorlati problémákat elemzem. Második kutatási kérdésem, azaz a vizsgált módszer és környezete közötti kölcsönhatások közül az alkalmazó környezet és a tudományos megrendelői környezet és a társadalmi környezet kölcsönhatásait vizsgálom részletesen.

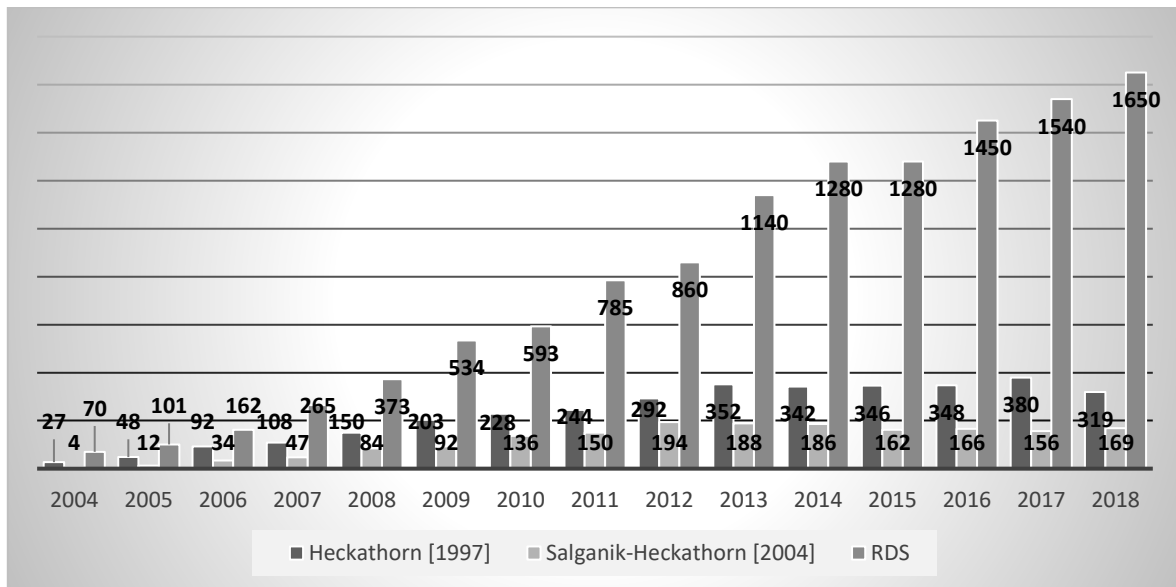
III.1. A válaszadó-vezérelt mintavétel alkalmazása nemzetközi kutatásokban

A tudományos megrendelő környezet alakulását a vizsgált módszer szempontjából egyszerű tudományometriai módszerrel vizsgálom. Vizsgálatom eszköze a Google Scholar válaszadó-vezérelt mintavétellel kapcsolatos találati listájának elemzése évenkénti bontásban.

Egy ilyen vizsgálat szempontjából meghatározó kérdés, hogy mely cikket tekintünk az adott módszerrel kapcsolatba hozhatónak. Tekintettel arra, hogy a vizsgálat célja a módszer elterjedési dinamikájának bemutatása és az érintett kutatási területek feltárása, egy szűk (konzervatív alulbecsléshez vezető) és egy tágabb (felülbecsléshez vezető) feltételt alkalmaztam. Szűkebb definícióm azon a premisszán alapul, hogy a módszert megfontoltan alkalmazó kutatások publikációi minden bizonnyal hivatkoznak a válaszadó-vezérelt mintavétel szempontjából meghatározó Salganik-Heckathorn [2004] cikkeire, illetve az módszer megalkotásához köthető első Heckathorn [1997] cikkeire. Azért tekinthető ez a feltétel konzervatívnak, mert előfordulhat ugyanakkor, hogy valóban a módszert alkalmazó kutatás leírása során nem hivatkoznak az említett a cikkekre, vagy nem az említett a cikkekre hivatkoznak, hanem esetleg más vonatkozó irodalomra. Adatgyűjtésem során tágabb definícióként a „respondent-driven sampling” kifejezés elfordulását vizsgáltam. A vizsgálat módszerének korlátja egyrészt a vizsgálat eszköze, a Google Scholar, amely esetleg a téma szempontjából releváns cikkeket nem tartalmaz vagy a releváns cikkeket esetenként többször, külön entitásként tartalmazza. Másrészt korlátként jelenik meg a nyelv, mivel csak azokról a cikkekről szerezhettem így információt, amelyek angol nyelven jelentek meg, vagy – és erre különösen sok kínai nyelvű cikk esetén volt példa – legalább az absztrakt angolul is elérhető, esetleg más nyelvű cikk tartalmazza a hivatkozást vagy a módszer angol nyelvű nevét, illetve a

szerzők maguk regisztrálták a cikket a Google Scholar rendszerében. A talált adatokat a **Hiba! A hivatkozási forrás nem található.** foglalja össze évenként 2004 és 2018 között bemutatva a módszer szempontjából meghatározó két hivatkozott cikk idézettségét, illetve a „respondent-driven sampling” kifejezés előfordulási gyakoriságát²⁷.

3. ábra A válaszadó-vezérelt mintavétel megjelenése publikációkban: Salganik és Heckathorn 2004-ben, illetve Heckathorn 1997-ben publikált cikkének idézettsége, valamint a „respondent-driven sampling” kifejezés megjelenése tudományos publikációkban 2004 és 2018



Forrás: Google Scholar, saját adatgyűjtés

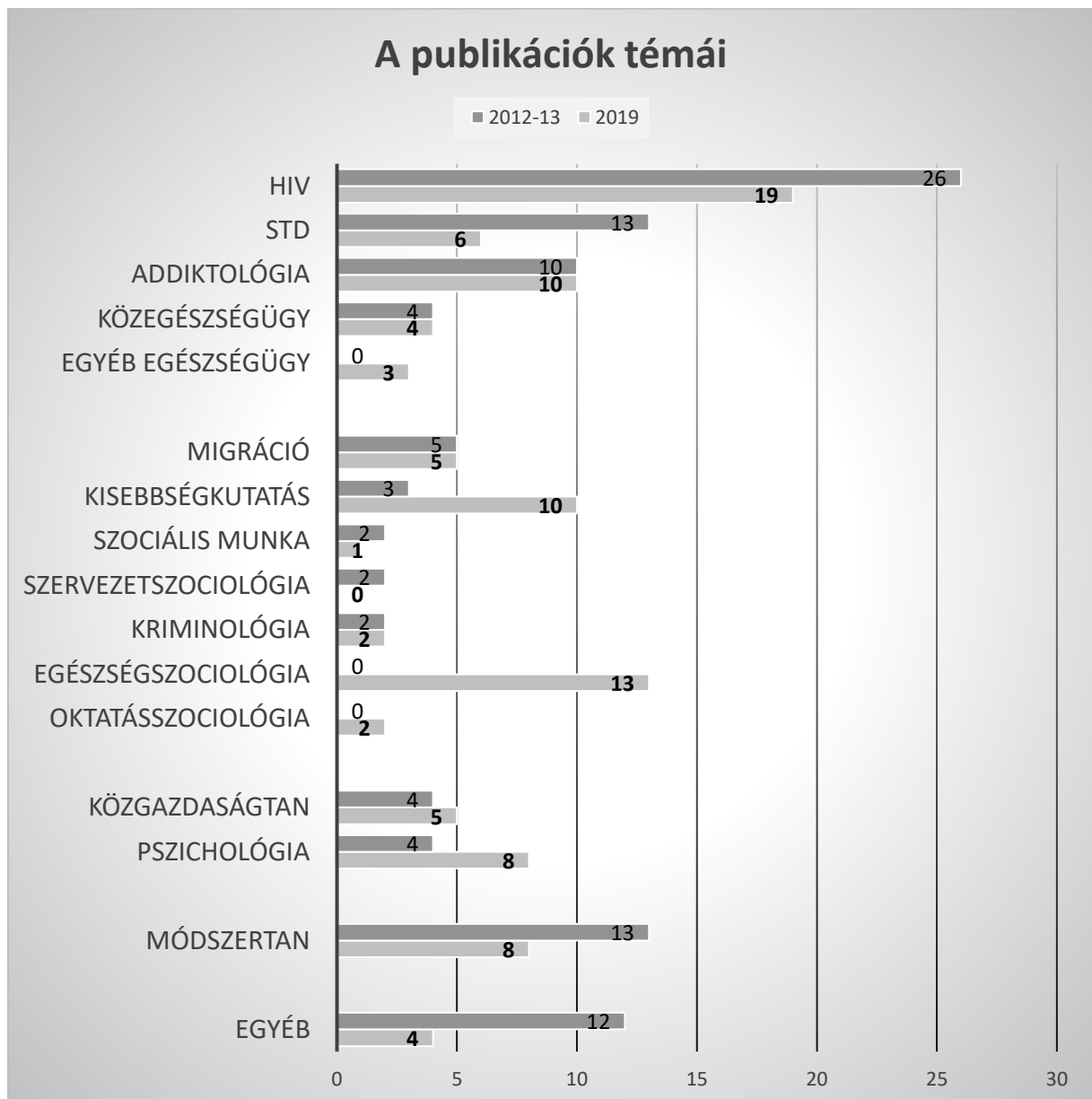
Mindhárom vizsgált adat alapján az látszik, hogy a módszer használata terjedőben van, a módszerre hivatkozó 1600 feletti éves megjelenés 2018-ra pedig azt jelzi, hogy jelenleg világszerte jelentős számú kutatás alkalmazza fő vagy kiegészítő módszerként a válaszadó-vezérelt mintavételt. Meg kell ugyanakkor jegyezni, hogy jelentős a különbség az általam meghatározónak gondolt források idézettsége és a módszer nevének előfordulása között, ami az abszolút számok tekintetében óvatosságra int. Mindazonáltal kimondható, hogy a tudományos megrendelői környezet növekvő érdeklődést mutat a vizsgált módszer iránt.

Természetesen az említések, illetve idézések gyakorisága nem jelenti azt, hogy minden cikk mögött válaszadó-vezérelt mintavétellel végzett kutatás állna (előfordulhat például, hogy csak fel- és elvetett módszerként szerepel az adott cikkben a vizsgált módszer). Emellett fontos azt is megvizsgálni, hogy vajon milyen területeken jellemző elsősorban a módszer használata. Mivel a válaszadó-vezérelt mintavétel alkalmazása 2010-es évek közepére jelentős mértékűvé

²⁷ A publikációk listájának elkészítéséhez a Google Scholart használtam, 2019 februárjában, a kereső kifejezés a „respondent-driven sampling”, a találatokat idő sorrendbe állítottam.

vált, a 2012-2013-ban, illetve a tendenciák feltárása érdekében 2019 első félévében megjelent, a válaszadó-vezérelt mintavételre („respondent-driven sampling”) hivatkozó publikációból a megjelenés ideje szerint rétegzett 100-100 elemű mintát vettem. A mintákba került cikkek alapján megvizsgálom a válaszadó-vezérelt mintavétel, mint eszköz elhelyezkedését a tudományos közbeszéd terében (4. ábra)²⁸.

4. ábra 2012-2013 között, illetve 2019 első félévében megjelent válaszadó-vezérelt mintavételre („Respondent-driven sampling”) hivatkozó publikációk közül véletlenszerűen kiválasztott 100-100 cikk témája (darab, illetve százalék)



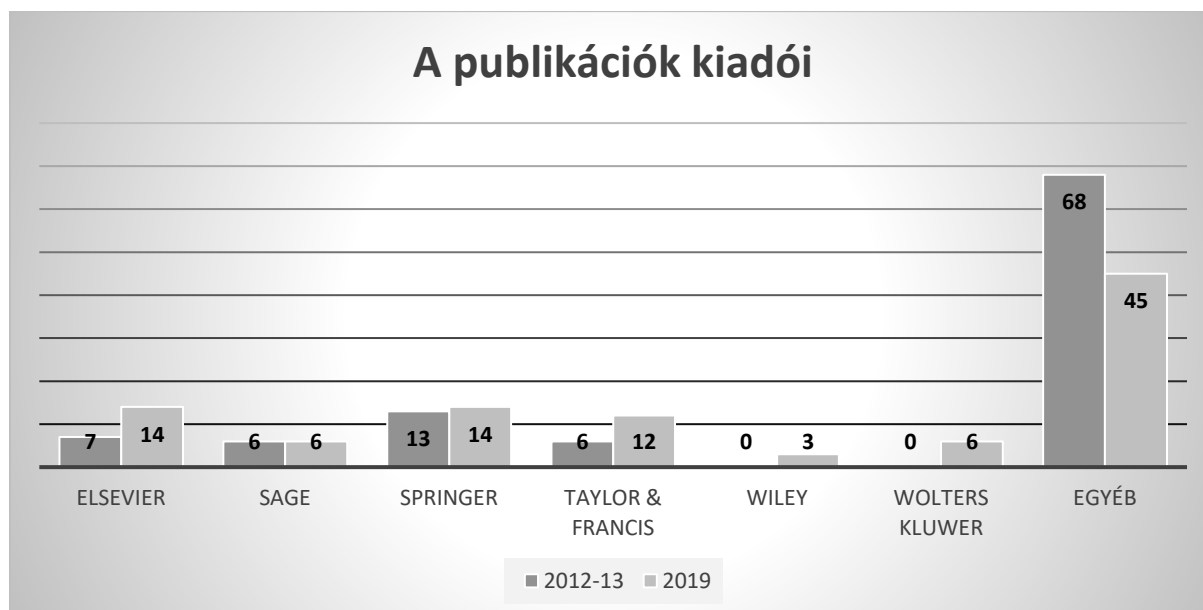
Forrás: Google Scholar, saját adatgyűjtés

²⁸ A cikkek témáját a cikkek címe, kulcsszavai, illetve amennyiben ez nem volt elegendő, a cikkek absztraktja alapján állapítottam meg.

Az ábrán tudományterületek szerint csoportosítottam a megjelenéseket. Jól látható, hogy a korábban említetteknek megfelelően a módszert legkiterjedtebben az egészségügyi területen alkalmazták mind 2012-2013-ban (52%), mind 2019-ben (42%), ezen belül is kiemelkednek a HIV fertőzöttel kapcsolatos kutatások (a vizsgált cikkek 26%, illetve 19%-a), illetve összességében a nemi betegségekkel foglalkozó kutatások (összesen 39% illetve 25%). Megfigyelhető ugyanakkor, hogy az egészségügyi alkalmazás aránya csökkenő tendenciát mutat. Nem elhanyagolható e mellett a szűkebben vett szociológiai tárgyú publikációk aránya sem: 2012-13-ban csaknem a vizsgált cikkek ötöde (összesen 19%-át), 2019-ben már harmada (összesen 33%-át) volt ide sorolható. A szociológiai tárgyú cikkek viszonylag sok témakörben jelentek meg, de valamelyest kiemelkedik a migrációs, kisebbségkutatás és 2019-ben az egészségszociológiai tárgyú cikkek. Viszonylag magas és növekvő tendenciát mutat a szociológia társtudományainak (közgazdaságtan, pszichológia) előfordulása is. A korábbi fejezetben idézett nagyszámú módszertani közlemény mellett nem meglepő, hogy a vizsgált cikkek mintegy tizedét tették ki a módszertani írások, csökkenő tendenciát mutatva. Viszonylag nagyarányú az egyéb kategóriában szereplő publikáció. Ezek között akadt többek között témájában iparművészeti (a módszert a releváns szakirodalmak kiválasztására használták), úrkutatási (mintavétel mérnökök közül), bioetikai (mintavétel egészségügyi szakszemélyzetből) is. A publikációk az előre vártak megfelelően nagyrészt angol nyelvűek voltak, ugyanakkor 2012-13 fordulóján hat, 2019-ben 10 ettől eltérő (francia, kínai, német, olasz, orosz, portugál, spanyol) nyelven íródtak – ez azonban a vizsgálat módszere miatt nem tekinthető jellemzőnek a megjelenő cikkekre általában.

A vizsgált publikációk jelentőségének megítélése érdekében elkészítettem azok kiadók szerinti megoszlását is (5. ábra).

5. ábra 2012-2013 között, illetve 2019 első félévében megjelent válaszadó-vezérelt mintavételre („Respondent-driven sampling”) hivatkozó publikációk közül véletlenszerűen kiválasztott 100-100 cikk kiadója (darab, illetve százalék)



Forrás: Google Scholar, saját adatgyűjtés

Az ábrán kiemeltem a leggyakrabban előforduló és egyúttal jelentősebbnek tekinthető kiadót. Az hat leggyakoribb kiadó együttesen 2012-13-ban a megjelenő publikációk valamivel több, mint két-ötödét (42%-át), 2019-ben már több, mint felé (55%-át) fedi le. Ez a hat kiadó a tudományos kiadás piacán a megbízhatóságot jelzi, ami a módszer bevettségét, elfogadottságát támaszthatja alá. A kiadónál pontosabb képet kaphatunk a folyóirat cikkek esetén, ha a folyóiratok tudományometriai mérőszámait is megvizsgáljuk. A tudományometriai mérőszámok közül a Scopus adatbázisára épülő SCImago rendszert²⁹ használtam. A mérőszámok közül a rendszer saját kialakítású Scientific Journal Ranking (SJR) indikátorát³⁰ vizsgáltam.

A vizsgált publikációk négy-ötöde volt cikk formátumú mindkét időszakban. A további publikációk között volt könyv fejezet, konferencia közlemény, szakdolgozat és disszertáció, illetve egyéb kiadvány. A 2012-13-ban megjelent cikk formátumú publikációk közül 63 (77,8%), 2019-ben 73 (91,3%) cikk megjelenési hely volt fellelhető a hivatkozott adatbázisban, ami egyértelmű növekvő tendenciát jelez.

Ha csak azokat a cikket vesszük figyelembe, amelyek esetén a tudományometriai adatok elérhetőek voltak, akkor az átlagos SJR érték 2012-13-ban 1,17, 2019-ben 1,53 volt. Ez az átlag

²⁹ <http://www.scimagojr.com/index.php>

³⁰ Az indikátor hálózatelemzéssel kiszámított presztizs mérőszám, amelynél a vizsgált hálózatot a folyóiratok, az irányított éleket pedig a hivatkozások jelentik (Guerrero-Bote - Moya-Anegón [2012])

2019-ben a szociológiai és politológiai folyóiratok közül sorban a 93. (Mobilities) és a 94. (Internet Research) közötti helynek feleltethető meg. Ezen belül a fent szociológiai tartalmúként definiált cikkek esetén az átlagos SJR érték 2012-13-ban 0,815, 2019-ben 0,979 míg az egészségügyi publikációk esetén 2012-13-ban 1,24, 2019-ben 1,70. Ez a különbség azonban megfeleltethető a két tudományterület átlagos pontszámkülönbségeinek.

A fentiek alapján elmondható, hogy az RDS módszert alkalmazó írások között a HIV témájú kutatások mind a korai (2004-es), mind a 2012-2013-as publikációk között jelentős gyakorisággal szerepelnek. Ennek köszönhető, hogy a módszer eredményeit összehasonlító jelentősebb elemzés ezen a területen jelent meg. A következőkben két tanulmányt foglalok össze, amelyek egy átfogó nemzetközi metaelemzés ismertetnek (Johnston és mtsai. [2008], Malekinejad és mtsai. [2008]). A metaelemzés során 123 Egyesült Államokon kívül végzett kutatást találtak, amelyek megfeleltek a felállított kritériumoknak (megfelelő kutatási kérdést vizsgáltak és RDS módszert alkalmaztak). A 3. táblázat foglalja össze a 123 kutatás főbb paramétereit.

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, 11 esetben nem volt információ
Interjúkészítés módja	110 esetben személyes kérdezés, 8 esetben számítógéppel segített strukturált interjú, 5 esetben önkitöltős kérdőív
A kérdezés helyszíne	92 esetben egy helyszín, maximum öt helyszín
Kezdőpontok kiválasztása	120 esetben demográfiai vagy magatartási kockázatok alapján diverzifikáltan történt, csupán 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 válaszadónak 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 kondomot, 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)
Design hatás	1-2,5 között változott, 34 esetben volt nagyobb 1,5-nél (32 esetben ilyen adatot nem közöltek) ³¹
Mintaméret	Az átlagos mintaméret 273 volt (59 és 963 között változott), 5 esetben nem közöltek ilyen adatot; a kitűzött mintaméret átlagosan 98%-át 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) ³²
Becslés módja	121 esetben az RDSAT szoftvert használták, 2 esetben más módszert

Forrás: Malekinejad és mtsai. [2008]

A Hiba! A hivatkozási forrás nem található. 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 alkalmazása, pénzjutalom a részvételért, az elemzéshez az

³¹ A szerzők nem közlik, hogy a design hatás kiszámítása milyen módon történt az egyes kutatások esetén.

³² A szerzők nem közlik, hogy mit fogadtak el egyensúlyi állapotként (pl.: hány változó esetén vizsgálták az egyensúlyi állapot meglétét az egyes kutatásokban).

RDSAT szoftver használata). Ugyanakkor más vizsgált szempontok mentén már jelentősebb eltérések voltak tapasztalhatók a vizsgálatba vont kutatások között (kezdőpontok száma, rekrutációs határidő megszabása, mintaméret). A vizsgált tényezők egy harmadik csoportja relatíve kis számban kerülnek említésre (jutalom a rekrutációért) a vizsgált publikációkban. Az 4. táblázat foglalja össze azokat a kvantitatív, illetve kvalitatív tényezőket, amelyek a vizsgált RDS mintavétellel készült kutatások eredményességét a metaanalízis alapján a leginkább befolyásolták.

4. táblázat RDS mintavétellel kapcsolatos siker és kudarc tényezők 128 nemzetközi HIV fertőzéssel kapcsolatos kutatás metaelemzése alapján

Siker tényezők	Kudarc tényezők
Egyes csoportok esetén nagyobb valószínűséggel teljesül a kitűzött mintaméret cél (pl.: intravénás droghasználók)	A rosszul meghatározott célcsoport a mintavétel során torzításhoz vezethet (pl.: amennyiben a célcsoport korhoz kötött és a kapcsolatszám kérdezésénél ezt nem veszik figyelembe, akkor a kapcsolatszámok torzítottak lesznek)
A rekrutálás határidőhöz kötése szignifikánsan növelte a tervezett és megvalósult minta méretének hányadosát.	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 kapcsolatfelvételben)
Pont és intervallumbecslések készítéséhez hasznos az RDSAT szoftver alkalmazása.	

Forrás: Johnston és mtsai. [2008]

III.2. A válaszadó-vezérelt mintavétel alkalmazása hazai kutatásokban

Az alábbi fejezetben a hazai irodalomban előforduló (kisszámú) válaszadó-vezérelt kutatást tekintem át a fent leírt módszertani fejtegetések tükrében³³. A fentiekben már említettem, hogy – a szerző és munkatársai két megjelent és fent már idézett cikkét nem számítva – a válaszadó-vezérelt mintavétel elmélete mindössze két helyen kerül említésre a magyar módszertani

³³ A folyóirat közléseken 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 banki ügynök kutatása (V. HUNNET 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.

szakirodalomban³⁴. Domokos Tamás és szerzőtársai [2010] több mint 600 oldalas módszertani összefoglalójuk „Rejtett populációk becslési módszerei” című fejezetében valamivel kevesebb, mint három oldalon foglalják össze. A szerzők Heckathorn (illetve szerzőtársai) három cikkére hivatkoznak, a mintavétel módszerét vázolják fel, annak becslési háttere nélkül (így az összefoglaló nem alkalmas arra, hogy kutatást tervezzenek a leírás nyomán). A szerzők egy pontatlanságát azonban érdemes korrigálni: a módszer nem valószínűségi mintát, hanem torzítatlan becslésre alkalmas mintát ad (bár a fentiek alapján látható, hogy a helyzet valójában bonyolultabb). A módszer másik magyar irodalomban megjelenő leírását Kapitány [2010] adja. A szerző cikkében a ritka populációk mintavételi módszereit foglalja össze, a 16 oldalas cikkben szűk egy oldalt szentel a módszer leírásának. A leírás a mintavétel módszerét vázolja fel röviden, két cikkre (Heckathorn [1997] és Salganik-Heckathorn, [2004]) hivatkozva. A szerző 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 a módszer azon feltétele, hogy bármely két célcsoporttagot összeköti az ismeretségi háló. Összefoglalva tehát a magyar módszertani szakirodalom eddig nem biztosított alapos áttekintést az angolszász területen egyébként jelentős mértékben használt módszerről.

A magyar kutatási irodalmat vizsgálva összesen hat kutatás kapcsán található nyoma a vizsgált módszer alkalmazásának. A 5. táblázat összefoglalja a módszert alkalmazó kutatások adatait.

5. táblázat *Válaszadó-vezérelt kutatások a hazai irodalomban*

	Vizsgált csoport	Mintaméret
Bernát [2006]	Romák	N=775
Márványkövi és szerzőtársai [2008]	Roma intravénás droghasználók	N=70
Horváth-Kmetty [2008]	Külföldön élő magyar diplomások	N=358
Hárs [2009]	Magyar munkavállalók Ausztriában	N=418
Simon [2010]	Budapesti romák	N=220
Blaskó [2015]	Magyar kivándorlók	N=125
Hárs – Simon [2016]	Külföldön dolgozó magyar orvosok	N=155

Forrás: saját adatgyűjtés

A magyar irodalomban szereplő válaszadó-vezérelt cikkek közös jellemzője, hogy viszonylag kevés információt tartalmaznak a mintavételről, különösen annak tényleges lebonyolításáról és

³⁴ Örvendetes, ugyanakkor, hogy az egyetemi tananyagokban is megjelent a módszer, még ha esetenként csak említés szintjén is (Balogh-Simon [2011]; Kmetty-Sztárayné [2011];)

az alkalmazott becslési eljárásokról. Az alábbiakban ismertetem az egyes cikkekben elérhető módszertant.

Az irodalomban fellelhető első valóban válaszadó-vezérelt mintavételt alkalmazó kutatást Bernát [2006] publikálta. A kutatás célja a roma és nem roma népesség migrációs potenciáljának vizsgálata, különös tekintettel annak hálózati hátterére. Válaszadó-vezérelt mintavételre a romákra vonatkozó minta kialakításánál került sor. A kutatók 775 fős mintát alakítottak ki. A kistérségek mintáit a 2001-es Népszámlálás roma válaszadóinak arányában súlyozták, így alakult ki az országos minta (erre a kutatásra a következő alfejezetben még visszatérek). A kutatás során a 150 kistérségből vettek mintát, majd kistérségenként alkalmazták a válaszadó-vezérelt mintavételt. Kistérségenként 3 kezdőpontból 42 fős minta kialakítására került sor úgy, hogy minden válaszadó 6 ajánlást adhatott, amiből egyet kerestek fel a kérdezőbiztosok (ez a design szerint 14 elemű láncokat jelentett).

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

Horváth és Kmetty [2008] kutatásának célja a Magyarországon és a külföldön élő magyar diplomások értékrendjének összehasonlítása volt. A hazai diplomások értékrendjének vizsgálatához a European Social Survey³⁵ magyarországi adatait (az 1500 fős országos minta diplomás almintáját) használták fel. A külföldön élő diplomások vizsgálatához használták a szerzők a válaszadó-vezérelt mintavételt. Az adatfelvétel interneten keresztül történt, az e-mail címet használták egyedi azonosítóként. Kiindulópontként több különböző forrást használtak (volt, aki közvetlen személyre szóló e-mailt kapott, 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

³⁵ A European Social Survey egy 2002 óta két évente végzett nemzetközi kutatás, amelyben európai államok kutató intézményei önkéntesen vesznek részt finanszírozva az adott ország adatfelvételének költségeit (a kutatás honlapja: <https://www.europeansocialsurvey.org/> Utolsó elérés: 2019. február 14.)

állapot, ami a minta reprezentativitásának indikátora”. Nem található azonban információ arról, hogy hány lépésben került kialakításra a minta, illetve mekkora volt az átlagos lánchosszúság.

A következő említhető kutatás Hárs Ágnes [2009] kutatása, melynek célja a rendszerváltás óta Ausztriában munkát vállaló magyarok attitűdjének, valamint munkavállalásukkal kapcsolatos adatainak felmérése volt. A minta mérete összesen 418 fő. A kutatás érdekében két almintát alakítottak ki (Nyugat-Magyarországi régióban, illetve az ország többi részében). A kutatók megyénként adott számú kezdőpontból indultak ki (a kezdőpontok száma a célcsoport korábbi kutatások alapján becsülhető arányának felelt meg). A kérdezőbiztosoknak kezdőpontként 14 elemű láncot kellett lekérdeznie 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 kialakítani, azonban „a lekérdezés során meghiusulások, alacsony kapcsolatszám és egyéb nehézségek miatt az országos minta 163 fős lett, míg a Nyugat-Magyarországi mintába 255-en kerültek”. A válaszadók mindössze 19%-a ajánlott kettő vagy több, 9% három vagy több személyt. A kérdezés négy hónapon keresztül zajlott. Az adatokat az ajánlott személyek számával (fokszám) fordítottan arányos súlyozással, valamint korábbi kutatás alapján NUTS 1³⁶ régió, iskolai végzettség és kor alapján súlyozták, oly módon, hogy a korábbi kutatásból csak a peremeloszlásokat használták, és iteratív arányos illesztéssel alakították ki a súlyt. A kutatók az adatok érvényességét független adatok (munkavállalási statisztikák, amelyek azonban nem tartalmazzák a teljes vizsgált csoportot) segítségével ellenőrizték, amely alapján nem tekintetében jól közelített a súlyozott minta, hasonlóan jól becsülte a minta a 25-44 év közötti korosztályt is. Alulbecsülte ugyanakkor a minta a 45 év felettieket és a burgenlandi munkavállalókat, míg felülbecsülte a 25 év alatti és a bécsi munkavállalókat.

Az ELTE Módszertani Kutatóközpontja és a Társadalmi Konfliktusok Kutatóközpont közös kutatásaként 2010-ben a budapesti romák körében került sor – félbe szakadt – válaszadó-vezérelt mintavétellel végzett kutatásra³⁷ (Simon [2010]). A kutatásról készült részletes beszámoló lehetővé teszi a kutatás az előkészítés és részben a lebonyolítás kapcsán felmerült körülmény (kérdezők felkészítése, adatkezelés problémái), illetve nehézségek bemutatását. A

³⁶ NUTS régiók az Európai Unió belül érvényes egységes, hierarchikus területi felosztási rendszer, első szintje a NUTS 1 régiók a fő társadalmi-gazdasági régiók. Magyarország három NUTS 1 régióra osztható: Közép-Magyarország, Dunántúl, Észak és Alföld. (NUTS régiók leírása az Eurostat honlapján található meg. Internet: <https://ec.europa.eu/eurostat/web/nuts/background> - utolsó elérés: 2019.február 14.

³⁷ Az itt szereplő

válaszadó-vezérelt mintavétel felhasználását az indokolta, hogy egyrészt a korábbi kvantitatív kutatások esetén igen magas volt a megíúsulások aránya (Kemény István vezette 1993-1994-es kutatásnál például Budapesten 28% volt; Kertesi-Kézdi [1998]). Másrészt a kutatás során felvetett problémák (a romák helyzete helyett elsősorban alkalmazkodási stratégiáik vizsgálata) azt indokolták, hogy a korábbi adatfelvételek döntő többsége esetén alkalmazott, a környezet megítélésén alapuló cigányság definíció helyett a cigány közösségek véleménye legyen a mintába kerülés alapja.

A kutatás célcsoportja a budapesti 18 éven felüli romák, cigányok voltak, akiket a roma/cigány közösség egy másik öt ismerő tagja romának/cigánynak tartott. Azaz általánosítva a kutatás célcsoportja a Budapesten élő azon csoport, amelynek tagjai a csoport más tagjait romának/cigánynak tartanak. Ez a megközelítés eltér a szociológiai cigánykutatások (például Kemény és munkatársai [1976] vagy Kertesi-Kézdi [1998]) szokásos megközelítésétől, ahol az önbesorolás vagy a nem cigány környezet besorolása adja meg a vizsgált populáció definícióját. A célcsoport ilyen módon történő definiálását a vizsgált kutatási kérdések indokolják, mivel azok inkább kötődnek a roma/cigány közösség kulturális sajátosságaihoz, működéséhez, mint a többségi-kisebbségi viszonyhoz, különösen a többségi előítéletekhez.

A kutatás lebonyolítását az ELTE Társadalomtudományi Kar Módszertani Kutatóközpontja (MKK) végezte. A kutatás oktató és szakértő résztvevői önkéntesként, a kérdezést végző hallgatók többsége szakmai gyakorlat keretében végezte. A kutatás előkészítése során az MKK munkatársai a kar hallgatói közül toboroztak kérdezőbiztosokat, akiknek egy kötelező 4 órás felkészítésen kellett részt venniük. A felkészítés során a következőkre került sor:

- a mintavétel és a kérdezőbiztosi tevékenység bemutatása (V. melléklet)
- a kérdőív bemutatása
- az adatkezelési terv ismertetése (IV. melléklet)
- egy roma származású kommunikációs szakember segítségével gyakorlati felkészítés a terepmunkára
- a hallgatók megkapták az első lekérdezendő (előre egyeztetett) interjúalanyuk elérhetőségét

A kutatás kapcsán érdemes külön kitérni az adatkezelési tervre. Ahogy azt a VI.1 fejezetben leírtam, a személyes adatok kezelése válaszadó-vezérelt mintavétel esetén különösen nagy figyelmet érdemel. Az itt bemutatott kutatás szempontjából a célcsoport specifikuma

(etnikai hovatartozás) miatt ráadásul a személyes adatok megsemmisítéséig különleges adatokat kezeltek a kutatásban résztvevők. Érdeemes megemlíteni, hogy a kutatás időpontjában a jelenleg (2019-ben) hatályos adatvédelmi rendelkezésektől eltérő jogszabályok voltak hatályban, melyek értelmében elegendő volt adatkezelési tervet létrehozni, azt az érintettekkel megismertetni és az adatok védelmét biztosító rendelkezéseket betartani (többek között a titoktartási nyilatkozatokat aláírni - III. melléklet). Ma egy hasonló kutatás valamivel jelentősebb adminisztrációt igényelne – ugyanakkor a védendő adatok köre, a védelem jellege a változó szabályozás mellett sem változott jelentősen.

Az adatfelvétel első lépcsőjében az első kérdezést követően a lekérdezett kérdőíveket összegyűjtésre kerültek és a hallgatók visszajelzést adtak arról, hogy milyen problémákkal szembesültek a kérdezés során. A beérkezett problémákat feldolgozva a kutatást lebonyolító oktatói team a kérdőív kérdéssorrendjének módosítását határozta el (a kérdőívben belüli sorrendváltogatás és módosított ugrás következtében az arra érzékeny interjúalanyok többségét nem kérdezték bizonyos munkalehetőségek végzéséről). A továbbiakban a kérdezést az új kérdőívvel bonyolították, ezt követően nem történt az első lépcsőhöz hasonló negatív visszajelzés.

A nehézkes időpontegyeztetések és a hallgatók elfoglaltságai miatt az adatgyűjtés meglehetősen hosszúra nyúlt a tervezett 2010 októbertől decemberig tartó időszak helyett 2011 májusáig. A kutatás első fázisában 220 interjú került lekérdezésre, részben szervezési okok, részben az ajánlások alacsony száma, illetve a hallgatói kérdezők alacsonyabb hatékonysága, valamint a hosszúra nyúló adatfelvételi idő miatt további lekérdezésre már nem került sor, a kutatás félbe szakadt.

Blaskó [2015] az Európai Unió SEEMIG projektje keretében végzett magyar kivándorlókat célzó kutatás kapcsán számol be a válaszadó-vezérelt mintavétel alkalmazási kísérletéről. Ebben az esetben – bár a szerző idézi a vonatkozó irodalmat – nem valósult meg a válaszadó-vezérelt mintavétel, sokkal inkább beszélhetünk egy véletlen mintavétel hálózati alapú kiterjesztéséről. A kutatás során a hazai rendszeres Munkaerő-felmérés³⁸ (MEF) háztartási mintájából indultak ki és az egészítették ki a mintába került háztartásokhoz kapcsolódó kivándorlók két csoportjával: a volt háztartástagokkal és a háztartástagok kivándorolt

³⁸ A Munkaerő-felmérés a Nemzetközi Munkaügyi Szervezet (az ENSZ szakosított szervezete) által kezdeményezett Labour Force Survey, nagymintás kérdőíves kutatás hazai megvalósulása, amelyet a KSH bonyolít le (A kutatás leírása megtalálható a KSH honlapján: <https://www.ksh.hu/emef> - utolsó elérés: 2019. február 14.)

testvéreivel³⁹. Az így kialakított minta válaszadóit kezdőpontként használva célozták meg a kutatás tervezői válaszadó-vezérelt minta kialakítását. Bár az első hullám (a mintába kerültek ismerőseinek) lekérdezése megtörtént, a mintavétel nem folytatódott tovább. A sikertelenség számos tényezőre vezethető vissza. A válaszadók számára nem volt lehetséges a következő mintatag közvetlen bevonása, csupán az elérhetőség megadása (ami adatvédelmi kérdéseket vetett fel lásd VI.1 alfejezet). A mintába került háztartásokkal kapcsolatban álló azonosított 1908 elvándorló közül 1362 esetén a háztartások nem adtak meg elérhetőséget. Az 546 személyből, akiknek az elérhetősége a kutatók rendelkezésére állt 125 esetében volt sikeres az interjú. A 125 sikeres interjúból a válaszadók 51 esetben adtak meg további elvándorló személyhez elérhetőséget. A kockázatokat és a leírt arányokat mérlegelve a kutatók a válaszadó-vezérelt mintavétel felfüggesztése mellett döntöttek.

A Kopint Alapítvány 2012-ben az OTKA támogatását elnyerve a hazai orvosok migrációjának kutatását indította el⁴⁰, amelyről Hárs és Simon [2016] számol be. E kutatás kapcsán részletesebben mutatom be az adatfelvétel során felmerült problémákat. A kutatás két célt fogalmazott meg:

- érvényes becslést kívánt adni a magyar orvosok kifelé irányuló migrációjának mértékére (*outmigration*) valamint a külföldi munkavállalás egyes jellemzőire (munkavégzés hossza, a visszatérés, élet és munkakörülmények) továbbá a Magyarországon munkát vállaló orvosok számára (*immigration*) (leíró rész)
- a külföldi munkavállalást érintő magyarázó modellek tesztelése (*push* és *pull* tényezők, hálózati tényezők, jövedelem különbségre épülő modellek, komplex magyarázó modellek)

Terjedelmi okok, illetve a dolgozat témája miatt a fenti kutatási célok közül a kifelé irányuló migráció mértékére és jellemzőire irányuló, illetve a kifelé irányuló migráció okait firtató célokkal foglalkozom. A fent meghatározott célok közül a csoportméret becslések esetén nem merül fel mintavételi probléma. A további célok többsége azonban a külföldön munkát vállaló orvosok jellemzőinek vizsgálatát igényli.

A kutatás kezdetén a kutatás fő célcsoportja a következő módon került definiálásra: kivándorló (külföldön munkát vállaló) orvosok azok az orvosok és fogorvosok, akik magyar

³⁹ Itt nem térek ki a mintavétel statisztikai megalapozottságára, erről Blaskó [2015] cikkében, illetve az általa idézett

⁴⁰ A kutatás címe: „Migráció az egészségügyben: jelenségek és magyarázatok” azonosító száma: 101067. A jelen rész a kutatási jelentés átdolgozott változata, amelyet a szerző készített a kutatás módszertani szakértőjeként.

állampolgársággal rendelkeznek és legalább egy hétig dolgoztak külföldön a kutatást megelőző tíz évben.

Az adatfelvételi mód szempontjából a célcsoport jellegéből adódóan költséghatékonyan kizárólag az online adatfelvétel valósítható meg, ugyanakkor fontos szempont volt az elérhető legnagyobb megbízhatóság és érvényesség.

A célcsoport egy tipikusan ritka és nehezen körülhatárolható csoport. Egyrészt az ismert és a kutatások számára elérhető regiszterek (például kamarai tagság vagy orvosi regisztráció) listázott tagjai közül a célcsoporthoz tartozók az előzetes ismeretek alapján alacsony arányt képviselnek (ritka csoport). Másrészt a külföldi munkavállalás – később körüljárandó adatvédelmi okok miatt – csak az adott orvos (önkéntes) nyilatkozata alapján derül ki, arról egyértelmű regiszter nem áll a kutatók rendelkezésére (nehezen körülhatárolható csoport).

E ritka és nehezen körülhatárolható célcsoportok vizsgálatára az egyik lehetséges – e kutatás során is választott – módszerként merül fel a válaszadó-vezérelt mintavétel alkalmazása. A kutatás során, a célcsoporthoz igazodva több mintavételi design alkalmazására került sor.

A válaszadó-vezérelt mintavétel alkalmazásához statisztikai (S), módszertani (M), szociológiai (SZ) és jogi (J) kérdéseket kell tisztázni. A mintavétel statisztikai háttere szempontjából a következő kérdések megkerülhetetlenek:

S1) a válaszadók egy ismeretségi hálózatot képeznek-e

S2) ez a hálózat nem (jelentősen) klaszterezett (azaz véletlen bolyongás esetén a mintavétel nem rendelkezik memóriával), legalábbis a vizsgált változók mentén

Módszertani szempontból a következő kérdések szorulnak tisztázásra:

M1) hajlandók-e a válaszadók a célcsoporthoz tartozókat ajánlani / bevonní a kutatásba

M2) a bevont válaszadók hajlandónak-e részt venni a kutatásban

Szociológiai szempontból a következő kérdés megválaszolása szükséges:

SZ1) ismerik-e a válaszadók egymás célcsoport hovatartozását

Jogi szempontból megválaszolandó a következő kérdés:

J1) kezelhetőek-e a válaszadó ajánlásai

A válaszadó-vezérelt mintavétel abban az esetben hatékony, ha a fenti több szempontú kérdéssor mindegyikére pozitív a válasz. Ugyanakkor a kérdések jelentős részére nem adható előzetes válasz, ráadásul célcsoportonként a válasz eltérő lehet.

A kiinduló mintavételi design (D0) szerint a tetszőlegesen kiválasztott célcsoporttagokból álló kiinduló minta tagjai az online rendszeren keresztül meghívhatták a célcsoport további tagjait (minden mintatag legfeljebb hat további célcsoporttagot hívhatott meg, amennyiben nem hívott meg senkit a kérdőív véglegesítése előtt figyelmeztetést kapott). A kutatás célkitűzése – összhangban a válaszadó-vezérelt mintavétel matematika-statisztikai elméletével - kisszámú kiinduló minta és hosszú ajánlási láncok megvalósítása volt.

A leírt feltételek közül a kialakított design a jogi feltételt (J) tudta kezelni, mivel ez – legalább is a kiinduló modellnél – egzakt módon kezelhető, megválaszolható. A magyar jogszabályok, ahogy erről a jogi, adatvédelmi kérdésekkel foglalkozó részben már írtam, nem teszik lehetővé, hogy a válaszadó más személyes adatát (így például e-mail címét vagy telefonszámát) megadja a kutatónak – kivéve külön törvényi felhatalmazás esetén. Ugyanakkor a Nemzeti Adatvédelmi és Információszabadság Hatóság állásfoglalása szerint a válaszadó felkérhető arra, hogy vonjon be másokat a kutatásba és az általa közvetített beleegyezés esetén a személyes adatok átadása is lehetséges. Mivel az orvos migrációs kutatás esetén a kiinduló mintavételi design szerint a válaszadót kértük fel arra, hogy vonjon be újabb válaszadót a célcsoportból a jogi feltétel teljesült.

6. táblázat A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott kiinduló design eredménye

	Meghívottak száma	Kapott felhívást	Kitöltötte	Meghívott másokat	Meghívott / meghívó
0. kör	18	18	12	8	3,25
1. kör	26	9	2	0	0

Forrás: saját adatok

A 6. táblázat alapján elmondható, hogy a kiinduló design alapján a leírt feltételek közül a módszertaniak nem, illetve részben valósultak meg: a kutatásba meghívottak alacsony arányban (0. körben 66%, az első körben 22,2% (2 válaszadó)) válaszoltak, illetve a válaszadók kevésé vonta be további célcsoporttagokat a mintába (0. kör 75%, 1. kör 0% (0 válaszadó a 2 válaszadóból)).

Tekintettel a kiinduló design sikertelenségének jellegére célszerűnek tűnt a kiinduló minta bővítése helyett a designt módosítani.

Az első módosított design (D1) új feltételezésekkel élt, amelyek a korábbi problematikus feltételezések (M1, M2) pótlására szolgáltak. Újabb feltételezéseink a következők voltak (a nem szereplő feltételezések azonosak voltak, a módosítást alpontként jeleztük):

S1.1) az orvosok (beleértve egyaránt a hazai és külföldön munkát vállaló orvosokat) egyetlen ismeretségi hálót képeznek

M1.1) a teljes orvosi körben a válaszadók hajlandósága arra, hogy további célcsoport tagokat vonjon be magasabb, mint a külföldön dolgozó orvosok esetén

M2.1) a teljes orvosi körben a válaszadási hajlandóság magasabb, mint külföldön dolgozó orvosok esetén

Az első módosított design esetén a célcsoportot kiterjesztettük az összes magyar orvosra (és fogorvosra) úgy, hogy azon belül a külföldön munkát vállaló orvosok bevonását preferáltuk (ezt kértük a válaszadóktól). A célcsoport kiterjesztése lényegesen megnövelte a kiinduló minta lehetséges méretét (számos nyilvánosan elérhető orvos regiszter létezik e-mail címekkel).

7. táblázat A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott első módosított mintavételi design

	Meghívottak száma	Kapott felhívást	Kitöltötte	Meghívott másokat	Meghívott / meghívó
0. kör	133	133	12	4	1,5
1. kör	6	6	1	0	0

Forrás: saját adatok

A 7. táblázat összefoglalja az első módosított mintavételi design eredményét. Az eredmények azt mutatják, hogy az M1.1 és M2,1 feltételek egyáltalán nem teljesülnek: mind a meghívási hajlandóság (33%), mind a válaszadási hajlandóság (~10%) tekintetében a kiinduló designhoz képest rosszabb eredményeket tapasztaltunk.

Az első módosított design tapasztalatai alapján a mintavételi designt jelentősen módosítani kellett, a feltételezések módosításával. A második módosított mintavételi design (D2) esetén a következő feltételezésekkel éltünk:

M1.2) A célcsoporttagok bevonása a közösségi portálon elérhető hálózati kapcsolatok segítségével jól közelíthető

M2.2) A válaszadási hajlandóság személyes megszólítás (a kutató személyének közvetlen azonosíthatósága) esetén magasabb, mint az egyszerű e-mail küldés esetén

SZ1.2) A célcsoporttagok (nagyreszt) azonosíthatók a vizsgált közösségi portálon

SZ2.2) A célcsoporttagok jelentős hányada tagja a vizsgált közösségi portálnak

J1.2) A közösségi portál jogi környezete lehetővé teszi a kutatási célú megkeresést

A második módosított mintavételi design a válaszadó-vezérelt mintavétel szimulációjára épül egy népszerű közösségi portálon, a Facebook-on. A kutatásba mintavételi munkatársként bevont diákok a külföldön munkát vállaló orvosok azonosítható kezdőpontjaiból kiindulva véletlen bolyongást végeztek az elérhető kapcsolathálózaton a célcsoporthoz tartozók (azaz az a külföldön munkát vállaló orvosok) körében. Az így kialakított minta tagjait a kutatók személyes facebook profiljukból kérték fel a részvételre.

8. táblázat *A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott második módosított mintavételi design mintakialakítási fázisának leíró jellemzői*⁴¹

Kiinduló minta (kezdőpontok)	Teljes minta	Hibás azonosítás	Kitöltötte	Átlagos lánc hossz
55	443	2	155	8,05

Forrás: saját adatok

A felkért 443 válaszadóból a kutatás során 155-en töltötték ki a kérdőívet, ami 35% körüli (34,98%) válaszadási hajlandóságnak felel meg. Ez a válaszadási hajlandóság már megközelíti a személyes megkeresés válaszadási hajlandóságát⁴². A létrejött minta mérete azonban továbbra sem volt elegendően nagy megbízható elemzés végzéséhez, így a mintavétel kiegészítésére volt szükség, amelyet azonban már nem válaszadó-vezérelt mintavétellel, hanem szűrőkérdőíves kérdéssel oldottunk meg széles körű orvosi e-mail adatbázisból kiinduló minta alapján⁴³. Erre

⁴¹ A szimulált mintavétel miatt az előző táblázatokkal azonos szerkezetű, felvételi körönként információt tartalmazó táblázat kevésbé volna informatív, ezért itt inkább a lebonyolított mintavétel tulajdonságait foglalom össze.

⁴² A nagyon körültekintő módszertannal végzett és a lehető legmagasabb válaszadási arányt célként kitűző European Social Survey 7. körének válaszadási rátája 31,4% és 67,9% között változott országonként (az országok többségében 50% körül volt).

⁴³ Az adatfelvétel ezen fázisában egy jelentős orvos e-mail adatbázissal rendelkező direktmarketing tevékenységet folytató céget bízunk meg 5000 majd további 2000 orvos és fogorvos számára (2013. évi hazai KSH statisztikáknak megfelelő korcsoportos, régiós és szakmai (kórházi orvos, háziorvos, fogorvos, egyéb orvos) bontásban) e-mailben kutatási meghívásunk kiküldésével. A felkérést követően egy emlékeztető levelet küldtünk

a döntésre azért került sor, mert a mintavétel során a minta mérete megközelítette a külföldön tartózkodó magyar orvosok becsült számának tizedét, ennek következtében a minta kialakítását végző diákoknak egyre gyakrabban szembesültek azzal, hogy a személy, akihez eljutottak a mintavétel során már szerepel a mintában, ami a további munka komoly pszichés gátjává vált.

Összefoglalva a hazai kutatások közül kevés alkalmazta a válaszadó-vezérelt mintavételt, ezek is sok esetben kompromisszumok mellett. A problémák hazai környezetben elsősorban: a jogi környezet (az ajánló nem adhatja át az ajánlott személy személyes adatait), az ajánlások alacsony száma, az célcsoport mérete (orvosmigrációs kutatás). Ugyanakkor biztató kezdeményezés – természetesen a populációs torzítás figyelembevétele mellett – a közösségi portálok bevonásának lehetősége, amely új perspektívát ad a válaszadó-vezérelt mintavételnek.

III.3. A válaszadó-vezérelt mintavétel alkalmazói, megrendelői és társadalmi környezete

A hatodik fejezet zárásaként érdemes megvizsgálni, hogy a fejezet fényében hogyan értelmezhető a két kutatási kérdés, a válaszadó-vezérelt mintavétel módszertani megfelelősége, illetve milyen kapcsolatok, hatások merültek fel a vizsgált módszerre tudományos konstrukcióként.

A I.1 alfejezetben megfogalmazott szempontokat (megbízhatóságot és érvényességet) közvetlenül ebben a fejezetben nem vizsgáltam, ugyanakkor a hazai és nemzetközi alkalmazási gyakorlat kapcsán bemutatott problémák erőteljesen befolyásolják a vizsgált módszer megbízhatóságát és érvényességét, így módszertani szempontból sem mellékesek. A I.2 alfejezetben megfogalmazott tudományos konstrukció modell számos környezeti eleme közül a megrendelői, az alkalmazói és a társadalmi környezet hatását vizsgáltam. A fejezet nemzetközi áttekintést tartalmazó részében bemutattam a vizsgált módszer dinamikus terjedését. A válaszadó-vezérelt mintavétel terjedésének vizsgálata során egy cikk mintán rámutattam a módszer alkalmazásának tudományági megoszlására. A tudományági megoszlás szerint mind 2012-2013 fordulóján, mind 2019-ben a társadalomtudományi határterületként értelmezhető népegészségügyi témák mellett a szociológiai témájú cikkek szerepeltek a

ki. Az első kiküldés során minden levelet kapó orvost arra kértünk, hogy töltsse ki kérdőívünket, illetve továbbítsa felkérésünket külföldi munkatapasztalattal rendelkező ismerősének. A második kiküldés során már csak azoktól vártunk választ, akik külföldi munkatapasztalattal rendelkeznek. Az első hullámban 579 választ kaptunk, ebből 102 válaszadó rendelkezett külföldi munkavállalási tapasztalattal. A második hullámban 46 választ kaptunk és ebből 39-en rendelkeztek külföldi munkavállalási tapasztalattal. Az így keletkezett külföldi munkavállalási tapasztalattal rendelkezőkből álló minta egy véletlen mintavételen alapuló minta egy tulajdonság alapján szűrt almintájaként kezelhető.

második helyen a válaszadó-vezérelt mintavétel alkalmazási területeként, a vizsgált időszakban emelkedő tendenciát mutatva.

Az alkalmazói környezet és a vizsgált módszer kapcsolatára jellemző a módszer alkalmazásának következetessége, amelyet egy konkrét területen végzett metaanalízis bemutatása segítségével ismertetek nemzetközi szinten, és a hazai kutatások részletesebb áttekintésével elemzek részletesebben. A vizsgált példák alapján kijelenthető, hogy a válaszadó-vezérelt mintavétel alkalmazói rendszeresen kötnek az eredményeket jelentősen befolyásoló kompromisszumokat az alkalmazás során. Ezekre a kompromisszumokra azonban általában a kivitelezhetőség miatt szükség van, hatásukat azonban nem feltétlenül mérik fel.

Végül a vizsgált módszer és a társadalmi környezet kapcsolata a már említett hazai kutatások részletesebb bemutatása kapcsán érhető tetten. Mivel a válaszadó-vezérelt mintavétel a hagyományos mintavételi és adatfelvételi módszereknél több kooperációt igényel, a mintavétel sikeressége nagy mértékben múlik a társadalmi környezet viszonyán a mintavétel során elvárt magatartáshoz (a célcsoporthoz tartozó személyek bevonása a kutatásban a válaszadó által). A részletesen bemutatott hazai példák (különösen a KSH migrációkutatása és az orvismigrációs vizsgálat) kapcsán demonstráltam, hogy a kooperáció alacsonyabb foka, ezt a módszert teljes mértékben ellehetetleníti, vagy a módszeren alapuló alternatív megoldások alkalmazására sarkall.

IV. A válaszadó-vezérelt mintavétel alkalmazása kapcsán felmerülő módszertani kérdések⁴⁴

A válaszadó-vezérelt mintavételt az angolszász irodalom tanulsága szerint már több tucat országban, több száz esetben használták⁴⁵, ennek ellenére csak a legutóbbi években jelentek meg olyan kutatások, amelyek a módszert valós adatokon is evaluálják. A fejezet első két alfejezetében a Salganik és Heckathorn által 2004-ben írt cikkre hivatkozó 1761⁴⁶ publikáció közül azokat ismertetem, amelyek a módszer értékelésére törekedtek, illetve a felmerülő etikai kérdésekre javasolnak megoldásokat. A fejezet harmadik alfejezetében bemutatok egy saját szimulációs vizsgálatot. Végül az utolsó alfejezetben tárgyalom a kutatási kérdés a módszertani fejezetben leírtak összefüggéseit.

IV.1. A válaszadó-vezérelt mintavétel tulajdonságaira irányuló kutatások

Az alfejezetben összefoglalom a legfontosabb módszertani cikkeket, amelyek eltérő adatokon vizsgálják a válaszadó-vezérelt mintavétel tulajdonságait. A fejezet első részében bemutatom, hogy az egyes szerzők milyen eszközökkel, milyen adatokon vizsgálták a válaszadó-vezérelt mintavétel sajátosságait, majd rendszerezve közlöm a legfontosabb eredményeket – kiemelve az egymással esetlegesen ellentmondó következtetéseket.

Az idézett kutatásokban használt mintákat és az alkalmazott módszereket a 9. táblázat foglalja össze. A kutatások három alapvető csoportba sorolhatók módszertani megközelítés szempontjából: valós RDS adatfelvételek (általában összehasonlító) módszertani vizsgálata, valós társadalmi hálózatokon szimulált RDS adatfelvételek vizsgálata, modellezett hálózat(ok)on szimulált RDS adatfelvételek vizsgálata. Értelemszerűen az első típusú vizsgálatokban lehetséges az RDS tényleges folyamatának – beleértve az ajánlás és foksám megadás – sajátosságainak elemzése, ugyanakkor ezekben a kutatásokban nincs lehetőség ezen körülmények kontrollálására. A második típusba tartozó kutatások alkalmasak eltérő feltételezett adatfelvételi körülmények, illetve becsléstípusok és valós hálózat együttes hatásának tesztelésére, de nem alkalmasak a hálózati sajátosságok kontrollálására.

⁴⁴ A fejezet a Statisztikai Szemlében megjelent cikk jelentősen bővített, átdolgozott változata (Simon [2012])

⁴⁵ Malekinejad és munkatársai csak az AIDS és HIV fertőzések kapcsán az Egyesült Államokon kívül 123 RDS módszert használó kutatást azonosítottak összesen 28 országban. A módszer megjelenéséről a tudományos publikációkban lásd az III.1 fejezetet.

⁴⁶ A cikkre hivatkozó cikkek áttekintése a Google Scholar segítségével történt és a 2018 novemberében aktuális állapotot tükrözi.

9. táblázat Az válaszadó-vezérelt mintavétel módszertani vizsgálata kapcsán idézett cikkeben használt módszerek sajátosságai

	Célcsoport és mintaméret	Matematikai levezetés	Modellezett sokaság, modellezett mintavétel	Valós sokaság (hálózat), modellezett mintavétel	Valós sokaság, valós mintavétel
Kendall és mtsai. [2008]	MSM, N=406, [N=127 – hólabda], [N=274 – tér-idő]				+
Goel-Salganik[2009]	magas HIV fertőzés kockázatú személyek hálózata (Project 90), illetve középiskolások hálózata, szimulált mintavétel, N=500 (10 000 ismétlés hálózatonként)				+
Wejnert [2009]	egyetemi hallgatók online hálózata, két valós mintavétel, N=378 és N=159	+		+	
Gile-Handcock [2009]	korábbi kutatás valós hálózata alapján modellezett hálózat (ERGM), N=1000 (1000 ismétlés)		+		
Goel-Salganik [2010]	4430 fős részhálón szimulált, N=500 (10 000 ismétlés)			+	
Tomas-Gile [2010]	1000 modellezett 1000 fős hálózat, N=200 és N=500 (hálózatonként egy-egy)		+		
Kral és mtsai. [2010]	intravénás kábítószer használók, N=534 [N=651 – célzott mintavétel]				+
Guntuboyina és mtsai. [2012]	HIV kutatás alapján szimulált két hálózat, N=373 (hálózatonként 10 000 ismétlés)		+		
Wejnert és mtsai. [2012]	43 korábbi RDS minta, eltérő mintaméret				+
Lu és mtsai. [2012b]	16 082 fős online közösségi hálón szimulált mintavétel, eltérő mintaelemszámok (10 000 ismétlés paraméter kombinációnként)			+	
Mills és mtsai. [2014]	korábbi két kutatás adatfelvétele és au alapján modellezett 10 000 fős alapsokaság, N=350 (ismétlések száma ismeretlen)		+	+	
Nesterko és Blitzstein [2015]	szimulált alapsokaság, eltérő mintaméret (500 ismétlés paraméter kombinációnként)			+	
Verdery és mtsai [2015]	2015 egyetemi ismerettség hálón, (Facebook) szimulált adatfelvétel, eltérő mintaméret (100 000 ismétlés)		+		
Khabbazian és mtsai. [2017]	random gráf modellen és 4 kétezer fő körüli iskolai hálózaton modellezett adatfelvétel eltérő mintaméret (10 000 ismétlés paraméter kombinációnként)			+	

Forrás: saját adatgyűjtés

Míg a harmadik típus esetén bármely körülmény kontrollálása lehetséges, ugyanakkor az alkalmazott hálózatok, illetve a vizsgált körülmények csak közelítik a tényleges adatfelvételt. Mindezeket a sajátosságokat érdemes figyelembe venni az egyes elemzések eredményeinek interpretálásánál.

A 10. táblázat összefoglalja, hogy az idézett tanulmányok közül melyik milyen tényezők hatásait vizsgálta. Az alábbiakban tematikusan tekintem át az egyes cikkek által vizsgált területeket és témánként közlöm az eredményeket kiemelve az esetleges ellentmondásokat.

(1) *Homofília hatásának vizsgálata.* Homofília alatt az értjük, ha az ajánlások során a hasonló tulajdonságú, azaz a kutatásban vizsgált változókon azonos értékkel jellemezhető személyek preferenciája jellemző az egyszerű véletlen választáshoz viszonyítva⁴⁷. Tomas és Gile [2010] vizsgálata szerint (I) *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 esetén a becslések kisebb szórása miatt az RDS II becslés alkalmazandó.* Gile és Handcock [2009] szimulációjuk során azt találták, hogy (II) *jelentős mértékű homofília esetén egyáltalán nem javasolt a válaszadó-vezérelt mintavétel alkalmazása, ezért (III) előzetes kutatást javasolnak a kutatási kérdés(ek)ben a homofília meghatározása érdekében.* Goel és Salganik [2010] valós hálózaton modellezve a válaszadó-vezérelt mintavételt azt találta, hogy (IV) *a becslések aszimptotikusan torzítatlanok, ugyanakkor (V) a becslések varianciája jelentős, a számított design hatás 5 és 10 között van, ami ötszörös, tízszeres mintát tenne szükségessé az egyszerű véletlen mintavételhez képest hasonló megbízhatóság eléréséhez.* Goel és Salganik [2009] a válaszadó-vezérelt mintavétel MCMC tulajdonságából levezetik, hogy korábban tévesen vizsgálták csupán a homofília hatását a teljes hálózatra nézve, a homofília eltérései lokálisan is fontosak lehetnek különösen rövid mintavételi láncok és a vizsgált változó egyensúlyi eloszlásától távol eső kiinduló minta esetén. Másként fogalmazva, (VI) *ha a vizsgált hálózat partícionálható olyan módon, hogy a képzett partíciókon belül a vizsgált változó eloszlása eltérő és a partíciók közötti kapcsolatok száma kicsi, akkor az a becslések szórásának jelentős növekedéséhez vezet.*

⁴⁷ A homofília RDS mintavétel esetén a mintába került szereplők tulajdonságai alapján nem mérhető, mivel olyan torzítást okoz a mintában, amely elfedi azt. Egonetwork kérdésekkel azonban becslés adható a homofília mértékre (Crawford és munkatársai [2017]). A homofília problémája szorosan kapcsolódik, de nem azonos a mintavétel alapját képező hálózat klaszterezettségével, azaz azzal, hogy milyen mértékben jellemzők a hálózatra olyan csoportok, amelyeken belül magas a kapcsolatsűrűség, azonban a csoportok között alacsony.

10. táblázat A válaszadó-vezérelt mintavétel módszerét vizsgáló idézett cikkek által figyelembe vett tényezők

	Kendall és mtsai. [2008]	Wejnert [2009]	Goel-Salganik [2009]	Gile-Handcock [2009]	Goel-Salganik [2010]	Tomas-Gile [2010]	Kral és mtsai. [2010]	Guntu-bovina és mtsai. [2012]	Wejnert és mtsai. [2012]	Lu és mtsai. [2012b]	Mills és mtsai. [2014]	Verdery és mtsai. [2015]	Nesterko és Blitzstein [2015]	Khabbazian és mtsai. [2017]
(1) Homofília (hasonló ajánlásának tendenciája)			+	+	+	+				+		+		+
(2) Egyéb hálózati sajátosságok				+				+		+				
(3) Kiindulópontokból származó torzítás				+						+			+	
(4) Ajánlás sajátosságai				+		+		+		+		+	+	
(5) Fokszám mérés hatása		+	+								+		+	
(6) Minta paraméterei (méret, visszatevés, egyensúlyon kívüli minta)		+		+						+			+	
(7) Válaszmeztagadás						+				+				
(8) Eltérő becslések összevetése		+		+		+			+				+	
(9) Eltérő módszertanú mintavételek összevetése	+						+							

Forrás: saját adatgyűjtés

Verdery és munkatársai [2015] kimutatták, hogy (VII) *azon változók esetén, amelyek jelentős klaszterezettséghez vezetnek (azaz a hálózaton belül olyan alcsoportok vannak, amelyekre jellemző a nagy számú alcsoporton belüli és kevés azon kívüli kapcsolat és ezen alcsoportok esetén a vizsgált változó szóródása kisebb) a design hatás igen jelentőssé válhat (egyes esetekben akár 85-re is növekedhet).* Ezt az eredményt Lu és munkatársai [2012b] is megerősítették egy valós hálózaton szimulálva, igaz, a négy vizsgált változó esetén a design hatás 5 és 13 között változott (a különbségek a hálózati hatás és a homofília közötti összefüggésre utalhatnak). Ez az állítás a (VI.) állítás egy speciális, szélsőséges formájának is tekinthető. A szerzők arra is rámutattak, hogy (VIII) *a tapasztalt design hatás és a becslések ismert varianciabecslései nem korrelálnak egymással.* Khabbazian és munkatársai [2017] az ajánlási utasítás módosítását javasolják, a homofíliából származó hatások csökkentése érdekében. A szerzők azt javasolják, hogy a véletlenszerű ajánlás helyett a válaszadókat a kutatás során arra kérjék, hogy olyanokat ajánljanak, akikkel kevés közös ismerősük van⁴⁸ A módszert antiklaszter RDS mintavételnek nevezték el. Szimulációs eredményeik alapján (IX) *az antiklaszter RDS mintavétel mind szimulált, mind valós hálózaton jelentős mértékben csökkentette a becslések szórását a hagyományos RDS mintavétellel összevetve.*

(2) *Egyéb hálózati sajátosságok hatása.* Gile és Handcock [2009] vizsgálták, hogy milyen hatása van annak, ha a vizsgált változó egyes értékei által kijelölt csoportok átlagos fokszáma eltér egymástól (például, ha nemi megoszlásra vagyunk kíváncsiak és a nők átlagos kapcsolatszámja jelentősen eltér a férfiakétól). A szerzők azt találták, hogy (I) *az átlagos fokszámban mutatkozó különbségek a nagyobb átlagos fokszámú érték gyakoriságának alulbecsléséhez vezet.* (II) *Ez a hatás jelentőssé válik, ha a gyakoriságok jelentősen eltérnek az egyenletestől* (vagyis a vizsgált változó eloszlása nem egyenletes). Guntuboyina és szerzőtársai [2012] azt mutatták ki, hogy (III) *az ajánlónkénti ajánlottak egyenlőtlen száma esetén a hálózati struktúra jelentős hatással van a bekerülési valószínűségekre, és így az alkalmazott becslésekre, amelyek fokszám arányos bekerülési valószínűséget feltételeznek.* Lu és munkatársai [2012b] a hálózati hatást több szempontból is vizsgálták. Azt találták, hogy (IV) *amennyiben a hálózat irányított kapcsolatokkal rendelkezik, az RDS II becslések (változónként eltérő mértékben) az irányítatlan hálózatnál – változatlan becslés szórások mellett – nagyobb torzítást mutatnak (bár a torzítás a ki- és be-fok számok ismeretében kiküszöbölhető).*

⁴⁸ Két alternatív megfogalmazást is javasolnak a szerzők: „ajánljon olyan személyt a célcsoportból, akiknek az ismerősei többsége nem ismeri őt”, illetve „ajánljon olyan személyt a célcsoportból, aki nem ismeri az ismerősei többségét”.

(3) *Kiindulópontokból származó torzítás hatása.* Bár mind az RDS I, mind az RDS II becslések aszimptotikusan torzítatlanok (nagy minta elemszám esetén a becslések nem függenek a kiinduló-pontoktól), az elméleti megfontolások is hangsúlyozzák a hosszú láncok fontosságát. Gile és Handcock [2009] kimutatták, hogy (I) *különösen nagy homofília esetén fontos a hosszú láncok alkalmazása (a modellezés során 4-ről 6-ra emelték a láncok hosszát).* Lu és munkatársai [2012b] valós hálózaton végzett szimulációikkal alátámasztják ezt az eredményt: (II) *azonos minta elemszámok mellett mind a nagyobb kezdőpontszám, mind az ajánlónkénti nagyobb ajánlott szám növeli a vizsgált RDS II becslés szórását (mindkettő lényegében rövidebb lánchosszokat eredményez a mintavétel sajátossága folytán).* Nesterko és Blitzstein (2015) eredményei alapján (III) *nem mutatott jelentős eltérést az RDS II becslés varianciája egyenletes vagy fokszám arányos véletlen kezdőpont kiválasztás esetén.*

(4) *Ajánlás sajátosságainak hatása.* A válaszadó-vezérelt mintavétel feltételezése az, hogy a válaszadók az ismerőseik közül véletlenszerűen választanak. Gile és Handcock [2009] vizsgálták az ettől eltérő ajánlás hatásait. A szerzők azt találták, hogy (I) *adott változó esetén az ajánlás során valamely érték felé mutató torzítás azonos irányba torzítja a becsléseket* (RDS II becslést vizsgálva, 20%-os ajánlási torzítás esetén a szimulált eredmények 2 százalékpont körüli torzítást mutattak az ismert design esetén). Goel és Salganik [2009] cikkükben szimuláció segítségével bemutatták, hogy az ajánlónként kiválasztott további mintatagok száma a vizsgált kutatás ajánlónként 1-3 ajánlottja mellett 3,7-szeresére növelte a szimulált varianciát az 500 fős mintán egy hasonló méretű egyszerű véletlen mintához képest, vagyis (II) *az ajánlónként nagyobb ajánlott szám jelentősen növelheti a becslések szórását* (ezt Lu és munkatársai [2012b] is alátámasztották). Guntuboyina és szerzőtársai [2012] szimulációjuk során (III) *a feltételként szabott ajánlónkénti egy (vagy legalább azonos számú) ajánlott feltételének megsértése esetén (ha az eltérő hálózati struktúrával párosul) a várt fokszám arányos bekerülési valószínűségtől való jelentős eltérésről számoltak be.* Ezt megerősítik Lu és munkatársai [2012b] eredményei, amelyek szerint (IV) *a preferenciális ajánlás (a közeli ismerősök nagyobb valószínűségű ajánlása) mind az RDS II becslés szórását, mind torzítását növeli.* Ugyanakkor mindennek ellentmondani látszanak a kutatás azon eredményei, amelyek szerint (V) *ha az ajánló bizonyos valószínűséggel „elfelejti” egyes kapcsolatait (azaz kihagyja az ajánlottak közül), az még akkor sincs jelentős hatással a becslésre, ha összefügg a vizsgált változóval.* Nesterko és Blitzstein [2015] eredményei szerint (VI) *a megkérdozettenkénti (2-6 közötti, de megkérdozettenként azonos) ajánlás szám nem befolyásolta az RDS II becslést.* Verdery és munkatársai [2015] azt találták, hogy (VIIa) *amennyiben az ajánlás nem memóriamentes (azaz a mintavétel folyamata során keletkező láncok nem tekinthetők elsőfajú Markov láncnak), az*

az RDS II becsléshez tartozó hibabecslésnél jelentősen nagyobb szórást eredményez. A szerzők egy egészségügyi kutatás (Add Health) és a Facebook 100 alhálóját vizsgálva azt találták, hogy (VIIb) *a vizsgált változók és hálózatok 90%-a esetén a kapott láncok nem elsőfajú Markov láncok.*

(5) *Fokszám mérés hatása.* Amint a fent 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 következő módokon vizsgálták a fokszámot (2004-ben és 2008-ban némileg eltérő módon): hány embert vett fel a válaszadó számítógépes csevegő programjába⁴⁹, hány embert tudnának ajánlani a kutatáshoz (2008-ban hány emberrel beszélnek meg fontos dolgokat), az elmúlt 30 napban (2008-ban az elmúlt 14 napban) hány emberrel leveleztek e-mailen, a válaszadónak hány ismerőse, hány barátja és hány közeli barátja van (2008-ban a mobiltelefonban szereplő ismerősök száma is szerepelt). Az eredmények azt mutatták, hogy (I) *a becslések nem tértek el jelentősen a különböző fokszám adatok felhasználásával, de kijelenthető, hogy (II) annak a fokszám adatnak a felhasználása adja a sokaságra vonatkozó legjobb becslést, amely a legjobban köthető az ajánláshoz, azaz elsősorban az, hogy hány embert tudna ajánlani, másodsorban az, hogy hány ismerőse van a megkérdezettnek a számítógépes csevegő programjában (általában így egyeztettek a válaszadók, mielőtt ajánlottak volna valakit).* Nesterko és Blitzstein [2015] kutatása alátámasztja, hogy a valós fokszámtól való eltérés nem feltétlenül rontja a becslés tulajdonságait. Eredményeik szerint (III) *a fokszám véletlenszerű eltérése (amennyiben az Poisson eloszlást követ) az RDS II becslést nem befolyásolja.* Ugyanakkor a fokszám mérésében bekövetkező szisztematikus torzítás már jelentős hatással bírhat. Mills és munkatársai [2014] kutatása szerint (IV) *a fokszám kerekítése (legközelebbi 5-re vagy 10-re) jelentősen növelheti a z RDS II becslés szórását és torzítását (a pontos mérés különösen alacsony fokszám esetén fontos).* McLaughlin és munkatársai [2015] a leírt bizonytalanságok, illetve a fokszámra vonatkozó adathiány kezelésére tesznek javaslatot a mért hálózati tulajdonságok és a riportált fokszám alapján maximum likelihood becslés elvégzésével helyettesítve a tényleges, illetve hiányzó fokszám adatot.

(6) *A minta paramétereinek (mintaméret, visszatevés, egyensúlyi állapoton kívüli esetek) hatásai.* Wejnert [2009] kimutatta, hogy (I) *kis mintákra az RDS I becslés bootstrap szórásbecslése alulbecsül – így a becslés gyakran a konfidenciaintervallumon kívül esik, ezzel*

⁴⁹ Pl: hány ismerőse van MSN vagy Skype, stb. programban.

szemben az RDS II becslés egzakt szórásbecslése felülbecsül (ami az intervallum-becslés pontatlanságát indokolatlanul növeli, illetve szükségtelenül nagy méretű mintát igényel). Gile és Handcock [2009] azt találta, hogy (IIa) amennyiben a minta mérete közelít a célcsoport méretéhez (ami nagyon kis méretű vizsgált célcsoportok esetén fordulhat elő), a vizsgált változó egyes értékeihez tartozó eltérő átlagos fokszám torzító hatás jelentős mértékben megnő⁵⁰. Ugyanakkor Nesterko és Blitzstein [2015] eredményei szerint (IIb) amennyiben a minta-populáció arány alacsonyabb (10%-33%) változása nem befolyásolja jelentősen az RDS II becslést. Fontos tényező (különösen az RDS II becslés esetén), hogy a minta mikor éri el az egyensúlyi állapotot. Egyes kutatások esetén az egyensúlyi állapot elérése előtti mintát elhagyták az elemzésnél. Wejnert [2009] azt találta, hogy (III) az egyensúlyi állapot elérése előtti adatok elhagyása nem javította a becslések pontosságát, viszont növelte a becslés szórását, ezért a szerző nem javasolja az egyensúlyi állapot elérése előtti adatok elhagyását (ezt megerősíti Gile és Handcock [2009] is). A válaszadó-vezérelt mintavétel modelljének feltételezése, hogy a mintavétel visszatevéses. Gile és Handcock [2009] vizsgálta, hogy milyen hatást okoz a gyakorlatban alkalmazott visszatevés nélküli mintavétel a becslések esetén. A szerzők azt találták, hogy (IV) a visszatevés nélküli mintavétel csökkenti a becslések szórását és így a design hatást (ezen túl a visszatevéses mintavétel esetén gyorsabban csökken a kiinduló minta torzító hatása is). Ezt az eredményt saját szimulációjuk során Lu és munkatársai [2012b] is megerősítették.

(7) *Válaszmehtagadás hatása.* Gile és Handcock [2009] vizsgálta, hogy a változó egyes értékeihez tartozó eltérő válaszmehtagadási arányok hogyan befolyásolják a becsléseket. A szerzők azt tapasztalták, hogy (I) a változó értékei szerint eltérő arányú válaszmehtagadás az alacsonyabb válaszmehtagadással jellemezhető érték felülbecsléséhez vezet (nincs szignifikáns különbség az RDS I és RDS II becslések között, azonban mindkettő hatékonyabb, mint a naiv becslés). Lu és munkatársai [2012b] valós hálózaton végzett szimulációjuk során azt találták, hogy (II) a válaszmehtagadás hatása mind a torzításra, mind a becslések szórására csekély abban az esetben is, ha a válaszmehtagadás valószínűsége összefügg a vizsgált változóval. Ugyanakkor kutatásuk azt mutatja, hogy (III) ha a válaszmehtagadás és az ajánló részéről történő negligálás összefügg, az a becslés jelentős hibájához vezethet.

(8) *Eltérő becslések összehasonlítása.* Tomas és Gile [2010] több becslés hatékonyságát is vizsgálja. Ezek közül azonban az úgynevezett naiv becslésre ismertetem a cikk eredményeit

⁵⁰ Ez a populációs kapcsolatháló éleinek egyenletestől eltérő valószínűségű mintába kerülési valószínűségével magyarázható (Ott és Gile [2016]).

(naiv becslés alatt azt a becslést 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 terjedelmi okok miatt nem foglalkozom. A szerzők szimulációjának eredményei azt mutatták, hogy (I) *egyik becslési eljárás sem jobb minden vizsgált körülmény esetén a másiktól*. Wejnert [2009] az RDS I és RDS II becsléseket hasonlítja össze. Az összehasonlítás alapján elmondható, hogy (II) *a két becslési módszer becslései nem térnek el jelentősen egymástól*. Gile és Handcock [2009] azt mutatta ki, hogy (III) *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 (IV) *rövid láncok, a kiinduló-pontok jelentős torzítása és az ajánlás torzítása (lásd (4)) esetén az RDS I becslés átlagos hibája kisebb*. Wejnert és munkatársai valós kutatások esetén azt találták, hogy a minták becsült dizájn hatása a korábban várt 2 helyett 3 körül mozgott (az interkvartilis terjedelem a változók többsége esetén 2-4 közé esett). Eredményük úgy interpretálható, hogy (V) *az alkalmazott módszertan mellett az egyszerű véletlen mintavétellel összevetve mintegy háromszoros mintára van szükség azonos megbízhatóság eléréséhez, másként fogalmazva egy RDS minta becslésének varianciája háromszor akkora, mint egy azonos méretű egyszerű véletlen mintáé*. Nesterko és Blitzstein (2015) eredményei szerint (VI) *homofil hálózatok esetén az egyszerű átlag mind torzítás, mind a becslés varianciája kisebb, mint az RDS II becslés, ugyanakkor (xx) foksám alapú preferenciális kapcsolódás alapján létrehozott hálózatban⁵¹ az RDS II becslés minden reális beállítás mellett jobban teljesít, mint az egyszerű átlag*.

(9) *Eltérő mintavételi módszerekkel vett minták összevetése az RDS módszerre vett mintákkal*. A két idézett cikk (Kendall és mtsai. [2008], Kral és mtsai. [2010]) egyaránt megerősítik az RDS mintavétel használhatóságát. Kendall és munkatársai hely-idő, illetve hólabda mintával összevetve (I) *jelentős eltérést találtak az RDS minta esetén gazdasági, társadalmi státusz eloszlása tekintetében, ugyanakkor (II) az RDS minta esetén kapott eredmények jobban közelítették a vizsgált város teljes népességének eloszlását⁵² a gazdasági, társadalmi státusz tekintetében, mint a hely-idő, illetve a hólabda minta*. Kral és munkatársai (III) *a célzott mintavétellel vett mintával összevetve az RDS mintavétellel vett mintát a legtöbb vizsgált változóra tett intervallumbecsléseket jelentős mértékben átfedőnek találták, kivételt képezett ez*

⁵¹ A hálózati modell úgy épül fel, hogy az új kapcsolat kialakulásának valószínűsége a már meglévő kapcsolatok számával arányos.

⁵² A vizsgált célcsoport (férfiakkal szexuális kapcsolatot részesítő férfiak) valós gazdasági-társadalmi jellemzői természetesen nem voltak ismertek.

alól az etnikai megoszlás (afroamerikaiak megoszlása), illetve egy többségében zárt közösséget alkotó etnikai csoport (afroamerikaiak) által lakott postai körzet eloszlása. Utóbbi eltérés alapján Kral és munkatársai azt javasolják, hogy (IV) *az RDS mintavétel esetén a kezdőpontok kiválasztását egyéb mintavételi technikák segítségével tegyék heterogénebbé.*

IV.2. A válaszadó-vezérelt mintavétel egyes feltételeinek vizsgálata és kezelésük

A válaszadó-vezérelt mintavétel tulajdonságait vizsgáló cikkek összefoglalását követően kitérek néhány további kutatásra is, amelyek a válaszadó-vezérelt mintavétel egyes feltételei (lásd II.3 fejezet) megsértése esetén további lehetőségeket jelentenek a módszer módosított kivitelezésére, illetve lehetőséget adnak a feltételek teljesülésének ellenőrzésére. Ezeket a tanulmányokat fenti tömör összefoglalásnál részletesebben idézem és kritikai észrevételekkel látom el.

Ahogy a II.3 fejezetben bemutattam, a válaszadó-vezérelt mintavételből történő becslés torzítatlansága számos feltétel teljesülése esetén lehetséges. A korábbi szakirodalom e feltételek megsértésének hatásait kiterjedten vizsgálta, de nem állított fel olyan indikátorrendszert, amely alkalmas lenne e feltételek teljesítésének, illetve megsértésének becslésére, tekintettel arra, hogy e feltételek a legtöbb esetben nem vizsgálhatók egyszerűen. Gile és munkatársai [2015] javaslatot tesznek az egyes feltételek teljesülésének ellenőrzésére használható eszközökre. Tekintettel a kérdés fontosságára, részletesen bemutatom a szerzők által javasolt indikátorokat, helyenként a szerzők javaslatait pontosítva, kiegészítve, kritikai értelmezésekkel ellátva.

A szerzők az egyes eszközöket egy Dominikai Köztársaság területén 12 különböző, HIV fertőzéssel veszélyeztetett közösségben párhuzamosan végzett adatfelvétel eredményén mutatják be. A válaszadó-vezérelt mintavétel szerzők által kiemelt alkalmazási feltételeit az általuk közölt tábla módosított változatával foglalom össze (11. táblázat).

11. táblázat A válaszadó-vezérelt mintavétel Voltz és Heckathorn [2008] által javasolt becslésének (saját hivatkozásomban RDS II becslés) feltételei

		Feltételek*
Matematikai-statisztikai kérdések	Mintavételi feltételek	<i>Visszatevéses mintavétel</i> (a hálózat mérete nagy a mintához képest ($N \gg n$))
		Egyetlen el nem ágazó lánc
		Megfelelő számú mintavételi hullám
Társadalomtudományi kérdések	Kapcsolathálózati feltételek	<i>Elegendően gyenge homofília</i>
		<i>Szűk keresztmetszetek limitáltak</i>
		<i>Kölcsönös kapcsoltok</i>
	Válaszadói magatartási feltételek	Összekapcsolt gráf
		<i>Pontosan mért fokszám</i> <i>Ajánlottak közül véletlen kiválasztás</i>

Forrás: Gile és munkatársai [2012]

A táblázatban szereplő feltételek megsértésének hatásait a fentiekben már bemutattam, illetve a IV.3. fejezetben egyes esetekben saját szimulációs kutatásommal is demonstrálom. A táblázatot a tanulmányban szereplőhöz képest átalakítottam. Ennek két oka volt. Egyrészt az eredeti táblázat kategorizálását nem tartottam konzekvensnek: a kategorizálás két dimenziója (véletlen séta modell, kiinduló minta hatásának eltűnése, válaszadói viselkedés / hálózati struktúra feltétel, mintavételi feltétel) valójában nem alkotott koherens dimenziót. Másrészt dolgozatomban egészének szellemében itt is jelezni kívántam, hogy melyek a módszer alkalmazhatóságának matematikai-statisztikai és társadalomtudományi kérdései.

A szerzők elsőként a visszatevéses mintavétel feltételének megsértését ellenőrző eszközökre tesznek javaslatot. Itt szükség van némi magyarázattal kiegészítenem a szerzők felvetését. A visszatevéses mintavétel feltételezése statisztikai szempontból szükséges a korábban leírt becslések torzítatlanságához, a gyakorlatban azonban a kutatások kivitelezése során soha nem történik visszatevéses mintavétel. A helyzet csak részben hasonló ahhoz, ahogyan a véletlen mintavétel esetén kezelhető ugyanez a probléma. A véletlen mintavétel esetén szükséges és elégséges, a válaszadó-vezérelt mintavételnél azonban szükséges, de nem elégséges feltétel az, hogy a vizsgált célcsoport (a populáció) mérete jelentősen haladja meg a minta méretét. A válaszadó-vezérelt mintavétel esetén ugyanis ezt a feltételt nem csak globálisan, hanem hálózati értelemben véve lokálisan is teljesíteni kell. Előfordulhat ugyanis, hogy a populáció jóval meghaladja a tervezett minta méretét, de a mintavétel során egy olyan csoport tagja kerül

kiválasztásra, amely csoporton belül sok kapcsolat található, azonban a csoportból csak kevés kapcsolat mutat kifelé. Ebben az esetben nagy az esélye annak, hogy egy-egy válaszadó ismerősei közül nagy számban már részt vettek a kutatásban, ami egyrészt torzításhoz, másrészt az adott mintavételi szál elakadásához vezethet (ezt a problémát röviden már a fejezet korábbi részében a hálózati struktúra torzító hatásaira vonatkozó kutatások ismertetésénél is érintettem, ott röviden hálózatelemzési kifejezéssel klaszterezettségnek neveztem). A válaszadó-vezérelt mintavétel tipikus célcsoportjai esetén sem a célcsoport teljes mérete, sem klaszterezettsége nem ismert, így fontos annak becslése, hogy vajon a probléma az adott kutatás során fellép-e vagy sem.

A szerzők három indikátort javasolnak a visszatevéses mintavétel feltétel teljesülésének ellenőrzésére: a tervezett mintaméret elérhetetlensége (a szálak elakadása miatt), az ismételt kiválasztás gyakorisága, illetve egy direkt kérdés arra vonatkozóan, hogy a mintába került személy célcsoportához tartozó ismerősei közül hányan vettek részt a kutatásban. Az első indikátor lényegében úgy értelmezhető, hogy amennyiben nem sikerül a tervezett mintaméretet elérni (a szálak elakadása miatt), akkor fennáll a kockázata annak, hogy – akár lokálisan, de több és heterogén kiindulópont esetén nagyobb valószínűséggel globálisan – megsérül a visszatevéses mintavétel elve. Ez az indikátor csak az adatfelvétel lezárultát követően értékelhető, így nem ad lehetőséget az adatfelvétel menetébe történő beavatkozásra, de lehetővé teszi alternatív becslési eljárás használatát (lásd később, Gile [2011]). A második indikátor úgy értelmezhető, hogy a válaszadók ismételt elérésének egy szálon mérhető vagy az összes szálon mérhető gyakorisága a vizsgált feltétel lokális, illetve a globális megsértését jelezhetik (a különválasztást a szerzők nem teszik meg, de logikusan következik az indikátor tulajdonságaiból). Végül a harmadik javasolt indikátor – amely a megkérdezettek egonetworkjében méri a kutatásba már bevontak arányát (az adatfelvételbe épített kérdéssel) – szálankénti vagy lépésenként bekövetkező növekedése az előbbi indikátorhoz hasonlóan a visszatevéses mintavétel feltételének lokális, illetve globális sérülését jelzi. A három indikátor közül az első kettő objektív mérésen (is) alapulhat, az utolsó mindenképpen a megkérdezett szubjektív percepcióján. Az első indikátor csak az adatfelvétel lezárultát követően értékelhető, míg a másik kettő folyamatosan. Az első indikátor csak a visszatevéses mintavétel feltételének globális sérülését jelezheti, a másik kettőből lokális következtetés is levonható.

A válaszadó-vezérelt mintavétel talán legjelentősebb korlátja, hogy csak abban az esetben juthatunk torzítatlan mintához, ha a célcsoport tagjai egyetlen komponenst alkotnak, azaz a célcsoport bármely tagjától a célcsoport bármely más tagjához eljuthatunk célcsoporttagokon

és közöttük lévő kapcsolatokon keresztül. Kuran és munkatársai [2011] egy online hálózatokból történő preferenciális mintavételről szóló tanulmányukban egy olyan eszközt mutatnak be, amely a válaszadó-vezérelt mintavétel esetén is alkalmazható, és amelynek segítségével az egy komponenshez tartozás kritériuma lazítható. A tanulmány az online hálózatokon (leginkább közösségi oldalakon) belül, bizonyos tulajdonságokkal rendelkező szereplők optimális mintavételét tűzi ki célul figyelembe véve azt is, hogy az adott tulajdonságok adott esetben ritkák. A probléma hasonlít az általunk tárgyalt problémához azzal a különbséggel, hogy az online hálózatok esetén a hálózat rögzített adatbázis, így a mintavétel technikai eszközökkel történhet és nem kell számolni a személyes adatfelvétel torzító tényezőivel (pl.: fókuszpontatlan becslése). A cikk szerzőinek javaslata a „séta a gráfon nagyítóval” nevet kapta. A módszer lényegében ötvözi a megnövelt kiválasztási valószínűséggel végzett rétegzett mintavétel és a gráfon végzett véletlen sétás mintavétel módszerét (a válaszadó-vezérelt mintavétel lényegében az utóbbi módosított változata). A mintavételi folyamat a lényegét kiemelve úgy zajlik, hogy a célcsoportnak kijelölt csoportok⁵³ a kapcsolathálózaton történő véletlen séta folyamán a kiválasztott szereplővel kapcsolatban álló csúcsok közül megemelt valószínűséggel kerülnek kiválasztásra, míg a célcsoporthoz nem tartozó szereplők alacsony valószínűséggel kerülnek kiválasztásra. A szerzők javaslata, hogy a megfelelő kiválasztási valószínűségek meghatározásához egy előzetes egyszerű véletlen sétás mintavételt érdemes végezni, amelyből megállapíthatók az egyes csoportok közötti átmenet valószínűségek (a csoportok tagjai közötti kapcsolatok száma), illetve a csoportok mérete. A válaszadó-vezérelt mintavétel egy komponensű hálózaton történő elvégezhetőségének problémájára lefordítva a kérdést érdemes a) végezni egy a teljes populációra⁵⁴ vonatkozó kismintás válaszadó-vezérelt mintavételt, majd b) ezek alapján meghatározni a populáció és a célcsoport közötti átmenet valószínűségeket, illetve a célcsoport relatív méretét⁵⁵ végül c) ezek ismeretében döntést hozni arról, hogy szükséges-e a célcsoporton kívüli személyeket is bevonni a kutatásba és mekkora valószínűséggel.

⁵³ A cikk feltételezése szerint több, egymáshoz viszonyítva egyenlőtlen méretű, de a teljes populációnak is csak kisebb részét kitevő csoportból történik a mintavétel.

⁵⁴ Populáció alatt értve azt a halmazt, amely teljes egészében magában foglalja a célcsoportot, és elvárásaink szerint egy komponensű képez.

⁵⁵ A becslésekhez a leírt becslő módszerek alkalmazhatók.

IV.3. A válaszadó-vezérelt mintavétel vizsgálata szimuláció segítségével⁵⁶

Egy saját korábbi kutatásomban különböző paraméterekkel rendelkező szimulált hálózatok, majd a hálózatokon a válaszadó-vezérelt mintavételi eljárás szimulálásán keresztül azt igyekeztem vizsgálni, milyen pontossággal lehet becsülni bizonyos sokasági paramétereket, különböző elméleti populáció esetében. Ezen felül azt is vizsgáltam, hogy a becslés egyes elemei milyen módon járulnak hozzá a becslések szórásához, amely a módszer használhatósága szempontjából döntő fontosságú. A kutatás célja az volt, hogy pontosítsa a módszer felhasználási feltételeit és alapot teremtsen a fent bemutatott problémák kiküszöböléséhez.

IV.3.1. Korábbi szimulációs kutatások összefoglalása

A válaszadó-vezérelt mintavétel megbízhatóságának és érvényességének vizsgálatával számos korábbi kutatás foglalkozott, melyeket korábban összefoglaltam, itt azokat az – elsősorban szimulációs – kutatásokat idézem, amelyek szimulációs módszertant alkalmaztak és az elemzés megértése szempontjából lényegesek.

Salganik és Heckathorn [2004] szimulációs vizsgálatának részletes bemutatását azért tartom fontosnak, mivel saját szimulációm is ezen alapul, de röviden kitérek más – az adott kutatási kérdéssel foglalkozó – elemzésekre is, amelyek ismertetése során elsősorban az eredményekre koncentrálok, amelyeket a fejezet utolsó részében összevetek saját eredményeimmel.

Salganik és Heckathorn [2004] a fent már idézett becslés mellett az általuk javasolt (RDS I) becslés tulajdonságait, elsősorban torzítatlanságát szimuláció segítségével is vizsgálta. Vizsgálatukban első lépésként létrehozták a szimulációban a célpopulációs szerepét betöltő alappopulációt. Az alappopulációt két részre osztották: A tulajdonsággal, illetve B tulajdonsággal rendelkező alappopulációra (ez megfeleltethető egy változó két értékének). Definiálták egy meghatározott eloszlás alapján, hogy mekkora legyen az egyes populáció tagok kapcsolatszáma. Definiálták, hogy mekkora legyen a kapcsolatok aránya a két csoport tagjai között a kisebb összkapcsolatszámmal rendelkező csoport összes kapcsolatához képest, ezt nevezzük a továbbiakban heterofiliának. Tehát a heterofília az A és B tulajdonságú csoportok összekötöttségét méri. Értéke 0, ha nincs kapcsolat a két csoport között, és 1, ha a kisebb összkapcsolatszámmal rendelkező csoport összes kapcsolata a nagyobb csoport felé irányul. Az alappopuláció mérete 10 000 volt Salganik és Heckathorn szimulációjában. Mind a két csoport

⁵⁶ A fejezet a Szociológiai szemle 2013/2 számában megjelent tanulmány (Kmetty-Simon [2013]) átdolgozásával készült. A felhasznált szöveg a szerző munkája.

esetén a kapcsolatszám exponenciális eloszlású volt, az A csoport esetén 10, a B csoport esetén 20-as foksám átlaggal. A heterofiliát pedig 0,6-ra állították be, tehát a kisebb kapcsolatszámú csoport kapcsolatainak 60 százaléka irányult a nagyobb csoport irányába.

A hálózat kialakítása után szimulációs módszerrel vizsgálták az RDS mintavételt, 5 kezdőpontból kiindulva. Minden lépésnél a már mintába került személyekhez a velük kapcsolatban állók közül két-két további populációtagot választottak ki véletlenszerűen, mellyel bővítették a mintát. Az eljárást addig folytatták, amíg el nem érték a kívánt mintaméretet. Az eljárás során visszatevéses mintavételt alkalmaztak, azaz egy-egy populációtag többször is a mintába kerülhetett. A mintavételt 1000-szer ismételték meg minden beállítás mellett. A szimulációk során elsősorban a torzítás mértékét (azaz a populációs és a mintából becsült arány különbségét) vizsgálták egyrészt a mintaméret, másrészt a populációs arányok, harmadrészt a kezdőpont kiválasztás szempontjai alapján. A szerzők azt találták, hogy a torzítás mértéke már 500 elemű minta esetén bármely populációs arányra 0,1%-nál kisebb. A kezdőpontok kiválasztása esetén azt találták, hogy egyszerű véletlen kezdőpont kiválasztás, kapcsolatszám arányos valószínűséggel történt kezdőpont kiválasztás, illetve a kapcsolatszám négyzetével arányos valószínűségű kezdőpont kiválasztás esetén is 500 elemű mintára a becslések torzítása 0,1% alatt marad. A cikk ábrái a becslések szórását is tartalmazzák, azonban ezeket sajnálatos módon a szerzők nem számszerűsítik.

A továbbiakban még négy tanulmányt ismertettek részletesebben, amelyek eltérő adatokon vizsgálják a válaszadó-vezérelt mintavétel tulajdonságait⁵⁷. A tanulmányok ismertetése során három tényező szerepére koncentrálok a becslések szórására és a torzításra vonatkozóan: a heterofília hatására, a hálózati tulajdonságokra és a mintavétel sajátosságaira.

Gile és Handcock (2009) 1000 modellezett hálózatot vizsgált, amelyek egyenként 1000 fős populációt szimuláltak (a nem változtatott hálózati paramétereket egy korábbi kutatás valós hálózati adatai alapján állították be, a szimulált hálózatok a fent leírthoz hasonló módon kerültek előállításra exponenciális foksám eloszlással és változó heterofília mértékkel). Az 1000 darab 1000 elemű hálózatból 500 elemű mintákat vettek. Szimulációjuk során azt találták, hogy nagyon alacsony heterofília esetén (tehát ha a kapcsolatok nagy része a vizsgált változó értékei által képzett csoportokon belülre mutat) az RDS II becslések szórása jelentősen megnő. Ugyanakkor azt is kimutatták, hogy torz (a populációs eloszlástól jelentősen eltérő) kiinduló

⁵⁷A cikkek eredményei a módszertani kutatásokat összefoglaló **Hiba! A hivatkozási forrás nem található.** fejezetben is szerepeltek, itt azonban részletesben ismertetem a kutatásokat

minta esetén a hosszabb láncok (kevesebb kezdőpont) kisebb torzításhoz vezetnek, összességben tehát a torz kiinduló minta nincs jelentős hatással az RDS II becslés szórására. E mellett a szerzők azt is vizsgálták, hogy milyen hatása van annak, ha a vizsgált változó egyes értékei által kijelölt csoportok átlagos kapcsolatszámuk eltér egymástól (például, ha nemek szerinti megoszlásra vagyunk kíváncsiak és a nők átlagos kapcsolatszámuk jelentősen eltér a férfiakétól). A szerzők azt találták, hogy az átlagos kapcsolatszámokban mutatkozó különbségek a nagyobb átlagos kapcsolatszámúval jellemezhető csoport értéke gyakoriságának alulbecsléséhez vezet. Ez a hatás még jelentősebb, ha a gyakoriságok nagymértékben eltérnek az egyenletestől. Azt tapasztalták továbbá, hogy amennyiben a minta mérete közelíti a populáció méretéhez, a vizsgált változó egyes értékeihez tartozó eltérő átlagos kapcsolatszám torzító hatása jelentős mértékben megnő. A szerzők végül azt is megállapították, hogy a visszatevés nélküli mintavétel csökkenti a becslések (mind az RDS I, mind az RDS II) szórását (valamint a kiinduló minta torzító hatását).

Goela és Salganik (2010) egy valós nagymintás kutatás hálózatának 4430 fős részhálóját használták (ezen belül 85 különálló komponens⁵⁸), amelyen az RDS I becslés feltételeinek megfelelő 500 fős mintákat szimuláltak. A mintavételt 10 ezer alkalommal ismételték meg. Eredményeik azt mutatták, hogy a becslések szórása számottevően meghaladta a várt mértéket, mértéke egy egyszerű véletlen mintával összevetve 5-10-szeres is lehet. Ez azt jelenti, hogy egy 1000 fős RDS mintához számolt becslési hiba akár olyan nagy is lehet, mint egy 100 fős véletlen minta esetében.

Tomas és Gile (2010) 1000 fős szimulált hálózattal modelleztek egy populációt (a hálózat tulajdonságait úgy paraméterezték, hogy az egy korábbi kutatás valós paramétereinek, például átlagos fokszámának feleljen meg). Vizsgálatuk szerint a célcsoporthoz viszonyítva kis minta és alacsony heterofília esetén az RDS I becslés kisebb torzítást eredményez a becslések nagyobb szórása mellett, míg nagyobb heterofília mellett, az RDS II becslések szórása kisebb.

Verdery és munkatársai (2015) a Facebook 215, egyetemalapú ismeretségi hálóján modellezett mintavétellel vizsgálták a válaszadó-vezérelt mintavétel tulajdonságait. A kutatás során 10 ezer mintavételt modelleztek. Elemzésük során kimutatták, hogy azon változók esetén, amelyek jelentős klaszterezettséghez vezetnek (azaz a hálózaton belül olyan alcsoportok vannak,

⁵⁸ A komponens a hálózatkutatás és gráfelmélet szóhasználatában olyan részhálózatot (részgráfot) jelent, amelyen belül bármely részhálózattagtól el lehet jutni bármely más részhálózattaghoz kapcsolatok és hálózattagok egymásutánján keresztül, de egyetlen további hálózattaggal bővítve sem teljesülne ez a feltétel (a feltételre nézve maximális a részgráf).

amelyekre jellemző a nagyszámú alcsoporton belüli és kevés az azon kívüli kapcsolat, valamint ezen alcsoportok esetén a vizsgált változó szóródása kisebb) a becslések szórása jelentősen megnövekedett. Ez gyakorlatilag megegyezik azzal a korábbi eredménnyel, miszerint az alacsony heterofília megnöveli a becsléseinkhez tartozó konfidencia intervallumot.

Összefoglalva a leírtak fontosabb megállapításait, a korábbi kutatások azt állapították meg, hogy a heterofília mértéke fordított összefüggést mutat a becslések szórásával (azaz a megbízhatósággal), az egyes csoportok átlagos kapcsolatszámuk közötti különbség ugyancsak növeli a becslések szórását, mindezek mellett a kiinduló minta nagysága (amely adott mintanagyság mellett a láncok hosszát is megszabja) a becslések szórását nem befolyásolta jelentősen.

IV.3.2. A szimulációs vizsgálat kutatási kérdése

A vonatkozó kutatásokat áttekintve nem találtam olyat, amely szisztematikusan vizsgálta volna a fontosabb tényezők együttes hatását. Ezért úgy döntöttem, hogy a válaszadó-vezérelt mintavétel két szakirodalomban legelterjedtebben használt becslésének (RDS I és RDS II) szóródását és torzítását vizsgálom különféle hálózati és populációs paraméterek kombinációja esetén. Emellett egy, a korábbi irodalomban egyáltalán nem szereplő tényezőt is vizsgállok, azt, hogy a változók populációs eloszlására adott becslések szórásának mekkora hányadát adják a becslés egyes tényezői, remélve, hogy így jobban megérthetjük, hogy az egyes paraméterek mekkora mértékben hatnak a becslések megbízhatóságára. A hálózati és mintavételi tényezők közül vizsgáltam az átlagos fokszámot, a heterofília nagyságának hatását, valamint a kiindulópontok számát. Mind a három paraméter esetében további három-három paraméter beállítását vizsgáltam. Az eredmények vizsgálatánál elsősorban az RDS I és RDS II becslések megbízhatóságára koncentráltam, de vizsgáltam az A és B tulajdonságú populációs tagok fokszám becslését, a teljes sokaságra adott fokszám becslést, valamint az A és B, illetve B és A csoportok közötti összekötöttség (heterofília) becslését is.

Mindezek alapján kutatási kérdéseimet a következő módon fogalmaztam meg: *hogyan befolyásolják a hálózati és mintavételi tényezők a becslések torzítását és szórását, a becslések szórását milyen mértékben határozzák meg a becslés egyes tényezői, valamint hatással vannak-e egyéb hálózati tényezők a becslések szórására, illetve torzítására?*

IV.3.3. A szimulációs kutatás módszertana

A kutatási kérdések megválaszolásához a szakirodalmi előzményekre támaszkodva egy hálózati szimulációs módszert választottam. A hálózatok szimulálásához Salganik és Heckathorn (2004) módszerének módosított változatát alkalmaztam. A szimulációk során fix paraméterként kezeltem a vizsgált változó eloszlását (dichotóm, 0,3-0,7 eloszlású, a továbbiakban A értékű és B értékű), a populáció méretét (10 000), valamint a minta méretét (1000). Ezek a paraméterek az összes futtatásban állandóak voltak. A szimulációk során a két hálózati és egy mintavételi paraméter esetén (átlagos fokszám, heterofília illetve kiindulópontok száma) három-három értéket határoztam meg. A különböző paraméter tényezők esetében több mintavételi szituációt lefuttattam, és megvizsgáltam minden esetben a kapott becslés érvényességét és megbízhatóságát, valamint azt, hogy az egyes tényezőknek mekkora a hatása becslések torzítására és szórására. A fokszám esetén a kisebbik csoport nagyobb átlagos fokszámát feltételeztem (ezzel Salganik és Heckathorn idézett modelljének feleltem meg). A kisebbik csoport esetén (A értékkel jellemezhető) a fokszámok elméleti várhatóértéke⁵⁹: 7,66; 12,11 és 21. A nagyobb csoport esetén (B értékkel jellemezhető) a fokszámok elméleti várhatóértéke: 4,33; 6,55; és 11. A heterofília tekintetében azt határoztam meg, hogy a kisebbik csoport (A értékkel jellemezhető) tagjainak összes kapcsolata milyen arányban irányul a nagyobbik csoport (B értékkel jellemezhető) felé. Az értékek a következők voltak: 15%; 30%; 60%⁶⁰. Végül a minta kiindulópontjainak száma tekintetében 20, 40 és 80 értékek mellett döntöttem. Ez utóbbi egyúttal befolyásolta a mintavétel során a mintamérethez szükséges egymást követő lépések számát is (lánc hossz): 20 kiindulópont esetén a várható átlagos lánc hossz 5,6, 40 kiindulópont esetén 4,6, míg 80 kiindulópont esetén 3,6. Mivel a mintavétel szimulálása során – az idézett szerzőktől eltérően – a gyakorlatban alkalmazott visszatevés nélküli modellt alkalmaztam, az itt megadott átlagos lánc hosszok kis mértékben eltérnek a tapasztalttól. A szimuláció során minden paraméter beállítással 10 hálózatot generáltam és minden hálózat esetén 1000 mintavételt szimuláltam, összesen 270 hálózaton 270 000, 1000 elemű mintavétel történt.

Az eredményeket paraméter kombinációnként hasonlítottam össze egymással, illetve egy azonos paraméterű egyszerű véletlen minta (EVM) várható adataival. Minden minta esetén

⁵⁹ A fokszámokat exponenciális eloszlásúnak feltételeztük, technikai okokból – a hálózat összefüggését garantálendő – minden populációtag esetén egyel megnöveltem, így nulla fokszám nem fordulhatott elő.

⁶⁰ Tekintettel arra, hogy a nagyobb csoport a populáció 70%-át teszi ki, átlagos fokszámuk viszont a kisebb csoport valamivel több, mint fele (52-56%-a), a 15%-os arány viszonylag erőteljes homofiliát, a 30%-os arány véletlenhez közeli kapcsolatválasztást, míg a 60%-os arány kifejezett heterofiliát jelent.

kiszámítottam az RDS I és RDS II becslést, a becslések hibájának szórását, valamint kiszámítottam az RDS I becslést a valós heterofília értékekkel és a becült fokszámokkal, illetve a becült heterofília értékekkel és a valós fokszámokkal. Az így kapott becslések segítségével paraméter kombinációként meghatároztam, hogy milyen mértékben járul hozzá a becslések szórásához a fokszám, illetve a heterofília. Az eltérő hálózat szimulációk hatását azonos paraméterek mellett a becslések populációs értéktől való eltérésére vonatkozóan ANOVA elemzéssel vizsgáltam.

IV.3.4. A szimulációs kutatás eredményei

Az alábbiakban a szimulációk eredményeit⁶¹ foglalom össze a kutatási kérdéseket követve.

Elsőként az RDS I és RDS II becslések torzítását vizsgálom a különböző paraméterek mentén. Az egyes paraméterek független (kontrollált) hatásának vizsgálata érdekében egy Általános Lineáris Modellt (GLM) alkalmaztam⁶². Modellemben a paraméter változók alacsony mérési szintűként szerepeltek, így az elemzés tulajdonképpen több utas, többszintű ANOVA elemzésnek feleltethető meg.

A modell alapján elmondható, hogy mindhárom paraméter és a hálózat szimuláció is szignifikáns hatással volt mindkét becslés esetén. A hatáserőségek értékelése érdekében megvizsgáltam a modell alapján becsülhető marginálisokat az egyes paraméterek esetén (**Hiba! A hivatkozási forrás nem található.**), amely megmutatja, hogy az egyes paraméterek, tehát a heterofília, fokszám és kiinduló minta méret külön-külön milyen hatást gyakorol a becslések eltérésére a tényleges populációs értékektől.

⁶¹ A hálózatok mérete a szimulációs sajátosságai miatt esetenként némileg kisebb az eredeti méretnél. A fokszám átlagok jól követik az előre rögzített értékeket.

⁶² Random faktorként figyelembe vettük azt is, hogy az egyes paraméter beállítások mentén 10-10 hálózatot generáltunk, tehát összesen 270 szimulált hálózatból vettük a hálózatonkénti 1000 darab mintát.

12. táblázat A heterofília, foksám és kiinduló mintanagyság becsült hatása (a modell alapján kiszámított marginális átlagok és standard hibájuk)

Heterofília	RDS I		RDS II	
	Átlag (torzítás)	Std hiba	Átlag (torzítás)	Std hiba
15%	0,018	<0,001	-0,007	<0,001
30%	0,014	<0,001	0,003	<0,001
60%	0,006	<0,001	0,007	<0,001

Foksám	RDS I		RDS II	
	Átlag (torzítás)	Std hiba	Átlag (torzítás)	Std hiba
21	0,014	<0,001	0,004	<0,001
12,11	0,013	<0,001	0,002	<0,001
7,66	0,012	<0,001	-0,003	<0,001

Kiinduló minta	RDS I		RDS II	
	Átlag (torzítás)	Std hiba	Átlag (torzítás)	Std hiba
20	0,013	<0,001	<0,001	<0,001
40	0,013	<0,001	0,001	<0,001
80	0,013	<0,001	0,002	<0,001

Forrás: Kmetty-Simon [2013]

A tábla adatai alapján látható, hogy mindkét becslés elsősorban a csoportok összekötöttségére, a heterofíliára érzékeny. Míg azonban az RDS I becslés a heterofília növekedésével szinte lineárisan egyre kisebb torzítású becslést ad, az RDS II becslés a közepes (véletlenhez közel álló) heterofília esetén adja a legpontosabb becslést. Kisebb mértékben, de mindkét becslésre hatással van a foksám várható értéke (a táblázatban a kisebb csoport foksámának beállított várható értékeit közöljük). Ebben az esetben is hasonló jelenséget tapasztalhatunk, mint a heterofília esetén: az RDS I becslés az egyre kisebb foksám várható értékű hálózatok esetén egyre pontosabb becslést ad, az RDS II viszont valahol a közepes tartományban adja a legpontosabb becslést. Végül a kiinduló minta nagyságának hatása – bár statisztikai értelemben szignifikáns – mégis gyakorlati szempontból nem tekinthető jelentősnek. A hálózat hatás hasonló módon nem értékelhető, hiszen itt random hatásról van szó. A megadott paraméterekkel véletlenszerűen szimulált hálózatok közötti eltérés magyarázóerejét mutató parciális eta-négyzet értéke RDS I becslés esetén 0,002, míg RDS II becslés esetén 0,008, ami a hálózatok – paraméteresen nem kontrollált – struktúrájának szignifikáns, de nagyon kis mértékű hatását mutatja. Végül nem hagyható figyelmen kívül az a tény, hogy az RDS I becslés torzítása kedvezőtlen paraméterezés esetén közel egy nagyságrenddel nagyobb, mint az RDS II becslésé. Összefoglalva az RDS I becslés, különösen kedvezőtlen paraméterek esetén jelentősebb torzítással bír, egyértelműen csökken a torzítása a heterofília növekedésével és kisebb

mértékben a fokszám csökkenésével, ezzel szemben az RDS II becslés általában kisebb torzítással bír, de mind a fokszám, mind a heterofília tekintetében az általam közepesnek tekintett értékektől való eltérés a torzítást növeli.

13. táblázat A heterofília (H), a fokszám várhatóérték ($E(D)$) és a kiinduló minta méretének (Nk) hatása az RDS I és RDS II becslés szórására (Levene-teszt eredményei alapján mindhárom tényező mindkét becslési módszer esetén szignifikáns eltérést hozott létre a becslések szórásában)

H	E(D)	Nk	A becslések szórása	
			RDS I	RDS II
15%	21	20	0,0352	0,0442
		40	0,0353	0,0391
		80	0,0367	0,0344
	12,11	20	0,0349	0,0436
		40	0,0350	0,0385
		80	0,0359	0,0337
	7,66	20	0,0345	0,0433
		40	0,0354	0,0381
		80	0,0362	0,0334
30%	21	20	0,0256	0,0269
		40	0,0259	0,0266
		80	0,0263	0,0257
	12,11	20	0,0246	0,0261
		40	0,0250	0,0253
		80	0,0257	0,0248
	7,66	20	0,0241	0,0255
		40	0,0243	0,0247
		80	0,0251	0,0238
60%	21	20	0,0201	0,0200
		40	0,0201	0,0200
		80	0,0201	0,0199
	12,11	20	0,0184	0,0183
		40	0,0185	0,0184
		80	0,0187	0,0184
	7,66	20	0,0171	0,0170
		40	0,0175	0,0173
		80	0,0176	0,0173

Forrás: Kmetty-Simon [2013]

A következő lépésben a becslések megbízhatóságát vizsgálom meg. Sajnos ebben az esetben nincs lehetőség komplex modell alkalmazására, így egyszerű szórás egyenlőség vizsgálatokra, illetve a paraméterkombinációk alkalmazása mellett a becslések szórás adataira

hagyatkozhattam. A három vizsgált paraméter hatását a becslések szórására variancia-homogenitás vizsgálattal teszteltem (13. táblázat).

A Levene tesztek adatai alapján mindhárom vizsgált paraméter szignifikánsan befolyásolta mindkét becslés szórását. Az RDS I becslés szórására a legjelentősebb hatást a heterofília gyakorolta, a nagyobb heterofília erősen csökkentette a becslések szórását (15% esetén 0,0345-0,0367 között, míg 60% esetén 0,0171 és 0,0201 között). A foksám és a kiinduló minta méretének hatása ennél jóval gyengébb volt: az alacsonyabb foksám várható érték és az alacsonyabb kiinduló mintaméret egy nagyságrenddel kisebb mértékben csökkentette az RDS I becslés szórását. Az RDS II becslés esetén is a heterofília hatása volt a legerősebb a becslés megbízhatóságára: a nagyobb heterofília kisebb szórással járt (15% esetén 0,0334-0,0442 között, 60% esetén 0,0173-0,0200 között). Ebben az esetben azonban a kiinduló minta mérete valamivel nagyobb hatást gyakorolt a szórásra: nagyobb kiinduló minta kisebb szórást eredményezett. E mellett csak gyenge hatása volt a foksám várható értékének: azonos körülmények között a kisebb foksám várható érték az RDS II becslés valamivel kisebb szórását eredményezte. A két becslés szórását összevetve azt láthatjuk, hogy az optimálistól eltérő körülmények esetén az RDS I becslés szórása alacsonyabb. Érdekes a tapasztalt szórás értékeket összevetni egy egyszerű véletlen minta esetén hasonló populációs eloszlású változóra adott becslés szórásával. Ez az érték 1000 elemű egyszerű véletlen minta és 0,3-0,7 eloszlású változó esetén 0,0145, amely értéket a legkedvezőbb paraméter értékek esetén (nagy összekötöttség, alacsony várható foksám) mindkét becslés szórása megközelíti. Ugyanakkor a legkedvezőtlenebb vizsgált paraméter beállítás esetén az RDS I becslés szórása 2,5 szerese, míg az RDS II becslés szórása valamivel több, mint 3-szorosa az azonos méretű EVM mintából származó becslésének. Tehát egy ugyanakkora EVM és RDS minta esetében, utóbbinál akár háromszor szélesebb konfidencia intervallum tartozhat egy változó várható értéknek becsléshez.

A hálózati szimuláció hatásának tesztelése érdekében a szóráshomogenitás vizsgálatot minden egyes paraméter beállítás mellett elvégeztem a 10-10 hálózatra. Az eredmények azt mutatták, hogy az eltérő hálózatok általában nem okoztak szignifikáns eltérést a becslések szórásában⁶³.

Végül a több hálózati paraméter (foksám és heterofília) alapján becslő RDS I esetén megvizsgáltam, hogy a becslés eltérései milyen mértékben függenek össze az egyes hálózati

⁶³ A 27 féle paraméter kombinációból az RDS I becslés szórása 4 esetben, míg az RDS II becslés szórása 5 esetben volt szignifikánsan különböző (5%-os elsőfajú hiba mellett) hálózatonként.

paraméterbecslések eltéréseitől. Ennek érdekében az RDS I becslést úgy is kiszámítottam, hogy ahhoz az egyik esetben a populációs heterofília adatokat és a foksám becsléseket (a megmaradt eltérések a foksámbecslések hatását tükrözik), a másik esetben a populációs foksám adatokat és a heterofíliára vonatkozó becsléseket használtam (a megmaradó hiba az összekötöttség becsléséből származik). A becslések hibájának átlagát tekintve megállapítható, hogy a heterofíliára vonatkozó becslések hibája okozza az RDS I becslések konzisztens pozitív torzítását: a populációs foksám átlag használatával a becslések átlagos eltérése minden paraméter beállítás mellett növekedett, míg a populációs heterofília adatok használatával negatívvá vált, abszolút értéke pedig általában lényegesen csökkent. A becslések szórása esetén az egyes paraméter beállításokra kiszámítottam a populációs foksám átlaggal, illetve populációs heterofília adattal adott becslések szórásnégyzeteinek összegét, amelyek jól közelítették az eredeti becslések szórásnégyzetét. Ezt az összefüggést használva a továbbiakban az egyes hálózati paraméterbecslések által hordozott hiba hányad a szórásnégyzet hányadával írható le. A szórásnégyzet hányadok és az egyes vizsgált paraméterek (foksám, heterofília, kiinduló minta mérete) hatását vizsgálva azt találtam, hogy a kiinduló minta mérete nem gyakorolt jelentős hatást, így e paraméter mentén átlagoltam a szórásnégyzet hányadokat. **Hiba! A hivatkozási forrás nem található..** tábla a heterofília és a foksám várható érték mentén mutatja meg a foksám becslés részesedését az RDS I becslések hibájából.

14. táblázat A heterofília és a foksám várható érték hatása a foksám becslés hiba részesedésére az RDS I becslés varianciája esetén (a kezdőpontok számának hatását átlagolva)

Heterofília	Foksám várható érték		
	21	12,11	7,66
15%	0,120	0,167	0,224
30%	0,260	0,331	0,424
60%	0,515	0,604	0,678

Forrás: Kmetty-Simon [2013]

A 14. táblázat alapján elmondható, hogy mind a heterofília, mind a foksám várható érték befolyásolta a foksám becslés hibája által magyarázható hiba arányát az RDS I becslés varianciáján belül. Látható, hogy a heterofília hatása ebben az esetben is nagyobb a vizsgált paraméter összeállítás mentén. Nagyobb heterofília és kisebb foksám várhatóérték esetén a foksám becslés hibája kevesebbet magyaráz az RDS I becslés varianciájából, ennek megfelelően ezekben az esetekben a heterofíliára vonatkozó becslés hibája magyaráz jóval nagyobb arányt a becslések szórásnégyzetéből. Ha ehhez az eredményhez figyelembe vesszük az RDS I becslés szórásának paraméterfüggésénél leírtakat is, akkor elmondható, hogy azokban az esetekben, amikor az RDS I becslés szórása nagy (elsősorban alacsony heterofília,

másodsorban, de kisebb mértékben magas fokszám esetén), a becslés szórásáért jelentős részben a heterofília becslésének hibája felelős.

IV.3.5. A szimulációs kutatás konklúziója és további kutatási lehetőségek

A következőkben igyekszem összefoglalni a szimulációs kutatás eredményeit, azokat összevetni az ismertett irodalom eredményeivel.

Vizsgálatom első része az RDS I és RDS II becslés összehasonlítására vonatkozott különféle paraméterek mellett. Az **Hiba! A hivatkozási forrás nem található.** összefoglalja a korábban részletesen leírt eredményeket.

15. táblázat Az összekötöttség, fokszám, kiinduló minta és egyéb hálózati hatások befolyása az RDS I és RDS II becslés torzítására és szórására

	RDS I becslés		RDS II becslés	
	Torzítás	Szórás	Torzítás	Szórás
Az RDS I és RDS II egymáshoz viszonyítva	Nagyobb (főleg kedvezőtlen esetben)	Legtöbbször kisebb	Legtöbbször kisebb	Nagyobb (főleg kedvezőtlen esetben)
Nagyobb heterofília	Jelentősen csökkenti	Jelentősen csökkenti	A heterofília közepes értéknél megszűnik	Jelentősen csökkenti
Nagyobb átlagos fokszám	Kismértékben növeli	Kismértékben növeli	A heterofília közepes értéknél megszűnik	Kismértékben növeli
Nagyobb kiinduló minta	Lényegében nem befolyásolja	Kismértékben növeli	Lényegében nem befolyásolja	Csökkenti
Egyéb hálózati hatás	Szignifikáns, de kicsi	Nem szignifikáns	Szignifikáns, de kicsi	Nem szignifikáns

Forrás: Kmetty-Simon [2013]

Összevetve a két becslésre vonatkozó szimulációs eredményeket azt mondható ki, hogy mindkét becslés erősen érzékeny a heterofília hatásra, azaz arra, hogy a vizsgált változó értékei szerint létrehozható csoportok között milyen gyakoriságúak a kapcsolatok. Amennyiben a változó értékei mentén a kapcsolathálózat nagyon szegmentált, mindkét becslés szórása magas lesz, azonban míg az RDS I becslés torzítása is megnő ezzel szemben megbízhatósága valamivel alacsonyabb marad, mint a kisebb torzítású, de nagyobb szórású RDS II becslésé. A vizsgált célcsoport átlagos fokszáma mindkét becslést csak kis mértékben befolyásolja (mind torzítás, mind szórás tekintetében). A nagyobb kiinduló minta közepes mértékben képes csökkenteni az RDS II minta szórását, míg az RDS I becslésre alig van hatással – meg kell

azonban jegyezni, hogy szimulációnk során foksám arányos véletlen kiindulópontokat használtunk, amit a gyakorlatban nem lehet pontosan reprodukálni (bár az idézett szerzők, elsősorban Heckathorn érvelnek amellett, hogy a kezdőpont keresésnél a nagyobb foksámú célcsoporttagok könnyebben fellelhetők, így kvázi megvalósulhat a foksám arányos kezdőpont kiválasztás). Végül úgy tűnik, hogy legalább is a jelen hálózati szimulációs modellben a nem kontrollált hálózati paraméterek nem játszottak jelentős szerepet a becslésekben. Erre utalt az azonos paraméterekkel szimulált hálózatok közötti különbség alacsony (bár szignifikáns) magyarázóereje a becslésekre nézve.

Az első kutatási kérdésem arra irányult, hogyan befolyásolják a hálózati és mintavételi tényezők a becslések torzítását és szórását, illetve mely paraméter kombinációk mellett melyik becslés (RDS I, RDS II) tekinthető hatékonyabbnak.

A heterofília, azaz a vizsgált változó mentén eltérő értéket választók közötti kapcsolatok aránya gyakorolja a legjelentősebb hatást mind a becslések szórására, mind torzítására, oly módon, hogy a kisebb heterofília nagyobb torzítással és a becslés nagyobb szórásával jár együtt. Ez az eredmény megfelel az idézett korábbi kutatásoknak, azzal a pontosítással, hogy elemzésem során a foksám várhatóértéket és a kiinduló minta elemszámát is kontrolláltam (azaz az összefüggések a két vizsgált paraméter minden értéke mellett igazak).

A heterofília mellett a foksám is szignifikánsan befolyásolta mind a becslések torzítását, mind szórásukat, azonban a hatás erőssége a heterofília hatásánál gyengébb volt. Általában az alacsonyabb foksám várhatóérték (kisebb sűrűségű hálózat) esetén a becslések valamivel kisebb torzításúak és szórásúak voltak (ez különösen az RDS II becslés esetén volt igaz). E paraméter hatására vonatkozó korábbi kutatási adatot nem találtam.

A kiinduló minta nagysága és ezzel párhuzamosan az átlagos lánchossz nem gyakorolt jelentős és konzisztens befolyást sem a torzításra, sem a becslések szórására. Ez az eredmény ismét megfelel a korábbi kutatások eredményeinek azzal a kiegészítéssel, hogy ebben az esetben is kontrolláltan történt a vizsgálat a többi paraméterre nézve (azaz a többi vizsgált paraméter különböző értékei mellett is igaznak bizonyult).

A második kutatási kérdésem a becslések megbízhatóságára vonatkozott, illetve ezzel összefüggésben arra, hogy az egyes paraméterek ezt mennyiben befolyásolják.

Kutatásom igazi újdonsága, hogy az RDS I becslés esetén - amely az átlagos foksám és a heterofília becslésén alapul - kimutattam, hogy e két becslés hibája különféle paraméter

beállítások mellett milyen mértékben járul hozzá a teljes becslés hibájához. Ezzel kapcsolatban azt állapítottam meg, hogy a heterofília változása határozza meg elsősorban, hogy melyik hálózati paraméter becslésének hibája gyakorolt nagyobb befolyást a becslés teljes hibájára. A nagyobb heterofília a heterofília becslés hibájának jelentősen kisebb hatásával, míg a kisebb heterofília a heterofília becslés jelentősen nagyobb hatásával jár együtt. E mellett a foksám várhatóérték eltéréseinek hatása jóval csekélyebb: a magasabb foksám várhatóérték a foksám átlag becslésének a teljes becslés szórására gyakorolt valamivel kisebb hatásával jár együtt. Ez az eredmény különösen annak fényében tűnik fontosnak, hogy Lu korábban idézett cikkében (Lu [2012]) egonetwork adatok felhasználására épülő módosított RDS I becslésével éppen az heterofíliára ad pontosabb becslést.

Végül, de nem utolsó sorban az, hogy egy-egy paraméter beállítással több hálózatot szimuláltam és egy-egy hálózaton több mintavételt végeztem, lehetővé tette azt is, hogy megvizsgáltam az azonos paraméterű, de eltérő hálózatokból származó minták esetén a becslések tulajdonságainak eltéréseit is. Ez alapján az mondható ki, hogy a hálózat egyéb, nem kontrollált tulajdonságai a modellben bár szignifikáns hatást gyakoroltak, de a hatás mértéke csekély volt, így lényegében az egyéb hálózati tényezők hatását elhanyagolhatónak tekinthetjük.

A szimulációs eredmények több gyakorlati javaslatot is implikálnak. Amennyiben azt feltételezhetjük, hogy az általunk leginkább fontosnak tartott változók mentén klasztereződik a sokaság (azaz kicsi a heterofília), akkor érdemesebb nagyobb mintaméretet tervezni már a kutatásunk elején. Összességben sem az RDS I, sem az RDS II becslés nem preferálható a másikkal szemben, ha azonban a mintavételi design több kiindulópontot tartalmaz és rövidebb láncokat, valamint a kiindulópontok esetén megközelíthető a foksám arányos bekerülési valószínűség, akkor az RDS II becslés valószínűleg pontosabb. További fontos eredmény, hogy nem kell figyelembe venni a becslés pontossága szempontjából a vizsgált célcsoport kapcsolathálózatának sűrűségét. A fent leírtakon túl nincs nagy jelentősége (legalábbis a vizsgált tartományon belül és a használt mintavétel mellett) a kiindulópontok számának. Gyakorlati szempontból további fontos eredmény, hogy érdemes minden vizsgált változó esetén az elemzést megelőzően becslést tenni a változó értékei szerint képzett csoportok összekötöttségére, és amennyiben az összekötöttség alacsony, akkor szélesebb konfidencia intervallumokkal érdemes számolni.

IV.4. Módszertani megfelelés és alkalmazói környezet

A harmadik fejezet zárásaként érdemes megvizsgálni, hogy a fejezet fényében, hogy értelmezhető a két kutatási kérdés, a válaszadó-vezérelt mintavétel módszertani megfelelése, illetve milyen kapcsolatok, hatások merültek fel a vizsgált módszerre tudományos konstrukcióként.

A I.1 alfejezetben megfogalmazott szempontokat a jelent fejezetben részletesen vizsgáltam mind a rendelkezésre álló szakirodalom, mind saját szimulációs vizsgálat segítségével. A fejezetben ismét több kutatás segítségével, illetve saját szimulációs vizsgálatom eredményével is sikerült alátámasztanom a válaszadó-vezérelt mintavételből adható becslések torzítatlanságát. E mellett számos kutatási eredményt bemutattam, amelyek a megbízhatóságot befolyásoló tényezőket vizsgálják. E téren a legtöbbet vizsgált a homofília (az egymással kapcsolatban állók hasonlóságának valószínűsége) és az ezzel kapcsolatban hozható hálózati klaszterezettség, amely általánosságban kedvezőtlenül befolyásolja a vizsgált módszer megbízhatóságát, de eltérő körülmények között eltérő becslést, illetve módosított mintavételi design alkalmazva a hatás csökkenthető. A fejezetben ismét több kutatás segítségével, illetve saját szimulációs vizsgálatom eredményével is sikerült alátámasztanom a válaszadó-vezérelt mintavételből adható becslések torzítatlanságát.

A I.2 alfejezetben megfogalmazott tudományos konstrukció modell számos környezeti eleme közül az alkalmazói környezetet vizsgáltam. A módszertani irodalom és saját kutatásunk alapján az alkalmazói környezet heterogenitása és nagy számú probléma megjelenése állapítható meg. Mindezt alátámasztja a nagy számú módszertani ajánlás, amelyet az idézett kutatások – beleértve saját kutatásunkat is – megfogalmaznak.

V. A válaszadó-vezérelt mintavétel és szociológiai elméletek⁶⁴

A válaszadó-vezérelt mintavételt ebben a fejezetben szociológiai szempontból közelítem meg a ritka és rejtett társadalmi csoportokat. Ezeket a csoportokat ebben a fejezetben nehezen körülhatárolható csoportként jelölöm meg. Azokat a társadalmi csoportokat tekintem elméleti szempontból nehezen körülhatárolhatónak, amelyek a cselekvéseiket összehangolják, vagy legalábbis hasonló tevékenységet végeznek (lásd alább Angelusz [2000]), ugyanakkor a hagyományos strukturáló változókkal (kor, nem, iskolai végzettség, jövedelem stb.) nem határolhatók el a társadalom más csoportjaitól, nem vagy kevéssé vizsgálhatók a szociológiai rétegződéselméletek hagyományos dimenziói segítségével. Azt állítom, hogy az így definiálható csoportok egyúttal speciális nehézségek elé állítják az empirikus kutatót, ha kvantitatív kutatást kíván végezni az adott célcsoporton belül – azaz módszertani értelemben véve ezek a csoportok nehezen elérhetőek abban az értelemben, ahogyan azt a II.1 alfejezetben definiálom.

Az alábbiakban két elméleti megközelítést és egy empirikus kutatást ismertetek, amellyel a fenti gondolatmenetet támasztom alá: Angelusz Róbert társadalmi látásviszonyok romlásáról szóló cikkét (Angelusz [2000]), Ulrich Beck kockázattársadalomról szóló elméletének a társadalmi strukturáló tényezők változásáról szóló részét (Beck [2003]) és Róbert Péter empirikus vizsgálatát az életstílus meghatározottságáról (Róbert [2000]).

V.1. Elméleti megközelítések

Angelusz Róbert cikkében (Angelusz [2000]) azt állítja, hogy a társadalmi nagycsoportok láthatósága csökken, és ennek hatása van a kollektív cselekvésekre is. A nehezen körülhatárolható társadalmi csoportok kutatása szempontjából ennek az állításnak az a jelentősége, hogy a társadalmi cselekvések magyarázatában a hagyományosan jól leírható (és jól kutatható) társadalmi csoportok kevesebb szerepet kapnak és helyettük olyan csoportok jelennek meg a cselekvések háttérében, amelyek a hagyományos módszertani eszközökkel nem kutathatók. Erre jó példa lehet bármelyik, a közösségi honlapokon szerveződő mozgalom (számos esetben megfigyelhető Magyarországon is, de a 2011 elején kirobban észak-afrikai felkelések szintén evidens példát szolgáltatnak erre a jelenségre). A részletek elemzése nélkül állítható, hogy egyrészt ezek a csoportok jelentős hatóerejűek és képesek a közös cselekvésre,

⁶⁴ A fejezet a szerző Balogh Anikóval közösen írt elektronikus tankönyve I. fejezet egyes részeinek felhasználásával készült (Balogh-Simon [2011]) A felhasznált tankönyv fejezetet a szerző írta.

másrészről tagságuk, működésük kvantitatív módon nem vizsgálható a hagyományos, véletlen mintavételen alapuló módszerekkel.

Angelusz Róbert a társadalmi nagycsoportok láthatóságának csökkenése mellett érvelve Blau [1997] jól ismert rétegződés-modelljét módosítja, egyúttal kritizálva Blau állításait.

Blau eredeti modelljében a társadalom vertikális rétegzettségéért az éles választóvonal nélküli, folytonos graduális változók felelnek, míg a diszkrét nominális változók éles határvonalat húznak meg és horizontális csoportokat alakítanak ki. Angelusz a horizontális vagy vertikális, illetve folytonos vagy diszkrét dimenziókat elválasztva hozza létre saját rendszerét, amely, mint alább láthatjuk, már alkalmas a társadalmi látásviszonyok változásának háttérében lezajló folyamatok leírására.

16. táblázat A társadalmi rétegződést meghatározó változó típusok Angelusz Róbert Duncan Blau nyomán létrehozott modellje szerint

	Emergens	Nominális	Hierarchikus	Graduális
Hierarchikus	+	-	+	-
Éles választóvonal	+	+	-	-

Forrás: Angelusz [2000] nyomán

Érdeemes az egyes változók jelentését és jelentőségének változását külön megvizsgálni:

Emergens változók: éles választóvonallal bíró (tehát nem folytonos), de társadalmi hierarchiát létrehozó változók, mint például a tulajdon. Angelusz érvelése szerint ezek jelentősége csökken, például a részvénytulajdonnal eltűnik az éles megkülönböztetés a tulajdonosi és tulajdon nélküli osztály között.

Nominális változók: éles választóvonallal bíró (tehát nem folytonos), de társadalmi hierarchikus különbséget nem okozó változók, mint például a vallási hovatartozás. Bár Angelusz nem foglalkozik e változótípus jelentőségének változásával, óvatos állítás tehető arra nézve, hogy növekvő tendenciát figyelhetünk meg a változótípus jelentősége terén.

Hierarchikus változók: folytonos változók, amelyek hozzájárulnak a társadalmi hierarchia kialakulásához (vertikálisak). Angelusz állítása szerint ezek jelentősége nő: egyes, korábban emergens változók hierarchikussá váltak, azaz megszűnt diszkrét jellegük, mint például a fent már említett tulajdonviszonyt jellemző változóknak.

Graduális változók: folytonos változók, amelyek nem járulnak hozzá a társadalmi hierarchia kialakulásához (horizontálisak). Angelusz ilyennek tekinti az életkort vagy a gyermekszámot – amely példákkal valószínűleg nem egy szociálpolitikus vitatkozna.

A fenti egyszerűsített érvelés mellett szól, hogy a társadalmi struktúrát kialakító változók közötti jelentőség eltolódása, illetve átalakulása a társadalmi struktúra láthatóságának csökkenéséhez vezethet. Angelusz egyúttal mellett is érvel, hogy a kevésbé átlátható struktúra a nagycsoportok kollektív cselekvésének valószínűségét is csökkenti.

A láthatóság és a kollektív cselekvés kapcsolatának vizsgálatára Angelusz egy újabb modellt állít fel. A modell a csoportok önreflexivitásának fokozatait veti össze a cselekvés típusaival.

A modellben a következő csoporttudat típusokat különíti el a szerző:

- Tömeg-csoporttudat: a csoport tagjai csak azonos státuszúak (státusekvivalencia), de ez nem tudatosul (nincs csoporthelyzet reflexió), ennek következtében nem alakítanak ki közös cselekvési mintázatokat (összehangolt reflexiókat)
- Reflexív-csoporttudat: az azonos státusz tudatosul a csoport tagjaiban (van csoporthelyzet reflexió), de nincs összehangolt cselekvés
- Kollektív tudat: tudatossá vált azonos státuszon alapuló közös cselekvés jellemző (kollektív cselekvés)

17. táblázat *A társadalmi nagycsoportok csoporttudatának tipológiája Angelusz szerint*

•	Státus- ekvivalencia	Csoporthelyzet reflexiója	Reflexiók összehangolása	Tipikus cselekvés
Tömeg-csoporttudat	+	-	-	Látens
Reflexív csoporttudat	+	+	-	Csoportspecifikus
Kollektív tudat	+	+	+	Kollektív

Forrás: Angelusz [2000] nyomán

A bemutatott modell alapján levezethető, hogy minél inkább reflektál egy csoport saját csoporthelyzetére és minél inkább képesek a csoporttagok összehangolt interakcióba lépni egymással, annál összehangoltabb a cselekvés.

A láthatóság fent leírt csökkenésével már a csoporthelyzet reflexiója (felismerése) is ellehetetlenül, nem beszélve a reflexiók összehangolásáról (amelyhez a csoporttagok felismerése egyértelműen szükséges).

Angelusz érvelése itt véget ér, de nehezen körülhatárolható társadalmi csoportok kutatása szempontjából a leírt érvelés még nem kielégítő: ahhoz elegendő, hogy bizonyítsuk, a hagyományos társadalmi nagycsoportok vizsgálata önállóan már nem képes magyarázni a társadalmi jelenségeket, de nem bizonyította a nehezen körülhatárolható társadalmi csoportok vizsgálatának szükségességét. Angelusz érveléséből az következik, hogy a társadalmi látásviszonyok csökkenésével a kollektív cselekvések egyértelműen csökkennek. Ez a folyamat talán egyes helyeken, időlegesen meg is figyelhető, de főként a hálózati kommunikáció terjedésével, jelentőségének növekedésével úgy tűnik, hogy olyan új, rendkívül gyorsan szerveződő és felbomló, de hatékony cselekvésre képes csoportok jönnek létre és válnak társadalmi aktorrá, amelyeket nyugodtan tekinthetünk nehezen körülhatárolhatóknak. A hálózati kommunikáció csoportformáló szerepén túl az olyan tényezők (nominális változók) szerepének erősödését is megfigyelhetjük, mint a nemzetiség (legyen szó akár helyben élő, akár bevándorló csoportokról). Ezek a csoportok pedig szintén a nehezen körülhatárolható csoportok körét bővítik.

Ulrich Beck a kockázat-társadalom koncepció részeként írja le a társadalmi struktúra átalakulására vonatkozó megfigyeléseit és elméletét (Beck [2003]), amelyből lehetséges jövőképeket vezet le. Az általa vázolt társadalmi változások közül a következő pontokat emelnénk ki:

- az egyenlőtlenségi viszonyok nem változtak, de csökkent a társadalmi problémák élessége, ezért csökkent az osztálytudat és az osztályfogalom társadalmi észlelhetősége
- csökkentek a megélhetés költségei és a munkaidő, növekedett viszont a tulajdonosok (pl.: lakástulajdonosok) száma, a megtakarítás mértéke, az életidő és munkajövedelem
- növekvő mobilitás: „az életút önállósul” a származáshoz képest a *személyes sors* jelentősége nőtt
- női munkaerő, a női szerep átalakulása, a család dezintegrációja – a nők egyenlőtlenségének tudatosulása
- nem mindenki osztja töretlenül a státus és jövedelem orientált felemelkedés gondolatát

E néhány pont alapján Beck növekvő individualizmusra és osztályvilágok helyett a fogyasztási stílus fontosságának növekedésére következtet.

Beck a fent – összefoglalva – leírt változások alapján három lehetséges jövőképet vázol fel:

- Új, modernizálódott osztályképződés: Beck az új osztályképződés alapjának elsősorban a kockázatközösségek (azonos kockázatokkal rendelkező társadalmi nagycsoportok) kialakulását tartja
- Új társadalmi hálók, mozgalmak, identitások jönnek létre: önkiteljesítés, én alakítás (pl.: nőmozgalmak)
- Új, individualizált életformák, életstílusok alakulnak ki
 - a társadalmi intézmények őrzik a rendi jellegű osztálystruktúrát, amely azonban már láthatatlan
 - a mobilitás, mint nagycsoportok közötti vándorlás élménye megszűnik
 - egyedi, esemény-orientált koalíciók születnek
 - tartóssá váló konfliktusok tulajdonított jegyek mentén (bőrszín, nem, etnikai hovatartozás)

Bár a fogalom nem kerül elő Beck elméletében, mégis azt mondhatjuk, hogy a becki elmélet közelebb visz bennünket alapállításunk igazolásához, azaz ahhoz, hogy bizonyítsuk: a nehezen körülhatárolható csoportok társadalmi jelentősége növekvőben van.

Beck állításai közül a legfontosabbak, amelyek elméleti szempontból alátámasztják a nehezen körülhatárolható társadalmi csoportok jelentőségének növekedését a következők:

- A fogyasztási stílus szerepének növekedése: mind tudományos, mind piackutatási szempontból felértékeli a fogyasztási stílus alapján kialakuló csoportok jelentőségét
- Kockázatközösségek létrejötte: e közösségek nem a hagyományos értelemben vett strukturáló változókon alapulnak, külső szemlélő számára nehezen körülhatárolhatók
- Új társadalmi hálók, mozgalmak, identitások létrejötte: e mozgalmak – még, ha egy jól látható társadalmi csoport, például a nők bázisán jönnek is létre – alapvetően önkéntesek, a résztvevők köre nehezen fellelhető, hatóerejük ugyanakkor jelentős

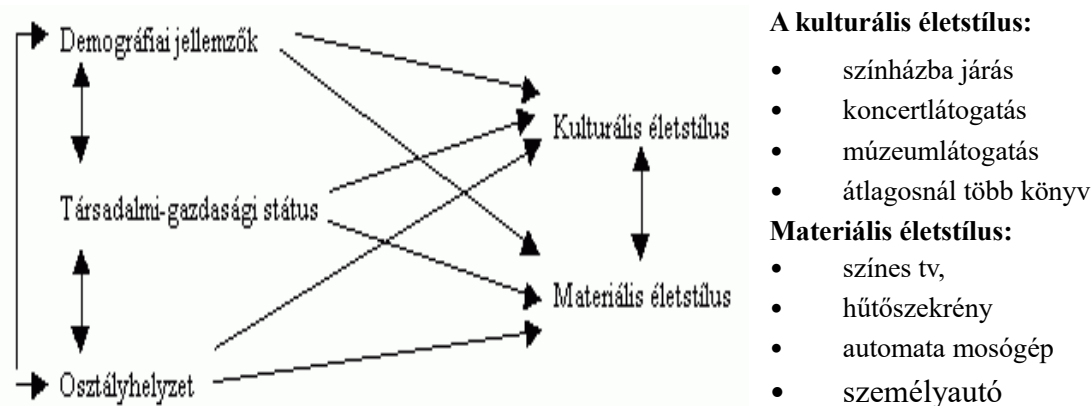
V.2. A válaszadó-vezérelt mintavétel szociológiai relevanciáját megerősítő empirikus kutatás

Végezetül a hazai empirikus kutatások közül hozok egyet példaként, amely szintén a társadalmi struktúra változásaival foglalkozik a magyar társadalomban. Róbert Péter kutatásában (Róbert,

[2000]) a rendszerváltás időszakában, 1982 és 1998 között vizsgálta az életstílus meghatározottságát. Elemzésében a szocio-demográfiai tényezők életstílusra gyakorolt hatását, valamint az anyagi és kulturális életstílus viszonyának változását vizsgálta. Kutatási kérdései és hipotézisei a következők voltak:

- „K1. Hogyan korrelál egymással a kulturális és az anyagi életstílus?
- H1. Feltételezhetően a kulturális és az anyagi életstílus közti korreláció 1982 és 1998 között növekedett.
- K2. Hogyan hatnak a társadalmi-demográfiai jellemzők a kulturális és az anyagi életstílusra?
- H2. A kulturális és az anyagi életstílus meghatározottsága 1982 és 1998 között növekedett.” (Róbert, [2000], a K a kutatási kérdéseket, míg a H a kérdések válaszaként feltételezett hipotéziseket jelenti)

6. ábra A demográfiai jellemzők, társadalmi-gazdasági státus és osztályhelyzet lehetséges összefüggései a kulturális és materiális életstílussal Róbert Péter szerint



Forrás: Róbert [2000]

A hipotézisek éppen a fejezet első részében ismertetett elméleti szempontokkal szemben a státus kikristályosodását (meghatározottságának növekedését) feltételezték, azonban az eredmények ezt csak részben erősítették meg. Az alábbi pontokban foglalhatók össze a kutatás eredményei:

- a kulturális és materiális életstílus közötti összefüggés 1982 és 1992 között növekedett, 1998-ra viszont csökkent, vagyis a két életstílus dimenzió csak a rendszerváltásig vált koherensebbé

- a kulturális és az anyagi életstílus szocio-demográfiai, státus és osztály meghatározottsága egyaránt hasonló mintázatot követett
- egyedül az iskolai végzettség hatása erősödött az egész időszak alatt (a hipotézisnek megfelelően)

A fentiek alapján az a következtetés vonható le, hogy a státuskikristályosodás inkább a rendszerváltás idejére volt jellemző, azt követően ellentétes változás indult meg.

A fentiekben már leírt elméleteket az ismertetett empirikus kutatás abban a vonatkozásban nem cáfolja, hogy bár a rendszerváltás időszakában az életstílus strukturális meghatározottsága nőtt, ez a folyamat a rendszerváltást követően megfordulni látszott. Ez arra utal, hogy a hagyományosan a társadalmi struktúrát meghatározó szocio-demográfiai, státus és osztály változók szerepe a rendszerváltást követően csökkent az életstílus meghatározásában.

Ez az empirikus eredmény megerősíti az elméleti részben leírt változásokat a társadalmi struktúrát meghatározó változók szerepének változása szempontjából, így alátámasztja azt a feltételezésünket, hogy legalábbis csökken a hagyományos társadalmi csoportok hatása. A fenti eredmények ugyanakkor nem nyújtottak támpontot abból a szempontból, hogy vajon milyen új csoportok léphetnek a korábban jól körülhatárolt társadalmi csoportok (rendek, osztályok), mint társadalmi aktorok helyére.

V.3. A válaszadó-vezérelt mintavétel szociológiai kutatási környezete (összefoglaló)

A negyedik fejezet zárásaként érdemes megvizsgálni, hogy a fejezet fényében hogyan értelmezhető a két kutatási kérdés, a válaszadó-vezérelt mintavétel módszertani megfelelése, illetve milyen kapcsolatok, hatások merültek fel a vizsgált módszerre tudományos konstrukcióként.

A I.1 alfejezetben megfogalmazott szempontokat ebben a fejezetben nem vizsgáltam. A I.2 alfejezetben megfogalmazott tudományos konstrukció modell számos környezeti eleme közül a megrendelői környezetet vettem szemügyre néhány a témához kapcsolódó kutatás, illetve elméleti írás kapcsán. A megrendelői környezetben belül a fejezet a szociológiai kutatási környezettel foglalkozott, nem tértem ki a piaci kutatási környezetre. A fejezetben igyekeztem igazolni, hogy mind az empirikus kutatások (életstílus és státusz összefüggése Róbert [2000]), mind a kortárs elméletek (kockázattársadalom kialakulásának hatása (Beck [2003], hagyományos társadalmi osztályképződés lehetőségének csökkenése (Angelusz [2000])) a nehezen körülhatárolható csoportok jelentőségének növekedését támasztják alá. Ez a

körülmény pedig kedvező környezetet termet az ilyen célcsoportok vizsgálatára alkalmas mintavételi eljárásoknak, így a válaszadó-vezérelt mintavételnek is.

VI. A válaszadó-vezérelt mintavétel jogi és etikai szempontjai

A válaszadó-vezérelt mintavétel matematikai és módszertani környezetének áttekintése után a társadalmi környezet két fontos elemével, a jogi (és adatvédelmi), illetve az etikai környezet vizsgálatával folytatom. Míg a jogi környezet a társadalmi normarendszer egyes részeinek általános érvényűvé válása és intézményesülése (Szilágyi [2011]), a tudomány számára és ezen belül természetesen a társadalomtudományi kutatás módszertana és a válaszadó-vezérelt mintavétel kivitelezői számára kötelezően betartandó kereteket jelent, addig az etikai kérdések egyrészt lelkiismereti jelentőségük, másrészt a tudomány társadalmi megítélését és ezen keresztül egy adott módszer gyakorlati alkalmazhatóságát befolyásolják elsősorban. A két környezet eltérő jellege miatt különböző megközelítésben vizsgálom meg e két tényező szerepét a válaszadó-vezérelt mintavétel módszertana szempontjából. A jogi keret esetében a hazai (és vonatkozó Európai Unió) helyzet ismertetésével mutatom be a jogszabályok és a jogértelmezés jelentette korlátokat, tehát egyfajta objektív leírást adok, amelyet értékelek a módszer szempontjából. Az etikai környezet kapcsán ezzel ellentétesen egy tudományos vitát idézek, amelynek a felvetéseit igyekszem általánosítani és összefoglalni anélkül, hogy a felmerült kérdésekben állást foglalnék.

VI.1. Jogi, adatvédelmi kérdések⁶⁵

A válaszadó-vezérelt mintavétel hazai közvetlen jogi környezetét négy jogszabály képezi, amelyből kettő, a 2016. évi CLV. törvény a hivatalos statisztikáról (továbbiakban: statisztikai törvény) és a 2012. évi C. törvény a Büntető törvénykönyvről (továbbiakban: Btk) csak kisebb részleteiben (bár éppen az utóbbi az, amely a másik két idézendő törvény szankciót tartalmazó háttérét teremti meg), míg kettő, a 2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról (továbbiakban: Infotv.), illetve az 1995. évi CXIX. törvény a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről (továbbiakban: kutatási törvény) részletesebben foglalkozik a témát érintő kérdésekkel. A hazai jogszabályi környezet mellett az adatvédelem területén mindenképpen meg kell említeni az Európai Tanács 2016/679 rendeletét a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról (továbbiakban: GDPR

⁶⁵ A fejezet Balogh-Simon [2011] tankönyvének IV. fejezetében szereplő vázlatos összefoglalás alapján, a jogszabályi változások követésével és a gyakorlati tapasztalatokkal kiegészítve készült. Az idézet fejezet a szerző írása.

rendeletet). Az utóbbi jogszabály elemzését a nemzetközi környezet figyelembevételével mellett az is indokolja, hogy közvetlenül is hatályos Magyarország területén is.

A statisztikai törvény jelenlegi hatálya nem terjed ki a nem hivatalos statisztikai tevékenységekre, azonban egyes, a jelen elemzés témáját érintő fogalmak, eljárások jogi meghatározására csak ebben a törvényben kerül sor, így ennek rövid áttekintése is hasznos, a konklúzióban pedig kitérek arra a furcsa jogi helyzetre, amely eltérő kutatóhelyek számára eltérő jogi környezete biztosít. A statisztikai törvény részben a statisztika feladatának definiálásával, részben az adatgyűjtés alapvető szabályainak meghatározásával érinti a dolgozatomban vizsgált területet. A törvény szerint a *„A hivatalos statisztikai tevékenység célja, hogy statisztikai információk nyilvánosságra hozatalával valósághű, tárgyilagos képet adjon a társadalom, a gazdaság, a környezet állapotáról és annak változásairól az állami szervek, az önkormányzatok, a gazdasági szervezetek - beleértve a pénzügyi piacokat -, a civil szervezetek, a tudományos tevékenységet végzők, a közvélemény, a média szereplői, valamint a nemzetközi szervezetek, különösen az Európai Unió intézményei számára.”* (idézett törvény 3.§ (1) bekezdése). A definíció értelmezése előtt meg kell jegyezni, hogy a statisztika ilyen értelmezése a statisztika szó tudományos értelmezésénél szűkebb, nem foglalja magába a statisztika adatok elemzésére, illetve az összefüggések vizsgálatára irányuló felhasználását. A definíció alapján három dolog válik azonnal nyilvánvalóvá: egyrészt az, hogy a statisztika vizsgálódásának tárgya a társadalom vizsgálatán túl terjeszkedik, másrészt célja az állapot és a változások valósághű, tárgyilagos – másként fogalmazva objektív – bemutatása, harmadrészt a statisztika célközönsége a társadalom (a definíció egyes társadalmi intézményeket is kiemel, de ez nem változtat ezen a tényen). A definíciónak megfelelően egy tudományos kutatásban a vizsgált módszer alkalmazásának célja éppen a társadalom (vagy inkább egyes csoportjainak) állapotbemutatása, illetve ezen állapot változásának követése, úgy, hogy arról objektív kép alakuljon ki, amelyet a tudományos közösség és a tágabb társadalom felé prezentál a kutató. Ilyen értelemben tehát kétséget kizáró módon a törvény definíciója alapján vizsgálatom tárgya statisztika (illetve annak része). A törvény másik kiemelendő része az adatok gyűjtésével kapcsolatban állít fel korlátokat. Általános rendelkezése szerint *„25.§ (1) Természetes személytől személyes adatára vonatkozó kötelező adatszolgáltatást - a 26. § (1) bekezdésben foglalt kivétellel [kivételt a gazdasági tevékenységet folytató természetes személy jelent] - csak törvény rendelhet el.”*. Vagyis adatközlő nem kötelezhető személyes adatainak⁶⁶ közlésére, ha a

⁶⁶ A személyes adat definícióját az idézett törvény nem tartalmazza, azt a fejezet későbbi részében részletesen tárgyalom.

törvény nem rendeli el – márpedig a tudományos kutatás esetén ez a törvényi kötelezettség általában nem áll fenn. Lényegesebb megkötés az a különleges megszorítás, amely a különleges adatokra⁶⁷ vonatkozik (statisztikai törvény 25. § (2)). Az idézett törvényi bekezdés jelentősége abban áll, hogy a válaszadó-vezérelt mintavétel célcsoportja gyakran éppen valamelyik különleges adattal jellemezhető csoport (például: nemzeti, etnikai vagy éppen szexuális kisebbség), a gyűjtött adatok, pedig szintén vonatkozhatnak ilyen területre (például: egészségügyi állapot, kóros szenvedély). A törvény rendelkezéseiből kutatásom tárgya szempontjából az a fontos, hogy mindkét pontban szereplő különleges adatok csak személyazonosításra alkalmatlan módon gyűjthetők. Ez a követelmény látszólag egyszerűen teljesíthető. Minden adatgyűjtésen alapuló társadalomtudományi kutatás esetén megtörténik a személyes adatok és a kutatási kérdésre vonatkozó adatok szétválasztása. Ez megtörténhet – és meg is történik – egy válaszadó-vezérelt mintavétellel történő kutatás esetén is. Az adatgyűjtési folyamatot részletesen végig követve azonban feltűnhet, hogy a tovább ajánlásokhoz (amennyiben a válaszadók bevonására úgy kerül sor, hogy egy kérdezőbiztos keresi fel őket) szükség van a következő bevonandó személy személyes adatára is. Ez pedig egy különleges adattal jellemezhető célcsoport esetén már sérti a fenti feltételt, hiszen a személyes adat megszerzésétől annak megsemmisítéséig a különleges adat és a személyes adat összekapcsolódik. A probléma feloldási lehetőségeire a fejezet végén térek vissza.

Ugyanezeket a kérdéseket jóval részletesebben szabályozza az Infotv., a GDPR rendelet, valamint a kutatási törvény.

Az Infotv. tartalmazza a fent már alkalmazott, de ott még pontosan nem definiált fogalmak, a személyes adat és a különleges adat definícióját, amelyeket fontosságuk miatt itt idézek:

„3.§ (...)

1. érintett: bármely meghatározott, személyes adat alapján azonosított vagy - közvetlenül vagy közvetve - azonosítható természetes személy;

2. személyes adat: az érintettel kapcsolatba hozható adat - különösen az érintett neve, azonosító jele, valamint egy vagy több fizikai, fiziológiai, mentális, gazdasági, kulturális vagy szociális azonosságára jellemző ismeret -, valamint az adatból levonható, az érintettre vonatkozó következtetés;

3. különleges adat:

⁶⁷ A különleges adat definícióját implicit módon az idézett törvénytörvény tartalmazza, de a pontos definiálásra – jelentősége miatt – később még a fejezet későbbi részében visszatérek.

- a) a faji eredetre, a nemzetiséghez tartozásra, a politikai véleményre vagy pártállásra, a vallásos vagy más világnézeti meggyőződésre, az érdek-képviseleti szervezeti tagságra, a szexuális életre vonatkozó személyes adat,
- b) az egészségi állapotra, a kóros szenvedélyre vonatkozó személyes adat, valamint a bűnügyi személyes adat;”

A GDPR rendelet vonatkozó definícióját (4. cikk 1. pont) itt nem idézem (a mellékletben szerepel), de területi hatálya miatt kutatás esetén érdemes azt is figyelembe venni (ez a jogszabály a különleges adat fogalmát nem ismeri).

A válaszadó-vezérelt mintavétel során személyes adat lehet a válaszadók neve, címe, e-mail címe, azok az adatok, amelyek a mintába került személy felkereséséhez szükségesek (a különleges adatokról korábban már írtam). Mindkét idézett jogszabály definiálja az adatkezelő, adatfeldolgozó és az adatkezelés fogalmát is (Infotv 3.§ 9-10.; GDPR 4. cikk 2. és 7.), ennek a kutatás lebonyolításánál szintén jelentősége van. A részletesebb meghatározást ezúttal a GDPR adja, így ezt idézem (az Infotv vonatkozó része a mellékletben található meg):

„4. cikk (...)

2. „adatkezelés”: a személyes adatokon vagy adatállományokon automatizált vagy nem automatizált módon végzett bármely művelet vagy műveletek összessége, így a gyűjtés, rögzítés, rendszerezés, tagolás, tárolás, átalakítás vagy megváltoztatás, lekérdezés, betekintés, felhasználás, közlés továbbítás, terjesztés vagy egyéb módon történő hozzáférhetővé tétel útján, összehangolás vagy összekapcsolás, korlátozás, törlés, illetve megsemmisítés;

(...)

7. „adatkezelő”: az a természetes vagy jogi személy, közhatalmi szerv, ügynökség vagy bármely egyéb szerv, amely a személyes adatok kezelésének céljait és eszközeit önállóan vagy másokkal együtt meghatározza; ha az adatkezelés céljait és eszközeit az uniós vagy a tagállami jog határozza meg, az adatkezelőt vagy az adatkezelő kijelölésére vonatkozó különös szempontokat az uniós vagy a tagállami jog is meghatározhatja;

8. „adatifeldolgozó”: az a természetes vagy jogi személy, közhatalmi szerv, ügynökség vagy bármely egyéb szerv, amely az adatkezelő nevében személyes adatokat kezel;”

E törvény vonatkozásában a válaszadó-vezérelt adatfelvétellel zajló kutatást végző kutató (az adatkezelés célját meghatározó, arra vonatkozó döntést meghozó személy) adatkezelő, az adatfelvételt végző (az adatkezelést végrehajtó) személyek adatfeldolgozók.

A két jogszabály rögzíti az adatkezelés alapelvét, amely fontos pontokon eltérést mutat, ezért itt mindkét törvény vonatkozó szövege idézésre kerül. Az Infotv. szerint:

„4. § (1) Személyes adatot kezelni csak meghatározott célból, jog gyakorlása és kötelezettség teljesítése érdekében lehet. Az adatkezelésnek minden szakaszában meg kell felelnie az adatkezelés céljának, az adatok felvételének és kezelésének tisztességesnek és törvényesnek kell lennie.

(2) Csak olyan személyes adat kezelhető, amely az adatkezelés céljának megvalósulásához elengedhetetlen, a cél elérésére alkalmas. A személyes adat csak a cél megvalósulásához szükséges mértékben és ideig kezelhető.

(3) A személyes adat az adatkezelés során mindaddig megőrzi e minőségét, amíg kapcsolata az érintettel helyreállítható. Az érintettel akkor helyreállítható a kapcsolat, ha az adatkezelő

rendelkezik azokkal a technikai feltételekkel, amelyek a helyreállításhoz szükségesek.

(4) Az adatkezelés során biztosítani kell az adatok pontosságát, teljességét és - ha az adatkezelés céljára tekintettel szükséges - naprakészségét, valamint azt, hogy az érintettet csak az adatkezelés céljához szükséges ideig lehessen azonosítani.

(5) A személyes adatok kezelését tisztességesnek és törvényesnek kell tekinteni, ha az érintett véleménynyilvánítási szabadságának biztosítása érdekében az érintett véleményét megismerni kívánó személy az érintett lakóhelyén vagy tartózkodási helyén felkeresi, feltéve, hogy az érintett személyes adatait e törvény rendelkezéseinek megfelelően kezelik és a személyes megkeresés nem üzleti célra irányul. A személyes megkeresésre a munka törvénykönyve szerinti munkaszüneti napon nem kerülhet sor.”

A GDPR szerint:

„5. cikk (...)

(1) A személyes adatok:

a) kezelését jogszerűen és tisztességesen, valamint az érintett számára átlátható módon kell végezni („jogszerűség, tisztességes eljárás és átláthatóság”);

b) gyűjtése csak meghatározott, egyértelmű és jogszerű célból történjen, és azokat ne kezeljék ezekkel a célokkal össze nem egyeztethető módon; a 89. cikk (1) bekezdésének megfelelően nem minősül az eredeti céllal össze nem egyeztethetőnek a közérdekű archiválás céljából, tudományos és történelmi kutatási célból vagy statisztikai célból történő további adatkezelés („célhoz kötöttség”);

c) az adatkezelés céljai szempontjából megfelelőek és relevánsak kell, hogy legyenek, és a szükségesre kell korlátozódniuk („adattakarékosság”);

d) pontosnak és szükség esetén naprakésznek kell lenniük; minden észszerű intézkedést meg kell tenni annak érdekében, hogy az adatkezelés céljai szempontjából pontatlan személyes adatokat haladéktalanul töröljék vagy helyesbítsék („pontosság”);

e) tárolásának olyan formában kell történnie, amely az érintettek azonosítását csak a személyes adatok kezelése céljainak eléréséhez szükséges ideig teszi lehetővé; a személyes adatok ennél hosszabb ideig történő tárolására csak akkor kerülhet sor, amennyiben a személyes adatok kezelésére a 89. cikk (1) bekezdésének megfelelően közérdekű archiválás céljából, tudományos és történelmi kutatási célból vagy statisztikai célból kerül majd sor, az e rendeletben az érintettek jogainak és szabadságainak védelme érdekében előírt megfelelő technikai és szervezési intézkedések végrehajtására is figyelemmel („korlátozott tárolhatóság”);

f) kezelését oly módon kell végezni, hogy megfelelő technikai vagy szervezési intézkedések alkalmazásával biztosítva legyen a személyes adatok megfelelő biztonsága, az adatok jogosulatlan vagy jogellenes kezelésével, véletlen elvesztésével, megsemmisítésével vagy károsodásával szembeni védelmet is ideértve („integritás és bizalmas jelleg”).

(2) Az adatkezelő felelős az (1) bekezdésnek való megfelelésért, továbbá képesnek kell lennie e megfelelés igazolására („elszámoltathatóság”).

A két hatályos jogi szabályozás alapelveiben ugyan hasonló (jogszerűség és átláthatóság, célhoz kötöttség, adattakarékosság, pontosság, korlátozott tárolhatóság, integritás, elszámoltathatóság), ugyanakkor éppen a vizsgált terület, a tudomány kutatás szempontjából lényeges eltérés is látható: míg a hazai jogszabály kivételt nem fogalmaz meg, a GDPR a tudományos kutatást több ponton kivételként kezeli és a nemzeti jogalkotó számára lehetővé teszi a speciális jogszabályok megalkotását a területen (lásd GDPR 89. cikk) Erre azonban a hazai jogalkotásban nem került sor.

Tudományos-, közvélemény- és piackutatás esetén más-más jogszabályban, de megtalálható a törvényi feltételben szereplő jog, illetve kötelezettség, amely alapján a személyes adatkezelés

lehetséges. Mindkét idézett jogszabály részletesebben is szabályozza a személyes adatok kezelésének jogalapját (feltételét). Az Infotv. szerint:

- „5. § (1) Személyes adat akkor kezelhető, ha
- a) ahhoz az érintett hozzájárul, vagy
 - b) azt törvény vagy - törvény felhatalmazása alapján, az abban meghatározott körben - helyi önkormányzat rendelete közérdeken alapuló célból elrendeli (a továbbiakban: kötelező adatkezelés).
- (2) Különleges adat a 6. §-ban meghatározott esetekben, valamint akkor kezelhető, ha
- a) az adatkezeléshez az érintett írásban hozzájárul, vagy
 - b) a 3. § 3. a) pontjában foglalt adatok esetében, az nemzetközi egyezményen alapul, vagy Alaptörvényben biztosított alapvető jog érvényesítése, továbbá a nemzetbiztonság, a bűnmegelőzés vagy a bűnüldözés vagy honvédelmi érdekből érdekeiben törvény elrendeli;
 - c) a 3. § 3. pont b) alpontjában foglalt adatok esetében azt törvény közérdeken alapuló célból elrendeli.”

„6. § (1) Személyes adat kezelhető akkor is, ha az érintett hozzájárulásának beszerzése lehetetlen vagy aránytalan költséggel járna, és a személyes adat kezelése

 - a) az adatkezelőre vonatkozó jogi kötelezettség teljesítése céljából szükséges, vagy
 - b) az adatkezelő vagy harmadik személy jogos érdekének érvényesítése céljából szükséges, és ezen érdek érvényesítése a személyes adatok védelméhez fűződő jog korlátozásával arányban áll.”

A GDPR szerint:

- „6. cikk (...)
- (1) A személyes adatok kezelése kizárólag akkor és annyiban jogszerű, amennyiben legalább az alábbiak egyike teljesül:
- a) az érintett hozzájárulását adta személyes adatainak egy vagy több konkrét célból történő kezeléséhez;
- (...)
- e) az adatkezelés közérdekű vagy az adatkezelőre ruházott közhatalmi jogosítvány gyakorlásának keretében végzett feladat végrehajtásához szükséges;
 - f) az adatkezelés az adatkezelő vagy egy harmadik fél jogos érdekeinek érvényesítéséhez szükséges, kivéve, ha ezen érdekekkel szemben elsőbbséget élveznek az érintett olyan érdekei vagy alapvető jogai és szabadságai, amelyek személyes adatok védelmét teszik szükségessé, különösen, ha az érintett gyermek.”

A látszólag indokolatlanul hosszú idézet oka az, hogy sok esetben ez a törvényi rendelkezés jelenti a legnagyobb kihívást a válaszadó-vezérelt mintavétel szempontjából. Abban az esetben ugyanis, ha nem a válaszadók keresik meg a kutatókat, hanem a kutatók (kérdezőbiztosok) a válaszadókat, illetve, ha az adatfelvétel esetleg elektronikusan, az interneten keresztül zajlik (erről a fejezet későbbi részében részletesen írok), akkor a válaszadó ismerőse adja meg az általa ajánlott következő személy elérhetőségét, ami a fenti definíciók alapján személyes adat. Ráadásul különleges adatok kezeléséhez írásbeli hozzájárulás szükséges, vagyis például egy romák körében folytatott kutatás esetén a lakcím, e-mail cím kezeléséhez (akár lejegyzéséhez is) az érintett írásbeli hozzájárulás szükséges. Ezt a hozzájárulást egy még fel nem keresett személy esetében meglehetősen nehéz megszerezni. Az Infotv. idézett 6.§-a lehetséges megoldást ad az érintett hozzájárulásának megszerzését előíró szabály alól. A kutató

hivatkozhat arra, hogy kutatói feladata teljesítéséhez kapcsolódóan jogos érdeke fűződik a kutatás elvégzéséhez és a kutatás lebonyolításának ez a leghatékonyabb, esetleg egyetlen elfogadható módja. A GDPR a különleges adatokra vonatkozó speciális szigorító szabályokat nem tartalmazza, így könnyebben alkalmazható az idézett 6. cikk f) bekezdésében található kutatói, illetve b) bekezdésben található közérdek. Természetesen kérdéses, hogy adott esetben az Adatvédelmi Hatóság vagy a bíróság milyen mértékig tartaná ezt az eljárást indokoltnak. A GDPR a jelzett általános rendelkezések mellett külön kitér a tudományos célú adatgyűjtésre vonatkozó általános beleegyezés lehetőségének megteremtésére is (sajnos csak a rendelet preambulumban):

- (33) *Gyakran nem lehetséges a tudományos kutatási célú személyesadat-kezelés célját az adatgyűjtés időpontjában teljes mértékben azonosítani. Ezért lehetővé kell tenni az érintettek számára, hogy a tudományos kutatás bizonyos területeire vonatkozóan hozzájárulásukat adják az adatkezeléshez, betartva a tudományos kutatásokra vonatkozó, elismert etikai előírásokat. Az érintettek számára biztosítani kell annak lehetőségét, hogy – annyiban, ha a célok ezt lehetővé teszik – csak egyes kutatási területekre vagy a kutatási projekteknek csupán bizonyos részeire vonatkozóan adjanak hozzájárulást.*

Bár a megfogalmazott lehetőség kecsegtető volna a válaszadó-vezérelt mintavételt alkalmazó kutatók számára, Magyarországon csupán a lakcím nyilvántartással kapcsolatos személyes adat kiadhatóságára vonatkozóan áll rendelkezésre a leírtakhoz hasonló (igaz opt out, azaz feltételezett beleegyezésen alapuló) lehetőség.

A kutatási törvény 4. §-s foglalkozik azzal, hogy milyen módon szerezhetők be kutatási célú személyes adatok:

- „3. § (1) A tudományos kutató, a közvélemény-kutató és a piackutató kapcsolatfelvétel és kapcsolattartás céljából, (...) illetőleg az általuk az adatok kezelésére, átvételére megbízott szerv név- és lakcímadatot a következő forrásból gyűjthet, illetve használhat fel:
- a) annak az érintettnek az adata, akivel korábban az adatkezelő szerv kapcsolatban állt (...);
 - b) a jogszerűen nyilvánosságra hozatal céljából készített és nyilvánosságra hozott adatállományban, név- és címjegyzékben, valamint kiadványban - így különösen telefonkönyv, szaknévsor, statisztikai névjegyzék - szereplő adat (...);
 - c) más, ugyanazon tevékenységet végző személytől vagy szervtől adat átvételével, amennyiben az érintett az adat átadását az erről szóló előzetes tájékoztatás után nem kifogásolta, vagy tiltotta meg;
 - d) adat igénylésével a polgárok személyi adatainak és lakcímének nyilvántartásáról szóló 1992. évi LXVI. törvény (a továbbiakban: Nytv.) hatálya alá tartozó nyilvántartásból az Nytv.-ben meghatározott feltételekkel, feltéve, hogy a polgár az adatainak kiadását nem tiltotta meg (...).
- (2) A tudományos kutatás, a közvélemény-kutatás, illetőleg a piackutatás megkezdéséhez szükséges minta kiválasztásakor, valamint kapcsolatfelvételi, illetve üzletszerzési lista összeállításakor nem lehet kiválasztási szempont olyan ismerv, amelyből egyértelműen különleges adatra lehet következtetni.”

A törvénytörvény érdeklődése, hogy kutatással kapcsolatban a közvetlen elérést csak személyes megkeresésen keresztül tételezi fel, így más adatokra (e-mail cím, telefonszám) ez a törvény

nem vonatkozik. Ha a fent leírt problémát, azaz a tovább ajánlás kérdését vizsgáljuk, akkor ez a törvény szigorúbb, mint az Infotv. A leírtak alapján név- és lakcímadatok szerzésére csak korábban a kutatásba már bevont (és az adatkezelésbe beleegyező) személytől, nyilvános címlista, hasonló tevékenységet végző más szervezettől, illetve a hatóságtól van lehetőség. Ez alapján – legalábbis lakcím adat – még szóbeli beleegyezéssel sem lehetne szerzhető, ami a válaszadó-vezérelt mintavétel lehetőségeit jelentősen korlátozza.

Érdemes megvizsgálni ezután a jogalkalmazási gyakorlatot is. Ezt egy példán keresztül szemléltetem. A KSH az Európai Unió által finanszírozott SEEMIG projekt keretében felvetette az RDS módszer alkalmazásának lehetőségét a külföldön élő, munkát vállaló magyarok elérése érdekében⁶⁸. Ebben az esetben a személyes megkeresés, vagy a válaszadók egy vagy néhány helyen történő megkérdezése (az RDS módszer eredeti leírása szerint) rendkívül költséges lett volna (valójában irreálisan költséges). A kutatás vezetői ezért a vegyes, elektronikus és telefonos eléréssel keresztül történő RDS mintavétel alkalmazását tervezték. A KSH az elvégzendő munkához kikérte a Nemzeti Adatvédelmi és Információszabadság Hatóság (NAIH) előzetes állásfoglalását. Az állásfoglalás teljes szövege a I. mellékletben olvasható. A Hatóság állásfoglalása szerint a telefonszám, az e-mail cím és a külföldről történő hazaérkezési időpontja is az Infotv. idézett 3. § 1. és 2. pontja alapján személyes adatnak minősül. Az állásfoglalás megállapítja továbbá, hogy ezeknek a személyes adatoknak az elkérése adatkezelésnek minősül (az Infotv. itt is idézett 3. § 10. pontja alapján). Személyes adat kezelésére pedig csak az érintett hozzájárulásával, vagy jogszabályi felhatalmazás alapján van lehetőség (az Infotv. itt is idézett 5. § (1) bekezdése alapján). Az állásfoglalás megállapítja, hogy a jogszabályi felhatalmazás a kutatás esetén nem áll fenn, így csak a külföldön élő személy hozzájárulásával kezelhető a személyes adat. Ennek megfelelően az állásfoglalás „nem kifogásolja” azt az eljárást, amikor az ajánló személy kapcsolatba lép az ajánlott személlyel és az ajánlott személy az ajánló személynek a kutatásban történő részvételi szándékát jelzi. Ugyanakkor aggályosnak tartja azt az eljárást, amikor az ajánló személy – az ajánlott személy elérhetetlensége esetén – a KSH részére megadja az ajánlott személy elérhetőségét (telefonszámát vagy e-mail címét) és az ajánlott személyt a KSH keresi meg. Az állásfoglalás körvonalazza az RDS kutatás lehetőségeit: csak az ajánlott személy beleegyezése esetén lehetséges a megkeresés, de az ajánló által „közvetített” beleegyezés elfogadható a törvény szerint megkívánt tájékozott beleegyezésnek. Meg kell jegyezni, hogy a NAIH ilyen

⁶⁸ A KSH a szerzőt és Kmetty Zoltánt kérte fel tanácsadónak a projekt ezen részéhez, a közölt információk a tanácsadói munkából származnak.

jogértelmezése meglehetősen korlátozó és a fent leírt és jogszabályban is megfogalmazott – tudományos – érdekeket lényegében nem veszi figyelembe, amelyek adott esetben az előzetes beleegyezéstől való eltekintést is lehetővé tennék. Ilyen módon tehát a hazai jogi környezet az RDS mintavételt korlátok közé szorítja, de nem teszi lehetetlenné.

VI.2. Etikai kérdések⁶⁹

Az RDS mintavétel etikai környezetét a vonatkozó irodalom értelmező áttekintésével mutatom be. 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).

Greg Scott 2007-ben végzett kutatása az intravénás droghasználók esetén vizsgálja a válaszadóvezérelt mintavétellel kapcsolatos problémákat, eredményei azonban más csoportok esetén is fontosak lehetnek (Scott [2008a]). Scott publikációja jelentős vitát váltott ki, melyet vázlatosan ismertetek az alábbiakban. Scott és munkatársai egy RDS módszerrel végzett kutatást követően a kutatás részvevőivel készítettek interjúkat, illetve a kutatás során végeztek résztvevő megfigyeléseket. Eredményeik alapján elsősorban a módszerhez kapcsolódó jutalmazási rendszer alkalmazása és annak következményei vezetnek olyan folyamatokhoz a vizsgált célcsoportban, amelyek etikai kérdéseket vetnek fel. Ezek a folyamatok a következők voltak:

- „Tülekedés a pénzért”: a jutalmazás célcsoportban elterjedő híre a pénzforrásért való harchoz vezetett, ami létrehozta a kutatáshoz kapcsolódó kuponok másodgazdaságát
- Kuponátadás, mint vállalkozási forma: a válaszadók ajánlásaikat 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. Ezeket a kapcsolatokat vizsgálva Scott és munkatársai arra jutottak, hogy ezekben gyakran

⁶⁹ A fejezet a szerző Statisztikai Szemlében megjelent cikke alapján készült (Simon [2012])

szerepeltek valamilyen fertőzést hordozó (HIV, hepatitis) személyek mellett nem fertőzött személyek is

- Élethelyzet különbségeinek növekedése: 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: azok, akik rendszeresen használt belövőhelyeket tartottak fenn 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 elhitetni 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, a mintavételt is károsan befolyásolhatták (pl.: a belövőhelyek tulajdonosai által szervezett rekrutálás).

A felvetett problémákra négy cikk válaszolt és egy viszontválasz született, ezeket egymáshoz kapcsolódóan foglalom össze.

Broadhead [2008] válaszcikkében elsősorban amellel érvel, hogy a fent leírt problémák nem egyediek az RDS módszer kapcsán, 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 Scott kutatómódszertana 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. Ezen túl elsősorban a Broadhead hangvételét és túlzó kijelentéseit kritizálja.

Lansky és Mastro [2008] válaszcikkükben szintén említik, hogy a Scott által leírt problémák jelentős része más módszerrel, de hasonló célcsoportra vonatkozó kutatás esetén felmerült már

korábban is. Ezen túl a szerzők kiemelik, hogy a Centers for Disease Control (CDC)⁷⁰ finanszírozásával folyó kutatás szigorú protokollt követett, amelynek során nagy hangsúlyt fektettek a tájékoztatásra, az informált beleegyezésre, valamint a vizsgált területen esetlegesen előforduló nem szándékolt események, incidensek folyamatos követésére. A szerzők (Broadheadhez hasonlóan) érvelnek amellett is, hogy a kutatás során alkalmazott jutalmak összege olyan csekély volt, hogy az nem lehetett alkalmas arra, hogy a vizsgált közösség gazdasági és társadalmi viszonyait jelentősen befolyásolja. Scott [2008b] viszontválaszában kiemeli, hogy cikkében nem állította, hogy a jutalmazási rendszer az egész közösségre jelentős hatást gyakorolt volna, csupán arra mutatott rá, hogy az általa ismertetett jelenségek megjelentek a közösségben és egyes résztvevők esetén jelentős hatást gyakoroltak.

Oullet [2008], mint Chicagóban a droghasználatot kutató etnográfus három ponton fogalmaz meg kritikát Scott kutatásával kapcsolatban: Scott csaknem teljesen figyelmen kívül hagyja a lokális kontextust, tévesen hiszi, hogy saját kutatása nem befolyásolja a közösségi folyamatokat, illetve, hogy a kutatása során folytatott interjúk esetén a válaszadók a teljes igazságot mondják el, végül Scott tévesen értelmezi a hatalmi viszonyokat. Scott viszontválaszában felhívja a figyelmet, hogy egyes jelenségek még azelőtt történtek, hogy a kutatókról kiderült volna, hogy részt vesznek a kutatásban, így megfigyelői hatás nem érvényesülhetett (Scott [2008b]).

Prachand és Benbow [2008] az RDS kutatást végző hatóság (Chicago Department of Public Health) részéről azt az etikai problémát vetették fel Scotttal szemben, hogy publikált cikke alapján olyan kutatást végzett az RDS kutatás résztvevői között, amelyet senki nem felügyelt és amely túlmutatott saját kutatási megbízásán (Scott az RDS kutatásban azt a feladatot kapta, hogy etnográfiai módszerekkel járuljon hozzá a mintavételi folyamat javításához). Viszontválaszában Scott kiemeli, hogy saját egyetemének intézményi ellenőrző testülete (IRB⁷¹) dokumentumokkal bizonyítható módon felügyelte kutatási tevékenységét, valamint azt, hogy a megfogalmazott kutatási célokat nem lépte túl (Scott [2008]).

A fenti vitától függetlenül Seeman [2008] az RDS módszerrel végzett AIDS kutatások során felmerülő etikai és szabályozási kérdéseket összefoglaló cikkében első helyen szintén a jutalmak kérdését emeli ki. Seeman szerint az RDS esetén alkalmazott alacsony jutalmak esetén általánosan elfogadott, hogy nem okoznak nem szándékolt hatásokat (ez alatt értve olyan

⁷⁰ A Centers for Disease Control az Egyesült Államok kormányának országos közegészségügyi intézménye.

⁷¹ Az Egyesült Államokban intézményi ellenőrző testületek felügyelik a humán kutatásokat.

cselekményeket, amelyeket a kutatás által érintett személyek a jutalom nélkül nem tennének meg). A módszer alkalmazásának etikai kérdéseit a 18. táblázat foglalja össze.

18. táblázat HIV-AIDS kutatások során felmerülő etikai problémák, kérdések és megoldások, válaszok

Terület	Problémák	Megoldások, válaszok
A résztvevők tiszteletben tartása	A jutalmat a résztvevők kábítószer vásárlásra használhatják.	Egy idézett kutatás (Festinger [2015]) 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 drog adag árának töredékének feleltek meg.
A kölcsönösség szociológiai és pszichológiai hatása	A jutalmak olyan érzetet kelthetnek a részvevőben, hogy azért cserébe teljesíteniük kell valamit.	A szerző érvei szerint a felhasznált időért és fáradtságért 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 teszt eredmény ismeretében vagy a teszt végzését megelőzően). Fel kell ajánlani segítséget a partnerekkel történő megbeszéléshez.

Forrás: Seemaan [2009]

Az itt összefoglalt problémák és válaszok jelentős része bármely célcsoport esetén fennáll, de fontos, hogy adott célcsoport esetén a kutatás tervezési fázisában hasonló lista készüljön arról, hogy a célcsoport kutatásban résztvevő tagjai esetén milyen etikai problémák merülhetnek fel, illetve terv készüljön azzal kapcsolatban, hogy hogyan lehet a nem szándékolt hatásokat elkerülni, illetve monitorozni a kutatás során.

VI.3. A válaszadó-vezérelt mintavétel jogi és etikai környezete

Az ötödik fejezet zárásaként érdemes megvizsgálni, hogy a fejezet fényében hogyan értelmezhető a két kutatási kérdés, a válaszadó-vezérelt mintavétel módszertani megfelelése, illetve milyen kapcsolatok, hatások merültek fel a vizsgált módszerre tudományos konstrukcióként.

A I.1 alfejezetben megfogalmazott szempontokat közvetlenül ebben a fejezetben nem vizsgáltam, ugyanakkor az itt leírtak erőteljesen befolyásolják a vizsgált módszer alkalmazhatóságát, így módszertani szempontból sem mellékesek.

A I.2 alfejezetben megfogalmazott tudományos konstrukció modell számos környezeti eleme közül jogi és etikai környezetet vizsgáltam a jelen fejezetben. A jogi környezet komoly problémát jelent a válaszadó-vezérelt mintavétel alkalmazása kapcsán a személyes adatok védelmére vonatkozó szabályok (Infotv., GDPR), illetve annak alkalmazása miatt. Míg a jogszabályok érvelésem alapján lehetővé tehetnék a válaszadó-vezérelt mintavétel lebonyolítását az idézett Adatvédelmi Biztos Hivatal állásfoglalás a jogszabály értelmezésével lényegesen megnehezíti azt. Az etikai környezet elsősorban a vizsgált célcsoportok rejtett jellege, illetve a kutatáshoz esetlegesen kapcsolódó jutalmazás miatt válik fontossá. Az idézett kutatások ugyan rámutatnak arra, hogy a módszer alkalmazása beavatkozást – esetenként negatív hatást – jelenthet a vizsgált célcsoport számára (a válaszadók aktív részvétele miatt ez a hatás jelentősebb, mint más hasonló mintavételek esetén), ugyanakkor ezek a hatások megfelelő odafigyeléssel csökkenthetők vagy kiküszöbölhetők.

VII. Következtetések

Dolgozatom zárófejezetében az I. fejezetben feltett kutatási kérdésekre adok – néhol nem végleges – választ. Az első alfejezetben a statisztikai módszertani szempontokra térek ki, azaz megvizsgálom, hogy az I.1 fejezetben megfogalmazott társadalomkutatási eljárásokkal szemben megfogalmazott elvárásoknak a II. és főként a IV. fejezetben ismertetett kutatások, illetve részben általam elvégzett vizsgálatok alapján a válaszadó-vezérelt mintavétel milyen mértékben felelnek meg. A második alfejezetben arra keresek választ, hogy a válaszadó-vezérelt mintavétel milyen módon áll kölcsönhatásban különböző társadalmi jelenségekkel és a szociológiai kutatással. Ez utóbbi vizsgálathoz az I.2 alfejezetben megfogalmazott elemzési keretet felhasználva az elemzés összes további fejezetének eredményeit felhasználom. Végül a fejezet harmadik alfejezetében rámutatok arra, hogy az első két alfejezetben adott válaszok közül mi tekinthető pusztán a dolgozat megszületésekor (vagy már akkor sem) érvényes válasznak és mit tekintek a dolgozat maradandó hasznának. Ebben az alfejezetben térek ki a további kutatási lehetőségekre is.

VII.1. A válaszadó-vezérelt mintavétel módszertani szempontból

A záró fejezet első alfejezetében a válaszadó-vezérelt mintavétel társadalomtudományi módszertani megfelelése kapcsán az I.1 alfejezetben megfogalmazott hipotézisek teljesülését vizsgálom. Az alfejezetben előbb a négy alhipotézis, majd a főhipotézis teljesülését értékelem.

H_{1a} : A válaszadó-vezérelt mintavétel alkalmas a vizsgált célcsoportra vonatkozó érvényes következtetések levonására.

Első alhipotézisem a válaszadó-vezérelt mintavétel érvényességére vonatkozott. Ezt a kérdést elsősorban dolgozatom II. és IV. fejezetében vizsgáltam. A II. fejezetben idézett tanulmányok alapján kijelenthető, hogy ideális körülmények között a válaszadó-vezérelt mintavétel alapján matematikailag igazolható módon adható torzítatlan becslés a populációs paraméterekre. A IV. fejezet módszertani kutatásai és saját szimulációs vizsgálatom is alátámasztotta ezt az állítást. Ugyanakkor a becslő eljárások nagy száma és eltérő – javasolt – alkalmazási feltételeik arra utalnak, hogy a valós körülmények kihívások elé állítják a vizsgált mintavételi módszert. Ezek a kihívások elsősorban azonban a megbízhatóság terén jelentkeznek.

H_{1b} : A válaszadó-vezérelt mintavétel megbízhatósága megbecsülhető és csak annyival kisebb, mint a hagyományos mintavételeké, hogy az még az alkalmazhatóságának nem gátja.

A IV. fejezetben számos kutatási eredményt bemutattem, amelyek a megbízhatóságot befolyásoló tényezőket vizsgálják. E téren a legtöbbet vizsgált a homofília (az egymással kapcsolatban állók hasonlóságának valószínűsége) és az ezzel kapcsolatban hozható hálózati klaszterezettség, amely általánosságban kedvezőtlenül befolyásolja a vizsgált módszer megbízhatóságát, de eltérő körülmények között eltérő becslést, illetve módosított mintavételi design-t alkalmazva a hatás csökkenthető.

H_I: Főhipotézisem szerint a válaszadó-vezérelt mintavétel a társadalomtudományi kutatás módszertani követelményeinek megfelel.

Összességében az alhipotézisek vizsgálata alapján elmondható, hogy a főhipotézis – megszorításokkal teljesül. Megszorítás alatt is a társadalmi környezet megfelelőségét értem, amelynek vizsgálata fontos minden konkrét kutatás esetén a módszer alkalmazhatósága szempontjából.

VII.2. A válaszadó-vezérelt mintavétel szociológiai kontextusa

A második alfejezetben a I.2 alfejezetben megfogalmazott válaszadó-vezérelt mintavétel társadalmi, elméleti kontextusához kapcsolódó hipotéziseit vizsgálom meg, ismét az alfejezet végére hagyva a főhipotézis teljesülésének ellenőrzését.

H_{IIa}: A válaszadó-vezérelt mintavétel az alkalmazói környezet matematikai ága szempontjából kevés kihívást jelent.

A II. fejezetben bemutattem, hogy matematikai szempontból a válaszadó-vezérelt mintavétel egy gráfon végrehajtott véletlen bolyongás probléma, amely Markov-lánc Monte Carlo mintavételként jellemezhető és mint ilyen ismert és jól közelíthető tulajdonságokkal rendelkezik. Matematikai kihívást elsősorban az eljáráshoz kapcsolódó becslő módszerek megbízhatóságának becslése jelent, amit sok esetben általános megoldásokkal (bootstrap tesztekkel) pótolnak.

H_{IIb}: A válaszadó-vezérelt mintavétel az alkalmazói környezet módszertani ága szempontjából jelentős kihívást jelent, ami az alkalmazási terület, célcsoport társadalmi sajátosságai miatt alakul ki.

A IV. fejezetben ismertetett módszertani irodalom és saját kutatásunk alapján az alkalmazói környezet heterogenitása és nagy számú probléma megjelenése állapítható meg. Mindezt alátámasztja a nagy számú módszertani ajánlás, amelyet az idézett kutatások – beleértve saját

kutatásunkat is – megfogalmazzuk. A III. fejezetben vizsgált nemzetközi és hazai kutatások alapján jellemző, hogy a válaszadó-vezérelt mintavétel alkalmazói rendszeresen kötnek az eredményeket jelentősen befolyásoló kompromisszumokat az alkalmazás során. Ezekre a kompromisszumokra azonban általában a kivitelezhetőség miatt szükség van, hatásukat azonban nem feltétlenül mérik fel.

H_{IIc}: A válaszadó-vezérelt mintavétel szociológiai kutatási környezete jelenlegi tendenciái alapján megrendelőként lép(het) fel a módszer kapcsán az elméletekben megfogalmazott és empirikusan igazolt társadalmi változások miatt.

A V. fejezetben igyekeztem igazolni, hogy mind az empirikus kutatások (életstílus és státusz összefüggése Róbert [2000]), mind a kortárs elméletek (kockázattársadalom kialakulásának hatása (Beck [2003], hagyományos társadalmi osztályképződés lehetőségének csökkenése (Angelusz [2000])) a nehezen körülhatárolható csoportok jelentőségének növekedését támasztják alá. Ez a körülmény pedig kedvező környezetet teremt az ilyen célcsoportok vizsgálatára alkalmas mintavételi eljárásoknak, így a válaszadó-vezérelt mintavételnek is, ami igazolja a hipotézist.

H_{IIId}: A jogi és etikai környezet a válaszadó-vezérelt mintavétel alkalmazásának komoly korlátjává válhat, akár el is lehetetleníti a módszer alkalmazását.

A jogi és etikai környezetet a VI. fejezetben vizsgáltam. A jogi környezet komoly problémát jelent a válaszadó-vezérelt mintavétel alkalmazása kapcsán a személyes adatok védelmére vonatkozó szabályok (Infotv., GDPR), illetve annak alkalmazása miatt. Míg a jogszabályok érvelésem alapján lehetővé tehetnék a válaszadó-vezérelt mintavétel lebonyolítását az idézett Adatvédelmi Biztosi Hivatal állásfoglalás a jogszabály értelmezésével lényegesen megnehezíti azt. Az etikai környezet elsősorban a vizsgált célcsoportok rejtett jellege, illetve a kutatáshoz esetlegesen kapcsolódó jutalmazás miatt válik fontossá. Az idézett kutatások ugyan rámutatnak arra, hogy a módszer alkalmazása beavatkozást – esetenként negatív hatást – jelenthet a vizsgált célcsoport számára (a válaszadók aktív részvétele miatt ez a hatás jelentősebb, mint más hasonló mintavételek esetén), ugyanakkor ezek a hatások megfelelő odafigyeléssel csökkenthetők vagy kiküszöbölhetők.

H_{IIj}: Főhipotézisem szerint a válaszadó-vezérelt mintavétel alkalmazását és alkalmazhatóságát befolyásolja a jogi, etikai, társadalmi, megrendelői és alkalmazói környezete.

Az alhipotézisek vizsgálatát követően kijelenthető, hogy a főhipotézis teljesült.

VII.3. Záró megjegyzések és további kutatási lehetőségek

Végezetül a záró fejezetben néhány összefoglaló megjegyzés mellett további kutatási irányokra hívnám fel a figyelmet szűken véve a válaszadó-vezérelt mintavétel, illetve tágabban a társadalomtudomány módszertani kérdéseinek vizsgálata szempontjából. Az alfejezet végén szeretnék köszönetet mondani mindazoknak, akik nélkül a dolgozat nem jöhetett volna létre.

Dolgozatomban módszertani szempontból talán több kérdését vetek fel a válaszadó-vezérelt mintavétel kapcsán, mint, amelyre választ adok. Ezek közül itt összefoglalom a számomra most legfontosabbnak tűnőket. Láttuk, hogy a megbízhatóságot jelentős mértékben rontja a hálózatok klaszterezettsége, illetve a jelentős mértékű homofília. Ígéretes megoldásnak tűnik a paraméterek figyelemmel kísérése a mintavétel során és adaptív mintavétel alkalmazása – azonban ennek megbízhatóságot növelő hatásáról még nem áll rendelkezésre kellő információ. Nagyobb társadalmi csoport, mint például a külföldre vándorlók esetén célravezetőnek tűnik a közösségi portálokon történő hálózati mintavétel. Milyen mértékű lefedettségi hiba jelentkezik ilyenkor és ez milyen mértékű torzítással jár? Vajon a vonatkozó – esetenként nemzetközi – jog keretek egyáltalán lehetővé teszik ezt a fajta mintavételt? Nem igazán használták fel eddig a válaszadó-vezérelt mintavételből származó minták hálózati adatait. Vajon milyen típusú becslések adhatók a célcsoport hálózati struktúrájára, illetve a hálózati kapcsolatok és a vizsgált szereplők tulajdonságai (jellemző, cselekvései) kapcsolatára? További lehetséges kutatási irány a bizalom és egyéb szociálpszichológiai tényezők szerepe a válaszadó-vezérelt mintavétel esetén.

Dolgozatom célja volt, hogy egy konkrét módszer vizsgálata mellett szerén példát adjak a társadalomtudományi módszertani kérdések társadalmi és társadalomtudományi beágyazottságának vizsgálatára. Reményeim szerint a dolgozat eredményei alátámasztják azt a véleményemet, amely szerint a társadalomtudományi kutatás módszertana elválaszthatatlan a vizsgált társadalmaktól és a társadalomtudományi elméletektől, ami természetesen nem jelenti a módszerek bármiféle elfogultságát.

A jelen dolgozathoz hozzájárultak az OTKA 101067 „Migráció az egészségügyben: jelenségek és magyarázatok”, illetve 120711 „Migráció – a valóság közelítése innovatív módszerekkel” kutatási eredményei, melyekben a szerző részt vett.

Végezetül szeretném megköszönni a segítséget mindazoknak, akik nélkül a dolgozat nem jöhetett volna létre. Mivel a számos különféle segítség módjának felsorolása túl hosszúvá nyúlna, így csak betűrendben felsorolom azokat, akik nélkülözhetetlen segítséget adtak: Balogh

Anikó, Barna Ildikó, Csepeli György, Dukay-Szabó Szilvia, Fokasz Nikosz, Hárs Ágnes, Katona Eszter, Kmetty Zoltán, Koltai Júlia, Németh Renáta, Örkény Antal, Pillók Péter, Rudas Tamás, Székelyi Mária, Tausz Katalin, Zerinváry Barbara.

VIII. Irodalom

- Aldrich, J. [2008]: Professor A.L. Bowley's theory of the representative method. Southampton, GB, University of Southampton (Discussion Papers in Economics and Econometrics 0801).
- Angelusz R. [2000]: Rétegződés és láthatóság. In: A láthatóság görbe tükrei. Budapest, Új Mandátum, 2000.
- Babbie, Earl [2008]: A társadalomtudományi kutatás gyakorlata. Balassi könyvkiadó, 2008
- Balogh A. – Simon D. [2011]: Nehezen körülhatárolható társadalmi csoportok kutatása. Digitális Tankönyvtár, 2011 Internet: <http://www.tankonyvtar.hu/>
- Barabási AL - Albert R - Jeong H.[1999]: Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*. 1999 Oct 1;272(1-2) pp 173-87.
- Bastos, L. S., Pinho, A. A., Codeço, C., & Bastos, F. I. [2012]: Binary regression analysis with network structure of respondent-driven sampling data. arXiv preprint arXiv:1206.5681.
- Beck, U. [2003]: A kockázat-társadalom. Budapest, Századvég Kiadó, 2003. pp 138-181.
- Bernát A. [2006]: A kapcsolathálózat jelentősége a migrációban – etnikai metszetben. In: Némédi D. – Somlai P. – Szabari V. – Szikra D. (szerk.): *Kötő-jelek 2005*. ELTE Társadalomtudományi Kar Szociológia Doktori Iskola. Budapest. pp 123–146.
- Biernacki P. – Waldorf P. [1981]: Snowball sampling. *Sociological Methods & Research* Vol. 10. No. 2. 1981. pp 141-163.
- 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* Volume 4, Issue 1, 2009 pp 96-109
- Blaskó, Zs. [2014]: Studying emigration by extending a large scale household survey. Methodology, evaluation and descriptive findings. Working Papers on Population, Family and Welfare, No. 21, Hungarian Demographic Research Institute, Budapest.
- Blau, Peter M. [1997]: Egyenlőtlenség és heterogenitás. Primitív elmélet a társadalmi struktúráról. (Ford.: Szalai Éva.) In Angelusz Róbert (szerk.): *A társadalmi rétegződés komponensei*. Budapest: Új Mandátum. 359–382. old. (Eredeti megjelenés: 1977.)
- Bolla Zs. [2011]: Online kutatási panelek Magyarországon. Szakdolgozat. ELTE Társadalomtudományi Kar, Survey Statisztika MSc. Konzulens: Kende Gábor
- Bozsonyi K. – Jelenfi G, Kmetty Z. [2008]: A Magyarországi aktív banki ügynökök, In: V. HUNNET kapcsolathálózati konferencia, 2008 Budapest
- Brewer K. – Timothy G. G. [2009]: Introduction to Survey Sampling. In: Pfeffermann D. – Rao C. R. (ed.) [2009]: *Sample Surveys: Design, Methods and Applications*. Elsevier, 2009. pp 9-37.
- Broadhead, R. S. [2008]: Notes on a cautionary (tall) tale about respondent-driven sampling: A critique of Scott's ethnography. In: *International Journal of Drug Policy* 19 (2008) pp 235–237.

- Crawford FW, Aronow PM, Zeng L, Li J. [2017]: Identification of homophily and preferential recruitment in respondent-driven sampling. *American journal of epidemiology*. 2017 Jun 9;187(1):153-60.
- Csanády M. – Kmetty Z. – Kucsera T – Személyi L. – Tarján G. [2008]: A magyar képzett migráció a rendszerváltás óta, *Magyar Tudomány* 2008/05 pp 603-616.
- Dávid B. – Snijders T. [2000]: A budapesti hajléktalanok számának becslése. In: *Szociológiai Szemle* 2000/3, Budapest. pp: 60-75.
- de Bussya, N. M. – Wolf, K. [2009]: The state of Australian public relations: Professionalisation and paradox. *Public Relations Review* Volume 35, Issue 4, November 2009, pp 376–381
- Domokos T. - Fábíán R. - Horváth G. Cs. - Márványkövi F. - Mervó B. - Rác J. [2010]: A szükséglet-meghatározás nemzetközi és hazai tapasztalatainak, az alkalmazott eszközök használhatóságának összefoglalása. *Szociálpolitikai és Munkaügyi Intézet*. Internet: <http://nbn.urn.hu/N2L?urn:nbn:hu-99192> (Utolsó letöltés: 2011. július 22.)
- Erdős P - Rényi A. [1959]: On random graphs. *Publicationes Mathematicae*, 1959. pp 6-290
- Festinger, D. S., Marlowe, D. B., Croft, J. R., Dugosh, K. L., Mastro, N. K., Lee, P. A., és mtsai. [2005]: Do research payments precipitate drug use or coerce participation? *Drug and Alcohol Dependence*, 78, 275–281.
- Frank O. [1977]: Survey Sampling in Graphs. *Journal of Statistical Planning and Inference* Volume 1, Issue 3, 1977 pp 235–264
- Frankfort-Nachmias Ch. – Leon-Guerrero A. [2011]: *Social Statistics for a Diverse Society*. SAGE 2011.
- Freedman, D. - Pisani, R. - Purves, R. [2005]: *Statisztika*. Typotex, 2005.
- Gile, K. J. – Handcock, M. S. [2009]: Respondent-Driven Sampling: An Assessment of Current Methodology. *Sociological Methodology*. Vol. 40. No. 1. pp. 285–327.
- Gile, K. J. (2011). Improved inference for respondent-driven sampling data with application to HIV prevalence estimation. *Journal of the American Statistical Association* 106 (493), pp 135-146.
- Gile, K. J. – Johnston, L. G. – Salganik M. J. [2015]: Diagnostics for Respondent-driven Sampling. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(1), pp.241-269.
- Goel S – Salganik M. J. [2009]: Respondent-driven sampling as Markov chain Monte Carlo. *Statistics in Medicine*. 2009;Vol. 28 pp. 2202-2229.
- Goel, S. – Salganik, M. J. [2010]: Assessing respondent-driven sampling . *PNAS* Volume 107 no. 15 pp. 6743-6747
- Goodman L. A. [1961]: Snowball sampling; *Annals of mathematical statistics* 32:223-258
- Guerrero-Bote, V. P. - Moya-Anegón F. [2012]: A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics* 6 (2012) pp 674–688

Guntuboyina, A., Barbour, R., & Heimer, R. [2012]: On the impossibility of constructing good population mean estimators in a realistic Respondent Driven Sampling model. arXiv preprint arXiv:1209.2072.

Hárs Á. – Simon D. [2009]: Ingázók, bevándorlók, munkaerő-migránsok? Munkaerő mobilitása konferencia – Országos Foglalkoztatási Alapítvány, Budapest

Hárs, Á. [2009]: Magyarok az osztrák munkaerőpiacon: Ingázók, bevándorlók, munkaerő-migránsok? KOPINT Konjunktúra Kutatási Alapítvány. Internet: http://www.ofa.hu/index.php?WG_NODE=WebIntRedirect&WG_OID=DSDf6ed153f74d5d7650 (Utolsó letöltés: 2011. július 22.)

Hárs Á. - Simon D. (2016): A magyarországi orvosok külföldi munkavállalását befolyásoló befolyásoló tényezők. In: Blaskó Zsuzsa, Fazekas Károly (szerk.) Munkaerőpiaci tükrök 2015. 279 p. Budapest: MTA Közgazdaság- és Regionális Tudományi Kutatóközpont Közgazdaságtudományi Intézet, 2016. pp. 96-103.

Heckathorn, D D [1997]: Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations. *Social Problems*, Volume 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*, Volume 37 pp. 151–207.

Heckathorn, D D [2011]: Comment: Snowball Versus Respondent-Driven Sampling. In: *Sociological Methodology* Volume 41, Issue 1, pp 355–366

de Heer, W. – de Leeuw E. D. – van der Zouwen J. [1999]: Methodological Issues in Survey Research : a Historical Review. *Bulletin de Méthodologie Sociologique*. 1999 64: 25 pp 25-48.

Hárs, Á., & Simon, D. [2016]: 2.5 FACTORS AFFECTING THE INTERNATIONAL LABOUR MIGRATION OF MEDICAL DOCTORS IN HUNGARY. *THE HUNGARIAN LABOUR MARKET 2016*, 97.

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 2008/1*, pp.153-171

Johnston, L. - Grazina és mtsai. [2008]: Implementation Challenges to Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance: Field Experiences in International Settings. In: Volume 12, Supplement 1 pp 131-141

Kapitány B. [2010]: Mintavételi módszerek ritka populációk esetén. *Statisztikai Szemle*, 88. évfolyam 7-8. pp 739-754

Kendall, C és mtsai. [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* Volume 12, Supplement I pp 97-104

Kemény I. - Rupp K. - Csalog Zs. - Havas G. [1976]: Beszámoló a magyarországi cigányok helyzetével foglalkozó 1971-ben végzett kutatásról. MTA Szociológiai Kutató Intézetének kiadványai, Budapest, 1976.

- Kertesi G. – Kézdi G. [1998]: A cigány népesség Magyarországon. Budapest, socio-typo, 1998.
- Khabbazian, M., Hanlon, B., Russek, Z. and Rohe, K., 2017. Novel sampling design for respondent-driven sampling. *Electronic Journal of Statistics*, 11(2), pp.4769-4812.
- Kmetty Zoltán - Simon Dávid (2013): A válaszadó-vezérelt mintavétel megbízhatóságának vizsgálata szimulációs módszerekkel. *Szociológiai Szemle* 2013/2.
- Kmetty Z. – Sztárayné Kézdi É. [2011]: Kutatásmódszertan és prezentációkészítés: 9. rész: Mintavétel a gyakorlatban. A Károli Gáspár Református Egyetem digitális tananyaga. Internet: http://igyk.pte.hu/files/tiny_mce/File/kari_projektek/informaciotudaservenyesules/tananyagok/1_alprojekt/KM/09_kutatasmodzertan9_mintavetel.pdf (utolsó letöltés: 2014. április 21.)
- Kurant, M. – Gjoka, M. – Butts, C. T. Markopoulou A. [2011]: Walking on a Graph with a Magnifying Glass: Stratified Sampling via Weighted Random Walks. *SIGMETRICS*, 2011 pp 281-292
- Kral Alex H. és mtsai. [2010]: Comparing respondent-driven sampling and targeted sampling methods of recruiting injection drug users in San Francisco. In: *J Urban Health*. 2010 szeptember; 87(5) pp 839-50.
- Lansky A. – Mastro T. D. [2008]: Using respondent-driven sampling for behavioural surveillance: Response to Scott. In: *International Journal of Drug Policy* 19 (2008) pp 241–243
- Liu, H., Li, J., Ha, T. and Li, J., [2012]: Assessment of random recruitment assumption in respondent-driven sampling in egocentric network data. *Social networking*, 1(2), p.13.
- Lohr, S. L. [2009]: Multiple-Frame Surveys. In: Pfeffermann D. – Rao C. R. (ed.) [2009]: *Sample Surveys: Design, Methods and Applications*. Elsevier, 2009. pp 71-88.
- Lu X. [2012]: Linked Ego Networks: Improving estimate reliability and validity with respondent-driven sampling. *Social Networks* Vol. 35 (2013) pp.: 669– 685.
- Lu X. – Malmros J. – Liljeros F. – Britton T. [2012a]: Respondent-driven sampling on directed networks. *Electronic Journal of Statistics* Vol. 7 2013 pp 292–322
- Lu X. – Bengtsson L. – Britton T. – Camitz M. – Beom J. K. – Thorson A. – Liljeros F. [2012b]: The sensitivity of respondent-driven sampling. *Journal of the Royal Statistical Society Series A* Vol. 175, No. 1, 2012. pp. 191–216
- Lunagómez, S – Airolidi E M [2012]: Bayesian Inference from Non-Ignorable Network Sampling Designs. Conference paper. *Social Network and Social Media Analysis: Methods, Models and Applications*, December 7, 2012 Nevada, Lake Tahoe Internet: <http://snap.stanford.edu/social2012/papers/lunagomez-airolidi.pdf> (Utolsó elérés: 2013. november 23.)
- Malekinejad, M és mtsai.[2008]: Using Respondent-Driven Sampling Methodology for HIV Biological and Behavioral Surveillance in International Settings: A Systematic Review. In: *AIDS and Behavior* Volume 12, Supplement 1. 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. In: *ADDIKTOLÓGIA – 2008. VII. ÉVFOLYAM* 3–4. szám

- McLaughlin K.R., Handcock M.S., Johnston L.G., Japuki X, Gexha-Bunjaku D, Deva E.[2015]: Inference for the visibility distribution for respondent-driven sampling. JSM Proceedings 2015, Statistical Computing Section. 2015.
- Mills, H.L., Johnson, S., Hickman, M., Jones, N.S. and Colijn, C., [2014]: Errors in reported degrees and respondent driven sampling: implications for bias. *Drug and alcohol dependence*, 142, pp.120-126.
- Muhib Farzana B, Lin LS, Stueve A, Miller RL, Ford WL, Johnson WD, Smith PJ [2001]: A venue-based method for sampling hard to reach populations. *Public Health Reports*. 2001;116 (1):216-22.
- Nesterko, S. and Blitzstein, J., [2015]: Bias–variance and breadth–depth tradeoffs in respondent-driven sampling. *Journal of Statistical Computation and Simulation*, 85(1), pp.89-102.
- Ott MQ, Gile KJ. [2016] Unequal edge inclusion probabilities in link-tracing network sampling with implications for respondent-driven sampling. *Electronic Journal of Statistics*. 2016;10(1):1109-32.
- Ott MQ, Gile KJ, Harrison MT, Johnston LG, Hogan JW. [2017]: Reduced Bias for respondent driven sampling: accounting for non-uniform edge sampling probabilities in people who inject drugs in Mauritius. arXiv preprint arXiv:1712.09149. 2017 Dec 26.
- Ouellet L. J. [2008]:Cautionary comments on an ethnographic tale gone wrong. In: *International Journal of Drug Policy* 19 (2008) pp 238–240
- Pfeffermann D. – Rao C. R. (ed.) [2009]: *Sample Surveys: Design, Methods and Applications*. Elsevier, 2009.
- Pintér, R. – Kátay B. [2010]: A hibrid adatfelvétel módszertani kihívásai. *Statisztikai Szemle*, 88(7-8), pp.723-738.
- 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. In: *International Journal of Drug Policy* 19 (2008) pp 244–245.
- Róbert P. [2000]: Az életstílus meghatározottságának változása, 1982-1998. In: *Szociológiai Szemle*, 2000. 2. szám.
- Salganik, M. J. - Heckathorn , D. D. [2004]: Sampling and Estimation in Hidden Populations Using Respondent-Driven Sampling . *Sociological Methodology*, Vol. 34, 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
- Schonlau M – Liebau E [2012]:Respondent-driven sampling. In: *The Stata Journal* Vol. 12, Number 1, pp. 72–93
- 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. In: *International Journal of Drug Policy* 19 (2008) pp 42-55

- Scott, G. [2008b]: Responses on a cautionary tale concerning the ethics of using respondent-driven sampling to study injection drug users. In: *International Journal of Drug Policy* 19 (2008) pp 246–247
- Semaan , S. [2009]: Ethical and regulatory considerations in HIV prevention studies employing respondent-driven sampling. In: *International Journal of Drug Policy* 20 (2009) pp 14–27
- Simon D. [2005]: Kapcsolathálózati mintavételi módszer a romák migrációs potenciáljának vizsgálatára: módszertani megfontolások. II. HUNNET kapcsolathálózti konferencia, 2005 Budapest
- Simon D. [2010]: A budapesti romák – kutatási jelentés. Társadalmi Konfliktusok Kutatóközpont, ELTE TáTK, Budapest, 2010 Internet: http://konfliktuskutato.hu/index.php?option=com_content&view=article&id=260:a-budapesti-romak-2010&catid=27:budapesti-romak (utolsó hozzáférés: 2019. február 14.)
- Simon D. [2012]: Válaszadó-vezérelt mintavétel: ritka és rejtett csoportok kvantitatív vizsgálata. *Statisztikai Szemle*, 90. évfolyam 4. szám pp 249-275
- Simmel, G. [1909]: The Problem of Sociology. *American Journal of Sociology* Vol. 15. No. 3. 1909 pp.: 289-320.
- Sirken, M. G.[1998]: “A Short History of Network Sampling.” North (1998) Internet: http://www.amstat.org/sections/srms/Proceedings/papers/1998_001.pdf
- Smith T. W. [2013]: Survey-Research Paradigms Old and New. *International journal of public opinion research* vol. 25, no. 2, 2013 pp. 218-229
- Snijders T. A. B. [1992]: Estimation on the Basis of Snowball Samples: How to Weight? *Buletin the Methodologie Sociologique* No. 36. 1992 pp 59-70.
- Spreen, M. [1992]: Rare populations, hidden populations, and link-tracing designs: what and why. In: *Bulletin de Méthodologie Sociologique* No. 36. September 1992, pp. 34-58.
- Szilágyi Péter [2011]: Jogi alaptan. ELTE Eötvös, Budapest, 2011.
- Tardos R. [1995]: A kapcsolathálózati megközelítés: Új paradigma? *Szociológiai Szemle*, 1995, 4. sz.
- Tomas, A. – Gile, K. J. [2010]: The Effect of Differential Recruitment, Non-response and Non-recruitment on Estimators for Respondent-Driven Sampling . Internet: <http://arxiv.org/abs/1012.4122v1> (Utolsó letöltés: 2011. július 22.)
- Velu, R – Naidu, G. M. [2009]: Survey Sampling Methods in Marketing Research: A Review of Telephone, Mall Intercept, Panel and Web Surveys. In: Pfeffermann D. – Rao C. R. (ed.) [2009]: *Sample Surveys: Design, Methods and Applications*. Elsevier, 2009. pp 513-538.
- Verdery, A. M., Mouw, T., Bauldry, S., & Mucha, P. J. [2015]: Network structure and biased variance estimation in respondent driven sampling. *PloS one*, 10(12), e0145296.
- Volz, E – Wejnert, C. – Degani, I. – Heckathorn, D. D. [2007]: Respondent-Driven Sampling Analysis Tool (RDSAT) Version 5.6. Ithaca, NY: Cornell University. Internet: <http://respondentdrivensampling.org/> (Utolsó letöltés: 2011. július 22.)

Volz, E – Heckathorn, D D [2008]: Probability Based Estimation Theory for Respondent Driven Sampling. *Journal of Official Statistics*, Vol. 24, No. 1, 2008, pp. 79–97

Wejnert C. [2009]: An Empirical Test of Respondent-driven Sampling: Point estimates, Variance, Degree Measures and Out-of-equilibrium data. *Sociological Methodology*. Volume 39 No. 1 pp. 73–116.

Wejnert C. – Pham H. – Krishna N. – Le B. – DiNenno E. [2012]: Estimating Design Effect and Calculating Sample Size for Respondent-Driven Sampling Studies of Injection Drug Users in the United States. *AIDS Behaviour* Vol. 16, No. 4, 2012 pp. 797–806

IX. Ábrajegyzék

1. ábra A válaszadó-vezérelt mintavétel, mint tudományos konstrukció modellje	10
2. ábra Mintavételi módszerek csoportosítása nehezen elérhető populációk vizsgálata esetén	16
3. ábra A válaszadó-vezérelt mintavétel megjelenése publikációkban: Salganik és Heckathorn 2004-ben, illetve Heckathorn 1997-ben publikált cikkének idézettsége, valamint a „respondent-driven sampling” kifejezés megjelenése tudományos publikációkban 2004 és 2018	37
4. ábra 2012-2013 között, illetve 2019 első félévében megjelent válaszadó-vezérelt mintavételre („Respondent-driven sampling”) hivatkozó publikációk közül véletlenszerűen kiválasztott 100-100 cikk témája (darab, illetve százalék)	38
5. ábra 2012-2013 között, illetve 2019 első félévében megjelent válaszadó-vezérelt mintavételre („Respondent-driven sampling”) hivatkozó publikációk közül véletlenszerűen kiválasztott 100-100 cikk kiadója (darab, illetve százalék)	40
6. ábra A demográfiai jellemzők, társadalmi-gazdasági státus és osztályhelyzet lehetséges összefüggései a kulturális és materiális életstílussal Róbert Péter szerint	88

X. Táblázatok jegyzéke

1. táblázat A nehezen elérhető társadalmi csoportok definiálása	15
2. táblázat Az RDS becslések evolúciója (Heckathorn [2011]) alapján kibővítve a fejezetben leírt becslésekkel)	32
3. táblázat RDS mintavétel főbb paraméterei 123 nemzetközi HIV fertőzéssel kapcsolatos kutatás metaelemzése alapján 2003-2007 között.....	42
4. táblázat RDS mintavétellel kapcsolatos siker és kudarc tényezők 128 nemzetközi HIV fertőzéssel kapcsolatos kutatás metaelemzése alapján	43
5. táblázat Válaszadó-vezérelt kutatások a hazai irodalomban	44
6. táblázat A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott kiinduló design eredménye.....	51
7. táblázat A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott első módosított mintavételi design	52
8. táblázat A magyarországi orvos elvándorlás válaszadó-vezérelt mintavétele kapcsán alkalmazott második módosított mintavételi design mintakialakítási fázisának leíró jellemzői	53
9. táblázat Az válaszadó-vezérelt mintavétel módszertani vizsgálata kapcsán idézett cikkekben használt módszerek sajátosságai	57
10. táblázat A válaszadó-vezérelt mintavétel módszerét vizsgáló idézett cikkek által figyelembe vett tényezők	59
11. táblázat A válaszadó-vezérelt mintavétel Voltz és Heckathorn [2008] által javasolt becslésének (saját hivatkozásomban RDS II becslés) feltételei	66
12. táblázat A heterofília, fokszám és kiinduló mintanagyság becsült hatása (a modell alapján kiszámított marginális átlagok és standard hibájuk)	75
13. táblázat A heterofília (H), a fokszám várhatóérték (E(D)) és a kiinduló minta méretének (NK) hatása az RDS I és RDS II becslés szórására (Levene-teszt eredményei alapján mindhárom tényező mindkét becslési módszer esetén szignifikáns eltérést hozott létre a becslések szórásában).....	76
14. táblázat A heterofília és a fokszám várható érték hatása a fokszám becslés hiba részesedésére az RDS I becslés varianciája esetén (a kezdőpontok számának hatását átlagolva)	78
15. táblázat Az összekötöttség, fokszám, kiinduló minta és egyéb hálózati hatások befolyása az RDS I és RDS II becslés torzítására és szórására.....	79
16. táblázat A társadalmi rétegződést meghatározó változó típusok Angelusz Róbert Duncan Blau nyomán létrehozott modellje szerint	84
17. táblázat A társadalmi nagycsoportok csoporttudatának tipológiája Angelusz szerint	85
18. táblázat HIV-AIDS kutatások során felmerülő etikai problémák, kérdések és megoldások, válaszok	102

XI. Összefoglaló

Dolgozatomban egy mintavételi eljárás, az úgynevezett válaszadó-vezérelt mintavétel (respondent driven sampling, RDS) alkalmazási kérdéseit járom körül. Dolgozatom célja azonban nem egyszerűen az eljárás bemutatása, a felmerülő szempontok (statisztikai, módszertani, etikai, jogi) áttekintése. A szűkebb módszertani megközelítésen túl dolgozatomban arra törekszem, hogy egy módszertani eszközt, mint a társadalomtudományi kutatás tárgyát, mint tudományos konstrukciót mutassam be, amely különféle társadalmi szereplőkkel és társadalmi jelenségekkel interakcióban működik, használható, értelmezhető.

Dolgozatomnak a leírtaknak megfelelően két olvasata van: egyrészt egy, a módszertan keretein belül értelmezhető elemzés egy konkrét módszer alkalmazhatósága kapcsán, másrészt egy módszer elhelyezése, kontextualizálása a társadalomtudományos, illetve szűkebben véve szociológiai fogalomtérben. Ennek megfelelően két kérdésre keresem a választ: mennyiben alkalmazható a válaszadó-vezérelt mintavétel a szociológiai kutatásokban a szociológiai módszertan elfogadott követelményei alapján és milyen módon áll kölcsönhatásban e módszer egyes társadalmi jelenségekkel és a szociológiai elméletekkel.

Dolgozatomban a felvetett kérdéseket hét fejezetben tárgyalom. Minden egyes fejezetben egyszerre próbálom megválaszolni mindkét fő kérdést, de visszautalásokkal, értelmezéssel igyekszem világossá tenni, hogy mely állításaim melyik kérdés szempontjából milyen jelentőséggel bírnak. Természetesen az egyes áttekintett témák különböző mértékben érintik a két feltett kérdést.

A bevezető fejezetben bővebben kifejtem dolgozatom fent leírt két kutatási kérdését. A második fejezetben a módszerrel kapcsolatos alapvető irodalomra támaszkodva magát az eljárást mutatom be. A harmadik fejezetben a módszer gyakorlati alkalmazását mutatom be a nemzetközi és hazai kutatások gyakorlatának összefoglalásán keresztül. A negyedik fejezet a válaszadó-vezérelt mintavétellel kapcsolatos szorosán vett módszertani kérdéseket tisztázza a vonatkozó irodalom összefoglaló, kritikai értelmezésén keresztül. E fejezet részeként egy saját szimulációs kutatás eredményeit is bemutatom a vizsgált módszer alkalmazhatóságához kapcsolódóan. Az ötödik fejezetben a vizsgált módszer és a társadalomtudományi elméletek között keresek kapcsolatot. A hatodik fejezetben a válaszadó-vezérelt mintavétel tágabb társadalmi kontextusaként a kapcsoló jogi és etikai kérdésekkel foglalkozom a vonatkozó irodalom, jogszabályok, illetve korábbi kutatásaim ilyen irányú „melléktermékei” kapcsán. Végül a hetedik fejezetben összefoglalom a feltett két kérdésre adott válaszaimat.

XII. Summary

In my dissertation I discuss the application issues of a sampling procedure called respondent driven sampling (RDS). However, the purpose of my dissertation is not simply to introduce the procedure, but to review the emerging aspects (statistical, methodological, ethical, legal) as well. Beyond the narrower methodological approach, in my dissertation I strive to present a methodological tool, as the subject of social science research, as a scientific construct that interacts, uses and interprets with various social actors and social phenomena.

As described above, my thesis has two readings: first, an analysis of the applicability of a particular method within the framework of the methodology, and the positioning and contextualization of a method in the social science and, more specifically, sociological terms. Accordingly, I seek answers to two questions: the extent to which respondent-driven sampling can be used in sociological research on the basis of accepted requirements of the sociological methodology and how this method interacts with certain social phenomena and sociological theories.

In my paper I discuss the issues raised in seven chapters. In each chapter, I try to answer both main questions at the same time, but I try to make clear by referring back and interpreting which of my statements are relevant to which issue. Of course, each of the topics covered affects the two questions to a different degree.

In the introductory chapter, I elaborate on the two research questions described above. In the second chapter, I will present the process itself, drawing on the basic literature on the method. In the third chapter, I present the practical application of the method through a summary of international and domestic research practice. The fourth chapter clarifies closely the methodological issues of respondent-driven sampling through a concise, critical interpretation of the relevant literature. As part of this chapter I also present the results of my own simulation research on the applicability of the method under investigation. In the fifth chapter I look for the connection between the method under investigation and social science theories. In Chapter Six, as a broader social context of respondent-driven sampling, I deal with the related legal and ethical issues of relevant literature, legislation, and by-products of my previous research in this regard. Finally, in Chapter Seven, I summarize my answers to the two research questions.

I. Melléklet

Hivatkozott, de a főszövegben nem idézett jogszabályi helyek

2016. évi CLV. törvény a hivatalos statisztikáról (statisztikai törvény)

„25. § (2) Különleges adat statisztikai célból a következők szerint gyűjthető:

- a) az érintett faji eredetére, nemzetiségi hovatartozására, politikai véleményére vagy pártállására, vallásos, világnézeti meggyőződésére, érdekképviselési tagságára vonatkozó adat csak személyazonosításra alkalmatlan módon és az érintett természetes személy önkéntes adatszolgáltatása alapján,
- b) az egészségi állapotra, kóros szenvedélyre, szexuális életre vonatkozó adat, valamint bűnügyi személyes adat csak személyazonosításra alkalmatlan módon, az érintett természetes személy önkéntes adatszolgáltatása vagy törvény rendelkezése alapján.”

2011. évi CXII. törvény az információs önrendelkezési jogról és az információszabadságról (Infotv.)

„3.§ (...)

9. adatkezelő: az a természetes vagy jogi személy, illetve jogi személyiséggel nem rendelkező szervezet, aki vagy amely önállóan vagy másokkal együtt az adatok kezelésének célját meghatározza, az adatkezelésre (beleértve a felhasznált eszközt) vonatkozó döntéseket meghozza és végrehajtja, vagy az általa megbízott adatfeldolgozóval végrehajtatja;

10. adatkezelés: az alkalmazott eljárástól függetlenül az adatokon végzett bármely művelet vagy a műveletek összessége, így különösen gyűjtése, felvétele, rögzítése, rendszerezése, tárolása, megváltoztatása, felhasználása, lekérdezése, továbbítása, nyilvánosságra hozatala, összehangolása vagy összekapcsolása, zárolása, törlése és megsemmisítése, valamint az adatok további felhasználásának megakadályozása, fénykép-, hang- vagy képfelvétel készítése, valamint a személy azonosítására alkalmas fizikai jellemzők (pl. ujj- vagy tenyérynnyomat, DNS-minta, iriszkép) rögzítése;”

Az Európai Tanács 2016/679 rendelete a természetes személyeknek a személyes adatok kezelése tekintetében történő védelméről és az ilyen adatok szabad áramlásáról (GDPR)

„4. cikk (...)

1. „személyes adat”: azonosított vagy azonosítható természetes személyre („érintett”) vonatkozó bármely információ; azonosítható az a természetes személy, aki közvetlen vagy közvetett módon, különösen valamely azonosító, például név, szám, helymeghatározó adat, online azonosító vagy a természetes személy testi, fiziológiai, genetikai, szellemi, gazdasági, kulturális vagy szociális azonosságára vonatkozó egy vagy több tényező alapján azonosítható;”

II. Melléklet

A Nemzeti Adatvédelmi és Információszabadság Hatóság állásfoglalása a válaszadóvezérelt mintavételről a SEEMIG kutatás kapcsán



Ügyszám: NAIH-6156-2/2012/M.
Ögyintéző: dr. Buzás Péter

AGDIA - 35/1/2012.

Dr. Lakatos Miklós részére
adatvédelmi felelős

Központi Statisztikai Hivatal

Budapest
Keleti Károly utca 5-7
1024

Tisztelt Adatvédelmi Felelős Úr!

Beadvánnyal fordult a Nemzeti Adatvédelmi és Információszabadság Hatósághoz (a továbbiakban: Hatóság), amelyben a Központi Statisztikai Hivatal (a továbbiakban: KSH) és a KSH Népeségtudományi Kutatóintézete által a SEEMIG délkelet-európai migrációs projekt keretében tervezett adatgyűjtésével kapcsolatban kérte a Hatóság állásfoglalását.

Elmondása szerint a KSH külföldön élő személyekről gyűjtene elérhetőségi adatokat. Az adatgyűjtés során a KSH munkatársai arra kérnék az érintetteket, hogy mondják meg, van-e olyan személy, aki a háztartásukban lakott, de jelenleg külföldön él. Amennyiben van ilyen személy, a KSH munkatársa arra kéri a megkérdezettet, hogy lépjen kapcsolatba a kivándorolt személlyel és tudakolja meg tőle, hogy részt venne-e a kutatásban. Amennyiben a kivándorolt személy részt szeretne venni a kutatásban, akkor a KSH munkatársa felveszi az érintett valamilyen elérhetőségét (például e-mail cím, telefonszám, következő hazalátogatás időpontja, valamint a hazai elérhetőségei). Amennyiben a megkérdezett nem vállalja, hogy kapcsolatba lép az elvándorolt személlyel, de megadja utóbbi elérhetőségét a KSH munkatársának, akkor a KSH keresi meg az érintettet. Az Ön tájékoztatása szerint minden esetben egy, a kutatás vezetője által aláírt adatvédelmi nyilatkozatot adnak át az érintetteknek, amely tartalmazza, hogy a kutatás során felvett adatokat bizalmasan kezelik. Amennyiben az érintett nem kíván abban részt venni, akkor a személyes adatait azonnal megsemmisítik, részvételi hajlandóság esetén pedig a személyes adatokat elkülönítetten tárolják, majd a kutatás végén megsemmisítik.

Beadványával kapcsolatban a Hatóság állásfoglalása a következő:

1. Az információs önrendelkezési jogról és az információszabadságról szóló 2011. évi CXII. törvény (a továbbiakban: Infotv.) 3. §-ának értelmező rendelkezései szerint:
„1. érintett bármely meghatározott, személyes adat alapján azonosított vagy – közvetlenül vagy közvetve – azonosítható természetes személy,
2. személyes adat az érintettel kapcsolatba hozható adat – különösen az érintett neve, azonosító jele, valamint egy vagy több fizikai, fiziológiai, mentális, gazdasági, kulturális vagy szociális azonosságára jellemző ismeret –, valamint az adatból levonható, az érintettre vonatkozó következtetés”

A fentiekből az következik, hogy egy természetes személy neve, lakcíme, elektronikus levelezési címe és telefonszáma, valamint a külföldről történő hazatérésének időpontja személyes adatnak minősül.

2. Adatkezelés az alkalmazott eljárástól függetlenül az adatokon végzett bármely művelet vagy a műveletek összessége, így különösen gyűjtése, felvétele, rögzítése, rendszerezése, tárolása, megváltoztatása, felhasználása, lekérdezése, továbbítása, nyilvánosságra hozatala, összehangolása vagy összekapcsolása, zárolása, törlése és megsemmisítése, valamint az adatok további felhasználásának megakadályozása, fénykép-, hang- vagy képfelvétel készítése, valamint a személy azonosítására alkalmas fizikai jellemzők (pl. ujj- vagy tenyérnyomat, DNS-minta, irizskép) rögzítése

Az Infotv. 5 § (1) bekezdése kimondja, hogy személyes adat akkor kezelhető, ha ahhoz az érintett hozzájárul, vagy azt törvény vagy – törvény felhatalmazása alapján, az abban meghatározott körben – helyi önkormányzat rendelete közérdeken alapuló célból elrendeli (a továbbiakban: kötelező adatkezelés). A hozzájárulás az érintett akaratának önkéntes és határozott kinyilvánítása, amely megfelelő tájékoztatáson alapul, és amellyel félreérthetetlen beleegyezését adja a rá vonatkozó személyes adatok – teljes körű vagy egyes műveletekre kiterjedő – kezeléséhez [Infotv. 3 § 7. pont].

Az Infotv. idézett rendelkezéseire figyelemmel, külön törvényi felhatalmazás hiányában a KSH és a KSH Népeségstudományi Kutatóintézet a SEEMIG délkelet-európai migrációs projekt keretében kizárólag az érintettek, vagyis a külföldre távozott természetes személyek hozzájárulásával kezelheti a személyes adatokat. E tekintetben a Hatóság nem kifogásolja az Ön beadványának 2. pontjában említett adatgyűjtési módot, amely során a megkérdezett személy veszi fel a kapcsolatot az érintettel, aki elődöntheti, hogy hozzájárulását adja-e ahhoz, hogy a kutatást végzők a későbbiekben felkeressék. A levélnek 3. pontjában említett adatgyűjtési mód azonban már aggályokat vet fel abból a szempontból, hogy ott nem az érintett tudtával és beleegyezésével történik meg az adatok átadása, aki így módon nem élhet megfelelő módon az információs önrendelkezési jogával.

3. A fentiekben túlmenően az adatkezelés során az adatkezelőnek eleget kell tennie az Infotv.-ben foglalt egyéb kötelezettségeinek, elsősorban is a célhoz kötöttség elvének, valamint az érintett tájékoztatása követelményének.

3.1. Az Infotv. 4 § (1)-(2) bekezdései értelmében „[s]zemélyes adat kizárólag meghatározott célból, jog gyakorlása és kötelezettség teljesítése érdekében kezelhető. Az adatkezelésnek minden szakaszában meg kell felelnie az adatkezelés céljának, az adatok felvételének és kezelésének tisztességesnek és törvényesnek kell lennie. Csak olyan személyes adat kezelhető, amely az adatkezelés céljának megvalósulásához elengedhetetlen, a cél elérésére alkalmas. A személyes adat csak a cél megvalósulásához szükséges mértékben és ideig kezelhető”. A célhoz kötöttség elvéből következően a KSH a felvett adatokat kizárólag a SEEMIG délkelet-európai migrációs projekt keretében használhatja fel az érintettekkel való kapcsolattartás céljából.

3.2. Az Infotv. 20 § (1)-(2) bekezdései szerint az érintettel az adatkezelés megkezdése előtt közölni kell, hogy az adatkezelés hozzájáruláson alapul vagy kötelező. Egyértelműen és részletesen tájékoztatni kell továbbá az adatai kezelésével kapcsolatos minden tényről, így különösen az adatkezelés céljáról és jogalapjáról, az adatkezelésre és az adatfeldolgozásra jogosult személyéről, az adatkezelés időtartamáról, arról, hogy kik ismerhetik meg az adatokat. A tájékoztatásnak ki kell terjednie az érintett adatkezeléssel kapcsolatos jogaira és jogorvoslati lehetőségeire is.

4 Tájékoztatom továbbá, hogy a polgárok személyi adatainak és lakcímének nyilvántartásáról szóló 1992. évi LXVI. törvény (a továbbiakban: Nytv.) 4. § (2a) bekezdés b) pontja szerint a Magyarország területét külföldi letelepedés szándékával elhagyó magyar állampolgár adatai akkor szerepelnek a nyilvántartásban, ha azt az érintett kéri.

Az Nytv. 19. § (1) bekezdés b) pontja értelmében a bármely polgár, illetve jogi személy és jogi személyiséggel nem rendelkező szervezet a felhasználás céljának és jogalapjának igazolása mellett jogosult kérnél tudományos kutatás céljából a nyilvántartott személyi lakcímadatokat. Az Nytv. 19. § (2) bekezdés a) pontja szerint az adatigénylésre jogosultak az adatokat tudományos kutatás céljából az alábbi kiválasztási szempontok szerint igényelhetik:

- nevét;
- magyar vagy külföldi állampolgárságát, illetve hontalanságát, magyar állampolgársága megszűnésének tényét (a továbbiakban: állampolgárság), menekült vagy oltalmazott, bevándorolt, letelepedett jogállását, a szabad mozgás és tartózkodás jogával rendelkező személy esetében a szabad mozgáshoz és tartózkodáshoz való joggal történő rendelkezés tényét, illetve az Nytv. 4. § (1) bekezdés c) pontja szerinti külföldön élő polgár esetén a honosítás vagy visszahonosítás tényét;
- nemét;
- születési helyét és idejét;
- anyja nevét;
- elhalálózása helyét és idejét (a holtá nyilvánítását vagy a halál tényének bírói megállapítását);
- lakcímét;
- családi állapotát, házasságkötése vagy bejegyzett élettársi kapcsolata létesítésének helyét.

A fentiek értelmében a kivándorolt állampolgárok adatait, amennyiben az adatigénylés feltételei fennállnak, a KSH is igényelheti a nyilvántartást vezető szervtől. Mindazonáltal felhívom szíves figyelmét arra is, hogy a külföldön élő magyar állampolgárok nyilvántartása nemcsak a Magyarország területét külföldi letelepedés szándékával elhagyó magyar állampolgárok, hanem például a honosított vagy visszahonosított állampolgárok adatait is tartalmazza.

4 Végezetül a Hatóság felhívja szíves figyelmét arra, hogy az Infotv. 12. §-a tartalmazza a tudományos kutatások során történő adatkezelésekre vonatkozó általános szabályokat. Eszerint tudományos kutatás céljára felvett személyes adat csak tudományos kutatás céljára használható fel. A személyes adat érintettel való kapcsolatának megállapítását - míhelyt a kutatási cél megengedi - véglegesen lehetetlenné kell tenni. Ennek megtörténtéig is külön kell tárolni azokat az adatokat, amelyek meghatározott vagy meghatározható természetes személy azonosítására alkalmasak. Ezek az adatok egyéb adatokkal csak akkor kapcsolhatók össze, ha az a kutatás céljára szükséges. A kutatás során személyes adatokat csak akkor lehet nyilvánosságra hozni, amennyiben ahhoz az érintett hozzájárult.

A személyes adatok védelmét és a közérdekű adatok nyilvánosságát érintő ügyben forduljon a Nemzeti Adatvédelmi és Információszabadság Hatósághoz továbbra is bizalommal.

Budapest, 2012. november 29.

Üdvözlettel:


Dr. Révész Balázs
főosztályvezető
Vizsgálati Főosztály

III. Melléklet

Adatkezelési terv

a *Budapesti roma kutatás 2010*

című kutatáshoz

1. Vonatkozó törvények és dokumentumok

- 1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról (Avtv.)
- 1995. évi CXIX. törvény a kutatás és a közvetlen üzletszerzés célját szolgáló név- és lakcímadatok kezeléséről (Aktv.)
- 2005. évi CXXXIX. törvény a felsőoktatásról (Ftv.)
- ELTE Szervezeti Működési Szabályzat (ELTE SzMSz)
- ELTE Társadalomtudományi Kar Szervezeti Működési Szabályzat (ELTE TáTK SzMSz)

Az alábbi adatkezelési terv a hivatkozott dokumentumok alapján és azok figyelembevételével készült.

2. Kutatási jogosultság

Az Ftv. 4.§ (1) bekezdése szerint a felsőoktatási intézmények, így az Eötvös Loránd Tudományegyetem (Egyetem) alaptevékenysége a tudományos kutatás. Az ELTE SzMSz 131. § (1) bekezdése szerint: „Az Egyetem az oktatók, kutatók, hallgatók számára biztosítja a tudományos kutatás (...) szabadságát, támogatja feltételeinek megvalósítását. ”. Továbbá az ELTE SzMSz 131. § (2) bekezdés szerint: „Az Egyetem oktatóinak, kutatóinak a jelen Szabályzat keretei között kötelessége tudományos munka folytatása, illetve az oktatók, kutatók jogosultak (...) a munkaköri feladataikból származó feladatok mellett a maguk választotta tudományos téma kutatására. ”.

3. A kutatás célja

A kutatás célja a cigányok gazdasági tevékenységével foglalkozó kvalitatív kutatások eredményeinek felhasználásával átfogó kvantitatív képet nyújtani a budapesti cigány lakosság gazdasági tevékenységéről és közösségi struktúrájáról. A kutatás tárgyát képezi a kvalitatív (elsősorban antropológiai) kutatások során feltárt jellegzetes megélhetési stratégiák elterjedtségének mérése. További célként fogalmazódik meg a megélhetési stratégiák szerepe a cigányságon belüli szűkebb csoportok identitásának kialakításában. A kutatás választ keres arra is, hogy milyen kapcsolatok vannak az egyes szűkebb csoportok között, és ennek milyen gazdasági vetületeit lehet leírni.

4. A kezelendő személyes adatok köre és azok forrása

A kutatás válaszadó-vezérelt mintavétel módszerével, személye kérdőíves interjúval kerül megvalósításra. A módszer lényege, hogy az egyes interjúalanyok adják meg a következő interjúalany megkereséséhez szükséges személyes adatokat. Az adatfelvétel során a következő személyes adatok

kezelésére kerül sor: név, lakcím, mobiltelefonszám. A kutatás jellegénél fogva implicit módon egy különleges adatot, az etnikai kisebbséghez tartozást is érinti.

5. Az adatkezelés folyamata

A kutatás során a kutatást lebonyolító kérdezőbiztosok (az egyetem hallgatói) a kérdőív különválasztható részén rögzítik a szükséges személyes adatokat. A rögzített személyes adatok alapján jutnak el a következő interjúalanyhoz. Az egyes kérdőívek adatait a kérdezőbiztosok a személyes és különleges adatok nélkül rögzítik elektronikusan. A kérdőíveket a rögzítést követően a kérdezőbiztosok haladéktalanul eljuttatják az Egyetem kijelölt titkárságára, ahol zárt szobában, zárt szekrényben kerülnek tárolásra. A személyes adatokat a kutatás szervezői a kérdés ellenőrzéséhez, valamint a fogás-visszafogás és a hólabda becslés alapján a teljes populáció nagyságának becslésére is felhasználják. Az ellenőrzés és a becslések elvégzését követően, legkésőbb az adatfelvétel befejezésétől számított 3 hónapon belül a személyes és különleges adatokat tartalmazó lapok a kérdőívekről leválasztásra, majd megsemmisítésre kerülnek. Ezt követően csak a személyes adatokat nem tartalmazó kérdőívek, valamint az anonimizált adatokat tartalmazó elektronikus adatbázist őrzik meg a kutatók.

6. Az érintett jogai gyakorlati érvényesíthetőségének biztosítékai

A kérdezőbiztosok az Egyetemről megbízólevelet kapnak, amely tartalmazza a kutatás szervezéséért felelős kutatók nevét és elérhetőségét (minta mellékelve). A kérdezőbiztosok az interjú elején a bemutatkozó részben ismertetik és bemutatják az interjúalany számára Aktv. szerinti jogait, melyeket a kérdőív első lapja rögzít (vonatkozó rész mellékelve).

7. Az adatvédelmet biztosító technikai és szervezési intézkedések

A kérdezőbiztosok a számukra szervezett felkészítés során megismerik a jelen dokumentumban valamint a vonatkozó törvényekben szereplő adatkezelési elveket, továbbá az adatkezelés során elkövetett mulasztásokra vonatkozó törvényi szankciókat. A felkészítést követően minden kérdezőbiztos együttműködési megállapodást ír alá, amellyel – büntetőjogi felelősségük tudatában – vállalják a személyes és különleges adatokra vonatkozó titoktartást, valamint a vonatkozó törvények és az adatkezelési tervben foglaltak betartását. A kérdőívek átvételére kijelölt titkárság munkatársát a kutatást szervező kutatók tájékoztatják a jelen adatkezelési terv részleteiről, különösen saját feladatairól. A kijelölt titkársági munkatárs ezt követően aláírásával veszi tudomásul a rá vonatkozó szabályokat.

IV. Melléklet

Feladatvállalás és titoktartási nyilatkozat

Én

(Anyja neve:.....)

Személyi igazolványszáma:

Lakcím:

Telefonszám, mobilszám:.....

Születési helye és ideje:, év hó nap)

az ELTE hallgatója vállalom, hogy 2010. december 15-ig db személyes megkereséses kérdőíves interjút a mellékelt szabályok betartásával lebonyolítom, a kérdőívek adatait spss .sav formátumban rögzítem és a kitöltött kérdőíveket, valamint a rögzített spss .sav formátumú adatbázist a rögzítést követő legrövidebb időn, de legfeljebb egy héten belül az ELTE TáTK Módszertani Kutatóközpontnak eljuttatom.

Elismerem, hogy részt vettem a kutatás lebonyolításáról és a kapcsolódó kérdésekről szóló felkészítésen, a kutatás szabályait megértettem és azoktól eltérni nem fogok.

Vállalom, hogy a rám bízott személyes és különleges adatokat titokban tartom, azokat másoktól elzártan kezelem, a Módszertani Kutatóközpont munkatársain kívül másoknak nem adom át. Tudomásul veszem, hogy a visszaélés személyes- és különleges adatokkal, illetve, ha azok hanyag kezeléséből másnak jelentős érdeksérelem keletkezik, a BTK 177/A paragrafusára szerint vétség és két évig terjedő szabadságvesztést vonhat maga után. Tanúsítom, hogy a felkészítő során megismertem a kutatás adatkezelési tervének rám vonatkozó részét és azokat maradéktalanul betartom.

Tudomásul veszem, hogy kreditelismerésre / hallgatói szerződés megkötésére csak a vállalt feladatok teljes körű és szabályszerű teljesítése esetén kerül sor.

Budapest, 2010. október 2.

.....
egyetemi hallgató, kérdezőbiztos

Előttünk, mint tanúk előtt:

tanú 1

Szig. sz.:
Lakcím:

tanú 2

Szig. sz.:
Lakcím:

V. Melléklet

Kérdőbiztosi oktató anyag a Budapesti Roma Kutatás 2010 című kutatás mintavételéhez

A mintavétel a sajátos célcsoport miatt speciális technikával zajlik, úgynevezett válaszadó-vezérelt mintavétellel. A mintavétel lényege, hogy előre nem áll rendelkezésünkre egy címlista, sem véletlen sétás eljárás, csupán néhány kiinduló cím. Az első lekérdezéseknek ezeken a címeken kell megtörténniük. A kérdőív utolsó oldalán a válaszadóktól azt kérjük, hogy adjanak meg további **budapesti 18 éven felüli** roma ismerősöket, akiknél folytatódhat a kutatás. Ezt követően a kapott neveket abc sorrendbe kell rendezni és kiválasztani közülük a kérdőívhez mellékelt kulcs alapján a következő felkeresendő személyt és két pótcímet. Ha több, mint 15 nevet kapunk, akkor az első 15-ből választunk a kulcs segítségével. Ezt követően a kiválasztott következő interjúalany és a két pótcímként kiválasztott személy elérhetőségét (telefonszámát és lehetőség szerint címét) kérdezzük meg. Ezután megkérjük az interjúalanyt, hogy a telefonunkról hívja fel a kiválasztott ismerőst és modja el neki, hogy keresni fogjuk a kutatással kapcsolatban. Ezt követően a következő megkérdezettől nyert listából kiválasztott személlyel folytatjuk a kérdezést, stb.

Az így nyert minta alapján a hagyományos technikákkal (véletlen sétás módszer, címkártyás módszer) készült mintákhoz hasonlóan torzítatlan becslés adható.

Fontos szempontok:

- Csak akkor adható torzítatlan becslés, ha
 - **minden lehetséges kapcsolatot összegyűjtünk minden megkérdezettől** (ezek száma fontos szempont, rögzíteni kell akkor is, ha több mint a kérdőívben szereplő 15 hely)
 - **az összegyűjtött kapcsolatok közül a kulcsot használva választunk** (ellenkező esetben a minta torz lesz)

Mintavétel illusztrációval

1. Megkérdezzük hogy az interjúalany kiket tudna ajánlani a budapesti roma ismerősei közül, akikhez elmehetnénk interjút készíteni. A kapott neveket rögzítjük a kérdőív utolsó oldalán

Ezt a kérdőívet szeretnénk más roma/cigány emberektől is megkérdezni. Kérem, adja meg azoknak a roma/cigány ismerőseinek a nevét, akikhez el tudnék menni az interjút elkészíteni.	
<i>ÍRD BE AZ ÖSSZES NEVET, AZUTÁN SORSZÁMOZD MEG NÉVSOR SZERINT, MAJD VÁLASZD KI A KÓDLAP SEGÍTSÉGÉVEL A KÖVETKEZŐ INTERJÚALANYT ÉS A KÉT <u>PÓTCÍMET!</u></i>	
<i>Név</i>	<i>Sorszám (Névsor szerint!)</i>
Jónás István	
Balogh sándor	
Jónás Ilona	
Kolompár János	
Kiss Károly	

2. Ha több, mint 15 nevet kapunk, akkor rögzítjük a kapott nevek számát!

3. A kapott neveket ABC szerint sorszámozzuk a második oszlopban:

Ezt a kérdőívet szeretnénk más roma/cigány emberektől is megkérdezni. Kérem, adja meg azoknak roma/cigány ismerőseinek a nevét, akikhez el tudnék menni az interjút elkészíteni.

ÍRD BE AZ ÖSSZES NEVET, AZUTÁN SORSZÁMOZD MEG NÉVSOR SZERINT, MAJD VÁLASZD KI A KÓDLAP SEGÍTSÉGÉVEL A KÖVETKEZŐ INTERJÚALANYT ÉS A KÉT PÓTCÍMET!

Név	Sorszám (Névsor szerint!)
Jónás István	3
Balogh Sándor	1
Jónás Ilona	2
Kolompár János	5
Kiss Károly	4

4. A kérdőívhez tartozó KIVÁLASZTÁSI KULCS segítségével kiválasztjuk a következő interjúalanyt és két pótcímet. Ezt úgy tesszük, hogy megnézzük, a válaszadónk hány nevet adott meg, ennek megfelelő oszlopot megkeressük a Kiválasztási Kulcsban. Ha több, mint 15 nevet kaptunk az elsőként megadott 15 névből választunk a 15 ajánlott személyhez tartozó számhármast szerint!

Ellenőrizd, hogy azonos-e a kérdőív sorszámaival!

Példánkban 5 nevet kaptunk, ezért ezt a számhármast használjuk

Kérdőív sorszám: 001

Kiválasztási kulcs

		Ajánlott személyek száma																				
		1			2			3			4			5			6			7		
kulcs		1	0	0	1	2	0	2	1	3	4	2	1	2	5	3	3	6	6	2	3	

		Ajánlott személyek száma																							
		8			9			10			11			12			13			14			15		
kulcs		3	7	2	6	3	8	3	7	4	8	3	7	9	8	3	5	10	9	2	6	11	8	2	6

5. A KIVÁLASZTÁSI KULCSBAN található számhármast **első száma** lesz a következő megkérdezendő személy ABC rend szerinti sorszáma. A számhármast **következő két száma** a két pótcímhez tartozó két személy ABC rend szerinti sorszáma.

Ezt a kérdőívet szeretnénk más roma/cigány emberektől is megkérdezni. Kérem, adja meg azoknak a roma/cigány ismerőseinek a nevét, akikhez el tudnék menni az interjút elkészíteni.

ÍRD BE AZ ÖSSZES NEVET, AZUTÁN SORSZÁMOZD MEG NÉVSOR SZERINT, MAJD VÁLASZD KI A KÓDLAP SEGÍTSÉGÉVEL A KÖVETKEZŐ INTERJUALANYT ÉS A KÉT PÓTCÍMET!

Név	Sorszám (Névsor szerint!)
Jónás István	3
Balogh Sándor	1
Jónás Ilona	2
Kolompár János	5
Kiss Károly	4

Példánk szerint a számhármas következő két száma: 5, 3. Ezért Jónás István és Kolompár János lesz a két pótcím.

Példánk szerint a számhármas első száma 2. ezért Jónás Ilona a következő interjúalanyunk.

6. A kiválasztott következő interjúalany nevét, valamint a pótcímekhez tartozó neveket beírjuk a következő táblázatba.

A KÖVETKEZŐ INTERJÚALANY ÉS A PÓTCÍMEKHEZ TARTOZÓ NEVEKET ÍRD AZ ALÁBBI TÁBLÁZATBA!

Az előre megadott szabály szerint kiválasztottam három ismerősét, akikhez elmegyünk interjút készíteni. Kérem, most adja meg a következő ismerősei mobilszámát, és lehetőség szerint címét is, hogy meg tudjam beszélni velek az interjú időpontját!

Név	Mobilszám	Cím
Jónás Ilona		
Kolompár János		
Jónás István		

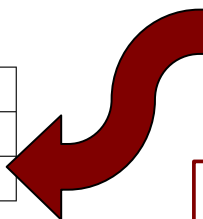
Sokat segítene, ha az első ismerősét felhívna a telefonomról és elmondaná neki, hogy keresni fogom az interjúval kapcsolatban! Segítségét nagyon köszönjük!

7. Ezután megkérjük az interjúalanyt, hogy adja meg a kiválasztott személyek mobilszámát és címét. Ezeket rögzítsük a táblázatba.
8. Ezt követően kérjük meg az interjúalanyt, hogy hívja fel a következő interjúalanyt kiválasztott személyt és mondja el neki, hogy a kutatással kapcsolatban keresni fogjuk.
9. Az interjú napján, vagy a következő napon hívjuk fel mi is a következő interjúalanyt és beszéljük meg vele a következő interjú időpontját.
10. Az interjú előtti napon ismét hívjuk fel az interjúalanyunkat és kérdezzük meg, hogy még megfelelő-e az időpont.
11. **NAGYON FONTOS!** A következő interjúalanyunk kérdőívének „Ajánló sorszáma” mezőjébe **az öt ajánló interjúalany kérdőívének sorszáma** kerüljön! Ha ez nem így történik, akkor nem tudjuk a kérdőív adatait felhasználni és sajnos neked sem tudjuk elfogadni.

A kérdőív sorszáma	1	2	2
Kérdező sorszáma	0	0	1
Ajánló sorszáma	1	2	1

Budapesti romák – 201

A kérdőív sorszáma	1	2	3
Kérdező sorszáma	0	0	1
Ajánló sorszáma	1	2	2



A következő interjúalanyunkat ajánló személy kérdőíve

A következő interjúalanyunk kérdőíve

Budapesti romák – 2010

ELTE Társadalomtudományi Kar Társadalmi Konfliktusok Kutatóközpontja
Lehívó: ELTE TáTK Módszertani Kutatóközpont

12. Az ajánló kérdőívére írjuk fel az első oldalra azoknak a kérdőívsorszámát, akiket az illető ajánlott és interjú is készült velük (nincs ilyen rovat, külön fel kell írni!).

Milyen nehézségek adódhatnak?

Még az interjú helyszínén kiderül, hogy a kijelölt következő interjúalany nem vállalja az interjút: kérjük meg a válaszadónkat, hogy hívja fel az első pótcímhez tartozó embert. Ha az egyik pótcím sem vállalja, akkor húzd ki azoknak a neveit, akiket nem lehetett elérni, sorszámozd újra és nézd meg újra a kulcsot a maradék nevet tekintve alapul (azaz az eredetnél 3-mal kevesebb névvel). Ha szükséges ezeket a lépéseket addig kell ismételni, amíg el nem fogy a teljes lista vagy sikerül valakit elérni!

Szál megszakadás: akkor fordulhat elő, ha az ajánlott és kiválasztott személy megtagadja a válaszadást, vagy nem érhető el, illetve egy válaszadó nem ad meg olyan személyt, aki még nem szerepelt a megkérdezettek között. Ezekben az esetekben **a szál utolsó ajánlója által ajánlott személyek közül kell kiválasztanod a pótcímek közül az elsőt (vagy azt, akinél még nem jártál!).** Ha ilyen nincs, akkor kérj tőlünk új címet! Ha itt találsz telefonszámot, akkor felhívod az adott személyt, ha csak címet találsz, akkor legalább három alkalommal meg kell kísérelni az interjút. **A kérdőívre minden esetben írd rá, hogy miért nem sikerült az interjú!**

A válaszadónk nem akar mobilszámot, címet megadni a nevekhez: mondjuk el újra, hogy ezekre az adatokra csak azért van szükségünk, hogy tovább folytathassuk a kérdezést. A pótcímekről mondjuk el, hogy csak a biztonság kedvéért kellenek. Mindenképpen hangsúlyozzuk, hogy a kérdőív utolsó oldalát a kérdőívekről leválasztjuk majd és meg semmisítjük. Ha ezzel kapcsolatban kétség merül fel, akkor felhívhatják a megbízó íven található számot. Ha csak mobilszámot vagy csak címet kapunk az elégséges lehet a tovább kérdezéshez.