
2. AZ ADATFELVÉTEL BEMUTATÁSA

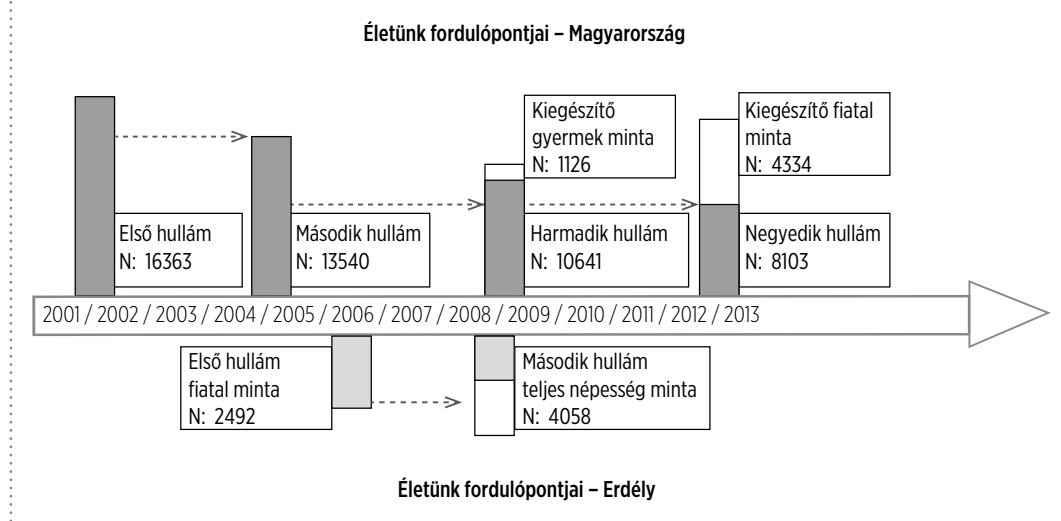
KAPITÁNY BALÁZS

BEVEZETŐ

Az *Életünk fordulópontjai* (ÉF) adatgyűjtési program keretében 2001 és 2015 között összesen négy magyarországi és két erdélyi magyar adatgyűjtésre került sor. A kutatási koncepciónak megfelelően arra törekedtünk, hogy egy olyan adatállomány-rendszert hozzunk létre, amely alapvetően ugyanazon személyek többszöri megkérdezésén alapul (longitudinális jellegű), ugyanakkor az egyes adatfelvételi hullámok időpontjában keresztmetszeti elemzésekre is lehetőséget nyújt. Ez az összetett módszertani igény nagyban befolyásolta a módszertani döntéseket a mintavétel módjától a terepmunka szervezésén és a kérdőívek kialakításán át a súlyozásig. Ez indokolta az adatgyűjtéshez kapcsolódó két kiegészítő minta (a 3. hullám gyerekmintája és a 4. hullám kiegészítő fiatal mintája) alkalmazását is. Az adatfelvétel főbb jellemzőit bemutató fejezetben elsősorban a négy magyarországi adatgyűjtést (ide értve a két kiegészítő mintát is) mutatjuk be, a fejezet végén csak röviden térünk ki az erdélyi adatgyűjtésekre.

A kötetnek ebben a fejezetében nem foglalkozunk a terepmunka megszervezésével és végrehajtásának módjával, mivel ezt önálló fejezet ismerteti majd (I. a kötet 4. fejezetét). A fejezet végén viszont röviden foglalkozunk a kérdőívek kialakításával. Itt azonban csak olyan szempontból vizsgáljuk a témát, hogy az adatfelvétel szerkezete hogyan hatott a kérdőívek kialakítására. A kérdőívekben vizsgált főbb kutatási kérdések és témakörök bemutatása az elméleti fejezetben már megtörtént, a kérdések katalógusa a kötet végén, a Mellékletben található.

1. ábra: Az adatfelvételi program áttekintése



A KÖVETÉSES ADATÁLLOMÁNY – A MINTAVÉTEL ELVEI

Az adatgyűjtési program 2001-es tervezésekor arra törekedtünk, hogy egy olyan sokaságot vonjunk be az adatgyűjtésbe, amely nemcsak keresztmetszetében reprezentálja jól az akkori magyar társadalmat, hanem hosszú távon is követhető és rábíráható lesz újabb adatszolgáltatásra.

A fenntarthatóságot szem előtt tartva döntöttünk úgy, hogy az alapsokaságot a Magyarországon élő magyar állampolgárokra korlátozzuk, mivel a Magyarországon élő külföldiek mobilitása igen nagy, sok esetben (pl. nyelvi okokból) nagyon nehéz az adatfelvétel megszervezése, lebonyolítása.

Szintén a fenntarthatóság állt a döntés hátterében, amikor a követéses vizsgálatba csak a 2001-ben 75 évesnél fiatalabb válaszadókat vontuk be. (Természetesen a mintába került személyeket később 75 éves koruk után is követtük.)

Részben szintén a fenntarthatóság miatt döntöttünk úgy, hogy „pótcímezés” helyett inkább a kiegészítő minta esetén is előzetesen „torzítjuk” a címlistát. Az eljárás lényege, hogy a megfelelő részsokasági elemszámok garantálása érdekében azokat a csoportokat, amelyek esetén a korábbi mintavételek tapasztalataira alapozva kisebb válaszadási hajlandóságot és nagyobb lemorzsolódási esélyt feltételeztünk (pl. fiatal nagyvárosi férfiak), a kiválasztás során kis mértékben túlreprezentáltuk. Így érhetjük el azt, hogy a megtörtént adatszolgáltatás után a tényleges adatbázis összetétele közelebb kerüljön az alapsokaságban kialakult arányokhoz. Az így kialakított sokaság megfelelőbb alapot jelent a követéses vizsgálatokhoz. (Ezt az eljárási módot nevezik „lemorzsolódós” mintavételnek.)

Végül a hosszú távú követhetőség igénye miatt döntöttünk amellett, hogy a kiinduló minta elemszáma a kutatás nemzetközi keretét jelentő *Generations and Gender Programme* (GGP) tízezer fős minimum-elvárásánál lényegesen nagyobb – az első adatgyűjtés esetén is legalább 15 ezer fős – legyen.

A kiinduló minta esetén a mintavétel fő jellemzői – a fenti elvek alapján – a következők lettek:¹ A mintavételi egység a Magyarországon magánháztartásban élő, 1926. január 1. és 1983. december 31. között született magyar állampolgárok köre volt. A válaszadók kiválasztása az interjúk elutasításának és meghíúsulásának a lehetőségével számolva előzetesen torzított, régiónként, településnagyság-kategóriánként és életkori kategóriánként rétegzett többlépcsős minta segítségével történt. Az elméleti kiválasztási arány: $15\,000 / 7\,454\,196$, tehát megközelítőleg 0,2% volt, ezt módosítottuk a korcsoportonként és településtípusonként eltérő „mintavételi biztonsági kvócienssel”.

Az adatok két különböző forrásból származtak: az alapsokaságra vonatkozó adatokat a Központi Statisztikai Hivatal, a válaszadó személyek tényleges adatait a népszámlálási nyilvántartó – akkori nevén: Belügyminisztérium Központi Adatfeldolgozó, Nyilvántartó és Választási Hivatala – biztosította.

Magának a településmintának a kiválasztásakor két rétegzési ismérv alkalmazására került sor.

Az első rétegzési szempont Magyarország regionális beosztása volt (Nyugat-Dunántúl, Közép-Dunántúl, Dél-Dunántúl, Központi régió, Észak-Magyarország, Észak-Alföld, Dél-Alföld). Az erre tekintettel levő rétegzés növeli a regionális adatok megbízhatóságát, s így megfelelő elemszám mellett előkészíti ezek jó színvonalú elemzésének a lehetőségét.

A második rétegzési szempont a településnagyság-kategória volt. A 2001-es mintavétel esetén tíz olyan településnagyság-kategóriát használtunk, amely kategóriákban az akkori összlakosság megközelítette az ország akkori összlakosságának egytizedét. A 6–10. kategóriába tartozó (vagyis 2001-ben 17 817 fő által lakott vagy népesebb) 70 önálló település és Budapest 23 kerülete ún. „önreprezentáló” települések voltak, vagyis ezek mindegyike a mintába került, és ezen településeken belül külön-külön történt a mintavétel.

Az 1–5. kategóriába 3063 település tartozott, amelyek lakossága 17 817 főnél alacsonyabb volt. Esetükben településnagyság-kategóriánként reprezentáns települések lettek kijelölve, hogy a megkérdezettek száma régiónként megfeleljen a kiválasztási valószínűség arányának, vagyis mintha egy régió egy településnagyság-kategóriájának összes települése egy önreprezentáló település lenne. A tényleges terepmunka-települések ekkor úgy kerültek kiválasztásra, hogy a kijelölt címek száma településenként minimum 14 fő legyen.

Végül a kérdésbe bevont települések száma 441 volt (Budapestet egy településnek számítva). Ezek egyfelől Budapestet és az ország 70 – akkor – legnépesebb vidéki városát jelentették, másfelől 370 kisebb vidéki települést, amelyek válaszadói a kistelepüléseket képviselték.

¹ A mintavétel részletes leírása (Kapitány 2003) megtalálható az első hullám dokumentációs kötetében, így itt csak a legfontosabb pontokat emeljük ki.

A terepmunka sikeressége, a válaszadási hajlandóság felülmúlta a tervezettet, így a válaszadók száma meghaladta a tervezett 15 ezer főt, a kiinduló tisztított mintanagyság végül 16 363 fő lett. Az adatgyűjtési program elsődleges célja a 2001–2002-es adatgyűjtés terepmunkájának befejezését követően az lett, hogy ezt a bő 16 ezer fős sokaságot életútja fordulópontjain át kövesse, és rendszeresen adatokat gyűjtsön tőlük.

A KÖVETÉS, ISMÉTELT FELKERESÉS ELVEI

Mivel az adatgyűjtési program alapvetően követéses jellegű, a későbbiekben alapelv volt, hogy az eredetileg a mintába került személyekhez mindenképpen térjünk vissza, tőlük ismét gyűjtsünk adatokat. Ebben az esetben viszont kulcskérdés, hogyan lehet minimalizálni a mintaelemszám csökkenését, illetve a csökkenés ellenére hogyan lehet garantálni, hogy a minta továbbra is a lehető legjobban reprezentálja az alapsokaságot.

Az adatfelvételi hullámok között természetesen maga az alapsokaság is csökken, a válaszadók egy része „kilép az alapsokaságból”. Ennek oka lehet a halálozás, a kivándorlás, illetve – ahogy a mi személyi háztartásban élőkre korlátozott mintánk esetén is – az intézményi háztartásba (kórházba, börtönbe, idősek otthonába stb.) kerülés. A mintaelemszám ilyen okokra visszavezethető csökkenése teljesen természetes, az eredmények megbízhatóságát nem csökkenti, hiszen ha például a mintát a 3. hullám idején vettük volna fel, ezek a személyek ekkor nem kerültek volna be az alapsokaságba, tehát hiányuk nem jelent torzítást keresztmetszeti értelemben sem. A gyakorlatban azonban problémát okoz, hogy míg a halálozásokról többé-kevésbé pontos adatok állnak rendelkezésre a népességnyilvántartás adatbázisából, addig a tartósan külföldre költözött válaszadók jelentős része „rejtve marad”, ezért egyszerűen csak az ismeretlen helyre költözött kategóriába kerülhet. A külföldre költözés és az intézményi háztartásba kerülés ráadásul ideiglenes kilépést is jelenthet: a kivándorló a következő adatgyűjtés idejére ismét visszatérhet az országba, az elítélt szabadlábra kerülhet stb. A fentiek ellenére – terepmunka-szervezési okokból – az alapsokaságból kilépők esetén a következő – harmadik vagy negyedik – adatfelvétel során már nem kíséreltük meg az újbóli felkeresést, vagyis azt feltételeztük, hogy aki egyszer kilép az alapsokaságból, az már nem lép vissza. A 2009 utáni kivándorlási hullám fényében ez a döntésünk nem tűnik szerencsésnek, de ne feledjük, az ország ideiglenes elhagyása az ezredforduló környékén jóval ritkább volt.

Problémát jelenthet, hogy az alapsokaságba történő utólagos belépést sem fedtük le a mintával, vagyis ebből a szempontból utólag már nem bővítettük a kiinduló mintát (pl. a bevándorló és állampolgárságot szerző személyekkel). Ennek gyakorlati okokból történő elmaradása – amely más hasonló külföldi vizsgálatok esetén is jellemző – azonban feltehetően nem okoz komoly torzítást az adatokban, mivel meglehetősen kis sokaságokról van szó.

A válaszadói létszám lemorzsolódásának másik fő típusa az, amikor az – egykori – adatszolgáltató az alapsokaság tagja marad, de mégsem sikerül tőle ismételt adatot gyűjteni. Ennek oka lehet például a válaszadás valamelyik későbbi adatfelvételi hullámban történő megtagadása, illetve az országon belül ismeretlen helyre való költözés, az „eltűnés”, vagy akár a rossz egészségi állapot (pl. magánháztartásban élő demensek stb.). Valójában ez a lemorzsolódásnak az a formája, amely kritikus veszélyt jelent az adatbázis megbízhatóságát illetően.

Itt a következő eljárást követtük: mind az adatfelvételi hullámok között, mind az adatgyűjtések alkalmával mindent megtettünk, hogy csökkentsük az ilyen típusú lemorzsolódást. (Ezeket a technikákat a kötet 4. fejezete tárgyalja.) Ha viszont egy válaszadót egy adatfelvételi hullámban végleg „elvesztettünk” – vagyis nem sikerült tőle adatokat gyűjtenünk –, már nem próbáltuk meg felkeresni a következő hullám idején. Vagyis aki lemorzsolódott, az „végképp” lemorzsolódott. Szűken vett módszertani szemmel nézve nyilván szerencsésebb lett volna megkísérelni a korábban elveszett válaszadók ismételt megkeresését a harmadik hullámtól (ez a lehetőség értelemszerűen ekkor adódhatott először). Ennek elmaradását az indokolja, hogy nagyon komoly idő- és anyagi ráfordításokat igényelt volna, amelyek jó eséllyel csökkentették volna a „hagyományos” válaszadók sikeres megkereséséhez szükséges erőforrásokat. Mindemellett a gyakorlatban ezen korábban elveszett, de utólag még visszanyert válaszadók esetén speciális kérdőívek alkalmazására is szükség lett volna (hiszen az előző hullám(ok) adatai hiányoztak). A tőlük begyűjtött adatokat ráadásul utólag igen nehéz lett volna harmonizált formában beilleszteni a „hagyományos” válaszadók válaszainak szerkezetét tükröző adatbázisokba.

AZ ADATFELVÉTELEK IDŐZÍTÉSE

Költségvetési és terepmunka-szervezési szempontból az egyes adatfelvételi hullámok minden esetben év végén kezdődtek, s áthúzódtak a következő évre (1. táblázat).

A 2001. évi adatfelvétel novemberben kezdődött, és 2002. február elején fejeződött be. A válaszadásra az esetek mintegy 3/5 részében 2001-ben került sor, a többiek jellemzően januárban válaszoltak, februárban már inkább csak pótlásokat, javításokat végeztünk. A felvétel eszmei időpontja 2001. december 31., tehát a válaszok (elvileg) az erre az időpontra érvényes helyzetet tükrözik.

A későbbi adatfelvételek esetén a szubjektív jellegű kérdések nagy száma és a terepmunka hosszú időtartama miatt eszmei időpontokat nem határoztunk meg, az információk így mindig a kérdés napjára vonatkoznak. A második adatfelvétel terepmunkája 2004. november és 2005. január 15. között zajlott volna, de a tervezettnél alacsonyabb elérés miatt részleges pótkérdezésre került sor 2005 májusában, így mintegy 700 fő ekkor válaszolt.

A harmadik adatfelvétel 2008 novemberében indult, és 2009. február 15-én fejeződött be, de a tényleges lekérdezések csaknem háromnegyede megtörtént november hónapban, a 2009-ben lefolytatott kérdezések aránya pedig alig 4% volt.

A harmadik és a negyedik adatfelvétel között ismét nem három, hanem négy év telt el, így a negyedik adatfelvétel 2012. november közepén indult. A kiegészítő minta miatti elemszám-növekedés és a kérdezők számának csökkenése miatt a terepmunka valamivel hosszabb volt a szokásosnál: általában február végéig tartott, de egyes területeken még márciusban is zajlottak utókérdezések.

Az adatbázisok az egyes adatgyűjtések esetén (az első, eszmei időponttal rendelkező adatgyűjtés kivételével) a terepmunka hónapját külön technikai változóknak tartalmazzák (a változók elnevezéséről és az adatbázisok felépítéséről lásd a kötet 6. fejezetét).

1. táblázat: Az ÉF kutatás magyarországi adatgyűjtéseinek időzítése

	A terepmunka		Kiegészítő adatgyűjtés
	kezdet	vége	
1. adatfelvételi hullám ^a	2001. november 8.	2002. február 15.	nem volt
2. adatfelvételi hullám	2004. november 1.	2005. január 15.	volt, 2005 májusában
3. adatfelvételi hullám	2008. november 1.	2009. február 15.	nem volt
4. adatfelvételi hullám	2012. november 17.	2013. február 28.	néhány megyében volt, 2013 márciusában

^a Az első adatfelvétel eszmei időpontja 2001. december 31.

KIEGÉSZÍTŐ MINTÁK

A követéses vizsgálatok általános jellemzője, hogy a kiinduló minta öregedése (és csökkenése) miatt az eredeti alapsokaság már egyre nehezkesebben használható keresztmetszeti elemzésekre. Ezt a problémát általában kiegészítő mintákkal szokták orvosolni. Az *Életünk fordulópontjai* adatgyűjtési program esetében is felmerült ez a probléma, amelyet a harmadik adatfelvételi hullám során egy speciális, „kettős használatú” gyermek kiegészítő mintával, a negyedik hullámban pedig egy általános – a kiinduló mintával csak a mintavétel módjában összehangolt – kiegészítő fiatal mintával orvosoltunk.

A 2007–2008-AS GYERMEKMINTA

A 2001–2002 fordulóján (eredetileg) 18–74 éves kor közötti populáció együtt öregedett a mintával, és hét év elteltével, a harmadik hullám idejére a legfiatalabb válaszadó is elérte a 25 éves kort. Egy esetleges keresztmetszeti elemzés esetén már egyre inkább kérdéses volt, hogy a 18–24 éves generáció adatainak híján mennyire kapunk érvényes képet a magyar népességre vonatkozóan.

Emiatt úgy döntöttünk, hogy egy kiegészítő mintát veszünk a 20–24 éves (1984–1988-ban született) fiatal felnőttek körében. Ezeket a fiatal felnőtt válaszadókat azonban nem a korábbi válaszadók mintájától független módon választottuk ki, hanem a mintát a válaszadók megfelelő korcsoportokhoz tartozó gyermekeivel töltöttük fel. Ez oly módon történt, hogy a 2008–2009-es harmadik adatgyűjtéskor minden válaszadót megkérdeztünk, van-e életben lévő, az 1984–1988 közötti években született vérszerinti vagy örökbe fogadott gyermeke. Ha az illetőnek volt ilyen gyermeke, bekerült a kiegészítő mintába. Ha a korábbi válaszadónak több, megfelelő korú gyermeke is volt, mindannyian belekerültek a kiegészítő mintába. Amennyiben a szülők mintája reprezentatív az adott népesség-

re – márpedig reményeink szerint az –, akkor a gyermekeik mintája is csaknem helyesen reprezentálja az adott korosztályokat, tehát ezen a módon egy reprezentatív kiegészítő mintához jutottunk.²

A nem teljesen pontos reprezentáció oka, hogy vannak olyan gyermekcsoportok, akiknek nincsenek életben lévő válaszadó szülei, és így a „szüleiken keresztül” mintába kerülni sincs esélyük. Azonban a 20–24 éves gyermekek között a teljesen árvák aránya még meglehetősen alacsony, így ennek a torzításnak a mértéke feltehetően nem jelentős.

Az 1580, feltételeknek megfelelő gyermekből 1126 gyermekkel sikerült interjút készítenünk. Ez a 71%-os válaszadói arány a mai magyarországi címlistas adatgyűjtések válaszadási hajlandóságához viszonyítva igen jó aránynak számít, főként mert egy alapvetően alacsony mértékben elérhető fiatal felnőtt csoportra vonatkozik. (Az elérési arányt viszont nagyban befolyásolták speciális tényezők: pl. az elvált apák külön élő gyermekei jóval alacsonyabb arányban voltak elérhetőek, a félárva bekerülési aránya a két élő szülővel rendelkezőkének a fele.)

A mintába bekerült új fiatal felnőtt válaszadók segítségével két szempontból is bővíthetővé váltak az elemzési lehetőségek. Egyfelől lehetőség nyílt a keresztmetszeti elemzés fiatal válaszadókkal való kibővítésére. (Ehhez speciális súlyok készültek.) Talán valamivel „izgalmasabb” a gyermekminta másik felhasználási módja: ezzel a speciális kiegészítő mintával ugyanis lehetségessé válik a szülők és a saját gyermekeik adatainak összekapcsolása, a szülő-gyermek válaszadó diádok létrehozása.³ Az így létrejövő speciális szülő-gyermek adatbázis pedig egy különösen izgalmas kutatási kérdéscsoport vizsgálatára alkalmas, segítségével lehetséges például az életkörülmények, értékek, beállítódások intergenerációs átörökítésének vizsgálata.⁴ Például, vajon a szülők 2001–2002-ben vallott családi értékei, a szülői ház anyagi viszonyai hogyan hatnak a későbbi fiatal felnőtt demográfiai viselkedésére, társadalmi helyzetére. Ezt az adatbázist nevezzük „Életünk fordulópontjai szülő-gyermek adatbázisnak”. (Az adatbázis kapcsán felmerülő további kérdésekről, módszertani kihívásokról (pl. súlyozás kérdése) lásd Kapitány 2012.)

A 2012–2013-AS „FIATAL ALMINTA”

A negyedik adatfelvétel idejére már 11 év telt el az első adatfelvétel óta, így már az első hullám legfiatalabb válaszadói is közelítettek a harminc éves korhoz.

Emellett a lemorzsolódás is egyre komolyabb problémává vált. Az adatok alapján előre jelezhető volt, hogy a negyedik adatgyűjtés idejére az elemszám jóval a 10 ezer fős

² Ennek a mérete azonban elvileg a keresztmetszeti vizsgálatokban arányosnak a duplája, mivel nem csak az anyák lányait, illetve a férfiak fiait kerestük fel.

³ Természetesen ebben az esetben a minta érdekessége, hogy noha minden válaszadó gyermekhez egy szülő adatai kapcsolódnak, a válaszadó szülők száma mégis alacsonyabb, mint a gyermekeké, hiszen az adott öt évfolyamban született több testvér esetén több fiatal felnőtt válaszadóhoz tartozik ugyanaz a szülői adat. (Vagyis több szülő-gyermek diádban a szülői adat megegyezik.)

⁴ Az ilyen szülő-gyermek vizsgálatok klasszikus mintája az Arland Thornton nevéhez köthető Intergenerational Panel Study of Parents and Children (<http://www.umich.edu/~cdmh/longitudinal/thornton.html>).

célérték alá süllyed majd. Mivel a válaszadói lemorzsolódásra vonatkozó adatok a fiatal korcsoport nagyobb lemorzsolódását mutatták, úgy döntöttünk, hogy az adatgyűjtési programot nemcsak a hiányzó 11 nagykorú évfolyammal, hanem egy teljes, a gyermekvállalási korban lévőkre önmagában is reprezentatív almintával egészítjük ki, amelynek elemszáma 5000 fő körüli.

Annak érdekében, hogy minimalizáljuk a megghiúsult és az elutasított interjúk számát, úgy döntöttünk, hogy ismét alkalmazzuk azokat a kérdezéstechnikai és mintavételi lépéseket, amelyeket a panelvizsgálat 2001–2002-es indulásakor (részletesen lásd Kapitány 2003), másfelől az előző mintához való igazodás miatt egy további mintatechnikai lépést is szükséges volt megtennünk.

A kiegészítő minta mintavételének fő jellemzői azonosak az eredeti, 2001-es mintáéval: ismét egy, az interjúk elutasításának és megghiúsulásának lehetőségével számolva előzetesen torzított, régióként, településnagyság-kategóriáinként és életkori kategóriáinként rétegzett, többlépcsős mintavételre kerül sor. A mintavételi egység a személy, az alapsokaság az 1963. január 1. és 1994. december 31. között született, magyarországi lakóhellyel rendelkező magyar állampolgárok köre volt.

A mintavételhez szükséges adatok két különböző forrásból származtak: az alapsokaságra vonatkozó adatokat a Központi Statisztikai Hivatal, a válaszadó személyek tényleges adatait a Közigazgatási és Elektronikus Közszolgáltatások Központi Hivatala (KEKKH) biztosította.

Az alapsokaságot a lakónépesség 2012. január 1-i állapotára vonatkozó, KSH által továbbvezetett becsült száma képezte, ugyanis a minta tervezésekor ez volt a legfrissebb továbbvezetett információ. A 2011. évi népszámlálás megfelelően részletes adatai még nem álltak a rendelkezésünkre.

Magának a településmintának a kiválasztásakor úgy jelöltük ki a mintát, hogy a mintatelepülések megegyezzenek az eredeti 2001-es kiinduló minta településeivel. Noha a követett válaszadók egy jelentős része 2001 óta elköltözött az eredeti mintatelepüléséről (a longitudinális minta „dekoncentrált”), ez a 441 eredeti mintatelepülés mindmáig a válaszadóink túlnyomó többségének lakóhelyéül szolgál. Ebből logikusan következett, hogy a kiegészítő minta címei szintén ezen a 441 településen kerüljenek kijelölésre, ezzel ugyanis optimalizálhatóak a terepmunka költségei.

E cél elérése érdekében azonban minimálisan módosítani kellett a mintavételi eljárást, nem lehetett pontosan megismételni a 2001-es kiválasztási folyamatot. Az eltelt 11 évben ugyanis egyfelől jelentősen csökkent hazánk lakossága, másfelől az általános csökkenés közepette nagyban átrendeződött a települések lakosságszám szerinti sorrendje is. Több tucat település nagyság szerinti sorrendjét tekintve kategóriát váltott, illetve akár ki- (Kalocsa, Csongrád, Szarvas stb.) vagy bekerült (Dunaharaszti, Fót stb.) a 70 legnagyobb lakosságszámú település közé. Amennyiben tehát egyszerűen megisméltük volna a 2001-es eljárást, jelentősen bővült volna a mintatelepülések száma. Ennek érdekében a települési nagyság szerinti rétegeinket átalakítottuk, és az önreprezentálás határát 20 000 főre növeltük. Így a kiegészítő mintában az önreprezentáló települések száma 71-ről 60-ra csökkent, ugyanakkor a lakosság területi koncentrációja miatt az itt élők aránya alig csökkent 2001-hez viszonyítva.

A nem önreprezentáló települések esetén megmaradt a korábban használt öt kategória, annyi változtatással, hogy nem népességtizedek, hanem adott kerek számok szerint húztuk meg a határokat, mivel így is meglehetősen jól tudtuk garantálni a hasonló méretű rétegek létrehozását. (Például 2001-ben 1214 főig terjedt a legkisebb települések csoportja, 2012-ben 999 fő volt a határ.)

Természetesen az általunk választott mintavételi mód esetén nemcsak a megghiúsulások, hanem a válaszmegtagadások torzító hatását is meg kellett becsülnünk. A megtagadások és a megghiúsulások arányának együttes becslésekor a 2001-es megvalósulási arányokból indultunk ki, remélve, hogy azóta nem romlott a válaszadási arány. (Ebben azonban tévedtünk.)

A becsült alapsokaság a kiegészítő minta esetén 4 472 618 fő volt. (Ebből 1 521 682 fő 18–29 éves, 2 950 936 fő 30–49 éves.) Az 5000 fő eléréséhez tehát a kiválasztási arány hozzávetőleg 1/895 lenne, ha minden kiválasztott személy válaszolna. Az egyes csoportokhoz tartozó – a kiválasztási arányból következő – elméleti esetszámokat ezután minden esetben felszoroztuk a tervezett elérési aránnyal.

A felkeresendő személyek településenkénti számának kijelölése a következőképpen történt:

Az önreprezentáló települések két csoportja esetén a címeket egyszerűen az adott településre egyenként kiszámolt esetszám szerint osztottuk el.

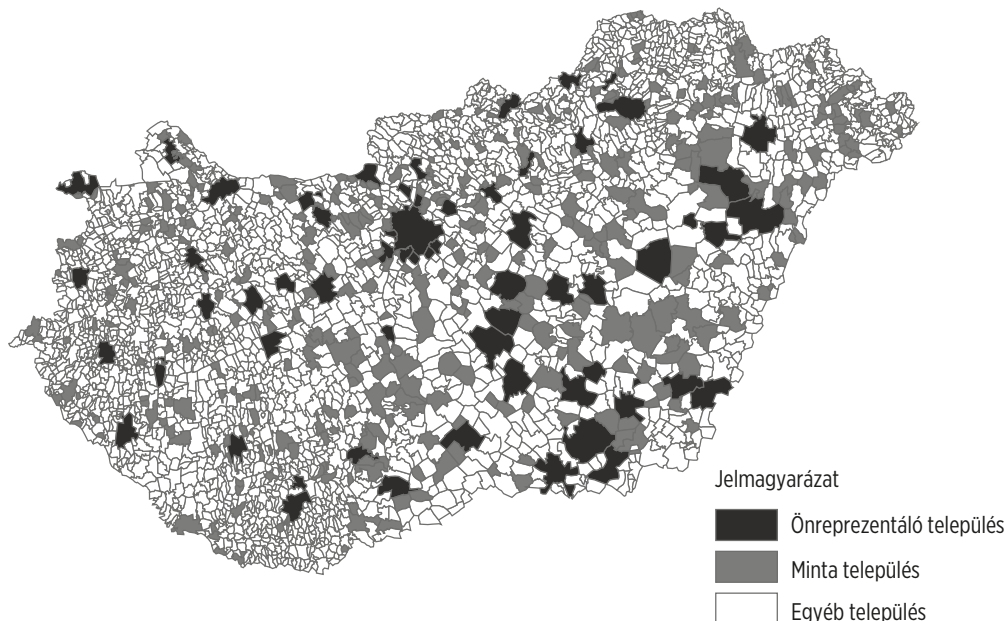
Budapest 23 kerülete esetében a kijelölendő címeket külön-külön számoltuk ki a kerület lakosságának korstruktúrája alapján úgy, hogy egységes (a legalacsonyabb) válaszadási hajlandósággal számoltunk.

A 20 ezer fő alatti településeknél területi és településkategóriánként (pl. Dél-Dunántúl, 1000–1999 fő közötti településeken lakók) állapítottuk meg a kikérendő személyi adatok számát a fenti módon, mintha egy területi és településkategória egyetlen település lenne. A kapott számot azután egyenletesen elosztottuk az adott településcsoportba tartozó eredeti mintatelepülések között. (Például a fenti településcsoport esetén Andocs, Görcsöny, Igal, Juta, Kajdacs, Kaposfő, Keszű, Madocsa, Mecseknádasd, Mosdós, Nagyberény, Somberek és Tevel községek között.)

Így megkaptuk mind a 441 településen felkeresendő 18–29, illetve 30–49 éves személyek számát, amely személyeket az adott településen élő megfelelő korú személyek közül véletlenszerűen jelölt ki számunkra a KEKKH 2012 augusztusában.

2012-ben a sikerességi arány lényegesen alacsonyabb volt, mint 2001-ben. Ez az elmaradás csak részben volt visszavezethető az alacsonyabb adatszolgáltatási hajlandóságra, a be nem jelentett kivándorlás legalább ilyen fontos ok volt. (A válaszadásra kijelölt válaszadók mintegy 6%-áról egyértelműen kiderült, hogy életvitelszerűen már nem Magyarországon él, másik 1,4%-ukról pedig ugyanez feltételezhető (Kapitány – Rohr 2013)). A kiegészítő mintában az adatszolgáltatók száma végül 4334 fő lett (a 8917 kijelölt személyből).

1. térkép: A vizsgálat magyarországi mintatelepülései (2011-es besorolás szerint)



AZ ÉLETÜNK FORDULÓPONTJAI - ERDÉLY VIZSGÁLAT

Mivel Románia is bekapcsolódott a GGP programba, ez felvetette annak lehetőségét, hogy egy majdani magyar–román összehasonlítás keretében megvizsgáljuk a Romániában élő magyarok demográfiai magatartását is. Azonban kiderült, hogy a romániai kutatás mintájában a magyar népesség elemszáma alacsony lesz az önálló elemzéshez, és hogy a romániai vizsgálat során kizárólag román nyelvű kérdőívvel dolgoztak. A fenti információk arra ösztönöztek, hogy az ÉF kutatás kérdőívét Erdélyben a magyar népesség körében magyar nyelven is lekérdezzük. Így került sor 2006-ban és 2008/2009 fordulóján két önálló adatfelvételi hullámra Erdélyben.

A 2006-ban induló *Életünk fordulópontjai - Erdély* (ÉFE) adatgyűjtés tehát egyszerre illeszkedik egy magyarországi adatfelvételhez és egy nagy nemzetközi vizsgálatosorozathoz, amelyben több európai ország, köztük Románia is részt vesz. Az *Életünk fordulópontjai* magyar és erdélyi adatgyűjtése során egységesen alkalmazott kérdésprogramnak köszönhetően összevethetjük az egyéni és a családi szintű demográfiai magatartás kulcsjellemzőit, az anyaországi és az erdélyi magyar társadalom meghatározó aspektusait. A GGP romániai adatbázisának felhasználásával pedig speciális magyarországi magyar–erdélyi magyar–romániai román összevetések is végezhetőek. Így megnyílik a lehetőség speciális kulturális/etnikai vs. társadalmi hatások vizsgálatára: Melyek azok a demográfiai/társadalmi viselkedési formák, amelyek etnikum-függők, tehát ahol az erdélyi és magyarországi magyarok mutatnak hasonlóságot, a románság pedig eltérést? Ezzel szemben milyen területeken húzódnak az országok közötti választóvonalak?

Emellett az ÉFE néhány lényeges Erdély-specifikus elemzésre is alkalmat nyújt, így például a kivándorlási szándékok mérésére, az etnikai kötődések részletesebb vizsgálatára. (További részletes információk az erdélyi kutatás koncepciójáról: Spéder 2009.)

A vizsgálat különlegessége a mintavételi eljárás, amely nem etnikai, hanem nyelvi alapon történt. Míg az erdélyi magyarokra vonatkozó reprezentatív vizsgálatok alapsokaságát jellemzően azok a személyek alkotják, akik magyar nemzetiségűnek vallják magukat, az ÉFE vizsgálat bekerülési szűrése nyelvi alapon történt. Azok vettek részt tehát a felvételen, akik magyar anyanyelvűek voltak, vagy nagyon jól beszéltek magyarul. (A mintavételről lásd részletesen: Kiss – Kapitány 2009.) Ennek következtében a vizsgálat egyedi elemzési lehetőségeket nyújt speciális becslések készítéséhez is: pl. magyar nyelvű romák száma Erdélyben, aktív és passzív magyar nyelvismeret az erdélyi magyar nagyvárosok lakói körében.

Természetesen az erdélyi adatgyűjtésnek, így az erre alapozott összehasonlításnak is megvannak a maga korlátai. A leglényegesebb ezek közül, hogy Erdélyben – anyagi okok miatt – az első adatfelvételi hullámban csak a fiatal (18–45 éves) népességet kérdeztük meg, s a minta csak a második hullámban terjedt ki az idősebbekre. Ez nyilván elemzési és tematikus megszorításokkal jár együtt.

AZ ADATGYŪJTÉSI PROGRAM ÉS A KÉRDŐÍVEK KIALAKÍTÁSÁNAK ELVEI

A kutatási koncepció összetettsége, és ehhez kapcsolódóan az adatgyűjtési program összetett céljai (lásd a kötet 1. fejezetét) nagyon sokféle módszertani kihívást jelentettek. Ezek a sokrétű igények ugyanis az adatgyűjtési program szinte minden módszertani elemét (mintavétel, súlyozás stb.) befolyásolták. A legnagyobb kompromisszumokat azonban egyértelműen a kérdőívek kialakítása, a kérdésprogramok kidolgozása vonatkozásában kellett kötni. A tervezett adatfelvétel technikailag adott keretei között olyan kérdőíveket kellett (volna) készíteni, amelyekből a kialakítható adatrendszerek a következő igényeknek egyszerre képesek megfelelni:

- az igen átfogó kutatási koncepciónak megfelelően áttekinthető képet nyújtsanak az életút fontosabb demográfiai fordulópontjairól, mind szubjektív, mind objektív adatok tekintetében,
- nemzetközi összehasonlításra alkalmasak legyenek (a *Generations and Gender Survey* (GGS) elvárásai),
- hazai összehasonlításokra alkalmasak legyenek (korábbi magyar vizsgálatok),
- keresztmetszeti elemzéseket lehessen készíteni a segítségükkel,
- az egyes adatgyűjtési hullámok eredményei egymással összehasonlíthatóak legyenek (változatlan kérdések),
- a longitudinális elemzések által nyújtott előnyök (oksági elemzések stb.) kihasználhatóak legyenek.

Az egyes tényezők fontossága ráadásul nagyban változott az adatgyűjtési program több mint egy évtizedes története során. Például ahogy kialakult a GGS kérdőívek szerkezete, és egyre több ország csatlakozott a programhoz, adatgyűjtésről adatgyűjtésre a magyar kutatók számára is fontosabb lett a nemzetközi összehasonlítás igényének való megfelelés.

Az adatfelvétel tervezésekor arra kellett ügyelnünk, hogy mindezeknek az igényeknek olyan módon feleljünk meg, hogy azok reális válaszadói terheket jelentsenek. Ez elsősorban – de nem kizárólag – a kérdezési idő hosszának a válaszadók számára még elviselhető szinten tartásában nyilvánult meg. Ugyanakkor ide tartozott a kérdezést könnyítő eszközök (pl. válaszlapok, élettörténeti kalen-dárium) alkalmazása, illetve az esetlegesen kényesnek, kínosnak számító adatok begyűjtésének módja is. (Egyes esetekben előre nyomtatott „nem kíván válaszolni” válaszkategória, anonim „borítékos kérdés” stb.)

A válaszadói terhek mellett fontos feladat volt a kérdezői munka összetettségének reális keretek között tartása is. Ahogy ugyanis egyre összetettebbé vált az adatgyűjtési program, a kijelölt válaszadók mintája, egyre nehezebbé vált a kérdezőbiztosok munkája is: a terepen célcsoportonként eltérő felkérő leveleket, marketing-anyagokat, rábeszélési technikákat és természetesen kérdőíveket kell párhuzamosan használniuk. Ami egy bizonyos határon túl már a kérdezői munka minőségének romlását eredményezhette volna. Ez is egy komoly érv volt például azon döntésünk mellett, hogy a harmadik hullám esetén nem keressük meg ismét az első hullámban válaszoló, de a második hullám idején ideiglenesen távol lévő mintatagokat.

A kérdezéstechnikai lehetőségek és adottságok is visszahatottak az adatgyűjtési programra, így a kérdésprogramra is (a terepmunkáról részletesen a 4. fejezetből tájékozódhatnak). Az első hullám idején egy egységes kérdőívvel dolgoztunk. A második hullám idején már lehetőség nyílt egyes – az első hullám idején felvett – adatok személyi szintű visszacsatolására, az élettörténeti kalendárium alkalmazására, amely csökkentette a válaszadói terheket, és megnövelte a begyűjtendő adatok mennyiségét. A harmadik adatgyűjtés idején már több különféle kérdőívvel dolgoztunk, és bevezettük az anonim önkéntes részt („borítékos kérdés”). Ez egyfelől csupán technikai kérdésnek tűnik, valójában azonban nagyban visszahat a begyűjtendő adatok körére (pl. a borítékos kérdés tette lehetővé a fogamzásgátlási szokások kérdésését, a több kérdőív alkalmazása pedig a gyermekminta lekérdezését). A negyedik hullám idején a laptopos kérdés bevezetése, illetve a weben keresztüli részleges önkéntes lehetőségének a megteremtése alapvetően hatott arra, hogy milyen adatokat lehet és érdemes gyűjteni. (Ezek a technikai változások lényegesen növelték a lehetőségeket, de azért bizonyos esetekben visszalépést is jelentettek.)

A fenti keretek között az egyes kérdőívek kialakítását minden adatfelvételi hullám esetén részletes, jellemzően több hónapos kérdőívviták előzték meg, amelyeken a kutató team tagjain kívül sok esetben külső szakértők is részt vettek.

Ezekon a vitákon egyszerre próbáltuk érvényesíteni a fent vázolt elveket, ami azonban értelemszerűen csak nagy kompromisszumokkal sikerült. Az eredmény egyfelől lehetőséget ad minden tervezett elemzési irány gyakorlati megvalósítására, másfelől viszont minden megközelítési mód szempontjából messze van a tökéletestől.

A Mellékletben található longitudinális kérdéskatalógus/témaadatbázis reményeink szerint jelentősen segíti majd az adatfelhasználók munkáját.

HIVATKOZÁSOK

- Kapitány Balázs (2003): A mintavétel, a nyers adatok megbízhatósága. In Kapitány Balázs (szerk.): *Módszertan és dokumentáció: Az „Életünk fordulópontjai” című demográfiai követéses vizsgálat első hullámának alapfelvételének ismertetése*. Műhelytanulmányok 2. KSH Népeségtudományi Kutatóintézet, Budapest, 29–39.
- Kapitány Balázs (2012): A hátrányos társadalmi helyzetek generációk közötti átörökítése: Egy magyarországi követéses vizsgálat eredményei. *Esély*, 23(2), 3–37.
- Kapitány Balázs – Rohr Adél (2014): Kivándorlás Magyarországról: Egy új becslési eljárás eredményei. In Spéder Zsolt (szerk.): *A család vonzásában: Tanulmányok Pongrácz Tiborné tiszteletére*. KSH Népeségtudományi Kutatóintézet, Budapest, 67–85.
- Kiss Tamás – Kapitány Balázs (2009): Magyarok Erdélyben: A minta kialakítása és az adatfelvétel. In Spéder Zsolt (szerk.): *Párhuzamok. Anyaországi és erdélyi magyarok a századfordulón*. KSH NKI Kutatási Jelentések 89. KSH Népeségtudományi Kutatóintézet, Budapest, 55–70.
- Spéder Zsolt (2009): Az anyaországi és az erdélyi magyar népeiséget összehasonlító kutatásról, az átmenet magyarországi és romániai kontextusáról. In Spéder Zsolt (szerk.): *Párhuzamok. Anyaországi és erdélyi magyarok a századfordulón*. KSH NKI Kutatási Jelentések 89. KSH Népeségtudományi Kutatóintézet, Budapest, 7–30.