

Výzkum životní dráhy a analýza sekvencí: možnosti studia rodinných drah*

Jana Chaloupková**
Sociologický ústav AV ČR, v.v.i.

Researching life course with sequence analysis: A new perspective for the study of family trajectories

Abstract: *The aim of this article is to present a specific method for the study of the life-course, which focuses on life-course trajectories as a whole through the use of sequence analysis. In the first part, two approaches for the quantitative analysis of the life-course are distinguished: an event-oriented perspective and a trajectory-based (holistic) perspective. The holistic perspective is based on sequence analysis and more specifically on optimal matching. The trajectory-based perspective does not focus on single life events, but on whole sequences of events. In the second part, using the Czech wave of the ISSP 2002 dataset, which includes partnership and family histories, this article presents several examples of the use of sequence analysis of family trajectories. This study shows that sequence analysis can help identify patterns associated with typical and distinctive life-course trajectories.*

Keywords: *sequence analysis, life-course, family trajectory, optimal matching*

*Data a výzkum - SDA Info 2009, Vol. 3, No. 2: 241-258.
(c) Sociologický ústav AV ČR, v.v.i., Praha 2009.*

Zdraví, rodinný stav, velikost rodiny, místo bydliště, zaměstnanecký status apod. nejsou statické veličiny, ale v průběhu života se mění v závislosti na individuálních i strukturních faktorech i předchozích událostech. Je proto užitečné nahlížet na ně dynamicky a sledovat jejich změny a následnost. Stále větší pozornost sociálních vědců se proto v posledních desetiletích namísto sledování stavu v jednom časovém bodě obrací ke studiu životních drah. Životní dráhu lze vymezit jako

* Tato stať vznikla v rámci projektu Proměny rodinných a pracovních drah v ČR podpořeného grantem GA AV ČR č. KJB700280802.

** Veškerou korespondenci pošlejte na adresu: Jana Chaloupková, Sociologický ústav AV ČR, v.v.i., Jilská 1, 110 00 Praha 1, nebo na e-mail: jana.chaloupkova@soc.cas.cz.

řetězec úzce propojených událostí nebo stavů v různých oblastech života, které člověk zažívá od narození do smrti [Alan 1989]. Na životní dráhu můžeme pohlížet jako na posloupnost profilů participace, tj. jako na posloupnosti stavů či rolí, kterými jedinec prochází v průběhu života v různých oblastech [Levy, Krüeger 2001]. Můžeme sledovat rodinné či pracovní dráhy, dráhy bydlení (rezidenční dráhy), dráhy vzdělání či zdraví apod. Obrat pozornosti ke studiu životních drah je úzce spojen s rozvojem teorie životní dráhy, rozšířením sběru longitudinálních dat a rozvojem metod jejich statistické analýzy. Dominantní metodou kvantitativního výzkumu životní dráhy je analýza historie událostí (event history analysis). Teorie životní dráhy naproti tomu usiluje o porozumění životní dráze jako celku a klade důraz na vzájemnou provázanost událostí či stavů, kterými jedinec v životě prochází. Zkoumání životní dráhy by se proto mělo zaměřovat nejen na časování událostí, ale i na jejich pořadí (následnost) a počet [Billari 2003]. Alternativou k perspektivě zaměřené na časování jednotlivých událostí je holistický přístup ke kvantitativní analýze životní dráhy – tzv. analýza sekvencí, zaměřující se na trajektorie jako celek. Metody sekvencí analýzy se rozvíjejí od 80. let 20. století, ale teprve v posledních letech se stávají součástí běžně užívaných statistických programů v sociálních vědách [Billari 2005, 2001].

Cílem tohoto článku je představit přístup ke kvantitativní analýze životní dráhy založený na sekvencí analýze. Nejprve stručně rozlišíme dva výše zmíněné metodologické přístupy ke studiu životní dráhy a poté se podrobněji zaměříme na samotnou sekvencí analýzu. Článek podá základní přehled možností analýzy sekvencí a za použití dat ISSP 2002 Rodina a gender role¹ představí několik příkladů jejího praktického využití při analýze rané fáze rodinné dráhy. Konkrétně se pokusíme identifikovat a charakterizovat typy rodinných drah mezi věkem 18 a 35 let. Pozornost přitom budeme věnovat zejména kohortnímu srovnání a zkoumání vývoje diverzity rodinných drah.

Přístup zaměřený na události

Nejrozšířenější přístup kvantitativní analýzy životní dráhy je založen na metodách analýzy přežití a analýzy historie událostí (*event history analysis*). Životní dráhu rozkládá na jednotlivé události či tranzice z jednoho stavu do druhého a snaží se odhalit, jaké faktory ovlivňují riziko (pravděpodobnost), že nastane v životě jedince určitá událost (např. narození prvního dítěte, vstup do manželství nebo vstup do zaměstnání)². Zajímají-li nás například vztahy práce a péče o rodinu, můžeme se zaměřit na zkoumání přechodu z pracovního zapojení k celodenní péči o dítě (mateřské/rodičovské dovolené), časování návratu do zaměstnání po rodičovské dovolené [např. Blossfeld, Drobnič 2001].

1 ISSP (International Social Survey Programme) je rozsáhlý mezinárodní výzkumný projekt, v jehož rámci probíhá od roku 1985 každoročně výzkumné šetření na určité téma. Témata se po určité době opakují. V současné době se ho účastní více než čtyři desítky zemí, v České republice toto výzkumné šetření koordinuje Sociologický ústav AV ČR, v.v.i. V tomto textu budeme z důvodů, které budou objasněny níže, využívat pouze data za ČR.

2 Základní přehled o metodě analýzy historie událostí podává např. Hamplová [2004] nebo Hendl [2004].

Tyto analýzy mohou odhalit, jaká skupina žen má tendenci zůstat kontinuálně v zaměstnání na plný úvazek či jaká skupina žen přeruší na určitou dobu své pracovní zapojení nebo zvolí dlouhodobý odchod z pracovního trhu do domácnosti. Jisté omezení ale spočívá v tom, že soustředěním se na jedinou tranzici (např. na návrat do zaměstnání z rodičovské dovolené) nebereme v úvahu, že kombinace pracovního zapojení a péče o rodinu je kontinuální proces a že situace v rodině či v zaměstnání se může v průběhu času měnit. Například matka může po návratu do zaměstnání zjistit, že se dítě špatně adaptuje na novou situaci nebo že je často nemocné, a rozhodnout se, že s ním ještě zůstane doma. Jinými slovy, sledovaný přechod z jednoho stavu do druhého nemusí být nevratný a výsledný stav nemusí být stabilní.

V mnoha situacích životní dráhy je možné přechod z jednoho stavu do druhého považovat spíše za pozvolně probíhající proces než za jednorázovou „událost“. Mnozí badatelé proto svoji pozornost přesunují od zkoumání časování jednotlivých událostí ke zkoumání následnosti událostí či stavů (tj. sekvencí). Poukazují například na to, že získání prvního zaměstnání neznamena ještě dosažení stabilního postavení na trhu a namísto zkoumání vstupu do prvního zaměstnání po ukončení studia se zaměřují na sledování případných změn zaměstnanického statusu v delším období [Brzinsky-Fay 2006; Scherer 2001]. Stejně tak studie zaměřující se na rodinné chování mohou vnímat formování partnerských vztahů a zakládání rodiny jako postupný proces, který není uzavřen vstupem do prvního partnerství.

Přístup zaměřený na trajektorie

Přístup zaměřený na trajektorie, který se v sociálních vědách rozvíjí od konce 80. let 20. století, se soustředí na životní dráhu jako na celek a bere v úvahu souběžné časování, počet událostí a jejich pořadí – odtud označení jako holistický [Billari 2001]. Trajektorie v tomto pojetí představuje konceptuální jednotku zkoumání. Na rozdíl od předchozího přístupu se nesoustředí na specifickou událost nebo tranzici, ale na individuální trajektorie reprezentované sekvencí stavů, které jedinec v jednotlivých časových jednotkách zaujímal. Dovoluje tak zkoumat heterogenitu existujících trajektorií a faktory, které ji ovlivňují.

Perspektiva zaměřená na trajektorie má přinejmenším dvojí zdůvodnění. Zaprvé tato perspektiva odpovídá představě životní dráhy jako projektu jedince. Umožňuje pohlízet na životní dráhu jako na smysluplný celek, který jedinec svým jednáním utváří. Za druhé i v případě, kdy nás zajímají primárně jednotlivé události životní dráhy, je třeba vzít v úvahu, že tyto události jsou ovlivněny předcházejícími událostmi a mají vliv na události následující. Je proto užitečné události zasadit do širšího kontextu životní dráhy a zkoumat nejen jejich časování, ale i to, co jim předcházelo [Billari 2001]. Výhodou tohoto přístupu je rovněž to, že dovoluje zkoumat paralelní trajektorie probíhající v různých oblastech života (např. rodinnou a pracovní dráhu) [Robette 2008; Aassve et al. 2007].

Metodologicky se holistický přístup opírá o sekvenční analýzu. Sekvenční analýza zahrnuje soubor metod, které se soustředí na komplexní popis trajektorií, identifikaci vzorců typických trajektorií a zkoumání faktorů, které ovlivňují

jejich diferenciaci, a popřípadě na zkoumání dopadů jednotlivých typů trajektorií na různé oblasti života (např. na životní spokojenost nebo zdraví) [Macindoe, Abbott 2004]. Cílem sekvenční analýzy je charakterizovat sekvence (trajektorie) v jejich celistvosti bez toho, aby byly redukovány na jednotlivé události. V sociálních vědách můžeme sekvenční definovat jako uspořádaný řetězec stavů nebo událostí. Stavů nebo událostí mohou být kódovány písmeny, řetězci znaků nebo čísly. Soubor všech stavů, které mohou v dané sekvenční nastat, se v sekvenční analýze označuje jako abeceda (*alphabet*).

Sekvenční data mají velmi komplexní povahu. V závislosti na délce sledovaných sekvencí a počtu možných stavů vzrůstá exponenciálně počet možných jedinečných sekvencí. Může se proto stát, že v datovém souboru nenalezneme dva jedince, jejichž trajektorie jsou naprosto identické. Například sledujeme-li pět stavů, kterými může jedinec opakovaně procházet v 10 časových bodech, můžeme teoreticky nalézt $5^{10} = 9\,765\,625$ rozdílných sekvencí. I když je pravděpodobné, že některé kombinace se v realitě nevyskytnou, je přesto možné očekávat vysokou rozmanitost pozorovaných sekvencí. Významnou součástí analýzy sekvencí je proto redukce komplexity dat a identifikace podobných sekvencí [Brzinsky-Fay et al. 2006; Abbott 1995]. Součástí sekvenční analýzy jsou tři typy úloh:

- popis a vizualizace trajektorií (charakteristika sekvencí a výpočet indikátorů jejich diverzity),
- srovnání podobnosti a nepodobnosti sekvencí a identifikace skupin podobných trajektorií,
- analýza příslušnosti do identifikovaných skupin sekvencí [Brzinsky-Fay et al. 2006].

Optimální seskupování

Patrně nejrozšířenější metodou sekvenční analýzy, která umožňuje porovnat podobnost sekvencí či trajektorií, je optimální seskupování (*optimal matching*)³. Jejím cílem je nalézt vzorce rytmů sociálních procesů, např. vzorce profesních trajektorií nebo vzorce použití času [Lesnard 2006]⁴. Metoda optimálního seskupování byla vyvinuta původně v molekulární biologii pro studium sekvencí DNA. Do sociálních věd byla uvedena Andrewem Abbottem a jeho kolegy [Abbott, Tsay 2000; Abbott 1995; Abbott, Hrycak 1990]. Východiskem optimálního seskupování je myšlenka, že čím jsou si dvě sekvence podobnější, tím snadněji je možné převést jednu do druhé.

3 K identifikaci skupin podobných trajektorií je možné použít i jiných metod – např. monotetický divizivní algoritmus [Piccarreta, Billari 2007] nebo harmonickou kvalitativní analýzu (AHQ), která analyzuje podíl času stráveného v jednotlivých stavech [Robette, Thibault 2008]. Alternativní metodu srovnávání sekvencí na základě hledání společných subsekvencí navrhl Elzinga [2003]. Základní informací o sekvenční analýze pro české čtenáře podávají Pakosta a Fučík [2009]. Podrobnější úvod do metody optimálního seskupování viz Macindoe a Abbott [2004] nebo Lesnard [2006].

4 Přehled aplikací optimálního seskupování v sociálních vědách uvádějí Abbott a Tsay [2000].

Podobnost (vzdálenost) dvou sekvencí tudíž závisí na tom, kolik operací je nutné provést, aby se tyto sekvence staly identickými. Srovnávat můžeme buď všechny trajektorie mezi sebou, nebo se stanovenou referenční trajektorií. Tato transformace může být prováděna prostřednictvím tří základních operací: vložením prvku (*insertion*), vypuštěním prvku (*deletion*), nebo nahrazením určitého prvku jiným (*substitution*). V praxi se vložení a vypuštění považuje za stejnou operaci (*indel*).⁵

Podobnost dvou sekvencí měříme podle počtu operací, které je nutné vykonat k jejich vzájemnému přiblížení. Cílem algoritmu je nalézt nejjednodušší (nejméně nákladný) způsob transformace. Jednotlivým operacím mohou být přitom přiřazeny různé váhy. Tyto váhy jsou označovány jako náklady transformace. Stanovení výše nákladů jednotlivých operací představuje klíčový krok optimálního seskupování [Gauthier et al. 2007; Lesnard 2006; Abbott, Tsay 2000]. Tento krok závisí na rozhodnutí výzkumníka. V nejjednodušším případě je operaci *indel* připsána váha jako 1 a substituci váha jako 2, protože při substituci je třeba provést vypuštění jednoho prvku a vložení druhého. Toto řešení odpovídá metodě nejdelší společné subsekvence [Gabadinho et al. 2009]. Pomocí různě vysokých nákladů je možné upřednostnit operace *indel* vůči substitucím nebo naopak⁶. Dále je možné stanovit různé náklady substitucím různých stavů. Můžeme totiž předpokládat, že některé stavy jsou si bližší než jiné, a proto je vhodné přiřadit substituci stavů, které jsou si v realitě vzdálené, vyšší náklady než substituci příbuzných stavů. V některých aplikacích této metody jsou náklady stanoveny fixně [Dijkstra, Taris 1995]. Ve většině případů jsou však substituční náklady rozlišeny. Substituční náklady mohou být stanoveny dvěma způsoby. Zaprvé, pokud je známa vzdálenost mezi jednotlivými kategoriemi, mohou být stanoveny ad hoc na základě teoretických předpokladů o povaze dat [Stovel, Bolan 2004; Scherer 2001; Halpin, Chan 1998]. Například při analýze profesní mobility mohou být přiřazeny nižší substituční náklady přechodu mezi kategoriemi nižší manuální a nižší nemanuální zaměstnání, zatímco substituci nižšího manuálního a vyššího manuálního zaměstnání mohou být přiřazeny vyšší náklady [Scherer 2001; Halpin, Chan 1998].

Pokud nejsou sledované stavy hierarchicky uspořádané a není předem zcela zřejmé, jaké tranzice mezi stavy jsou méně časté či obtížnější, doporučuje se stanovit substituční náklady na základě samotných dat a vypočítat pravděpodobnosti tranzice mezi různými stavy (prvky sekvence) [Rohwer, Pötter 2005]. Tato strategie určování nákladů patří mezi nejčastěji používané. Čím menší je pravděpodobnost přechodu z jedné kategorie do druhé, tím jsou náklady na substituci těchto dvou kategorií stanoveny výše. Metoda optimálního seskupování se postupně rozvíjí v reakci na předchozí kritiky a strategie určování nákladů se

5 Např. pokud srovnáváme dvě sekvence ABA a AA vypuštěním B z první sekvence získáme stejného výsledku jako vložení B do druhé sekvence.

6 *Indel* operace vložení nebo vypuštění stavu narušují průběh času a zpomalují nebo zrychlují studovaný sociální proces. Pokud nás zajímá časování událostí, je vhodné *indel* operacím připsat vyšší náklady. Naopak nižší náklady *indel* operacím je vhodné přiřadit při analýze nestejně dlouhých sekvencí [Lesnard 2006].

dále zdokonalují. Lesnard navrhuje ve výpočtu substitučních nákladů zohlednit i časování [Lesnard 2006]. Gauthier s kolegy doporučují stanovit substituční náklady na základě výpočtu pravděpodobností, že se dané dva prvky (stavy sekvence) budou vyskytovat na stejné pozici ve dvou různých sekvencích [Gauthier et al. 2007].

Matice tranzicí, ze kterých jsou substituční náklady odvozeny, je možné chápat jako reprezentaci makro vztahů v datech („kolektivní rytmy“) ukazující, které stavy jsou si bližší a které jsou si vzdálenější. Pomocí optimálního seskupování jsou matice tranzicí převáděny na inter-individuální distance mezi trajektoriemi [Lesnard 2006]. Výstupem optimálního seskupování je matice párových vzdáleností mezi zkoumanými sekvencemi. V případě, že jsou sekvence různé dlouhé (např. z důvodu neúplných dat), je třeba výsledné párové vzdálenosti standardizovat tím, že je vydělíme délkou delší sekvence z daného páru [Macindoe, Abbott 2004]. Vypočtené vzdálenosti mohou být dále analyzovány samostatně (např. průměrné vzdálenosti v určitých skupinách [Scherer 2001]), nejčastěji jsou ale použity jako základ pro klasifikaci pomocí shlukové analýzy či jiných klasifikačních metod. Takto vytvořené skupiny či typy trajektorií jsou většinou předmětem další analýzy.

Software pro sekvenční analýzu

Sekvenční analýza je v současnosti jednou z nejrychleji se rozvíjejících metod v sociálních vědách. Teprve v nedávné době se metody sekvenční analýzy staly součástí statistických programů běžně rozšířených v sociálních vědách, jako je STATA [Brzinsky-Fay et al. 2006]. Nástroje pro popis a vizualizaci sekvencí i pro identifikaci podobných trajektorií obsahuje také modul TraMineR (Trajectory Mining in R) v programu R, vyvinutý na univerzitě v Ženevě [Gabadinho et al., 2008]⁷. Jeho výhodou proti dříve zmíněnému je to, že je volně stažitelný. Optimální seskupování a řada deskriptivních statistik sekvencí je součástí freewarového programu TDA (Transitional Data Analysis) [Rohwer, Pötter 2005], který ale už není v současné době dále rozvíjen.⁸ Optimální seskupování je pochoitelně součástí řady programů určených pro analýzu sekvencí v biologii. Vedle toho existuje několik dalších programů specializovaných na měření podobnosti sekvencí v sociálních vědách, např. program Optimize vyvinutý v 90. letech ve spolupráci s Andrewem Abbottem⁹, program T-COFFEE/SALTT¹⁰ [Notredame et al. 2005], program CHESA [Elzinga 2003]. Tyto programy ale většinou nezhruňují prostředky pro deskripci a vizualizaci sekvencí.

7 Volně stažitelný na webové stránce <http://cran.r-project.org>, dokumentaci je možné najít na <http://mephisto.unige.ch>.

8 Program TDA (Transition Data Analysis) vytvořili Goetz Rohwer and Ulrich Poetter. Tento statistický program je možné si bezplatně stáhnout na <http://www.stat.ruhr-uni-bochum.de/tda.html>.

9 Volně dostupný ke stažení na: <http://home.uchicago.edu/~aabbott/om.html>

10 Program SALTT (Search Algorithm for Life Trajectories and Transitions) je volně stažitelný na <http://www.tcoffee.org/saltd>. Výhodou tohoto programu je to, že umožňuje pracovat s vícedimenzionálními trajektoriemi [Gauthier et al. 2008].

Použitá data a konstrukce partnerských a rodinných drah

Následující ukázka možností sekvenční analýzy využívá dat výzkumného šetření ISSP Rodina a gender role 2002¹¹. Toto výzkumné šetření bylo v České republice rozšířeno o modul otázek týkajících se partnerské a pracovní historie, které je možné využít pro sekvenční analýzu. V tomto článku se zaměříme na data partnerské (rodinné) historie. Prvním krokem sekvenční analýzy je uspořádání dat obsahujících informace o jednotlivých událostech a datu, kdy nastaly, do podoby uspořádaného řetězce událostí či stavů. Respondenti byli v české verzi výzkumného šetření ISSP 2002 dotazováni na to, zda žili v době sběru dat s partnerem/partnerkou a dále retrospektivně na případné předchozí korezidenční partnerské vztahy¹². Ve všech případech byly sbírány informace o počátku partnerského soužití (měsíc a rok) a s výjimkou současného partnerského soužití i datum ukončení partnerského soužití. Data obsahují i informace o tom, zda se jednalo o nesezdané soužití nebo manželství, popřípadě datum pozdějšího vstupu do manželství. Dále data obsahují údaje o datu narození dětí.

V této stati budeme sledovat pohyb mezi následujícími 6 situacemi: 1) nežije s partnerem(kou), bezdětný(á) (So); 2) nežije s partnerem(kou), dítě (SD); 3) žije v nesezdaném soužití, bezdětný (á) (NSo); 4) žije v nesezdaném soužití, dítě (NSD), 5) žije v manželství, bezdětný (á) (Mo); 6) žije v manželství, dítě (MD). Abychom zaručili, že všichni respondenti prošli sledovaným obdobím, omezili jsme sledovanou životní dráhu na věkový interval mezi 18 a 35 lety a pracujeme pouze s lidmi, kterým v době sběru dat bylo 35 let a více (530 žen a 299 mužů). Počet lidí, kteří byli ve 35 letech bezdětní, je poměrně nízký (7,7 žen a 13,7 % mužů). Rodinná dráha byla rekonstruována z dat ve dvou krocích. Nejprve byl vypočítán věk, ve kterém došlo ke sledovaným událostem v rodinné dráze (počátek soužití s partnerem/kou, vstup do manželství, případný konec prvního partnerského soužití a narození dítěte¹³). Poté byla vytvořena sada 35 nových proměnných reprezentujících půlroční věkové intervaly od 18 do 35 let a každé z nich byla přiřazena příslušná rodinná situace.

Data obsahují informace o rodinné dráze ve věku mezi 18 a 35 lety pro 471 žen a 263 mužů. V případě 9 mužů a 9 žen obsahuje rodinná dráha chybějící údaje. Celkově mohly být rodinné dráhy zkonstruovány jen pro 88 % respondentů, protože někteří nevedli informace o datech sledovaných událostí. V celku 734 sekvencí se nachází 398 odlišných trajektorií. Nejběžnější formou zápisu sekvence je záznam o stavu v každé sledované časové jednotce v řádku za sebou. Tabulka 1 uvádí příklad rodinné dráhy ženy, která začala žít ve 22 letech v nesezdaném soužití, o rok později se provdala a ve 26 letech se jí narodilo dítě. Na základě informací o

11 Výběr respondentů byl proveden metodou stratifikovaného pravděpodobnostního výběru. Původní datový soubor obsahuje údaje o 1289 respondentech ve věku od 18 let.

12 Respondenti byli dotazováni na předchozího a před-předchozího partnera(ku). Z technických důvodů byli do analýzy zařazeni pouze lidé, kteří žili s maximálně dvěma partnery a 38 lidí (2,8 %), kteří žili dříve s minimálně dvěma partnery, bylo z analýzy vyloučeno [srv. Hamplová 2003]. Důvodem je to, že v případě před-předchozího partnerství nevíme, zda mu předcházelo ještě jiné nebo zda se jednalo o první partnerství respondenta.

13 Pokud chyběl údaj o měsíci, byl použit pouze údaj o roku.

Tabulka 1. Příklad rodinné dráhy respondentky mezi 18 a 35 lety věku

Věk	18	18,5	19	19,5	20	20,5	21	21,5	22	22,5
Rodinný stav	So	So	So	So	So	So	So	So	NSo	NSo
Věk	23	23,5	24	24,5	25	25,5	26	26,5	27	27,5
Rodinný stav	Mo	Mo	Mo	Mo	Mo	Mo	MD	MD	MD	MD
Věk	28	28,5	29	29,5	30	30,5	31	31,5	32	32,5
Rodinný stav	MD	MD	MD	MD	MD	MD	MD	MD	MD	MD
Věk	33	33,5	34	34,5	35					
Rodinný stav	MD	MD	MD	MD	MD					

Poznámka: So – nežije s partnerem, bezdětná; NSo – žije s partnerem v nesezdaném soužití, bezdětná; Mo – žije v manželství, bezdětná; MD – žije v manželství s dítětem

věku při počátku partnerského soužití a vstupu do manželství a při narození dítěte můžeme zkonstruovat průběh rodinné dráhy. Pro analýzu sekvenčních dat byl použit modul TraMineR v programu R [Gabadinho et al. 2009].

Další možností zápisu je zkrácený záznam (tzv. state-permanence), který uvádí posloupnost jednotlivých stavů spolu s informací o délce trvání daného stavu [Elzinga, Liefbroer 2007; Aassve et al. 2007]. Ve zkrácené formě můžeme tuto životní dráhu zapsat jako (So,8)-(NSo,2)-(Mo,6)-(MD,19), kde číslice udává dobu strávenou v daném stavu (v našem případě počet půlroků), případně S-NS-M-MD, pokud nás zajímá pouze pořadí odlišných stavů [Gabadinho et al. 2009].

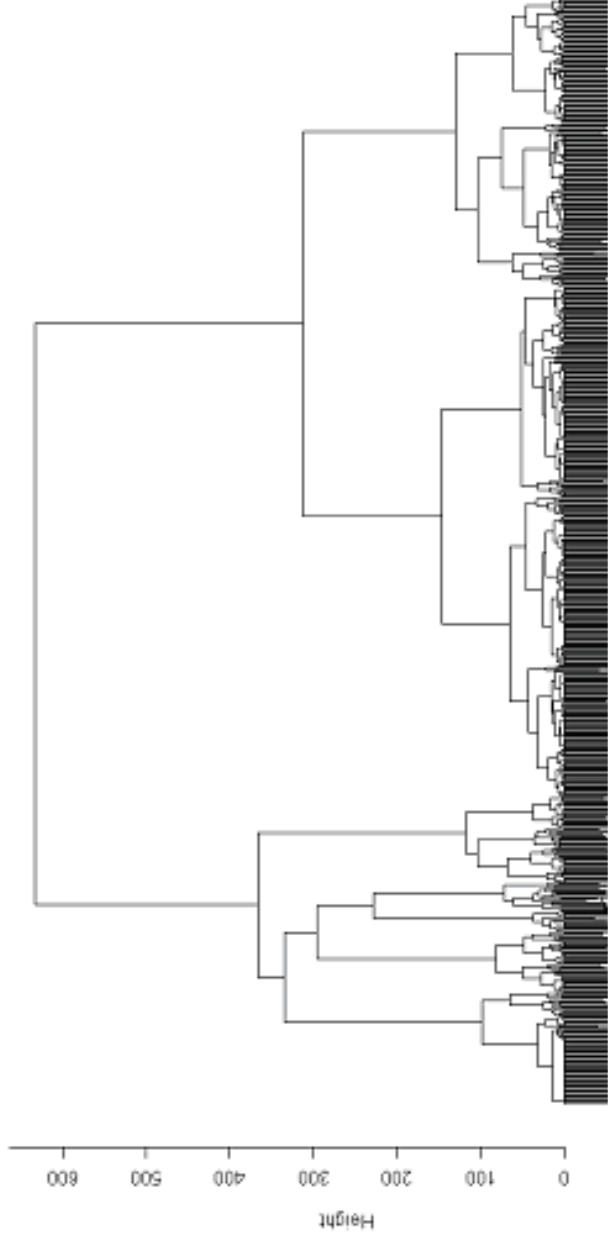
Typy rodinných drah ve věku mezi 18 a 35 lety

Na základě vzdáleností vypočtených metodou optimálního seskupování se substitučními náklady odvozenými empiricky z dat¹⁴ je možné identifikovat tři hlavní typy rané fáze rodinné dráhy (viz Graf 1). Pro shlukování byla použita Wardova metoda, která je založena na minimalizaci vnitroshlukového rozptylu [Macindoe, Abbott 2004].

Nejpočetnější skupinu 3, do které spadá 76 % žen a 65 % mužů, můžeme označit jako klasickou rodinnou dráhu. Charakterizuje ji poměrně časný vstup do manželství bezprostředně následovaný narozením prvního dítěte. Jako specific-

14 Matice substitučních nákladů je uvedena v příloze.

Graf 1. Hierarchická shluková analýza. Typy rodinné dráhy



Zdroj: ISSP 2002, podsoubor 35+, N=734.

Poznámka: Do hierarchického shlukování vstupovaly vzdálenosti mezi rodinnými dráhami vypočtené na základě optimálního seskupování (Agglomerative Nesting), Wardova metoda.

Tabulka 2. Průměrný věk při vstupu do prvního partnerského soužití, manželství a při narození dítěte

		věk při počátku 1. partn. soužití	věk při vstupu do 1. manželství	věk při narození 1. dítěte
Skupina 1 Singles (N=146)	ženy	24,5	25,9	30,8
	muži	29,2	30,1	33,0
	N	105	93	70
Skupina 2 Osamělí rodiče (N=57)	ženy	21,6	21,8	22,7
	muži	30,8	26,6	24,2
	N	28	25	57
Skupina 3 Klasic. rodinná dráha (N=531)	ženy	21,6	21,9	23,0
	muži	23,5	23,8	25,0
	N	529	529	526

Zdroj: ISSP 2002, podsoubor 35+.

Pozn.: Průměry jsou vypočteny jen pro ty, kteří danou tranzicí prošli.

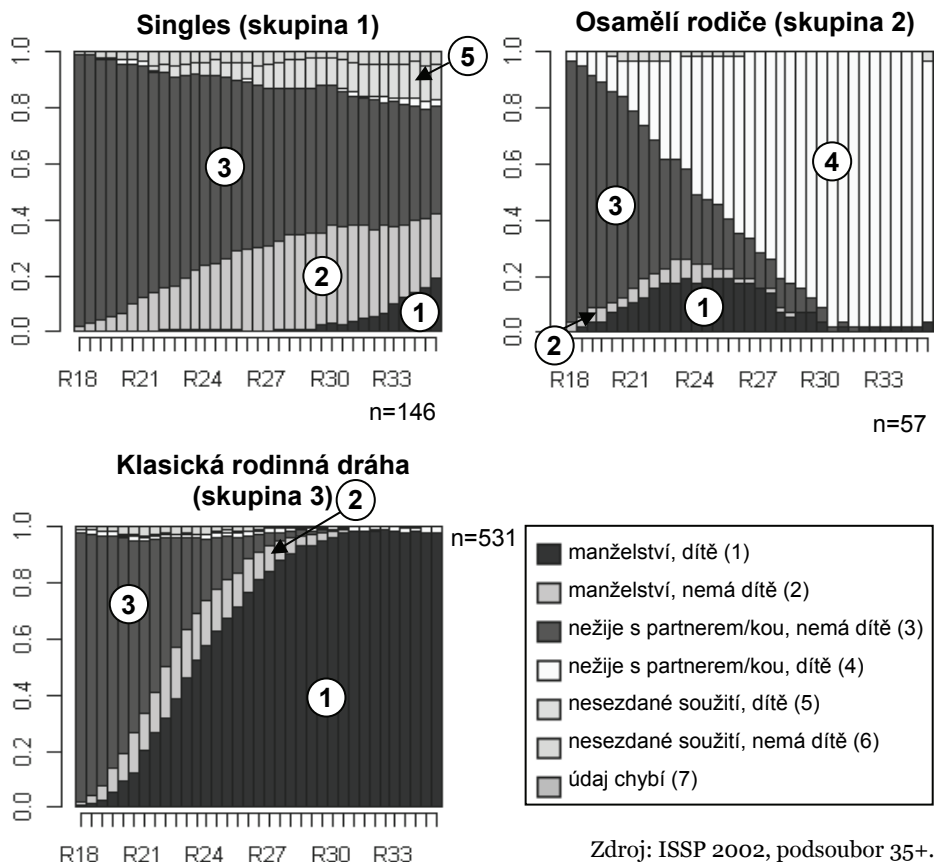
ký klastr se vydělují singles (Skupina 1), kteří po velkou část sledovaného období žili bez partnera, popřípadě žili v nesezdaném soužití a charakterizuje je pozdější vstup do rodičovství (srov. Tabulka 2). Tento typ rané fáze rodinné dráhy zaujímal 29 % mužů a 15 % žen. Vyšší zastoupení mužů v této skupině je dáno rozdílným časováním rodinných událostí u mužů a žen a odráží skutečnost, že muži vstupují do manželství a rodičovství v průměrně o několik let vyšším věku než ženy.

Poslední, početně nejmenší skupinu zahrnující 9 % žen a 6 % mužů tvoří osamělí rodiče (Skupina 2). Tato skupina zahrnuje jednak nikdy neprovdané rodiče, a dále ty, kteří uzavřeli v poměrně nízkém věku sňatek a brzo nato se rozvedli. Na tomto místě je třeba připomenout, že zde bereme v úvahu pouze partnerskou situaci a rodičovství a nikoli to, zda rodič žije s dítětem ve společné domácnosti. Je možné předpokládat, že matky žijící bez partnera žijí častěji společně se svým dítětem než svobodní či rozvedení otcové.

Následující grafy ukazují rozložení rodinných stavů v jednotlivých časových bodech podle věku (Graf 2). Tyto grafy prezentují informace o rodinných drahách v agregované podobě. Ukazují, jaký podíl mladých lidí se v daném věku nacházel v určité rodinné situaci. Oproti předchozí sadě grafů, která ukazuje jakousi sérii momentek rodinné situace v jednotlivých letech věku, druhá sada grafů (Graf 3) zobrazuje průběh individuálních trajektorií. Každý z horizontálních sloupců ukazuje rodinné situace, kterými mezi 18 a 35 lety jedinec procházel. Z důvodu zachování přehlednosti grafy ukazují pouze deset nejčastěji se vyskytujících individuálních trajektorií v každém ze tří typů rodinných drah. Šířka horizontálních sloupců odpovídá jejich relativnímu zastoupení v daném typu.

Doplňující charakteristiku jednotlivých typů rodinných drah nám může přinést informace o průměrné době trvání jednotlivých stavů (Tabulka 3). Zatímco singles (Skupina 1) strávili ve sledovaném období nejdelší čas ve stavu „nežije

Graf 2. Typy rodinných drah mezi 18 a 35 lety věku: rodinná situace v jednotlivých letech věku (agregovaná data)

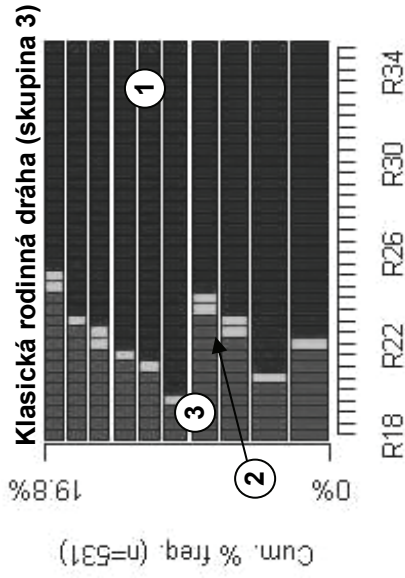
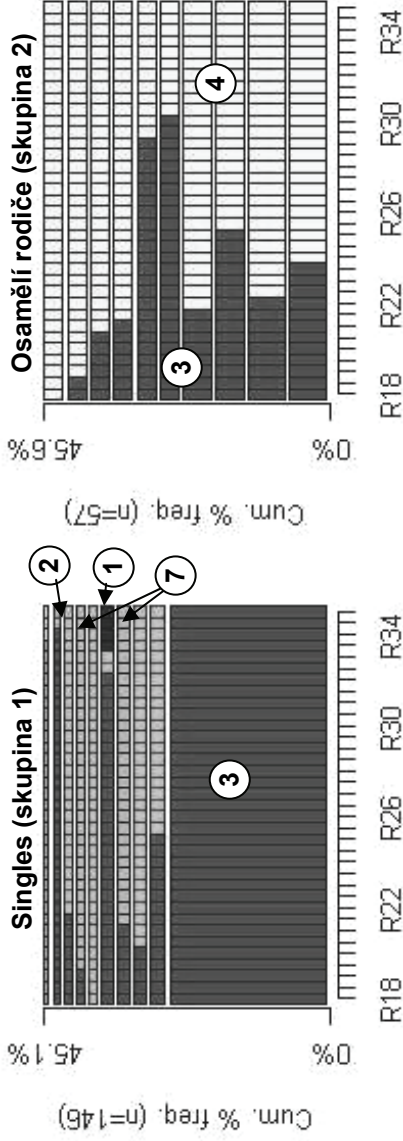


Zdroj: ISSP 2002, podsoubor 35+.

s partnerem, partnerkou, bezdětný“ – v průměru více než polovinu sledované doby – téměř 10 let v období mezi věkem 18 a 35 let, v rodinné dráze osamělých rodičů dominuje stav „nežije s partnerem(partnerkou), dítě“ a v klasické rodinné dráze stav „žije s manželem nebo manželkou a dítětem“ (Tabulka 3).

Dále nás zajímalo, zda existují rozdíly podle vzdělání a kohorty narození (Tabulka 4). Typ rodinné dráhy souvisí se vzděláním a délkou studia. Téměř třetinu lidí, kteří dosáhli vysokoškolského vzdělání, můžeme zařadit mezi singles, kteří odkládají založení rodiny do pozdějšího věku. Ve skupině lidí s maturitou zaujímalo dráhu singles do 35 let věku jen necelých 15 % a mezi lidmi bez maturity přibližně jedna pětina. Mezi lidmi s vyšším než maturitním vzděláním je tudíž relativně nejméně zastoupena klasická rodinná dráha charakterizovaná raným sňatkem a rodičovstvím a je mezi nimi rovněž nejméně zastoupena rodinná dráha osamělého rodiče (pouze 3,1 % ku přibližně 8 % lidí s nižším vzděláním).

Graf 3. Deset nejčastějších individuálních rodinných drah mezi 18 a 35 lety věku v jednotlivých typech



Zdroj: ISSP 2002, podsoubor 35+.

Tabulka 3. Průměrná doba strávená v jednotlivých stavech mezi 18 a 35 lety věku

	Skup. 1 Singles	Skup. 2 Osamělí rodiče	Skup. 3 Klasická rodinná dráha	Celkem
Manželství s dítětem (MD)	0,99	3,04	22,83	16,95
Manželství, bezdětný(á) Mo	7,25	0,91	2,53	3,34
Nežije s partnerem(kou), bezdětný(á)So	19,62	9,39	8,58	10,84
Nežije s partnerem(kou), dítě (SD)	0,25	21,23	0,44	2,02
Nesezdané soužití, dítě (NSD)	2,12	0,04	0,15	0,53
Nesezdané soužití, bezdětný(á) (NSo)	1,16	0,30	0,43	0,57
N	146	57	531	734

Zdroj: ISSP 2002, podsoubor 35+.

Pozn.: Hodnoty udávají počet půlroků.

Tabulka 4. Sociálně-demografické charakteristiky typů rodinných drah (v řádkových %)

		Skup. 1 Singles	Skup. 2 Osamělí rodiče	Skup. 3 Klasická rodinná dráha	Celkem (N=100 %)
Gender ¹⁾	ženy	14,9	8,7	76,4	471
	muži	28,9	6,1	65,0	263
Nejvyšší dosažené vzdělání ²⁾	bez maturity	20,9	7,9	71,2	445
	SŠ s maturitou	14,5	8,2	77,3	220
	VŠ	31,3	3,1	65,6	64
Kohorta narození ³⁾	nar. do r. 1951	19,3	8,3	72,4	362
	nar. 1952-61	21,5	5,2	73,4	233
	nar. 1962-1967	18,7	10,8	70,5	139

Zdroj: ISSP 2002, podsoubor 35+.

Poznámky: Chi2 Test: 1) $p < 0.001$, 2) $p < 0.05$, 3) není signifikantní.

Dostupná data umožňují srovnávat počáteční fázi rodinných drah lidí narozených mezi roky 1919–1967 (tj. těch, kterým bylo při sběru dat minimálně 35 let). V následujících analýzách budeme rozlišovat tři kohorty: lidi narozené do r. 1951, narozené mezi rokem 1952–1961 a narozené v letech 1962–1967. Přestože se příslušníci nejmladší kohorty narodili a vyrůstali za socialismu, do dospělosti vstoupili krátce před pádem komunistického režimu a většinu ze sledovaného období tudíž prožili v nových sociálních, politických a ekonomických podmínkách po roce 1989. I když můžeme pozorovat mírné zvýšení zastoupení rodinných drah osamělých rodičů, rozdíly mezi charakterem rodinných drah mezi sledovanými kohortami nejsou příliš výrazné. Důvodem je, že tato data neobsahují informace

o mladších kohortách lidí narozených v 70. letech, kteří byli hlavními nositeli změn rodinného chování po roce 1989.

Srovnání variability trajektorií

Jednou z významných otázek zkoumání životní dráhy je otázka, do jaké míry se životní dráhy individualizují a diferencují. Teorie individualizace předpokládá, že se v současné společnosti biografie stávají komplexnější a rozmanitější, což se odráží v diferenciaci životních situací a rozrůžňování jejich časování o oslabování modelu univerzální životní dráhy. Rodinné dráhy obsahují ve vzrůstající míře nové formy partnerského a soukromého života a rozšiřuje se nesezdané soužití nejenom jako předstupeň manželství, ale i jako alternativa a roste počet dětí narozených mimo manželství. Sekvenční data umožňují testovat hypotézu o de-standardizaci životní dráhy a zkoumat, zda se míra de-standardizace životní dráhy rozlišuje v různých sociálních skupinách [Widmer, Ritschard 2009; Elzinga, Liefbroer 2007; Fussell 2005; Scherer 2001].

Variabilitu trajektorií můžeme měřit pomocí indexu entropie [Fussell 2005], a to ze dvou úhlů: variabilitu stavů v jednotlivých časových bodech, např. letech věku (transverzální přístup) nebo variabilitu stavů v rámci individuální životní dráhy (longitudinální přístup) [Widmer, Ritschard 2009; Gabadinho et al. 2009].¹⁵

Ukazatel transverzální entropie podává informaci o tom, nakolik jsou životní situace jedinců v určitém věku homogenní – nebo naopak nakolik jsou rozdílné. Pokud by se všichni jedinci stejného věku nacházeli ve stejném stavu, index entropie by byl nulový. Index entropie můžeme využít pro srovnání variability životní dráhy v různých sociálních skupinách nebo pro srovnání vývoje variability životní dráhy v čase.

Naproti tomu index longitudinální entropie dovoluje zkoumat míru stability či nestability individuální životní dráhy. V tomto případě by nulové entropie dosahovali jedinci, kteří by po celou sledovanou dobu setrvali ve stejné životní situaci. Dalším konceptem měřícím diverzitu životní dráhy je koncept turbulence, formulovaný C. Elzingou [Elzinga, Liefbroer 2007], který vychází z výpočtu longitudinální entropie, ale na rozdíl od výpočtu entropie zohledňuje i počet subsekvencí dané trajektorie¹⁶. Pojem turbulence pochází z hydrodynamiky, kde označuje vlastnost toku, jehož rychlost a směr je nestabilní a tedy se rychle a nepravidelně mění. Vztáhneme-li tento pojem k životní dráze, jako turbulentní můžeme označit životní dráhu, pokud zahrnuje mnoho odlišných subsekvencí a pokud je rozptýl doby trvání jednotlivých stavů nízký [Elzinga, Liefbroer 2007].¹⁷

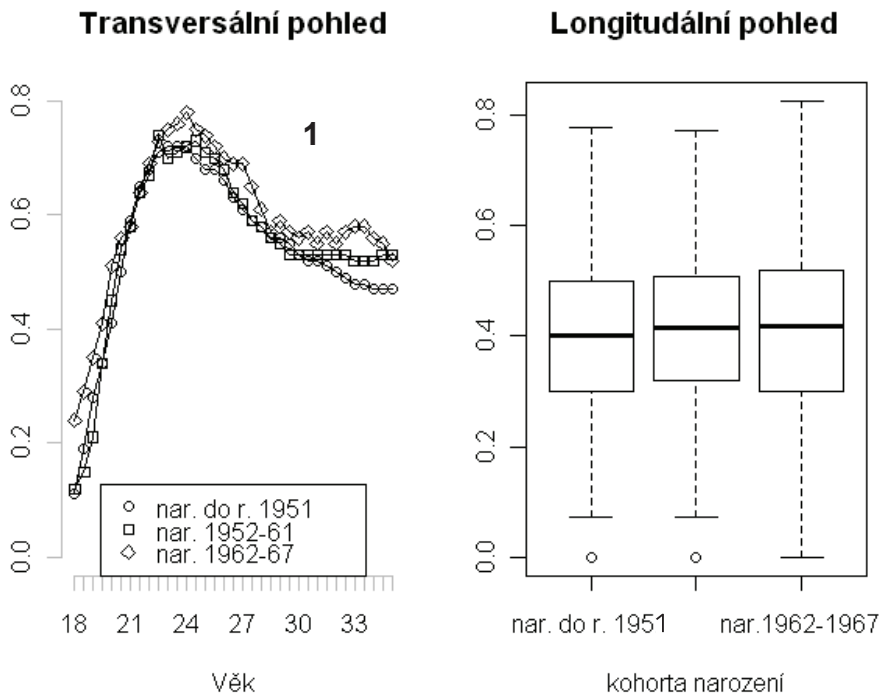
15 Entropii lze vypočítat následujícím způsobem:
$$E_t = -\sum_{j=1}^q p_j \log(p_j)$$

p_{ij} označuje podíl jedinců nacházejících se v čase t ve stavu j (stavy $j=1, \dots, q$) [Rohwer, Potter 2005].

16 Subsekvence sekvence A obsahuje stavy obsažené v sekvenci A , a to ve stejném pořadí, ale nemusí být bezprostředně následující, např. AC je subsekvencí sekvence ABC .

17 Vzorec výpočtu turbulence je následující:
$$T(x) = \log_2 \left(\Phi(x) * \frac{s_{r,\max}^2(x) + 1}{s_r^2(x) + 1} \right)$$

Graf 4. Míry diverzity rodinných drah mezi 18 a 35 lety podle kohorty (index entropie)



Zdroj: ISSP 2002, podsoubor 35+.

Poznámka: Rozdíly mezi průměry longitudinální entropie mezi kohortami nejsou statisticky významné.

Pro bližší vysvětlení konceptu turbulence Elzinga a Liefbroer [2007] uvádějí příklad tří sekvencí rodinných drah: (1) S-NS-M-MD; (2) S-NS-S-M a (3) S-NS-M-S. Všechny tři sekvence obsahují po čtyřech stavech, ale v sekvenci 2 a 3 se opakuje stav S. Nejvíce jedinečných stavů obsahuje trajektorie 1, a proto ji můžeme označit jako nejturbulentnější. Sekvence 2 a 3 mají stejný počet odlišných stavů, ale v sekvenci 3 se jedinec do stavu S vrací po dvou předchozích změnách, zatímco v sekvenci 2 pouze po jedné. Sekvence 3 je proto turbulentnější (zahrnuje více subsekvencí) než sekvence 2.

Výpočet entropie na datech ISSP 2002

Graf 4 ukazuje srovnání míry entropie podle kohort v transversálním a longitudinálním pohledu. Při výpočtu bylo použito normované míry entropie (hodnoty byly

$s_{t,\max}^2 = (n-1) (1-\bar{t})^2$, $\Phi(x)$ označuje počet subsekvencí sekvence x , s_t^2 je rozptyl délky trvání stavů sekvence x a $s_{t,\max}^2$ je maximum rozptylu, který může nabývat vzhledem k celkové délce sekvence (n). \bar{t} značí průměrné trvání stavu sekvence x (podíl doby trvání sekvence a počtu odlišných stavů v sekvenci) [Elzinga, Liefbroer 2007].

vyděleny maximální hodnotou, která může v daném případě entropie nastávat¹⁸). Index entropie tudíž nabývá hodnot mezi 0 a 1, kdy 0 označuje úplnou homogenitu rodinných stavů (popř. stabilitu rodinné dráhy) a 1 maximální heterogenitu (nestabilitu). Z prvního grafu je patrné, že se heterogenita rodinných situací v průběhu mladé dospělosti mění – zatímco kolem 18 let je heterogenita rodinných drah poměrně nízká, postupně se zvyšuje a vrcholu dosahuje kolem 25 let a poté se opět snižuje. Tento graf potvrzuje, že v mladších kohortách jsou rodinné dráhy více rozrůzněny. Druhý graf přináší srovnání míry nestability individuálních trajektorií. Výpočty nicméně nenaznačují, že by mezi kohortami narozenými mezi 1919-1967 docházelo k výraznému zvyšování nestability v rané fázi rodinných drah.

Závěr

Cílem tohoto článku bylo podat základní přehled možností holistického přístupu ke studiu životní dráhy. Sekvenční analýza představuje přínosný nástroj pro analýzu longitudinálních dat jakožto sekvenci stavů nebo událostí a představuje komplementární přístup k analýze historie událostí. Sekvenční analýza umožňuje exploraci dat životní dráhy – systematický popis trajektorií a identifikaci jejich vzorců. Za pomoci dat o rodinné a partnerské historii mezi 18 a 35 lety věku byly představeny některé postupy analýzy trajektorií v jejich celku. Tyto analýzy si nekladly za cíl vysvětlit změny rodinných drah ani testovat teorie o jejich diferenciaci, ale jejich hlavním cílem bylo představit holistickou perspektivu ke zkoumání životní dráhy a poukázat na typy otázek, na které je možné tento typ metod využít. Tyto otázky si jistě zaslouží větší pozornosti a hlubší analýzu. Použitá data obsahují informace jen o lidech, kterým bylo v roce 2002 nejméně 35 let. Pokud bychom chtěli rozšířit sledovaný věkový interval, musela by být dolní hranice věku ještě vyšší. Tato data tak neumožňují zachytit změny rodinného chování, ke kterým dochází od počátku 90. let 20. století. Studium životních trajektorií je možné až z většího časového odstupu.

Analýza sekvencí klade vysoké nároky na kvalitu dat a je vysoce citlivá na chybějící pozorování. Vzhledem k tomu, že můžeme předpokládat, že méně nestabilní trajektorie mohou obsahovat více chybějících údajů, může být heterogenita trajektorií v datech podreprezentována [Aassve et al. 2007]. Navíc se pod zdánlivě stabilními trajektoriemi mohou v závislosti na časové jednotce pozorování skrývat změny. Například pokud daný stav trval méně než šest měsíců, nemusí být v našich datech zaznamenán (např. pokud krátce po rozpadu prvního partnerského soužití začal respondent žít s novou partnerkou, nemusí být tato změna vůbec zaznamenána a jeho situace se bude jevit jako stabilní).

Analýzy dat ISSP 2002 zahrnujících informace o partnerské a rodinné historii identifikovaly tři typy rané fáze rodinné dráhy a jejich charakteristiky a představily některé možnosti popisu a vizualizace sekvencí dat. Ukázaly, že součástí analýzy sekvencí nejsou jen metody zaměřené na odhalování typických trajektorií, ale že sekvencí analýza přináší i nástroje pro zkoumání diverzity životních drah a jejího vývoje v čase.

18 Maximální entropie by bylo dosaženo, pokud by v dané sekvenci byly rovnoměrně rozloženy všechny sledované stavy.

Literatura

- Aassve, A., F. Billari, R. Piccarreta. 2007. „Strings of Adulthood: A Sequence Analysis of Young British Women’s Work-Family Trajectories“. *European Journal of Population* 23 (3-4): 369-388.
- Abbott, A. 1995. „Sequence analysis: new methods for old ideas“. *Annual Review of Sociology* 21: 93-113.
- Abbott, A., A. Hrycak. 1990. „Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians’ Careers“. *The American Journal of Sociology* 96 (1): 144-185.
- Abbott, A., A. Tsay. 2000. „Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect“. *Sociological Methods Research* 29 (3): 3-31.
- Alan, J. 1989. *Etapy života očima sociologie*. Praha: Panorama.
- Billari, F. 2003. „Life course analysis.“ Pp. 588-590 in ed. P. G. Demeny, G. McNicoll. *Encyclopedia of Population*. New York: Macmillan Reference.
- Billari, F. C. 2001. „Sequence Analysis in Demographic Research“. *Canadian Studies in Population* 28 (2): 439-458.
- Billari, F. C. 2005. „Life course analysis: two (complementary) cultures? Some reflections with examples from the analysis of the transition to adulthood.“ Pp. 261-282 in ed. R. Levy, P. Ghisletta, J.-M. Le Goff, et al. *Towards an Interdisciplinary Perspective on the Life Course*. *Advances in Life Course Research*. London: Elsevier.
- Blossfeld, H.-P., S. Drobnič (eds.). 2001. *Careers and couples in contemporary societies: From male breadwinner to dual earner families*. Oxford: Oxford University Press.
- Brzinsky-Fay, C. 2006. *Lost in Transition: Labour Market Entry Sequences of School Leavers in Europe*. *Discussion Paper*. Berlin: WZB.
- Brzinsky-Fay, C., U. Kohler, M. Luniak. 2006. „Sequence analysis with Stata“. *Stata Journal* 6 (4): 435-460.
- Dijkstra, W., T. Taris. 1995. „Measuring the agreement between sequences“. *Sociological Methods & Research* 24 (2): 214-231.
- Elzinga, C., H., A. Liefbroer, C. 2007. „De-standardization of Family-Life Trajectories of Young Adults: A Cross-National Comparison Using Sequence Analysis.“ Pp. 225.
- Elzinga, C. H. 2003. „Sequence Similarity: A Nonaligning Technique“. *Sociological Methods Research* 32 (1): 3-29.
- Fussell, E. 2005. „Measuring the early adult life course: an application of the entropy index.“ in ed. R. Macmillian. *The Structure of the life course: Standardized? Individualized? Differentiated?* London: Elsevier.
- Gabadinho, A., G. Ritschard, M. Studer, et al. 2009. *Mining sequence data in R with the TraMineR package: A user’s guide for version 1.2*. Geneva: University of Geneva.
- Gauthier, J.-A., E. Widmer, P. Bucher, et al. 2007. „How Much Does it Cost? Optimization of Costs in Sequence Analysis of Social Science Data“. *Social Science Research Network*. Dostupné na http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1288012 [cit. 24. 8. 2009].
- Gauthier, J.-A., E. Widmer, P. Bucher, et al. 2008. „Multichannel Sequence Analysis Applied to Social Science Data“. *Social Science Research Network*. Dostupné na http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1327725 [cit. 24. 8. 2009].

- Halpin, B., T. W. Chan. 1998. „Class Careers as Sequences: An Optimal Matching Analysis of Work-Life Histories“. *European Sociological Review* 14 (2): 111-130.
- Hamplová, D. 2003. *Vstup do manželství a nesezdaného soužití v České republice po roce 1989 v souvislosti se vzděláním. Sociologické texty/Sociological Papers*. Praha: Sociologický ústav Akademie věd ČR.
- Hamplová, D. 2004. „Výzkum životní dráhy a event-history analýza“. *SDA Info* 6 (1): 8-10.
- Hendl, J. 2004. *Přehled statistických metod zpracování dat*. Praha: Portál.
- Lesnard, L. 2006. *Optimal Matching and the Social Sciences. Document de travail du Crest*. Paris: INSEE.
- Levy, R., H. Krüeger. 2001. „Linking Life Courses, Work, and the Family: Theorizing a not so visible Nexus between Women and Men“. *Canadian Journal of Sociology* 26 (2): 145-166.
- Macindoe, H., A. Abbott. 2004. „Sequence Analysis and Optimal Matching technique for Social Science Data.“ Pp. 387-406 in ed. M. Hardy, A. Bryman. *Handbook of Data Analysis*. London: Sage Publications.
- Notredame, C., P. Bucher, J.-A. Gauthier, et al. 2005. *T-Coffee/SALTT: User's Guide and Reference Manual*. Lausanne: Centre National de la recherche scientifique and PAVIE.
- Pakosta, P., P. Fučík. 2009. „Vybrané metody analýzy panelových dat“. *Data a výzkum - SDA Info* 3(1): 77-96.
- Piccarreta, R., F. C. Billari. 2007. „Clustering work and family trajectories by using a divisive algorithm“. *Journal of the Royal Statistical Society* 170 (4): 1061-1078.
- Robette, N. 2008. *The Diversity of pathways to adulthood in France: a holistic approach*. Paris: Institut national d'études démographique.
- Robette, N., N. Thibault. 2008. „L'analyse exploratoire de trajectoire professionnelle: analyse harmonique qualitative ou appariement optimal?“. *Population* 63 (4): 621-646.
- Rohwer, G., U. Pötter. 2005. „TDA User's Manual.“ Bochum: Ruhr-Universität Bochum, Fakultät für Sozialwissenschaften.
- Scherer, S. 2001. „Early Career Patterns: A Comparison of Great Britain and West Germany“. *European Sociological Review* 17 (2): 119-144.
- Stovel, K., M. Bolan. 2004. „Residential Trajectories: Using Optimal Alignment to Reveal the Structure of Residential Mobility“. *Sociological Methods & Research* 32 (4): 559-598.
- Widmer, E. D., G. Ritschard. 2009. „The de-standardization of the life course: Are men and women equal?“. *Advances in Life Course Research* 14 (1-2): 28-39.