

Vysoká škola ekonomická v Praze

Fakulta informatiky a statistiky



Rekonstrukce časových řad počtu zemřelých a stavu
populace v Českých zemích před rokem 1920

HABILITAČNÍ PRÁCE

Obor habilitačního řízení: Statistika

Autor: Ing. Petr Mazouch, PhD.

Praha, říjen 2020

Prohlášení

Prohlašuji, že jsem předkládanou habilitační práci na téma „Rekonstrukce časových řad počtu zemřelých a stavu populace v Českých zemích před rokem 1920“ vypracoval samostatně a s použitím literatury a pramenů uvedených v závěru práce.

V Praze 14. října 2020

Ing. Petr Mazouch, PhD.

Obsah

Předmluva a poděkování.....	5
Úvod	8
1 Stav poznání a cíl práce	12
1.1 Cíl práce	18
2 Dostupnost jednotlivých ukazatelů.....	19
2.1 Teoretické vymezení rekonstruovaných ukazatelů.....	19
3 Metodika úpravy původních zdrojových dat.....	25
3.1 Rozdělení událostí a populace z kategorie „věk neznámý“.....	25
3.2 Rozdělení III. hlavních souborů událostí do struktury DES a HES.....	26
3.3 Rozdělení počtu událostí z víceletých věkových skupin do jednoletých skupin.....	28
3.4 Odhad koncového stavu populace ze sčítání konaného během roku.....	29
3.5 Rekonstrukce struktury populace v intercenzálním období	31
4 Dostupné datové zdroje (prameny dat)	35
4.1 Počet událostí.....	35
4.2 Stav populace	35
4.3 Shrnutí datových zdrojů	36
5 Rekonstrukce jednotlivých ukazatelů.....	39
5.1 Rozdělení III. hlavních souborů událostí do struktury DES a HES v období 1870–1894.....	39
5.2 Rekonstrukce ukazatelů pro období 1911–1920.....	41
5.3 Rekonstrukce koncového stavu roku 1920 a odhad středního stavu 1920	42
5.4 Rozdělení zemřelých do DES a HES ve věkové skupině 70 a více let v roce 1913	44
5.5 Odhad počtu zemřelých nezaznamenaných v civilních matrikách v období 1. sv. války	44
5.6 Odhad rozdělení migrace a neregistrovaných událostí v období 1911–1920.....	49
5.7 Rekonstrukce koncových stavů v období 1911–1919	52
6 Doplněk – data po roce 1920	53
6.1 Počet událostí.....	53

6.2	Stav populace	53
6.3	Využití dat pro transversální a kohortní analýzu.....	54
6.4	Odhad úrovně úmrtnosti během 2. světové války	54
7	Struktura publikovaných dat	58
7.1	Zemřelí v členění na elementární soubory DES a HES od roku 1870	58
7.2	Koncové stavy obyvatel v období 1870–1937 a 1945 do současnosti	58
7.3	Střední stavy obyvatelstva od roku 1870	59
7.4	Pravděpodobnosti úmrtí pro kohorty narozené 1820–1922	59
8	Využití rekonstruovaných dat.....	61
8.1	Základní výsledky kohortní analýzy	61
8.2	Základní výsledky transversální analýzy	64
8.3	Diskuse výsledků.....	67
	Závěr	75
	Seznam použité literatury	76

Předmluva a poděkování

Do uceleného textu, jako je kniha či závěrečná práce se autor vždy snaží zhmotnit myšlenky, které se formovaly po určitou dobu a z hlediska jeho vědeckého života je to určitý způsob, jak může zrekapitulovat, co ho vedlo právě k tomu, že objem oněch myšlenek nazrál úrovně, že je možné jej uzavřít a pokusit se ho vydat ve formě uceleného díla. Takové dílo pak většinou znamená milník v dosavadní cestě autora, v tomto případě vědecko-výzkumným životem.

Práci, ke které se pokouším psát předmluvu, lze zařadit do několika oblastí, záleží na úhlu pohledu, kterým budeme věc posuzovat. Jasným cílem je vytvořit kompletní datovou sadu věkově-pohlavní struktury obyvatel Českých zemí v období 1870–1919 a počtu zemřelých dle věku a pohlaví za stejné období. Taková ucelená sada zde zatím neexistuje a je určitou výzvou zařadit Českou republiku do skupiny zemí, které považujeme za statisticky rozvinuté nebo dokonce nejrozvinutější.

Na první pohled můžeme říci, že jde o práci demografickou, protože se snaží rozšířit data demografických řad a využívá základní demografické datové zdroje (a zčásti i nástroje). Druhý pohled je datový, kdy vzniká datový zdroj a demografie je pouze oblastí, které se tento datový zdroj (asi nejvíce) dotýká. Podstatný je zde obecný požadavek na kvalitu datových zdrojů, jejich vlastnosti a další tradiční nároky na datové zdroje, které bychom měli uplatňovat při jakékoliv datové analýze. Třetí pohled je statistický, protože zde zakládáme další potenciální zdroj pro uplatnění statistických analýz, ale také při rekonstrukci dat bylo využito určitých statistických postupů a nástrojů.

Rád bych pro tento moment zanechal rozboru, který z pohledů je podstatnější, určitou genezi vzniku práce nabídne následující kapitola uvozující odbornou část práce, a raději bych popsal vlivy, které mě od samého začátku studia na vysoké škole formovaly až do této doby a stojí nemalou měrou za důvody, proč vznikla tato práce.

Předtím, než jsem začal studovat statistiku, vnímal jsem tento obor spíše jako zdroj dat a informací, které následně mohu zpracovávat (s limitací znalostí základních popisných statistik ze střední školy). Mezi základní data, která jsem měl možnost pozorovat a která byla dostupná a pochopitelná, patřila data demografická. Demografická data se dotýkají každého z nás, jsme jejich součástí, a i proto jsou často demografická data spojována s počátky samotné statistiky.

Měl jsem nesmírné štěstí, že jsem na začátku svého studia mohl navštěvovat kurzy demografie vedené prof. Roubíčkem, člověkem s ohromným nadhledem a přehledem, který dokázal demografii podávat poutavě, s citem pro detail a mnoho jeho poznámek a výkladů si stále pamatuji a doufám, že pamatovat budu. V té době působily na katedře další osobnosti demografie, které vedly navazující odborné kurzy. K nejvýraznějším patřil doc. Felix Koschin, který se také ve svém výzkumu věnoval různým (na první

pohled vzdáleným) oblastem. Vyučoval kurz aktuárské demografie, která se orientovala zejména na analýzu úmrtnosti a vnášel do obecných pravidel demografie silnější matematické vnímání souvislostí. Cvičení bylo zajištěno kolegou RNDr. Tomášem Fialou, který studentům vysvětlil praktické aplikace teoretických funkcí, jež představil doc. Koschin v přednáškách. Oběma vděčím za otevření obzorů vnímání procesů úmrtnosti v různých aspektech, ale také třeba za to, že mi poskytli ucelený datový soubor stavů populace a počtu zemřelých za dostupné roky (o dostupnosti dat viz následující kapitolu 2).

Na druhé straně se doc. Koschin věnoval i dalším aspektům demografie, jako byla třeba statistika vzdělávání a její dopad na úroveň lidského kapitálu jedince a společnosti. Na toto téma jsem také u doc. Koschina psal svoji závěrečnou (diplomovou) práci, a protože téma bylo natolik zajímavé a zatím nerozvinuté, dohodl jsem se s doc. Koschinem na pokračování ve studiu na doktorském stupni s tématem právě v oblasti lidského kapitálu. To mi umožňovalo vnímat určitý nesoulad mezi jasně definovanými oblastmi, které chceme měřit a na druhé straně konkrétními ukazateli, jež máme k dispozici, což je zřetelný přesah do obecného pojetí práce s daty, který je doménou Katedry ekonomické statistiky. Když rok před dokončením mého doktorského studia doc. Koschin zemřel, převzala jeho roli doc. Jitka Langhamrová a i díky její pomoci jsem byl schopen disertační práci dokončit a také obhájit.

Již v té době jsem působil částečně na Katedře ekonomické statistiky, kde jsem se věnoval právě společným tématům lidského kapitálu. Postupně jsem z Katedry demografie přesunul svoje působení sem a zaměřil se na výzkum obecných požadavků na datové zdroje, jejich kvalitu, relevanci a další vlastnosti, které by měla každá data použitá pro analýzu splňovat. Mezi největší vzory, jež formovaly můj přístup k datům, patřil a patří prof. Jakub Fischer, který byl také členem řešitelského týmu projektu Reprodukce lidského kapitálu, pod vedením doc. Koschina. S ním jsem strávil nespočet hodin diskusemi nad způsobem zjišťování, zajišťováním kvality či vhodnou analýzou dat.

Stále mě však lákala i demografie, proto jsem se rozhodl přihlásit se k magisterskému studiu demografie na Přírodovědecké fakultě UK. Zde jsem potkal další velké české osobnosti demografie, jako je prof. Zdeněk Pavlík, prof. Jitka Rychtaříková či RNDr. Klára Hulíková Tesárková. Právě prof. Rychtaříková mě v roce 2011 oslovila s nabídkou, zda bychom nechtěli být součástí připravovaného grantu na tvorbu kohortních úmrtnostních tabulek pro ČR. Spolu s RNDr. Klárou Hulíkovou Tesárkovou zaštiťovaly tým za PŘF UK a já s Ing. Pavlem Zimmermannem naopak tým za FIS VŠE. A projekt kohortních úmrtnostních tabulek dal základ pro vznik této práce. Již při jeho řešení byla neocenitelnou pomocí také vstřícnost doc. Ludmily Fialové, která nás podporovala poskytnutím historických datových zdrojů a zejména velmi kvalitními a rozsáhlými diskusemi, které jsme s RNDr. Klárou Hulíkovou

Tesárkovou absolvovali s nadšením a patřičnou zvědavostí a tuto spolupráci dále rozvíjíme i po ukončení grantu.

Během svého působení na obou katedrách jsem samozřejmě potkal celou řadu dalších skvělých osobností, které formovaly (a některé stále ještě formují) můj názor na demografii, data či statistiku. Na Katedře ekonomické statistiky to je prof. Stanislava Hronová či Ing. Kristýna Vltavská, z ostatních kateder FIS VŠE například prof. Richard Hindls, doc. Luboš Marek, Ing. Libor Svoboda a mnoho dalších.

Rád bych všem jmenovaným, z VŠE i UK, ale i ostatním, které jsem nejspíše díky své roztržitosti opomněl zmínit, touto cestou poděkoval. Byli a většina z nich stále je součástí mého vědeckého života a u některých z nich mám tu radost říci, že jsou součástí i mého osobního života. Formovali moje úsudky po dlouhou dobu, dávali mi nahlédnout do jejich specializací a názorů a kladli mi argumenty, které bez otevřené diskuse nikdy nemohou zaznít. Děkuji za ně i za tu diskusi.

Kromě kolegyň a kolegů, se kterými mohu, či jsem mohl spolupracovat, bych pochopitelně rád poděkoval své rodině a blízkým. Bez jejich pomoci a zázemí by nebylo možné dokázat vůbec nic. Děkuji svým rodičům, že mě podporovali při studiu a hlavně své manželce, že při mně stojí i v těch náročných chvílích a také svým dvěma dětem, že mi dávají alespoň trochu prostoru k seberealizaci v této oblasti.

Pevně věřím, že následující stránky dokáží, že podpora všech, které jsem zde zmínil, nebyla zbytečná a že se podařilo vytvořit něco, co může pomoci dalším v rozvoji jejich projektů a aktivit. Přeji všem, aby výsledek následujících stran užili v co největší míře a sloužil k obohacení současného stavu poznání.

Úvod

Demografická statistika na území dnešní České republiky vždy patřila mezi nejlepší na světě. Existují ucelené a dlouhé časové řady zejména ze sčítání lidu. Tuto akci, pokrývající veškeré obyvatelstvo a kompletní území, považujeme za náročnou na provedení a zpracování i v dnešní době a uvědomíme-li si, že cenzy byly realizované i v době bez jakéhokoli počítačového vybavení, můžeme je právem považovat za projev rozvinuté oficiální statistiky. Také zajištění evidence demografických událostí bylo již od 19. století organizováno s vysokou mírou preciznosti, a i zde můžeme říci, že současná statistika stále nese znaky vtisknuté jí před sto i více lety.

Soudobá demografická analýza aplikovaná na data ze sčítání i z evidence respektovala jejich strukturu a omezení. I v pozdější době, kdy již byla k dispozici výpočetní technika a byly rozvinuty i teoretické metody pro využití pokročilejších analytických metod, se nicméně analýzy na našem území pouštěly do rekonstrukce dat a na nich postavených hlubších rozborů pouze výjimečně a stále využívaly původní ukazatele. Proti tomu státy s podobnou úrovní demografické statistiky rekonstrukci dat v období druhé poloviny 19. století a začátku 20. století ve velké míře realizovaly a mají také zpracovány podrobné analýzy daného období odpovídající úrovni současného stavu poznání v demografické analýze, což české demografii stále chybí.

Postupný rozvoj demografické statistiky vytvářel tlak na starší datové zdroje, aby bylo možné využívat srovnatelných metod demografické analýzy a získat tím delší časové řady ukazatelů. Díky tomuto tlaku vznikly metodické postupy, jak historická data rekonstruovat, jak pracovat s obdobími s horší kvalitou či spolehlivostí dat a jak tímto způsobem vytvořit koherentní časové řady ukazatelů, které je možné bez větších potíží analyzovat.

Mé osobní zaujetí analýzou úmrtnosti, zejména v kohortním pohledu, a úzká spolupráce s UK vyústily v meziuniverzitní projekt „*Generační úmrtnostní tabulky České republiky: data, biometrické funkce a trendy*“ s cílem konstrukce prvních kompletních generačních úmrtnostních tabulek v ČR¹. Cíl projektu byl ambiciózní, protože pro konstrukci kohortní úmrtnostní tabulky jsou třeba kompletní data za celý život konkrétní kohorty, což je přibližně 100 let (je třeba vyčkat, až kohorta téměř vymře, abychom byli schopni popsat řád jejího vymírání). Pokud chceme zkoumat více ročníků kohort, pak pochopitelně datová náročnost roste.

Projekt si ke svému zkoumání vymezil kohorty od roku 1870 do roku 1920. Tyto kohorty však bylo třeba sledovat po celou dobu jejich života, tedy muselo být analyzováno celé období od roku 1870 do

¹ Více o projektu zde: <https://www.natur.cuni.cz/geografie/demografie-a-geodemografie/veda-a-vyzkum/vybrane-projekty/generacni-umrtnostni-tabulky-ceske-republiky-data-biometricke-funkce-a-trendy>

současnosti. V tomto období existuje v demografických datech několik nekompletně nebo dokonce vůbec nepopsaných období. Jde zejména o období dvou světových válek, ale také o období před první světovou válkou, které se sice datově dochovalo, ale nebylo možné jej jednoduše zpracovat pro potřeby analýzy.

Bylo zřejmé, že než dojde k samotné konstrukci tabulek, musely se připravit datové podklady do podoby, která umožní bezproblémovou aplikaci jednotlivých kroků konstrukce úmrtnostních tabulek a následně analýzu výsledků. Před tím bylo nutné zkoumat základní datové zdroje, jejich validitu a způsoby zjišťování dostupných údajů oficiální statistikou. Po vyčištění dat, formátování do použitelné podoby, což bylo v rámci malého řešitelského týmu zejména mým úkolem, bylo možné zkonstruovat úmrtnostní tabulky, jejichž funkce jdou následně analyzovat.

Již v průběhu realizace projektu bylo jasné, že být je rekonstrukce dat nutným krokem ke konstrukci úmrtnostních tabulek, stále je to jen mezikrok, který je sice nutné metodicky dobře popsat², ale uživatele nejvíce zajímají právě výsledky v podobě tabulkových funkcí, které následně mohou analyzovat a zpracovávat. Proto byl při publikaci výsledků brán zřetel zejména na finální úmrtnostní tabulky (viz webové stránky projektu nebo publikaci Mazouch, Hulíková Tesárková, 2018) a vedle toho byla jako součást příloh publikována kompletní původní zdrojová data (údaje o počtech událostí a stavu populace – za období po vzniku ČSR střední a koncové stavy za každý kalendářní rok, za období před vznikem ČSR pouze údaje ze sčítání lidu). Popis postupu úpravy zdrojových dat byl součástí publikace, ale úmrtnostní tabulky obsahovaly již jen vlastní tabulkové funkce, založené na upravených a zrekonstruovaných datech do podoby, která umožňuje sestavení kohortních úmrtnostních tabulek; samotné vyčištěné datové sady publikovány nebyly. Specifikům formátu dat pro jednotlivé přístupy (transverzální a longitudinální) se věnuji v kapitole 2.1.,

Výsledky kohortních úmrtnostních tabulek ukázaly, že práce se základními daty dává kvalitní a koherentní výsledky s výstupy, které jsou již publikovány (viz kapitolu 1 a 8 dále). Povedlo se tak konečně zařadit ČR mezi země, které mají vedle transversálních tabulek také kohortní analýzu (Mazouch a Hulíková Tesárková, 2018, kap. 2.1).

Na druhou stranu vyvstávaly otázky, zda by data nemohla sloužit i jiným analýzám, než pouze kohortním úmrtnostním tabulkám. V úvahu připadaly například tabulky transversální, které v ucelené podobě existují pouze od roku 1920 (ČSÚ, 2020), ale před rokem 1920 již ne. Vedle analýzy úmrtnosti by bylo možné data použít i na jiné analýzy, jako například plodnosti či sňatečnosti, protože všechny

² Metodiku a základní výsledky popsali v publikaci Mazouch a Hulíková Tesárková (2018).

tyto procesy se váží na počet a strukturu obyvatel, jež byla v období před rokem 1920 publikována pouze za roky, kdy proběhlo sčítání lidu (Srb, 2004, s. 18).

Absence komplexně rekonstruovaných dat před rokem 1920 je o to zajímavější, že techniky zjišťování statistických ukazatelů v tomto období nebyly na špatné úrovni. Spíše naopak, rakousko-uherská statistika patřila k těm nejlepším na světě (podrobněji viz kapitolu 1) a ani území Českých zemí nebylo nijak zásadně odlišné od současné České republiky (opět podrobněji viz dále). Teoretické metody a přístupy, jak s takovými nekompletními daty pracovat, jak je upravovat a rekonstruovat, jsou také dobře popsány³ a tak je spíše otázkou, proč ještě stále nebyl žádný ucelený datový zdroj období před rokem 1920 pro území Českých zemí zpracován a publikován.

Za zmínku stojí také zvláštní postavení datových zdrojů vhodných (nejen) pro analýzu úmrtnosti v ČR obecně, ne pouze před rokem 1920. Na oficiálních stránkách Českého statistického úřadu (www.czso.cz) jsou například data o struktuře a stavu populace dostupná v pramenných dílech⁴ (a to pouze od roku 1982, v předchozích letech se počet obyvatel v tomto pramenném díle nepublikoval – viz kap. 6.2), souhrnně za více let jsou data k dispozici pouze v Demografické příručce⁵, a to v podobě třídění dle pohlaví a 5letých věkových skupinách od roku 1920 (k datovým zdrojům viz kapitolu 4). Jednoleté věkové skupiny, které jsou pro podrobnou analýzu důležité, hromadně publikovány nejsou. Od roku 1950 jsou k nalezení (v několika možných způsobech třídění, i po jednotkách věku) v Human Mortality Database (www.mortality.org [cit. 15. 6. 2020]).

K zaplnění vzniklé mezery lze využít základů z rekonstrukce dat pro konstruované kohortní úmrtnostní tabulky, které však musí být dále upraveny a rozvinuty tak, aby byla vytvořena ucelená datová báze, zahrnující jak stavy populace, tak i počty zemřelých. Jde o prodloužení stávající časové řady čítající 100 let o 50 % a vzniká tak 150letá časová řada úplných údajů, která bude zpřístupněná na jednom místě⁶. Tímto je vymezen základní cíl celé práce (podrobněji je specifikován a rozpracován v návaznosti na řešerši zdrojů v následující kapitole 1.1) – dát k dispozici základní datový materiál, který může pomoci podrobněji a lépe prozkoumat nejen historické období, ale také jeho dopady do současnosti.

Mezi konkrétní možnosti návazných analýz patří (vedle již zmíněných zpracovaných kohortních úmrtnostních tabulek) konstrukce transversálních úmrtnostních tabulek, získání vhodnějších dat pro

³ Například na webu www.mortality.org je zveřejněna metodika rekonstrukce dat pro taková období (WILMOTH, J., R., ANDREEV, K., JDANOV D., GLEI, D., A., RIFFE, T. 2017. *Methods Protocol for the Human Mortality Database*. [online]. Verze metodiky z 27. 11. 2017. 2017. [cit. 30.6.2020]. Dostupné z www: <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>)

⁴ Demografická ročenka dostupná z https://www.czso.cz/csu/czso/casova_rada_demografie [cit. 30.6.2020]

⁵ Demografická příručka elektronicky dostupná zde <https://www.czso.cz/csu/czso/demograficka-prirucka> [cit. 30.6.2020]

⁶ Data budou uceleně publikována na webu mortality.vse.cz

vymezení například 1. demografického přechodu, dopadů epidemií, analýzy plodnosti apod. Dochází tak ke zmenšení adekvační mezery, která nyní existuje, protože různé teoreticky vymezené jevy jsou zachyceny s pomocí méně dokonalých ukazatelů. Publikací kompletních datových sad se možnosti adekvace výrazně rozšíří.

Následující kapitola vymezuje stav poznání v této oblasti. Stručně rekapituluje pokusy o různé typy analýz, které byly na datech ve stávající podobě realizovány a na závěr podrobněji představuje cíl práce. Druhá kapitola obsahuje teoretický výčet a vymezení základních ukazatelů, které je nutné rekonstruovat (střední a koncové stavy, počty událostí v různých skupinách třídění), aby bylo možné připravit datové podklady nutné pro všechny způsoby návazných analýz.

Třetí kapitola představuje obecný metodický postup při řešení jednotlivých úprav dat. Dostupné datové zdroje jsou následně představeny a popsány ve čtvrté kapitole, na kterou navazuje již konkrétní řešení jednotlivých období, což je předmětem páté kapitoly. Na původní data jsou aplikovány metody a postupy obecně popsané ve třetí kapitole, s vymezením případných problémů.

Šestá kapitola popisuje návaznost na data po roce 1920, protože je nutné dodržet požadavek na koherenci datových zdrojů. Také je v této kapitole diskutován problém překlenutí období 2. světové války v případě kohortních dat. Následující sedmá kapitola stručně popisuje strukturu publikovaných dat, jednotlivé soubory ukazatelů a případné úpravy, které byly ještě nutné s daty provést.

Poslední, osmá kapitola hodnotí kvalitu provedených rekonstrukcí dat na základě srovnání výsledků spočtených s použitím nových datových podkladů a předchozích analýz, které jsou provedeny na původních datových zdrojích. Dosavadní výsledky ukazují vysokou míru koherence s předchozími výsledky, což dává velmi dobrý základ pro další návazné analýzy.

Celá práce vychází z původních postupů při rekonstrukci dat určených pro projekt kohortních úmrtnostních tabulek, které jsou popsány v publikaci Mazoucha a Hulíkové Tesárkové (2018). Jednotlivé kroky rekonstrukce však dále rozvíjí, ještě více zpřesňuje a podrobněji metodologicky popisuje a rozšiřuje. Datová sada, která nebyla nikdy kompletně publikovaná, tak obsahuje nejen nová data vhodná k transverzální analýze, ale umožní kohortní analýzu i pro starší kohorty (ne však v celém věkovém spektru, protože data začínají rokem 1870). Využití práce spočívá také v možnosti aplikovat obecně popsané metodické postupy na další datové zdroje, například za menší územní celky nebo jiné státy, které nemají data v potřebném formátu k dispozici.

1 Stav poznání a cíl práce

Vladimír Srb, jeden z předních českých demografů, ve svém souhrnném díle „1000 let obyvatelstva Českých zemí“ (Srb, 2004, s. 19) uvádí, že začátek statistického zjišťování událostí, tedy sňatků, narozených a zemřelých, můžeme sledovat na území Českých zemí od roku 1785. Byť existovaly záznamy starší, díky ztrátám původních podkladů, se zachovaly jen ve zlomkové podobě a rok 1785 je zlomovým zejména díky císařskému patentu z roku 1784 o civilních matrikách, účinnému od 1. ledna 1785 (Srb, 2004, s. 19).

Během následujících více než 100 let se princip zaznamenávání událostí příliš neměnil, reformním rokem je až rok 1895, kdy se změnil dosavadní tiskopis hlášení a byly vydány nové instrukce k jejich vyplňování (Srb, 2004, s. 19). Rok 1895 již zasahuje do námi zpracovávaného období a dopady této reformy budou vidět na změně třídění údajů o zemřelých, které bylo od tohoto roku podrobnější – viz kapitolu 4.2 dále.

Další reforma byla datována až do období po vzniku samostatného Československa, do roku 1925, která opět znamenala změnu metody zápisu, a to směrem od hromadných výkazů k jednotlivým „sčítacím lístkům“ (Srb, 2004, s. 19). Tyto změny se podepsaly spíše na rozšíření možností kombinačního třídění, tedy došlo k rozšíření pramenných publikací o pohybu obyvatelstva. Do souhrnných ukazatelů, které budou využity v následujících kapitolách, nebyl dopad téměř žádný a není zde, na rozdíl od roku 1895, patrná žádná výrazná změna.

Pro úplnost proběhla také třetí reforma, která byla v roce 1950, ta se však týkala pouze přenesení zpravodajské povinnosti na národní výbory, které převzaly matriky z duchovních správ (Srb, 2004, s. 19). Tato změna neměla opět vůbec žádný efekt na analyzované časové řady.

O vysoké úrovni statistiky na území Českých zemí píše Srb s Kučerou v roce 1954 ve své stati „Vývoj obyvatelstva českých zemí v XIX. století“, kde označují českou statistiku za jednu z nejstarších na světě, zejména díky délce časových řad (Srb, Kučera, 1954, s. 109). Zamýšlí se také nad otázkou, proč v dané době stále není k dispozici podrobnější studie o demografickém vývoji v 19. století, když jsou k dispozici takové prameny, byť k nim autoři mají drobné výhrady v podobě úplnosti či kvality dat (Srb, Kučera, 1954, s. 109).

Chvála statistiky se vztahovala také na další nutný ukazatel, který je nezbytný pro analýzu jednotlivých demografických procesů, a to zdroj informace o stavu a struktuře obyvatelstva. Nejprve se jednalo o soupisy obyvatelstva, a to od roku 1754 (Srb, Kučera, 1954, s. 109 nebo Srb, 2004, s. 16) a následně o sčítání lidu, za které se jako první počítá sčítání v roce 1857. To sice bylo realizováno dle standardů novodobého populačního cenzu, ale za první moderní sčítání se označuje až sčítání v roce 1869 (Srb,

2004, s. 18). Následovala sčítání v roce 1880, 1890, 1900, 1910 a po vzniku Československa v roce 1921, a pak každých 10 let s výjimkou období 2. světové války (Srb, 2004, tamtéž).

Byť nelze než souhlasit a obdivovat tehdejší schopnost realizovat census (jehož principy se budou využívat i v nadcházejícím cenzu 2021, tedy o 160 let později), je jasné, že informace o stavu obyvatelstva jednou za deset let není pro analýzu demografických procesů dostatečná. Pro podrobné zkoumání je nutné konstruovat (nejlépe specifické) míry a pro ně je třeba znát nejenom počet obyvatel v jednotlivých letech, ale i jeho strukturu. Pro potřeby tzv. intercenzálních období se vychází z jednoduché bilanční rovnice (viz například Roubíček, 1997, s. 34), která odhaduje stav obyvatelstva na konci období tak, že k počátečnímu stavu přičte narozené, odečte zemřelé (což tvoří přirozený přírůstek/úbytek), přičte přistěhované a odečte vystěhované (přírůstek/úbytek stěhováním).

Evidence přirozeného pohybu byla popsána výše a bylo zmíněno, že (až na drobnosti) jeho evidence byla po celou dobu konzistentní a bez větších problémů. Naopak evidence přírůstku či úbytku stěhováním, kterou Srb označuje jako „*statistiku mechanického pohybu obyvatelstva*“ (Srb, 2004, s. 19), nebyla na takové úrovni, jako evidence narozených a zemřelých. Srb (2004, tamtéž) uvádí, že stěhování se na území českých zemí sledovalo již od 19. století, a to jak mezi zeměmi v rámci Rakouska-Uherska, tak i přeshraniční migrace rakouského mocnářství. „Údaje však byly zřejmě velmi nepřesné a nebudí důvěru“, píše Srb (2004, tamtéž) ke kvalitě dat.

Proto je nutné pro odhad migrace použít tzv. „*Diferenční metodu*“, která využívá informace o stavu populace na začátku období a na jeho konci, dále znalosti o přirozeném pohybu obyvatelstva (narození a zemřelí) a nezjištěný rozdíl je považován za výsledek migrace (Srb, 2004, s. 209). Tento problém neplatí jen pro období 19. století, ale lze ho pozorovat i v dalších obdobích (například Srb, 2004, s. 211) a koneckonců podobnými problémy se špatnou evidencí zejména vystěhovaného obyvatelstva trpí i současná statistika (Srb, 2004, s. 20 a 212).

Do migrace odhadnuté diferenční metodou pak vedle migrace vstupují i další vlivy, jako například nepřesnosti populačního cenzu nebo nejistoty kolem registrace zemřelých nezapsaných v civilních matrikách (Srb, 2004, s. 209). K problematice podhodnocení počtu událostí registrovaných matrikami se vyjadřuje ve svém článku také Korčák (1946), který analyzoval počty nezapsaných (či opožděně zapsaných) událostí do matrik v období 2. světové války.

Diferenční metodu však lze aplikovat pouze ex post, tedy až v okamžiku, kdy známe stav populace na konci období, můžeme odhadnout saldo migrace a pro jednotlivé časové úseky ve sledovaném období (většinou pracujeme s obdobím mezi sčítáním, tedy 10 kalendářních let) uvažujeme nejčastěji rovnoměrné rozdělení stěhování (Wilmoth et al., 2017, s. 17). To však nemusí platit vždy, protože dle

platné metodiky ČSÚ,⁷ je případný stav zjištěný ve sčítání promítnut do korekce koncových stavů, ale ne nutně rovnoměrně.⁸

Vrátíme-li se k analýze období tzv. dlouhého 19. století (tj. s přesahem do století dvacátého, až po začátek 1. světové války), existuje několik prací, jež se jím uceleněji zabývají. Předně je to již výše uvedená a citovaná práce Srba a Kučery (1954) a jimi zmiňované práce Dvořáčka a Boháče (Srb, Kučera, 1954, s. 109), následované o několik let později publikacemi Kárníkové (1965), či Stříteského (1976).

Žádná z prací však nevyužívá potenciál k ucelenější rekonstrukci datové základny, která by umožnila hlubší a komplexnější analýzu demografických procesů a hlavně by umožnila tuto analýzu dále rozvíjet dalším autorům. Autoři využívají v drtivé většině publikovaných údajů, bez zásadnějších zásahů a základní ukazatele jsou vyjadřovány nejčastěji v absolutních hodnotách, výjimečně v hrubých mírách, jak to lze vidět u Srba a Kučery (1954, například s. 129–130) nebo u Doukoupila a Nesládkové (1987). Otázku použití hrubých měr, které jsou citlivé na věkovou strukturu (Roubíček, 1997, s. 38), rozebírá ve svém článku Růžička (1959b), který upozorňuje na riziko zejména mezinárodního srovnání. Naopak Dokoupil a Nesládková (1987) považují věkovou strukturu v Českých zemích v období konce 19. století a začátku 20. století za natolik stabilní a neměnnou, že v použití hrubých měr nevidí problém.

Zaměříme-li se více na práce, které se bez věkově-pohlavní struktury obyvatelstva neobejdou, jedná se zejména o díla obsahující analýzu funkcí z úmrtnostních tabulek. Srb a Kučera (1954, s. 150–151) publikují ve svém díle úmrtnostní tabulky za období kolem jednotlivých sčítání lidu. Využívají informace o stavu a struktuře obyvatelstva v okamžiku sčítání, která v letech 1869, 1890, 1900 a 1910 proběhla vždy s rozhodným okamžikem 31. prosince příslušného roku (Srb, 2004, s. 18). Tyto stavy pak lze považovat za střední stavy pro příslušné dvouletí (čímž se také rozloží možné náhodné výkyvy jednoletých údajů) a Srb s Kučerou tak zkonstruovali úmrtnostní tabulky za dvouletí vždy obklopující zmíněná sčítání lidu. Vznikly úmrtnostní tabulky za období 1869/1870, 1880/1881, 1890/1891, 1900/1901 a 1910/1911 (Srb, Kučera, 1954, s. 150–151), které však nemají silnější metodologický popis, s výjimkou zdůvodnění dvouletých období a ujasnění, že se jedná o tabulky zkrácené (Srb, Kučera, 1954, s. 129).

Podobný přístup, jako Srb a Kučera (1954) měl o téměř 20 let dříve i Státní úřad statistický, který vydal v rámci své „*Mimořádné zprávy*“ (SÚS, 1935) mimo jiné úmrtnostní tabulky za období 1899–1902 a 1909–1912 (které srovnával s úmrtnostními tabulkami 1929–1932, ale nekonstruoval tabulky za období kolem sčítání 1890 a 1869). Tyto úmrtnostní tabulky byly na rozdíl od tabulek Srba a Kučery

⁷ Metodika odhadu počtu obyvatel dostupná zde https://www.czso.cz/csu/czso/pocet_obyvatel_m

⁸ Viz vývoj počtu obyvatel ČR, kdy v letech sčítání je patrný vždy pokles počtu obyvatel, tedy korekce koncového stavu dle sčítání. Graf zde <https://www.czso.cz/documents/10180/121739322/13007020g01.xlsx/5ac32c4e-194e-419e-89e1-19ffb03e004f?version=1.1>

(1954) úplné a v rámci popisu konstrukce jsou velmi dobře popsány jednotlivé kroky konstrukce tabulky a také vyrovnání dat (SÚS, 1935).

Za předchozí období existovaly úmrtnostní tabulky, které Srb a Kučera dokonce označují za nejstarší úmrtnostní tabulky pro České země, tabulky z období 1869–1880⁹ (Srb, Kučera, 1954, s. 128), jež byly zkonstruovány V. Kitzem a vydány v pramenném díle ke sčítání 1880¹⁰ a v roce 1941 je vyrovnal Josef Talacko (Talacko, 1941). Problém Kitzových tabulek je také spojen s absencí popisu metodiky konstrukce (Matesová, 1983).

Důvod vzniku tabulek Srba a Kučery byl (dle jejich tvrzení ze s. 128–129) v metodické nesrovnalosti předchozích výstupů, a proto přistoupili k jednoduchému zkonstruování tabulek vlastních. Přesto se našli i po publikování výsledků Srba a Kučery autoři, kteří pracovali dále s původními tabulkami, jako například Musil (1959), který se pokoušel odhadnout průměrné koeficienty růstu tabulkových funkcí (pravděpodobností úmrtí a střední délky života) v intercenzálním období a srovnával je s následujícími obdobími. Podobně k problému přistupoval také Vávra (1960), který se soustředil opět na pravděpodobnosti úmrtí avšak s důrazem na kojeneckou úmrtnost. Později se tématu věnovala také Matesová (1983), která se pokoušela odhadnout trend mezi jednotlivými sčítáními pomocí regresní analýzy.

Jde vidět, že všechny analýzy se pokoušely o využití stávajících datových zdrojů. Buď přímo konstruovaly úmrtnostní tabulky v obdobích sčítání lidu (SÚS, 1935 nebo Srb a Kučera, 1954) nebo využívaly již zpracovaná data a dále je analyzovaly. Nikdo se však nepokusil o odhad v období mezi sčítáními (pokud za odhad nepočítáme interpolační odhady Musila (1959), Vávry (1960) nebo Matesové (1983)), pro něž by bylo třeba zrekonstruovat věkově-pohlavní strukturu populace.

Důležité je také zmínit, že existovaly samozřejmě i další demografické analýzy období před rokem 1920, které ho hodnotily z různých hledisek. Byly to například publikace zmiňované Kárníkové (1965), či Strítěského (1976), ale zřejmě i mnohé další. Ty však využívají k analýze úmrtnosti pouze hrubých měr a nepokouší se o využití pokročilejších metod demografické analýzy zejména kvůli absenci informace o struktuře obyvatelstva v intercenzálním období.

Bariérou pro rekonstrukci dat mohla být také 1. světová válka, která trvala více než 4 roky a po jejím skončení probíhaly na území Českých zemí výrazné politické změny. Státní úřad statistický sice vznikl již na začátku roku 1919 (ČSÚ, 2015, s. 24), ale počátky jeho existence nebyly nijak jednoduché, a tak toto

⁹ Existují ještě starší úmrtnostní tabulky, avšak ne za samostatné České země, ale za území celého Předlitavska, které jsou za období 1819–1828, avšak jejich konstrukce je výrazně odlišná od současných přístupů.

¹⁰ Oesterreichische Statistik, Svazek V., Sešit 3, Vídeň 1884.

období mohlo být jakousi bariérou pro napojení dat z Rakouska-Uherska na data nově vznikajícího Československa.

Přesto pro datově velmi náročné období 1. světové války vznikla na Českém statistickém úřadě publikace, která demograficky popisuje stav Českých zemí v období 1. světové války, která se jmenuje „*Přirozená měna obyvatelstva v zemích Koruny české v letech 1. světové války*“ (ČSÚ, 2005). Data byla zpracována až v období po 2. světové válce (nejspíše kvůli kapacitním a finančním důvodům), což způsobilo, že velká část materiálů, jež byly předány tehdejšímu Československu Ústřední statistickou komisí ve Vídni, byla ztracena či zničena a publikace je založena na rekonstruovaných torzech původních datových zdrojů (ČSÚ, 2005, s. 1). K publikování dat pak došlo až v roce 2005. Druhým nedostatkem je absence dat z vojenských matrik, které zachycují válečné ztráty (ČSÚ, 2015, s. 1). Data z civilních matrik tak podhodnocují úroveň úmrtnosti mužů a je nutné počty zemřelých odhadnout i se zahrnutím válečných ztrát (odhadům válečných ztrát se věnuje kapitola 5.5 dále).

Z důvodu absence rekonstruovaného datového zdroje o stavu a pohybu obyvatel v období před rokem 1920 existují určitá omezení, která se v demografické analýze Českých zemí dlouhodobě projevují. Do okamžiku publikování výsledků projektu Generačních úmrtnostních tabulek neexistoval za naše území ucelený materiál hodnotící kohortní úmrtnost (výjimkou jsou pokusy o odhady kohortní úmrtnosti pro vybrané kohorty učiněné Růžičkou, 1959a, Pavlíkem et al., 1984 a Rychaříkovou et al., 1994 – více viz Mazouch, Hulíková Tesárková, 2018). Výstupy projektu, zejména publikace Mazoucha a Hulíkové Tesárkové (2018), popisují kvalitu a kompletnost datových zdrojů z období konce 19. a začátku 20. století. Ukazují také základní postupy, které byly při rekonstrukci dat pro kohortní analýzu využity a které jsou v této práci dále rozpracovány a rozšířeny o postupy nutné pro další, nejen kohortní, využití dat. Samotné datové sady, jak již bylo řečeno v úvodu práce, publikovány nejsou, výstupy jsou ve formě úplných kohortních úmrtnostních tabulek, což bylo základním cílem projektu.

V mezinárodním srovnání, kdy většina rozvinutých zemí datové sady (a na ně navázané analýzy) před první světovou válkou rekonstruované má, je Česká republika znevýhodněna, protože ji nelze jednoduše mezinárodně srovnat v základních ukazatelích úmrtnosti, ale nejen v nich. Jde například o lepší vymezení demografického přechodu, či lepší odhad dopadu epidemií, které se v té době ještě v Českých zemích objevovaly.

Země, které mají k dispozici datové soubory za delší časové období, než má ČR, jsou (dle Human Mortality Database¹¹) Belgie, Dánsko, Finsko, Francie, Island, Itálie, Nizozemsko, Norsko, Švédsko a Švýcarsko. Existují i další země, které mají data k dispozici, ale nejsou zveřejněny na zmiňovaném webu,

¹¹ www.mortality.org [cit. 1.7.2020]

jejich výčet lze najít například v Mazouch, Hulíková Tesárková (2018, s. 13–14). Je zřejmé, že ve všech těchto zemích došlo v historii k určitým geopolitickým změnám, často výraznějším než na našem území. A i přesto tyto země disponují dlouhými rekonstruovanými demografickými řadami. Proto je cílem zařadit se do skupiny těchto demograficky rozvinutých zemí a nabídnout další analýzy období před rokem 1920, které stále ještě nebyly, díky absenci datových zdrojů, za České země zpracovány.

1.1 Cíl práce

Z předchozích analýz období před rokem 1920 je vidět, že zde existuje seriózní základ pro rekonstrukci demografických datových řad. Jako základní milník (podobně, jako Srb a Kučera, 1954) lze vzít sčítání v roce 1869 a od něj na základě běžných postupů (publikovaných například ve Wilmoth et al., 2017) začít rekonstruovat časové řady jednotlivých ukazatelů tak, aby byly použitelné pro další demografickou analýzu.

Základním cílem práce je rozšířit datovou základnu o 50 let (1870–1919) v těchto ukazatelích:

1. koncové stavy obyvatelstva;
2. střední stavy obyvatelstva;
3. zemřelí v členění na elementární soubory.

Součástí cíle je také popsat jednotlivé kroky, které byly při rekonstrukci použity a popsat původní datové zdroje.

Vzniklá datová řada bude napojena na stávající řady a vznikne tak celkový 150letý kontinuální datový podklad pro návazné analýzy.

Byť je cílem prodloužení datové řady před rok 1920, jako doplněk bude publikována také řada pravděpodobností úmrtí dle věku a pohlaví pro jednotlivé kohorty (od roku 1820 do současnosti), z důvodu datové nekonzistence v období kolem 2. světové války, což v současnosti znemožňuje kohortní analýzu.

Veškerá data budou publikována na webu mortality.vse.cz a budou volně přístupná všem uživatelům.

2 Dostupnost jednotlivých ukazatelů

Soubor ukazatelů, které se následně budou rekonstruovat, byl již vymezen v předchozí části s cílem celé práce. Bude však vhodné naznačit i vztahy mezi jednotlivými ukazateli, aby bylo zřejmé, že jsou zde možnosti, jak je kombinovat a získat tak podkladová data pro různé návazné analýzy. Součástí této části je i vymezení teoretických přístupů k rekonstrukci dat, pokud nejsou příslušná data (například v konkrétním členění) k dispozici, na což navazuje volně kapitola 3, která metodiky popisuje konkrétní kroky při rekonstrukci dat.

2.1 Teoretické vymezení rekonstruovaných ukazatelů

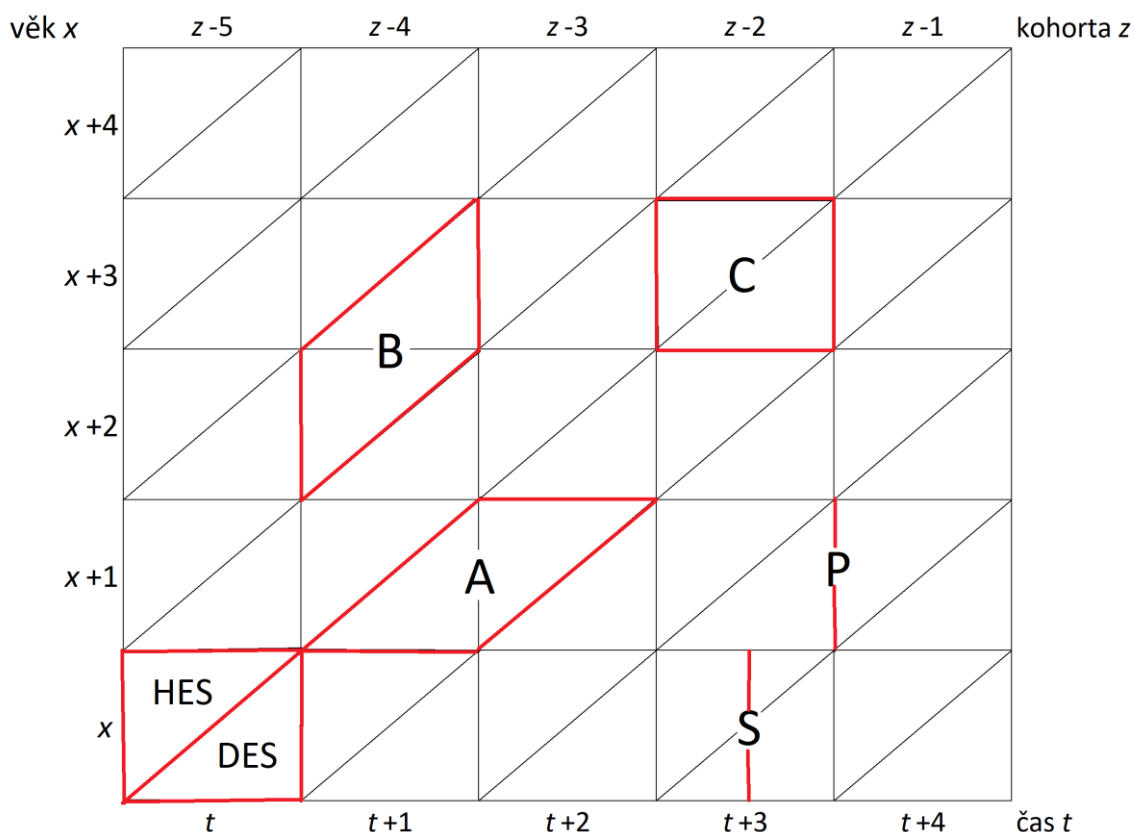
Pro jednodušší ilustraci ukazatelů a vazeb mezi nimi můžeme využít demografické sítě (označované též jako Lexisův diagram). Do sítě lze zaznamenávat demografické události dle okamžiku jejich nastání a sledovat vztah mezi klasickým časem (osa x) a věkem (nebo také chronologickým věkem) na ose y . Podrobněji o demografické síti píše například Pavlík et al. (1986), Koschin (2005), Kalibová et al. (2009) nebo kniha Hulíková Tesárková, Kurtinová (2018), které ve své knize popisují Lexisův diagram, včetně jeho různých modifikací.

Síť lze rozdělit do různých ploch, které jsou vymezeny některými z časových intervalů (kromě času t a věku x je možné využít i vymezení kohortou z , která je kombinací času a věku) a v takto vymezených plochách sledovat počty příslušných událostí (například úmrtí, sňatek, apod.), na Obr. 1 reprezentovány například plochami A, B či C (Pavlík et al., 1986 nebo Kalibová et al., 2009).

Dále je možné některé body reprezentující události propojovat. Například propojením bodu značícího narození osoby (takový bod se v klasickém Lexisově diagramu vyskytuje vždy na ose x , protože narození osoby je v přesném věku 0 příslušné osoby) a úmrtí osoby vytvoří úsečku pod úhlem 45 stupňů (někdy nazývanou čára života), a pak lze zkoumat nejen plochy, ale také průsečíky těchto úseček a čar vymežujících demografickou síť (horizontální i vertikální), což reprezentuje například úsečka P a S v Obr. 1 (Pavlík et al., 1986 nebo Kalibová et al., 2009).

Pokud se vrátíme k plochám A, B a C, tak mají jedno společné, všechny jsou tvořeny ze dvou trojúhelníků, které k sobě přiléhají některou ze stran. Ať již tedy chceme konstruovat jakkoliv vymezenou skupinu událostí, je vhodné mít k dispozici události rozdělené do jednotlivých (elementárních) trojúhelníků. Ty jsou na Obr. 1 označeny jako HES (horní elementární soubor událostí) a DES (dolní elementární soubor událostí). Jednotlivé plochy pak získáme jednoduše sečtením příslušných HES a DES (plocha A, B i C je vždy tvořena jedním HES a jedním DES).

Obr. 1: Lexisův diagram s vyznačením hlavních skupin událostí a stavů populace



Zdroj: Vlastní úprava dle Kalibová et al. (2009)

Dolní elementární soubor událostí na Obr. 1 je možné vyjádřit jako počet událostí U v roce t , v dokončeném věku x a generace z :

$${}_tU_x^z. \quad (1)$$

Horní elementární soubor událostí na Obr. 1 je pak počet událostí U v roce t , v dokončeném věku x a generace $z - 1$:

$${}_tU_x^{z-1}. \quad (2)$$

Při analýzách demografických událostí je nejčastěji využíván soubor událostí, který je na Obr. 1 označen jako C, tvořící čtverec a vymezující události jednou věkovou skupinou, jedním kalendářním rokem a dvěma kohortami, nazývaný také často III. hlavní soubor událostí (Pavlík et al., 1986 nebo Kalibová et al., 2009). Na Obr. 1 je to pak konkrétně součet elementárních souborů:

$${}_{t+3}U_{x+3}^{z-1} + {}_{t+3}U_{x+3}^z. \quad (3)$$

Tento přístup respektuje běžnou praxi při evidenci (nejen) demografických událostí, které jsou standardním *intervalovým* ukazatelem (Fischer et al., 2019, s. 11), kde je používaným intervalem (nejčastěji) kalendářní rok, ve kterém sledujeme počty událostí, které třídíme dle věku. Informace o

roku narození (kohortě) se pak může jevit jako podružná, ale pokud bychom ji nesledovali, nelze využít jiných třídění, než je III. hlavní soubor událostí (čtverec C na Obr. 1). To je problém například u sledování počtů zemřelých do roku 1895, kde se rok narození nesledoval (viz kap. 4.1 dále).

Pro III. hlavní soubory událostí také platí, že jsou základem transversální analýzy, kdy sledujeme počty událostí v daném kalendářním roce v jednotlivých věkových skupinách, bez rozdílu roku narození (Pavlík et al., 1986 nebo Kalibová et al., 2009). Jednotlivé čtverce tak na sebe jednoduše „navazují“ a tvoří sloupec, vymezený jedním kalendářním rokem. Data vhodná pro transversální analýzu nalezneme například v každoročně vydávaném pramenném díle Demografická ročenka¹².

Naopak pro kohortní analýzu III. hlavní soubory událostí vhodné nejsou (Mazouch, Hulíková Tesárková, 2018, s. 19). U jednotlivých kohort můžeme sledovat jejich vývoj v čase, což na Obr. 1 reprezentují diagonály (vedoucí z levého dolního rohu do pravého horního rohu pod úhlem 45 stupňů), a pokud bychom uspořádali čtverce v tomto směru, nenavazovaly by na sebe a navíc pokrývají vždy 2 kohorty (Mazouch, Hulíková Tesárková, 2018, s. 19).

Nabízí se proto uspořádání do jiných souborů, které jsou pro kohortní analýzu vhodnější. Soubor vyznačený plochou A v Obr. 1 se nazývá I. hlavní soubor událostí (Pavlík et al., 1986 nebo Kalibová et al., 2009) a na příkladu Obr. 1 je tvořen součtem elementárních souborů:

$${}_{t+1}U_{x+1}^z + {}_{t+2}U_{x+1}^z. \quad (4)$$

Při kohortní analýze je pak možné I. hlavní soubory jednoduše „navázat“ jeden na druhý a sledovat vývoj počtu událostí v dané kohortě dle jednotek věku. Drobnou nevýhodou je však překryv každého takového souboru přes dva kalendářní roky.

Poslední soubor událostí, který je na Obr. 1 označen písmenem B, je tzv. II. hlavní soubor událostí (Pavlík et al., 1986 nebo Kalibová et al., 2009), Na Obr. 1 je tvořen součtem elementárních souborů:

$${}_{t+1}U_{x+2}^{z-2} + {}_{t+1}U_{x+3}^{z-2}. \quad (5)$$

Byť proti I. hlavním souborům, které lze využít pouze ke kohortní analýze a III. hlavním souborům, které lze použít naopak pouze k analýze transversální, nabízí soubor II. hlavních souborů možnost jak kohortní, tak transversální analýzy (v obou případech na sebe vymezené oblasti hladce navazují – tvoří buď diagonálu nebo sloupec), je tento přístup v analýze velmi ojedinělý (Mazouch, Hulíková Tesárková, 2018, s. 19), zejména z důvodu překryvu vždy dvou věkových kategorií (v příkladu plochy B na Obr. 1 je to dokončený věk $x + 2$ a $x + 3$).

¹² Dostupné například zde https://www.czso.cz/csu/czso/casova_rada_demografie [cit. 15. 6. 2020]

Vzhledem k cíli práce, jehož základ tvoří časová řada událostí pro transversální i kohortní analýzu, je vhodné data ponechat v nejpodrobnějším třídění na základní elementární soubory, které se následně dají jednoduchou operací (viz vzorce (3), (4) a (5)) vyjádřit ve formě hlavních souborů událostí. Z hlediska publikační praxe oficiální statistiky je zvykem každoročně publikovat data o událostech ve dvou tříděních – III. hlavní soubory událostí a ve větším detailu rozdělené na elementární soubory – DES a HES (Koschin, 2000, s. 21). Pro I. a II. hlavní soubor událostí musí autor využít právě třídění na DES a HES (viz vzorce (4) a (5)).

Dosavadní pojetí ukazatelů bylo založeno na jejich absolutních hodnotách (počtech událostí). To však není vhodné nejen pro mezinárodní srovnání, ale ani pro srovnání v delším časovém rozmezí, jak píše například Roubíček (1997, s. 35) nebo Koschin (2000, s. 8). Počet událostí je nutné vztáhnout k velikosti populace, která byla „riziku“ vzniku události vystavena. Velikost populace je však ukazatel *okamžikový* (Fischer et al., 2019, s. 11), protože počet obyvatel je nutné zjišťovat k jednomu okamžiku. Otázkou je pak vhodné zvolení onoho okamžiku, zda volit spíše začátek intervalu, jeho střed nebo konec intervalu, protože velikost populace se během sledovaného období může měnit, a to má pak přímý vliv na úroveň výsledného ukazatele (Koschin, 2000, s. 8).

Velmi častým způsobem (ale ne jediným, jak se ukáže dále) je vyjádření intenzity daného procesu, například úmrtnosti, plodnosti, migrace, tzv. mírou. Jedná se o poměr počtu událostí a doby expozice, což je čas, po který byli členové populace vystaveni riziku nastání události (Koschin, 2000, s. 8). To vede k potřebě mít informaci o velikosti populace v konkrétní okamžiky, jež je možné znázornit v Obr. 1 svislicemi (P či S), které číselně vyjadřují počet průsečíků čar života s danou svislicí. Svislice P pak značí počet žijících v dokončeném věku ($x + 1$) na konci roku $t + 3$ (nebo na začátku roku $t + 4$). Druhá svislice vyznačená v Obr. 1 je úsečka S, která značí počet osob v dokončeném věku x let přesně v polovině roku $t + 3$ (k 30. 7. nebo k 1. 7.).

Teoreticky by bylo možné uvažovat takovou svislici v jakémkoliv bodě, protože jak píše Koschin (2000, s. 25), můžeme definovat intenzitu události pro libovolně malé intervaly (délku intervalu Koschin značí h), za něž sledujeme počty událostí, ale oficiální statistika publikuje data v dostatečném detailu pouze v ročních intervalech, což znamená, že *„všechny teoretické úvahy jsou k ničemu, protože jediné h , které můžeme reálně použít, je jeden rok“* (Koschin 2000, s. 25).

Jelikož je doba expozice sice formálně vyjádřitelná *„jako součet individuálních dob, po které jednotliví členové v populaci setrvali“* (Koschin, 2000, s. 8), v praxi se používá zjednodušujícího předpokladu, že události mající vliv na velikost populace, tedy úmrtí, narození a migrace, mají v čase rovnoměrné rozdělení, což není při úvaze o délce intervalu limitně se blížící nule nijak omezující. Doba expozice určíme jako součin stavu uprostřed intervalu (tzv. střední stav) a délky intervalu (Koschin, 2000, s. 9).

Otázkou pak (dle Koschina (2000, s. 25)) zůstává, zda jednoletý interval svojí délkou již nezkruskuje výsledky.

Pro různá vymezení souborů událostí pak existují rozdílné možnosti odhadu doby expozice (středního stavu). Pro I. hlavní soubory událostí, znázorněné na Obr. 1 plochou A, je vymezení středního stavu asi nejjednodušší, protože lze za střední stav považovat přímo stavy koncové, označené v Obr. 1 jako P . U II. hlavních souborů je situace složitější a střední stav je nutné odhadnout, protože svislice rozděluje plochu B na dvě poloviny se nejen nenachází v okamžiku konce (či začátku) kalendářního roku (což takové omezení není, jak uvidíme dále), ale také zahrnuje dvě věkové kategorie a každou jen částečně (v Obr. 1 to jsou věkové kategorie $x+2$ a $x+3$), proto by jednou z možností bylo vypočítat průměr z dvou koncových stavů:

$$\frac{{}_tP_{x+2}^{z-2} + {}_{t+1}P_{x+3}^{z-2}}{2}. \quad (6)$$

Pro III. hlavní soubory je situace podobná. Přestože je již v Obr. 1 střední stav vyznačen svislicí S, tak jeho odhad je většinou založen na průměru dvou sousedních koncových stavů (pro plochu C v Obr. 1):

$${}_{t+3}S_{x+3}^{z,z-1} = \frac{{}_{t+2}P_{x+3}^{z-1} + {}_{t+3}P_{x+3}^z}{2}. \quad (7)$$

Alternativní možnosti odhadu středního stavu, které uvažují například exponenciální či logaritmický vývoj lze nalézt v knize Roubíčka (1997, s. 36), avšak v posledních letech jsou stavy obyvatel ke středu roku publikovány na základě bilance i v rámci výstupů oficiální statistiky¹³.

Pokud se zaměříme na analýzu úmrtnosti, pak může být intenzita úmrtnosti vyjádřena mírou úmrtnosti základem analýzy sama o sobě. Pokud však uvažujeme o konstrukci úmrtnostních tabulek, je nutné míry úmrtnosti (m_x) převést na pravděpodobnosti úmrtí (q_x) například dle Roubíčka (1997, s. 89):

$$q_x = \frac{2m_x}{2+m_x}, \quad (8)$$

pro $x > 0$. Tento vztah předpokládá opět rovnoměrné rozdělení událostí (Roubíček, 1997, s. 90).

Pravděpodobnost úmrtí (či obecně události) však může být spočtena také přímo a v některých případech je to dokonce postup snadnější než výpočet přes míry. Vzorec (6) ukazoval postup odhadu středního stavu v případě II. hlavních souborů událostí. Pro tyto soubory je právě pravděpodobnost události odhadnutelná přímo z koncových stavů tak, jak jsou již publikovány (pro plochu B v Obr. 1 by platilo):

$${}_{t+1}q_{x+2,x+3}^{z-2} = \frac{{}_{t+1}D_{x+2,x+3}^{z-2}}{{}_tP_{x+2}^{z-2}}, \quad (9)$$

¹³ <https://www.czso.cz/csu/czso/vekove-slozeni-obyvatelstva-2019> (cit. 27. 7. 2020)

kde ${}_{t+1}D_{x+2,x+3}^{z-2}$ je počet zemřelých v roce $t + 1$ z kohorty $z - 2$ a v dokončeném věku $x + 2$ a $x + 3$. Ze vzorce (9) je patrné, že proti míře došlo u pravděpodobnosti pouze k „výměně“ jmenovatele, kde se místo středního stavu použije počáteční (koncový) stav populace.

Pro I. hlavní soubory je situace mírně složitější. Byť by zde bylo možné využít koncového stavu populace přímo ke zjištění míry (svislice rozděluje soubor A na obr. 1 na dvě stejné části a lze ji považovat za odhad středního stavu), je vhodnější použít přístup popsany Wilmothem et al. (2017, s. 40), který odhaduje pravděpodobnost úmrtí (pro soubor vymezený jako plocha A v Obr. 1) jako:

$$q_{x+1}^z = \frac{{}_{t+1}D_{x+1}^z + {}_{t+2}D_{x+1}^z}{{}_{t+1}P_{x+1}^z + {}_{t+1}D_{x+1}^z}. \quad (10)$$

Prostým přičtením počtu událostí v dolním elementárním souboru ke koncovému stavu populace získáme ukazatel, který není běžně publikovaný v oficiální statistice a je nazýván *počet žijících osob v přesném věku x v roce t* (Mazouch, Hulíková Tesárková, 2018, s. 20), což lze přeložit jako „počet osob, které v daném kalendářním roce oslavily x. narozeniny“. V úmrtnostních tabulkách je tento ukazatel označován jako $l(x)$ či l_x a nazývá se *počet dožívajících se věku x* (Koschin, 2000, s. 21).

Na Obr. 1 počty dožívajících se reprezentují vodorovné čáry, které zde však nejsou nijak zvýrazněny či označeny. Pro úplnost je vhodné dodat, že prostřednictvím vodorovných čar bychom také mohli odhadovat střední stavy populace, ale zde se již jedná o alternativu k výše popsaným postupům, a to není cílem této práce (více viz například Koschin, 2000, s. 42). Poslední III. hlavní soubor událostí je v případě odhadu hodnot pravděpodobností odkázán výhradně na přepočet dle vzorce (8) (Mazouch, Hulíková Tesárková, 2018, s. 20).

3 Metodika úpravy původních zdrojových dat

Přestože lze datové zdroje o událostech i struktuře populace považovat za konzistentní, bylo nutné při jejich přípravě pro následnou analýzu realizovat několik základních úprav. V následující části je představena metodika jednotlivých kroků, které byly na data aplikovány tak, aby byla data v celém období ve stejné struktuře – události rozděleny na DES a HES a populace publikovaná ve formě koncových i středních stavů za každý kalendářní rok po jednotkách věku.

3.1 Rozdělení událostí a populace z kategorie „věk neznámý“

V případě, že se v datech pro konkrétní rok objevila nenulová hodnota v kategorii „věk neznámý“, bylo pro korekci počtů událostí a počtu obyvatel v jednotlivých věkových skupinách nutné neznámé počty rozdělit, aby celkový počet událostí (nebo počet obyvatel) odpovídal součtu jednotlivých hodnot ve věkových skupinách.

Pro elementární soubory zemřelých pak taková korekce (dle metodiky Wilmoth et al., 2017, s. 11) vypadá následovně:

$${}_t^*D_x^z = {}_tD_x^z + {}_tD_{nezn} \cdot \frac{{}_tD_x^z}{\sum_x \sum_z ({}_tD_x^z)} = {}_tD_x^z \cdot \frac{\sum_x \sum_z ({}_tD_x^z)}{\sum_x \sum_z ({}_tD_x^z) + {}_tD_{nezn}}, \quad (11)$$

kde ${}_t^*D_x^z$ je elementární soubor zemřelých v roce t dokončeném věku x z kohorty z navýšený o podíl této skupiny na počtu zemřelých v neznámém věku. ${}_tD_x^z$ je elementární soubor zemřelých v roce t dokončeném věku x z kohorty z , a dále ${}_tD_{nezn}$ je počet zemřelých v neznámém věku v roce t a $\sum_x \sum_z ({}_tD_x^z)$ je součet zemřelých ve všech DES a HES v roce t .

Pro rozdělení neznámých věků ve stavech populace by vzorec vypadal následovně (pro koncové stavy):

$${}_t^*P_x = {}_tP_x + {}_tP_{nezn} \cdot \frac{{}_tP_x}{\sum_x ({}_tP_x)} = {}_tP_x \cdot \frac{\sum_x ({}_tP_x)}{\sum_x ({}_tP_x) + {}_tP_{nezn}}, \quad (12)$$

kde ${}_t^*P_x$ je koncový stav populace v roce t v dokončeném věku x navýšený o podíl této skupiny na populaci v neznámém věku. ${}_tP_x$ je počet obyvatel v dokončeném věku x v roce t , ${}_tP_{nezn}$ je počet obyvatel v neznámém věku v roce t a $\sum_x ({}_tP_x)$ je součet počtu obyvatel ve známém věku. Pro střední stavy bylo postupováno analogicky.

Pokud v publikovaných datech nekorespondoval součet počtu událostí přes jednotlivé věkové kategorie (vč. neznámého věku) a celková hodnota publikovaná v pramenném díle (podrobněji viz Mazouch, Hulíková Tesárková, 2018, s. 30), v takovém případě byl rozdíl mezi celkovou hodnotou a součtem dílčích hodnot přičten (nebo v případě záporného výsledku odečten) do skupiny neznámých věků a následně distribuován dle vzorce 11 mezi jednotlivé věkové kategorie.

V případě struktury událostí v III. hlavních souborech událostí (při absenci podrobnějšího členění na DES a HES) byl nejprve počet z kategorie neznámých věků distribuován do základních kategorií (III. hlavních souborů) a až následně bylo provedeno rozdělení III. hlavních souborů na DES a HES (viz následující podkapitola 3.2). Zvolená posloupnost kroků však nemá na výsledky žádný vliv, a pokud by byly počty ze skupiny neznámých věků distribuovány až na závěr, po rozdělení na DES a HES, byl by výsledek identický (Wilmoth et al., 2017, s. 11).

3.2 Rozdělení III. hlavních souborů událostí do struktury DES a HES

Pokud je uspořádání dat ve struktuře III. hlavních souborů událostí, je nutné data rozdělit na DES a HES pro každou jednotku věku v kalendářním roce, aby bylo možné data použít pro další analýzy (včetně například odhadu struktury populace v intercenzálním období, jak bude popsáno v podkapitole 3.5).

Jednoduchým přístupem by mohlo být rozdělení III. hlavního souboru událostí na poloviny a předpokládat, že HES a DES jsou si rovny. Jak píše Wilmoth et al. (2017, s. 12), tento přístup je pro určité věkové skupiny bez větších problémů, protože nadhodnocení úmrtnosti v jedné části je vykompenzováno podhodnocením úmrtnosti v části druhé a odchylky nejsou navíc většinou nijak výrazné. Ze dvou důvodů je však doporučeno použít způsob více sofistikovaný (Wilmoth et al., 2017, s. 12):

1. Zemřelí jsou během prvního roku života výrazněji koncentrováni v dolním elementárním souboru a prosté rozdělení na poloviny by mělo vliv na odhad intenzity úmrtnosti, protože průměrná doba expozice zde není 0,5 roku, jak by předpokládalo rozdělení DES a HES na polovinu, ale výrazně kratší (viz například Roubíček 1997, s. 244 nebo Koschin, 2000, s. 32).
2. V každém věkovém intervalu platí, že rozdělení zemřelých na DES a HES je ovlivněno velikostí jednotlivých kohort, které mohou výrazně fluktuovat díky historickým okolnostem jako jsou neočekávané poklesy nebo nárůsty počtu narozených například díky válkám.

Podíl zemřelých v dolním elementárním souboru v roce t ve věku x (z kohorty z) na celkovém počtu zemřelých v roce t v dokončeném věku x (tedy z kohort z a $z - 1$) můžeme vyjádřit jako:

$$\pi_d(x, t) = \frac{{}_tD_x^z}{{}_tD_x} = \frac{{}_tD_x^z}{{}_tD_x^z + {}_tD_x^{z-1}}. \quad (13)$$

DES a HES ve vzorci 13 nejsou známy, je znám jen celkový počet zemřelých ${}_tD_x$ v roce t ve věku x . Proto je nutné podíl DES na celkovém počtu zemřelých v daném věku (součet DES a HES) odhadnout prostřednictvím $\hat{\pi}_d(x, t)$. Díky tomuto odhadu pak budeme moci jednoduše odhadnout také DES:

$${}_t\hat{D}_x^z = \hat{\pi}_d(x, t) \cdot {}_tD_x \quad (14)$$

a HES:

$${}_t\hat{D}_x^{z-1} = (1 - \hat{\pi}_d(x, t)) \cdot {}_tD_x. \quad (15)$$

Pro odhad $\hat{\pi}_d(x, t)$ byla Wilmothem et al. (2017, s. 13) na datech 3 států (Francie, Švédsko a Japonsko) odhadnuta regresní funkce, která obsahuje několik parametrů a má rozdílný tvar pro muže a pro ženy. Pro ženy jsou podíly DES na celkovém počtu zemřelých v dané věkové skupině odhadnuty jako (Wilmoth, 2017, s. 13 a 14):

$$\begin{aligned} \hat{\pi}_d(x, t) = & 0,4710 + \hat{\alpha}_x^F + 0,7372 \cdot [\pi_b(x, t) - 0,5] + 0,1025 \cdot I(t = 1918) - 0,0237 \\ & \cdot I(t = 1919) - 0,0112 \cdot \log \text{IMR}(t) - 0,0688 \cdot \log \text{IMR}(t) \cdot I(x = 0) \\ & + 0,0268 \cdot \log \text{IMR}(t) \cdot I(x = 1) + 0,1526 \cdot [\log \text{IMR}(t) - \log(0,01)] \\ & \cdot I(x = 0) \cdot I(\text{IMR}(t) < 0,01) \end{aligned} \quad (16)$$

a pro muže:

$$\begin{aligned} \hat{\pi}_d(x, t) = & 0,4838 + \hat{\alpha}_x^M + 0,6992 \cdot [\pi_b(x, t) - 0,5] + 0,0728 \cdot I(t = 1918) - 0,0352 \\ & \cdot I(t = 1919) - 0,0088 \cdot \log \text{IMR}(t) - 0,0745 \cdot \log \text{IMR}(t) \cdot I(x = 0) \\ & + 0,0259 \cdot \log \text{IMR}(t) \cdot I(x = 1) + 0,1673 \cdot [\log \text{IMR}(t) - \log(0,01)] \\ & \cdot I(x = 0) \cdot I(\text{IMR}(t) < 0,01). \end{aligned} \quad (17)$$

Rovnice obsahuje několik logických proměnných $I(x=.)$, které nabývají hodnoty 1, pokud je tvrzení v závorce platné, nebo hodnotu 0, pokud tvrzení platné není. Dvě logické proměnné se týkají let 1918 a 1919, které by měly zachytit vliv epidemie Španělské chřipky do rozložení zemřelých (Wilmoth et al., 2017, s. 13). Hodnoty parametru $\hat{\alpha}_x^F$ pro ženy a jeho ekvivalentu $\hat{\alpha}_x^M$ pro muže jsou dány v tabulce A.1 a A.2, přílohy A, dokumentu Wilmoth et al. (2017, s. 48 a 49). Ty jsou rozděleny (s výjimkou věku 0 a 1) do víceletých věkových skupin, ve kterých je hodnota platná pro všechny věky v rámci věkové skupiny.

Parametr $\pi_b(x, t)$ zohledňuje velikost kohorty, neboť vyjadřuje podíl narozených v roce $(t - x)$ na počtu narozených v letech $(t - x)$ a $(t - x - 1)$ bez rozdílu pohlaví:

$$\pi_b(x, t) = \frac{{}_tB}{{}_tB + {}_{t-x-1}B}, \quad (18)$$

kde ${}_tB$ je počet živě narozených v roce $(t - x)$, a v případě, že počet narozených není znám, odhaduje se $\pi_b(x, t) = 0,5$.

Regresní rovnice dále obsahuje parametr kojenecké úmrtnosti $\text{IMR}(t)$, který je opět bez rozlišení pohlaví a je tedy pro obě rovnice shodný. Wilmoth et al. (2017, s. 13) odkazuje na vzorec odvozený Pressatem (1972), kde:

$$\text{IMR}(t) = \frac{{}_tD_0}{1/3 \cdot {}_{t-1}B + 2/3 \cdot {}_tB}. \quad (19)$$

Pokud neznáme ${}_{t-1}B$, předpokládáme pro výpočet $\text{IMR}(t)$ ${}_{t-1}B = {}_tB$.

Pro více detailů k odhadům regresní funkce, bližší specifikaci zdrojových dat a hodnocení spolehlivosti této metody doporučuji přílohu A dokumentu Wilmoth et al. (2017, s. 44–49).

3.3 Rozdělení počtu událostí z víceletých věkových skupin do jednoletých skupin

Některá data, zejména o počtech událostí, mohou být tříděna v menším detailu a místo jednoletých věkových kategorií jsou počty uvedeny za víceleté věkové kategorie (pak pochopitelně bez rozlišení DES a HES). V takovém případě je nutné nejprve odhadnout distribuci součtů do jednoletých věkových skupin a následně provést rozdělení těchto III. hlavních souborů událostí do struktury DES a HES.

Wilmoth et al. (2017, s. 14 a následně příloha B s. 50–52) uvádí, že pro rozdělení agregovaných dat do jednoletých skupin je vhodné použít kubických křivek odpovídajících distribuci kumulativních počtů zemřelých pro každý kalendářní rok. Postup není vhodné aplikovat na celé věkové spektrum, ale první věkový interval (často jednoletý) a druhý (většinou čtyřletý) je nutné ošetřit jinak. Na zbývajících věkových intervalech již lze postup bez problémů aplikovat až na poslední otevřený interval (Wilmoth et al., 2017, s. 50–52).

Pro tento postup je však také nutné rekonstruovat „dostatečný“ počet let, ve kterých došlo k agregaci počtu událostí do víceletých věkových skupin. V datovém souboru událostí za České země v období 1870–1919 existuje pouze rok 1913, který má údaje tříděné do agregovaných věkových skupin, a to ještě ne v celém věkovém spektru, ale pouze nad 70 let, které tak tvoří III. hlavní soubory událostí sdružující více jednoletých věkových skupin (konkrétně pětileté věkové skupiny až do věku 95 let, na který navazuje otevřená kategorie 95+) .

Přestože se jedná pouze o jeden kalendářní rok a v něm jen o omezený počet věkových skupin, bylo nutné data upravit, tak aby odpovídala struktuře všech ostatních let. Z tohoto důvodu byla zvolena jednodušší metoda rozdělení agregovaných počtů dle průměrných podílů jednoletého věkového intervalu na agregovaném počtu v předchozích třech letech, a to i pro nejvyšší věkovou kategorii 95+ let. Pro každý věkový interval byl spočten koeficient:

$${}_t k_x = \frac{\frac{{}_{t-3} D_x}{{}_{t-3} D_{x,x+4}} + \frac{{}_{t-2} D_x}{{}_{t-2} D_{x,x+4}} + \frac{{}_{t-1} D_x}{{}_{t-1} D_{x,x+4}}}{3}, \quad (20)$$

kde ${}_{t-3} D_{x,x+4}$ je počet zemřelých v roce $t-3$ v dokončeném věku $(x; x+4)$ ¹⁴. Dále byl díky koeficientům odhadnut počet událostí v jednoletých věkových intervalech až do věkové kategorie 99+ jako:

$${}_t \hat{D}_x = {}_t k_x \cdot {}_t D_{x,x+4}. \quad (21)$$

¹⁴ Pro přehlednost je zde použito kulatých závorek. Oba krajní věkové intervaly $-x$ i $x+4$ do intervalu patří.

Odhady tvoří III. hlavní soubory událostí, které je ještě nutné rozdělit na DES a HES, aby byla respektována struktura celé datové sady. Zde se nabízí možnost využít přístupu z předchozí podkapitoly 3.2, která popisuje způsob rozdělení III. hlavních souborů na DES a HES, ale vzhledem k tomu, že již hodnoty ${}_t\hat{D}_x$ jsou odhady, jeví se tento přístup jako zbytečně složitý. I Wilmoth et al. (2017, s. 12) uvádí, že pro věky nad 80 let je možné použít zjednodušení v podobě odhadu $\pi_d(x, t) = 0,5$, a to dokonce i pro situaci, kdy počty zemřelých ve III. hlavních souborech nejsou odhady, jako je tomu v tomto případě. Na straně 52 stejného dokumentu pak píše, že pro rozdělení v případě odhadů je možné použít rozdělení 50/50 pro všechny věkové skupiny (s výjimkou věkové skupiny 0).

Pro odhad podílu DES na III. hlavních souborech pro rok 1913 byla použita metoda odhadu na základě průměrů z předchozích tří let, podobně, jako tomu je ve vzorci (20):

$$\hat{\pi}_d(x, 1913) = \frac{\frac{{}_{1910}D_x^z}{{}_{1910}D_x^z + {}_{1910}D_x^{z-1}} + \frac{{}_{1911}D_x^z}{{}_{1911}D_x^z + {}_{1911}D_x^{z-1}} + \frac{{}_{1912}D_x^z}{{}_{1912}D_x^z + {}_{1912}D_x^{z-1}}}{3} \quad (22)$$

Elementární soubory událostí jsou pak pro jednoleté věkové skupiny pro 70 a více let odhadnuty v případě DES jako:

$${}_{1913}\hat{D}_x^z = \hat{\pi}_d(x, 1913) \cdot {}_{1913}D_x, \quad (23)$$

a HES:

$${}_{1913}\hat{D}_x^{z-1} = (1 - \hat{\pi}_d(x, 1913)) \cdot {}_{1913}D_x. \quad (24)$$

Vedle roku 1913, u kterého v pramenném díle chybí údaje od věku 70 let, jež jsou k dispozici pouze v souhrnných pětiletých věkových skupinách, jsou v datech pouze dvě období, které mají otevřený interval jiný než 99+ či 100+ let (období 1895–1897 a 1914–1918 mají poslední věkový interval 90+). Byť Wilmoth et al. (2017, Appendix C) doporučuje pro rozdělení otevřených věkových intervalů aplikovat Kannistův model, vzhledem k malému počtu let s tímto problémem a také nízkému počtu rozdělovaných událostí (jedná se o nízké stovky událostí v jednotlivých kalendářních letech), byl i pro tuto korekci využit analogicky postup představený ve vzorcích (20)–(24) s posledním věkovým intervalem 99+ let.

3.4 Odhad koncového stavu populace ze sčítání konaného během roku (rozhodné datum jiné než 31. 12.)

V případě konání sčítání lidu k jinému rozhodnému okamžiku, než je konec/začátek roku (půlnoc 31. prosince), je nutné provést odhad koncového stavu populace. Wilmoth et al. (2017, kap. 5) popisují různé situace, kdy sčítání proběhlo před či po konci roku, ve kterém chceme odhadnout koncový stav, ale pro potřeby tohoto materiálu stačí varianta, která popisuje odhad při konání sčítání po konci

sledovaného období. Během sledovaného období totiž všechna sčítání (1869, 1880, 1890, 1900 a 1910) proběhla s rozhodným okamžikem 31. 12. a pouze jedno sčítání v roce 1921 proběhlo v jiném termínu, s rozhodným okamžikem 15. února 1921 (Srb, 2004, s. 18).

Odhad konečného stavu populace je pak možný (dle Wilmoth et al., 2017, s. 20) za dvou předpokladů – rovnoměrného rozdělení osob ve věkových intervalech a počtu událostí během sledovaného období a také je zde zanedbána migrace. Tyto předpoklady jsou přijatelné, pokud je interval mezi okamžikem konání sčítání lidu a okamžikem, ke kterému odhadujeme strukturu populace (na Obr. 2 označený jako f a který v případě roku 1921 činí $1/8$ roku), dostatečně krátký, aby v něm bylo možné zmíněné předpoklady považovat za dodržené.

Pro odhad koncového stavu je nejprve nutné odhadnout velikost kohorty, jež byla na začátku kalendářního roku v dokončeném věku x . Taková kohorta se v okamžiku sčítání, které nemá rozhodný okamžik shodný se začátkem roku, rozprostírá nejen v části intervalu x (část o velikosti $1 - f$), ale také v intervalu $x + 1$ (část o velikosti f). Na Obr. 2 je tento počet znázorněn jako $C_{x'}$, a je odhadnut na základě vztahu (Wilmoth et al., 2017, s. 30):

$$C_{x'} = (1 - f) \cdot C_x + f \cdot C_{x+1}. \quad (25)$$

Následně je nutné odhadnout počty zemřelých, kteří zemřeli v části roku f , a to jak podíl zemřelých z DES (na Obr. 2 značeno jako D_a):

$${}_tD_a^z = f^2 \cdot {}_tD_{x+1}^z \quad (26)$$

a pro HES (na Obr. 2 značeno jako D_b):

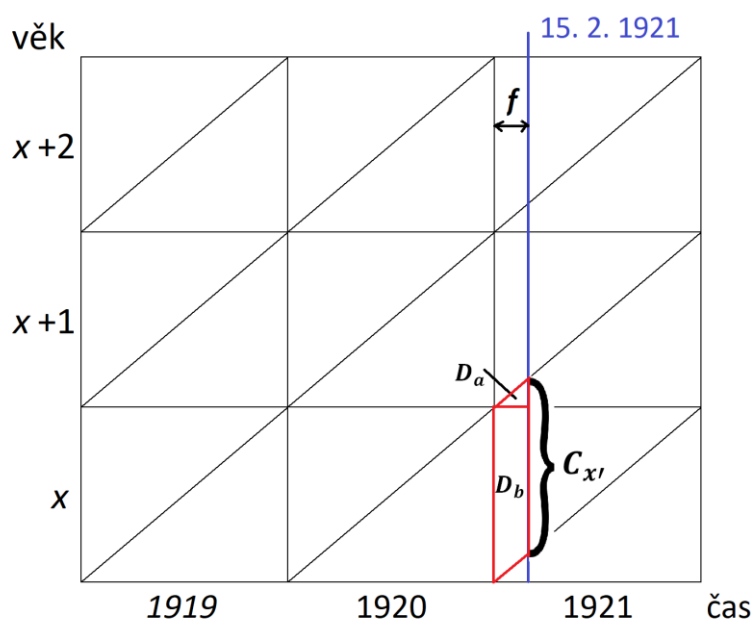
$${}_tD_b^z = (2f - f^2) \cdot {}_tD_x^z. \quad (27)$$

Odhad koncového stavu populace je pak založen na vztahu (Wilmoth et al., 2017, s. 24):

$${}_{t-1}P_x^z = C_{x'} + {}_tD_a^z + {}_tD_b^z. \quad (28)$$

Analogicky lze odhadnout stav populace k libovolnému datu kolem sčítání lidu, ne nutně pouze ke konci nebo začátku roku. Pouze v případě odhadu stavu populace k okamžiku následujícím po sčítání lidu je nutné postupovat odlišně v případě kohorty v dokončeném věku 0 let, jak popisují Wilmoth et al. (2017, s. 22).

Obr. 2: Odhad koncového stavu populace při sčítání konaného během roku



Zdroj: vlastní úprava dle Wilmoth et al. 2017

3.5 Rekonstrukce struktury populace v intercenzálním období

Pro období, ve kterém nejsou publikovány koncové ani střední stavy populace, je nutné stavy odhadnout. Všechna taková období musí být ohraničena sčítáním lidu (nebo okamžikem, ke kterému známe stav a strukturu populace) a v rámci těchto období musíme mít informaci o počtu událostí v co nejpodrobnějším členění na DES a HES. V souladu s postupem Wilmoth et al. (2017, s. 15–19) bude i následující postup rozdělen na dvě základní situace a) kohorty, které na začátku sledovaného období již žijí a b) kohorty, které se během sledovaného období narodí. Metodika bude představena na jednodušším modelu, kdy jsou sčítání (nebo známé stavy populace) vždy ke konci nebo začátku roku.

a) Kohorta žijící na začátku intercenzálního období

V případě žijící kohorty je přístup k odhadu intercenzálních stavů populace založen na bilančním principu (Roubíček, 1997, s. 34), kdy odečtením počtu zemřelých od počátečního stavu získáme počet osob na konci roku a tedy nový počáteční stav. Toto postupné načítání (či spíše odčítání) by v posledním kroku mělo vést k získání stavu na konci sledovaného období, který se rovná stavu zjištěnému v následujícím sčítání. Bohužel tento přístup by platil jen za předpokladu uzavřené populace (bez migrace) a s absolutně dokonalým způsobem evidence událostí a dokonalým způsobem sčítání lidu. Protože s takovými modelovými předpoklady nelze pracovat, jsou výsledky vždy rozdílné, a to z důvodu zmíněné migrace a chyb v záznamech o zemřelých, případně ve sčítání lidu. Přestože předpokládáme nízký vliv těchto faktorů vzhledem k počtům obyvatel v kohortě, nelze je zanedbat, a

proto budou tyto rozdíly rovnoměrně rozděleny do jednotlivých let, ve kterých odhadujeme koncovou populaci (Wilmoth et al., 2017, s. 16).

Označme počty osob v dokončeném věku x zjištěné v cenzu na počátku období, jako ${}_tC_x^z$, pak odhad počtu osob z kohorty z za n let můžeme vyjádřit jako (Wilmoth et al., 2017, s. 17):

$${}_{t+n}\hat{C}_{x+n}^z = {}_tC_x^z - \sum_{i=1}^n ({}_{t+i}D_{x+i-1}^z + {}_{t+i}D_{x+i}^z). \quad (29)$$

Pokud je však v populaci během intercenzálního období nějaká migrace, musí nutně dojít k situaci, kdy odhadované počty dle vzorce (29) nebudou korespondovat se skutečně zjištěnými stavy populace dle sčítání lidu:

$${}_{t+n}C_{x+n}^z = {}_{t+n}\hat{C}_{x+n}^z + \Delta^z, \quad (30)$$

kde Δ^z je zmiňovaná diference způsobená migrací a chybami v evidenci počtu událostí, případně ve sčítání lidu.

Při předpokladu ${}_tP_x^z = {}_tC_x^z$ (koncový/počáteční stav na začátku intercenzálního období je roven stavu ze sčítání) a rovnoměrného rozdělení zmíněných chyb lze odhadnout koncové stavy obyvatel v jednotlivých intercenzálních letech dle vztahu (Wilmoth et al., 2017, s. 17):

$${}_{t+a}P_{x+a}^z = {}_tC_x^z - \sum_{i=1}^a ({}_{t+i}D_{x+i-1}^z + {}_{t+i}D_{x+i}^z) + \frac{a}{n} \cdot \Delta^z, \quad (31)$$

pro $a = 1, \dots, n$.

V případě $a = n$ platí:

$${}_{t+n}P_{x+n}^z = {}_{t+n}C_{x+n}^z, \quad (32)$$

čímž je zajištěno, že koncový stav v posledním období je roven stavu zjištěnému v cenzu ohraničujícím intercenzální období shora.

V případě, že dojde při aplikaci vzorce (31) k situaci, že by byl odhadovaný počet obyvatel záporný, je tato hodnota nahrazena nulou. Tento problém může nastat zejména u nejvyšších věkových skupin, kde je již nízký počet osob a data nemusí být úplně kvalitní – viz komentář Mazoucha a Hulíkové Tesárkové (2018, s. 45), který ukazuje na problém s nepřesností nejvyšších věkových skupin. Ty před sčítáním vykazují často vysoké hodnoty stavu (nejspíše díky nezaznamenaným událostem) a po realizaci sčítání lidu jsou data výrazně korigována (Mazouch, Hulíková Tesárková, 2018, s. 45).

b) Kohorta narozená během intercenzálního období

Postup představený výše lze aplikovat pouze na kohorty, které již v době prvního sčítání lidu žily, ale nelze aplikovat na kohorty, které se během intercenzálního období narodily. Přesto není postup výrazně rozdílný, pouze zde dochází ke zkrácení sledovaného období (neboť kohorty neprožijí v intervalu celých n let) a jako „vstupující“ počet je zde ne stav v okamžiku prvního sčítání, ale počet živě narozených.

Postup popsany Wilmothem et al. (2017, s. 18) uvažuje kohorty narozené v intercenzálním období v roce $t + j$, pro $j = 1, \dots, n$. K je pak počet celých prožitých let žijícího člena narozené kohorty v okamžiku sčítání lidu a spočte se jako $(n - j)$ a odpovídá také dokončenému věku členů kohorty v okamžiku druhého sčítání.

Odhad stavu kohorty narozené v roce $t + j$ v okamžiku druhého sčítání na konci sledovaného období se pak provede analogicky se vzorcem (29) dle Wilmoth et al. (2017, s. 18):

$${}_{t+n}\hat{C}_K^{t+j} = {}_{t+j}B - {}_{t+j}D_0^{t+j} - \sum_{i=0}^{K-1} ({}_{t+j+i+1}D_i^z + {}_{t+j+i+1}D_{i+1}^z), \quad (33)$$

kde ${}_{t+n}\hat{C}_K^{t+j}$ je odhadovaný počet osob v dokončeném věku K na konci roku $t + n$ z kohorty narozené v roce $t + j$. ${}_{t+j}B$ je počet narozených v roce $t + j$ (v tomto případě je $z = t + j$), ${}_{t+j}D_0^{t+j}$ je DES zemřelých v roce $t + j$ z kohorty narozené ve stejném roce $(t + j)$.

Podobně jako ve vzorci (30) i zde platí, že:

$${}_{t+n}C_K^{t+j} = {}_{t+n}\hat{C}_K^{t+j} + \Delta^{t+j}, \quad (34)$$

kde Δ^{t+j} je zmiňovaná difference způsobená migrací a chybami v evidenci počtu událostí, případně ve sčítání lidu pro kohortu $(t + j)$.

Odhad koncového stavu populace v jednotlivých letech pro kohorty narozené v intercenzálním období (v roce $t + j$) pak odhadneme dle vztahu (modifikace Wilmoth et al., 2017, s. 18):

$${}_{t+k}P_k^{t+j} = {}_{t+j}B - {}_{t+j}D_0^{t+j} - \sum_{i=0}^{k-1} ({}_{t+j+i+1}D_i^z + {}_{t+j+i+1}D_{i+1}^z) + \frac{2k+1}{2K+1} \cdot \Delta^{t+j}, \quad (35)$$

pro $k = 0, \dots, K$.

Opět pro $k = K$ platí ${}_{t+n}P_K^{t+j} = {}_{t+n}C_K^{t+j}$, tedy že odhad koncového stavu posledního roku je roven stavu zjištěnému v rámci sčítání lidu.

Pro příklad předpokládejme kohortu narozenou v roce 1904. Sčítání se konala 31. 12. 1900 a 31. 12. 1910, tedy $t + 4$. Pak $j = 4$ a $K = 10 - j = 6$. Kohorta narozená v roce 1904 má v okamžik sčítání lidu v roce 1910 (31. 12.) dokončený věk 6 let. Odhady koncových stavů do $t + 10$ (1910) této kohorty budou vypadat následovně:

$${}_{1904}P_0^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} + \frac{1}{13} \cdot \Delta^{1904}, \quad (36)$$

$${}_{1905}P_1^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - ({}_{1905}D_0^{1904} + {}_{1905}D_1^{1904}) + \frac{3}{13} \cdot \Delta^{1904}, \quad (37)$$

$${}_{1906}P_2^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - \sum_{i=0}^1 ({}_{1905+i}D_i^{1904} + {}_{1905+i}D_{i+1}^{1904}) + \frac{5}{13} \cdot \Delta^{1904}, \quad (38)$$

$${}_{1907}P_3^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - \sum_{i=0}^2 ({}_{1905+i}D_i^{1904} + {}_{1905+i}D_{i+1}^{1904}) + \frac{7}{13} \cdot \Delta^{1904}, \quad (39)$$

$${}_{1908}P_4^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - \sum_{i=0}^3 ({}_{1905+i}D_i^{1904} + {}_{1905+i}D_{i+1}^{1904}) + \frac{9}{13} \cdot \Delta^{1904}, \quad (40)$$

$${}_{1909}P_5^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - \sum_{i=0}^4 ({}_{1905+i}D_i^{1904} + {}_{1905+i}D_{i+1}^{1904}) + \frac{11}{13} \cdot \Delta^{1904}, \quad (41)$$

$${}_{1910}P_6^{1904} = {}_{1904}B - {}_{1904}D_0^{1904} - \sum_{i=0}^5 ({}_{1905+i}D_i^{1904} + {}_{1905+i}D_{i+1}^{1904}) + \frac{13}{13} \cdot \Delta^{1904}. \quad (42)$$

V rámci metodiky popsané Wilmothem et al. (2017, s. 19–24) je popsána i metodika odhadu stavů populace v intercenzálním období pro případ, kdy sčítání nemají rozhodný okamžik 31. 12. Tento přístup by bylo možné využít v případě odhadu stavů v období 1911–1920, protože sčítání lidu v roce 1921 bylo až 15. 2. (Srb, 2004, s. 18). Vzhledem k tomu, že ekvivalentním přístupem proti tomuto postupu je využití nejprve odhadu koncového stavu v roce 1920 dle metodiky v předchozí podkapitole 3.4 a až následně aplikování postupů prezentovaných vzorci (29)–(35), byla pro období 1911–1920 zvolena metoda ve dvou krocích, a proto obecný přístup pro rozhodný okamžik jindy než 31. 12., zde není prezentován (pro zájemce odkazují na zmiňovanou část metodiky Wilmoth et al., 2017, s. 19–24).

4 Dostupné datové zdroje (prameny dat)

Podrobný výčet datových zdrojů pro období od roku 1870 do současnosti lze nalézt v publikaci Mazoucha a Hulíkové Tesárkové (2018) včetně elektronických odkazů na příslušná pramenná díla. Zde je proto výčet datových zdrojů pouze shrnující a jen do roku 1920. V případě, že publikace neobsahovala souhrnný údaj za celé území Českých zemí, byl údaj dopočten jako součet hodnot za Čechy, Moravu a Slezsko.

4.1 Počet událostí

V období 1870–1879 jsou zdrojem dat publikace „*Statistisches Jahrbuch Für das Jahr ...*“, které vycházely každoročně, s přibližně 2-3letým zpožděním. Na tuto řadu navazuje od roku 1880 „*Bewegung der Bevölkerung der im Reichsrathe vertretenen Königreiche und Länder*“, která opět vychází každoročně až do roku 1913 (s výjimkou 3 svazků, které obsahovaly dva roky 1881/1882, 1893/1894 a 1908/1909) a také s podobným zpožděním 2-3 let (Mazouch, Hulíková Tesárková, 2018, s. 27 a 29). Výjimkou je poslední rok 1913, který (nejspíše díky počátku 1. světové války) vychází až v roce 1918 a v některých pasážích není dodržena struktura z předchozích publikací jako například u počtu událostí, který je zde členěn do DES a HES pouze do věku 69 let a od věku 70 let jsou publikovány již pouze pětileté věkové skupiny.

Pro období 1. světové války je k dispozici publikace „*Přirozená měna obyvatelstva v zemích Koruny české v letech 1. světové války*“ (ČSÚ, 2005), která, jak již bylo uvedeno v kap. 1, vznikla na základě podkladů předaných Ústřední statistickou komisí ve Vídni, ale zpracovány byly až po 2. světové válce a publikace je tak založena na rekonstruovaných torzech původních datových zdrojů, které se týkají pouze údajů z civilních matrik (nejsou zde zahrnuty vojenské oběti 1. světové války). Roky 1919 a 1920 jsou pak již zpracovány přímo Státním úřadem statistickým pod názvem „*Pohyb obyvatelstva v republice Československé*“ a publikace vychází za oba roky v jednom svazku v roce 1929.

Všechny datové zdroje 1870–1894 jsou tříděny do III. hlavních souborů událostí a od roku 1895 na DES a HES událostí. V publikaci Mazoucha a Hulíkové Tesárkové (2018, kap. 4.1) jsou podrobně popsány také počty zemřelých v neznámých věcích a struktura dat v nejvyšších věcích (věk, pro který je již údaj publikován otevřeným věkovým intervalem).

4.2 Stav populace

Stav a strukturu populace v období 1870–1920 lze zjistit pouze v období konání sčítání lidu. V roce 1871 vychází publikace „*Bevölkerung und Viehstand der im Reichsrathe vertretenen Königreiche und Länder, dann der Militärgränze nach der Zählung vom 31. December 1869*“ obsahující výsledky sčítání lidu v roce 1869. Podobně v roce 1882 vychází publikace „*Die Ergebnisse der Volkszählung und der mit*

derselben verbundenen Zählung der häuslichen Nutzthiere vom 31. December 1880 in den im Reichsrathe vertretenen Königreichen und Ländern“, která obsahuje výsledky sčítání lidu v roce 1880. Datovým pramenem pro sčítání 1890 je publikace *„Die Ergebnisse der Volkszählung vom 31. December 1890 in den im Reichsrathe vertretenen Königreichen und Ländern*“, která vyšla v roce 1893, pro sčítání v roce 1900 publikace *„Die Ergebnisse der Volkszählung vom 31. December 1900 in den im Reichsrathe vertretenen Königreichen und Ländern*“ vycházející v roce 1904 a poslední rakousko-uherské sčítání v roce 1910 vychází v publikaci *„Die Ergebnisse der Volkszählung vom 31. Dezember 1910 in den im Reichsrathe vertretenen Königreichen und Ländern*“ v roce 1914. První sčítání realizované po 1. světové válce Státním úřadem statistickým nebylo s rozhodným okamžikem 31. 12. 1920, ale až 15. 2. 1921 a výsledky byly publikovány v roce 1924 v publikaci *„Sčítání lidu v Republice Československé ze dne 15. února 1921*“.

Pro rok 1920 je také k dispozici odhad středního stavu, který je prvním rokem časové řady středních stavů obyvatelstva. Tato řada pokračuje bez přerušení až do současnosti, pochází z interních materiálů úřadu a slouží například pro konstrukci úmrtnostních tabulek (ČSÚ, 2020). Střední stav 1920 však vykazuje určité známky nekonzistence s ostatními datovými zdroji (viz kap. 5.3).

Veškerá data o stavu populace jsou tříděna dle jednotek věku a pohlaví. Publikace Mazoucha a Hulíkové Tesárkové (2018, kap. 4.2) obsahuje podrobný výčet počtů osob v neznámém věku v jednotlivých letech a také informaci o věku, pro který je již údaj publikován otevřeným věkovým intervalem.

4.3 Shrnutí datových zdrojů

Pramenná díla, s výjimkou sčítání lidu 1921 a středního stavu obyvatel 1920, jsou k dispozici v elektronické podobě nejčastěji formou skenovaných původních materiálů. Bohužel nelze vše nalézt na jednom místě, internetové odkazy jsou kompletně uvedeny v seznamu datových zdrojů publikace Mazoucha a Hulíkové Tesárkové (2018, s. 69–80). Publikace také, vedle internetových odkazů na původní zdroje, obsahuje soubory, které kompilují jednotlivé prameny a umožňují přístup k datům na jednom místě a v zpracovatelné elektronické podobě.

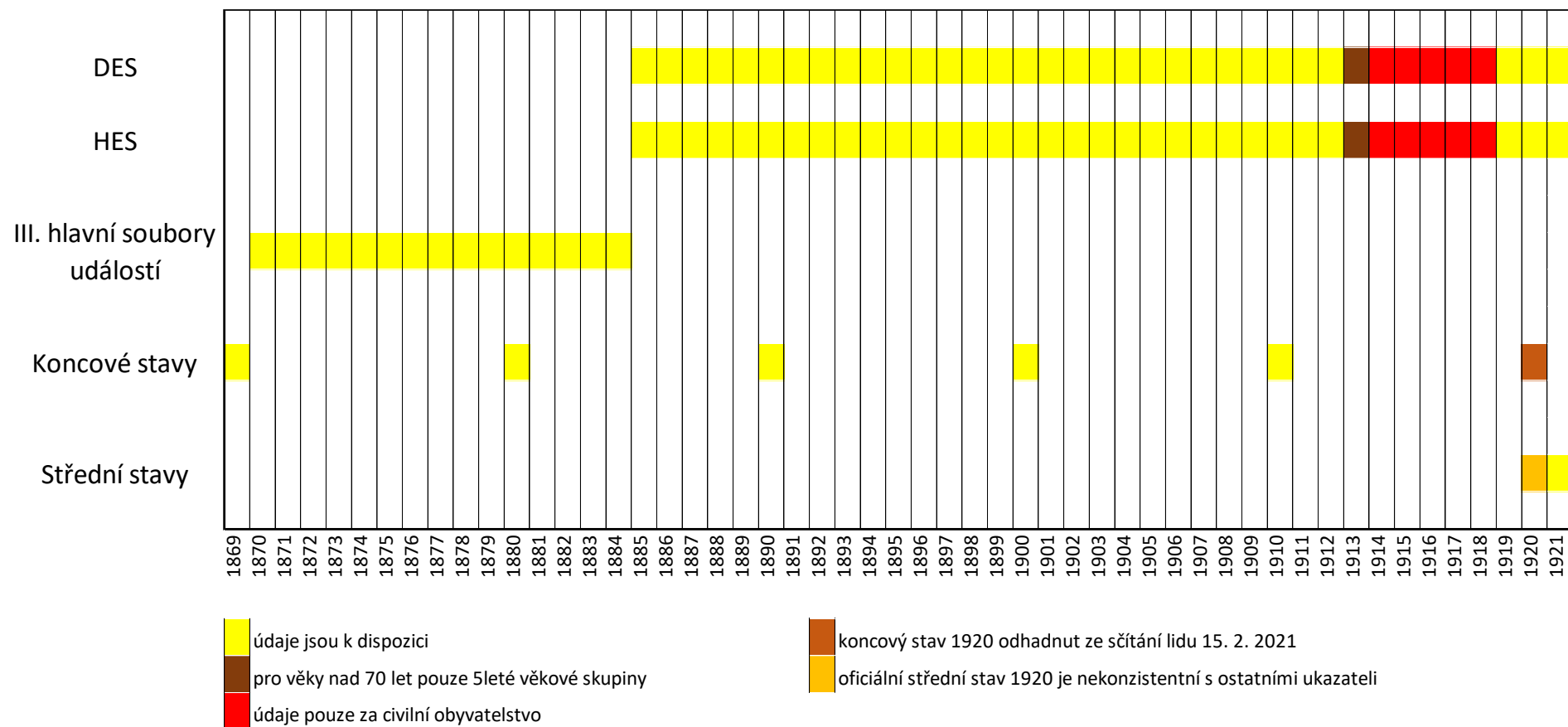
Vedle dat o zemřelých a stavu populace, jsou pro analýzu třeba i data o počtu živě narozených, která jsou také v rámci příloh publikace Mazoucha a Hulíkové Tesárkové (2018) k dispozici. Údaje o živě narozených jsou součástí pramenných děl uvedených v kapitole 4.1.

Pro lepší ilustraci dostupných dat může sloužit Obr. 3, který ukazuje jednotlivé indikátory (DES, HES, III. hlavní soubory událostí, koncové stavy a střední stavy) v perspektivě celého sledovaného období. Je zde patrné, že žádný indikátor není k dispozici za celé období bez nutnosti určité modifikace či

úpravy. Cílem je, aby v rámci kompletního sledovaného období došlo k prodloužení (případně doplnění) jednotlivých indikátorů do takové podoby, kdy budou všechny řady k dispozici v konzistentní podobě od roku 1870 do roku 1920. Z hlediska vazeb mezi jednotlivými ukazateli je nutné rekonstruovat zejména řady DES, HES a koncové stavy. III. hlavní soubory událostí a střední stavy obyvatelstva jsou pak jen jednoduše dopočteny z předchozích ukazatelů (viz vztahy popsané vzorcem (3) a (6) výše).

Schéma na Obr. 3 končí v roce 1921, ale pro další období platí, že jsou k dispozici všechny indikátory, a proto rekonstrukce řady 1870–1920 (pro DES a HES 1870–1918, protože data za rok 1919 a 1920 již v tomto třídění publikována jsou a jsou bez problémů) prodlouží stávající řady plynule do historie pro všechny sledované ukazatele.

Obr. 3: Znáznornění dostupnosti ukazatelů (počtů událostí a stavů populace) v jednotlivých kalendářních letech v období 1969–1921



Zdroj: Vlastní zpracování

5 Rekonstrukce jednotlivých ukazatelů

Kapitola 3 obsahuje obecný metodický popis přístupů k úpravám jednotlivých ukazatelů tak, aby odpovídaly požadavkům konzistence. Při aplikaci na reálných datech došlo k mírným úpravám, které budou představeny v následující části. Postup rekonstrukce dat bylo nutné, jak je patrné i z Obr. 3, začít distribucí III. hlavních souborů událostí v období 1870–1894 do struktury DES a HES, což je podmínka pro následný krok, který umožňuje zrekonstruovat také koncové stavy obyvatelstva.

Následně byly odhadnuty koncové stavy obyvatelstva v intercenzálních obdobích 1870–1879, 1881–1889, 1891–1899, 1901–1909. Specifický přístup bylo nutné zvolit pro období 1911–1920, které bylo postiženo 1. světovou válkou a díky tomu bylo porušeno několik základních předpokladů, stanovených v kapitole 3.

Před zahájením jednotlivých kroků byl však nejprve u datových zdrojů vyřešen problém s případnou skupinou událostí či osob, které nemají určený věk. Proto byl použit vzorec (11) v případě počtu událostí a vzorec (12) pro stav populace, aby v následujících krocích již nebylo nutné s touto kategorií pracovat, a počty událostí a osob bez udání věku byly zahrnuty do dalších analýz.

5.1 Rozdělení III. hlavních souborů událostí do struktury DES a HES v období 1870–1894

Aplikace postupu Wilmoth et al. (2017, kap. 4.2), popsaná v kapitole 3.2 ve formě vzorců (13) až (19), zajistila odhad podílů DES na III. hlavních souborech událostí pro všechna kalendářní léta a věkové skupiny. Následná validace výsledků ukázala, že většina odhadovaných parametrů $\hat{\pi}_d(x, t)$ pro $x > 2$ skutečně odpovídá předpokladu, že by hodnota měla být blízká 0,5, pokud není rozdělení ovlivněno relativní velikostí kohort (Wilmoth et al., 2017, s. 12).

Pro nejmladší věkové skupiny, zejména pro dokončený věk 0, by hodnota $\hat{\pi}_d(x, t)$ měla být vyšší než 0,5 a díky poklesu kojenecké úmrtnosti v čase by tato hodnota měla pro dokončený věk 0 růst a pro dokončené věky 1 a 2 naopak konvergovat k hodnotě 0,5 (Wilmoth et al., 2017, s. 13). Výsledky aplikace vzorců (13) až (19) na data 1870–1894 tyto podmínky splňovaly, ale problém nastal v navázání odhadů na časovou řadu od roku 1895, kdy jsou již hodnoty DES a HES známy.

Přechod z odhadovaných hodnot na hodnoty empirické by měl být hladký, protože není důvod předpokládat v časové řadě nějaký výrazný zlom. Obr. 4 však ukazuje, že odhadované hodnoty v některých případech úroveň podhodnocují či nadhodnocují. Ideální ilustrací hladké návaznosti je $\hat{\pi}_d(1, t)$ pro ženy (oranžová linka v Obr. 4). V průběhu řady nedochází k žádnému výraznému zlomu, pouze model odhaduje nižší úroveň variability hodnot, než je následně patrné z období po roce 1895. Trend také respektuje předpoklad postupného poklesu hodnoty v čase.

$\hat{\pi}_d(0, t)$ pro muže i pro ženy vykazuje podhodnocení hodnot, které shodně pro obě pohlaví nabývají po roce 1895 hodnot těsně nad 0,70, přičemž odhad v období 1870–1894 je na úrovni 0,55 pro ženy a 0,60 pro muže. $\hat{\pi}_d(1, t)$ pro muže je naopak mírně nadhodnocen, když po roce 1895 klesá tato hodnota postupně pod 0,6 a vyskytuje se nejčastěji v rozmezí 0,55–0,60 a odhad pro roky 1870–1894 je na úrovni 0,65.

Z těchto důvodů bylo nutné upravit odhad $\hat{\pi}_d(x, t)$ pro $x = 0$ a 1 pro muže a pro $x = 0$ pro ženy tak, aby hladce navazovaly na řady od roku 1895. K tomuto byl využit vztah popsáný Mazouchem a Hulíkovou Tesárkovou (2018, s. 32):

$$\hat{\pi}_d(0, t) = \frac{\frac{{}_tB}{{}_{t-1}B} \cdot 0,7}{\frac{{}_tB}{{}_{t-1}B} \cdot 0,7 + 0,3}, \quad (43)$$

a

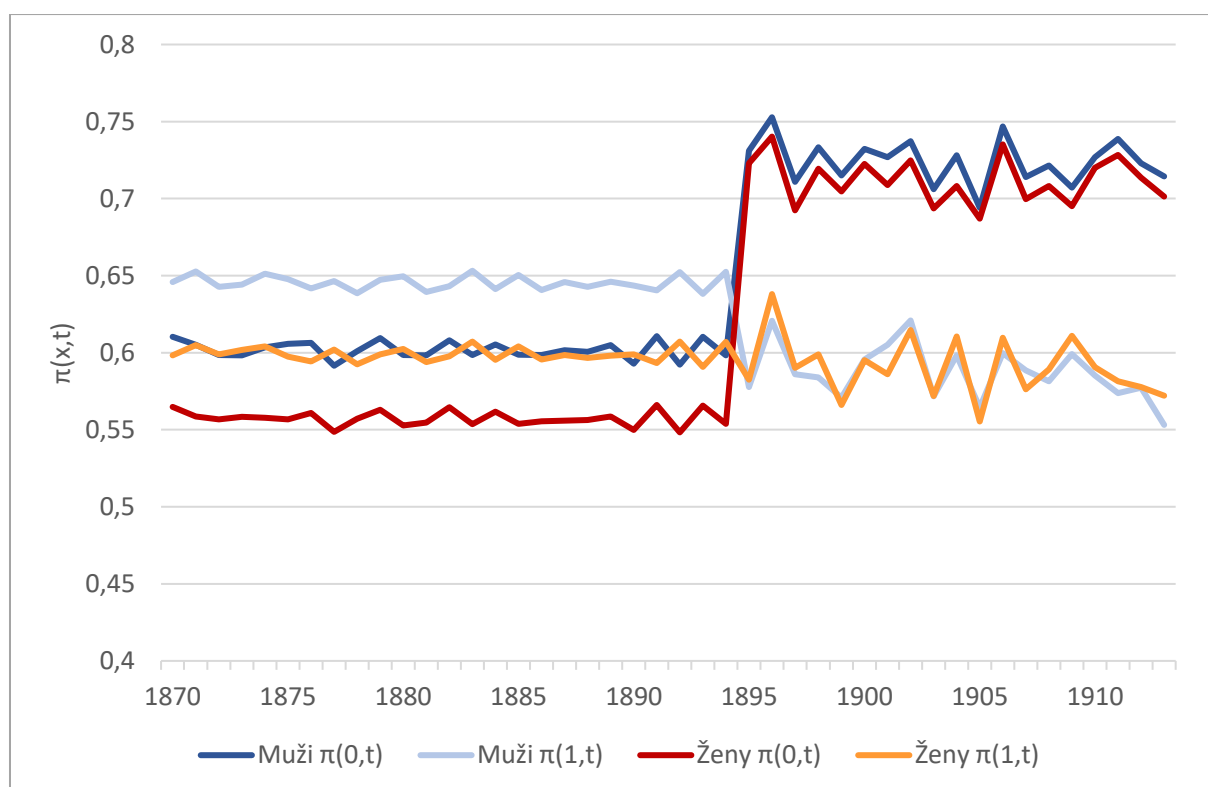
$$\hat{\pi}_d(1, t) = \frac{\frac{{}_{t-1}B}{{}_{t-2}B} \cdot 0,6}{\frac{{}_{t-1}B}{{}_{t-2}B} \cdot 0,6 + 0,4}, \quad (44)$$

kde ${}_tB$ je počet živě narozených v roce t .

Tímto způsobem odhadnuté $\hat{\pi}_d(x, t)$ pak hladce navazují na empirické hodnoty $\pi_d(x, t)$ pro $t > 1894$. Graficky lze výsledek vidět v knize Mazoucha a Hulíkové Tesárkové (2018, s. 33, obr. 5 a 6).

Po rozdělení III. hlavních souborů událostí do DES a HES v období 1870–1894 a rozdělení otevřených věkových intervalů 90+ v letech 1895–1897 již nic nebránilo aplikovat na data postup představený v kap. 3.5, který umožňuje odhadnout stav (a strukturu) populace v obdobích mimo roky, kdy se konalo sčítání lidu. Bezproblémová aplikace postupu byla provedena na data pro období počínající rokem 1869 (resp. jeho koncovým stavem, protože sčítání v roce 1869 mělo rozhodný okamžik 31. prosince) a končící rokem 1910, kdy opět rozhodný okamžik sčítání lidu připadal na 31. prosince. Zrekonstruována tak byla struktura obyvatelstva za období 1870–1879, 1881–1889, 1891–1899, 1901–1909, a to vždy rozdělená na jednotky věku a dle pohlaví. Tímto způsobem bylo doplněné schéma na Obr. 3 do podoby souvislých žlutých čar pro DES, HES a koncové stavy obyvatelstva v období 1870–1910. Aplikací vzorců (3) pro počet událostí a (7) pro stav populace byly doplněny stejně tak i řady událostí uspořádané do III. hlavních souborů událostí, resp. střední stavy populace.

Obrázek 4: Odhad (1870–1894) a reálná hodnota (1895–1913) $\pi_d(x,t)$ pro věky 0 a 1



Zdroj: Vlastní zpracování

5.2 Rekonstrukce ukazatelů pro období 1911–1920

Zmíněná 1. světová válka a následný vznik Československé republiky měly vliv na návaznost a koherenci datových zdrojů (nejen) demografické statistiky v období 1911–1919. Toto období si proto vyžaduje podrobnější popis, protože zde muselo být, před prostým odhadem stavů populace v intercensálním období, realizováno několik dílčích kroků, které umožnily odhadnout stavy populace lépe s přihlédnutím k různým specifikům sledovaného období. Došlo také ke korekcím samotných počtů událostí, jak bude popsáno v jednotlivých částech níže.

V knize Mazoucha a Hulíkové Tesárkové (2018, s. 27–28 a 41) je toto období také rekonstruováno, ale vzhledem k zaměření knihy pouze na kohortní analýzu byl zvolen jednodušší přístup. Některé kroky byly buď úplně ignorovány (například pro rok 1913 nebylo nutné rekonstruovat počty událostí pro věk 70 a více let, protože tyto věky již nebyly součástí analýzy) nebo byl zvolen výrazně zjednodušující postup, než je tomu v případě rekonstrukce dat nyní. Příprava dat, mimo jiné také pro transversální analýzu, vyžaduje přesnější přístup například k rozvržení zemřelých nezaznamenaných v civilních matrikách v rámci 1. světové války či migrace.

Následující kapitola proto začíná prvním krokem, kterým je rekonstrukce koncového stavu roku 1920, následuje část věnující se korekci počtů událostí v letech, které nejsou v požadované struktuře, dále jsou počty evidovaných událostí v civilních matrikách rozšířeny o odhady počtů zemřelých mužů za 1. světové války a je odhadnuto také rozdělení migrace v rámci sledovaného období a až na závěr je konečně možné rekonstruovat i koncové stavy pro roky 1911 až 1919.

5.3 Rekonstrukce koncového stavu roku 1920 a odhad středního stavu 1920

Pro odhad koncového stavu v roce 1920 bylo využito postupu popsaneho v kapitole 3.4 vzorci (25)–(28), což znamená, že koncový stav je odhadnut na základě dat ze sčítání lidu konaného v roce 1921 s rozhodným okamžikem o půlnoci 15. února (Srb, 2004, s. 18). Vzdálenost mezi rozhodným okamžikem a koncem předchozího roku je jen 1,5 měsíce, což je interval, který stále ještě naplňuje předpoklady o rovnoměrném rozdělení osob ve věkových intervalech a počtu událostí během sledovaného období a také zanedbává migraci (Wilmoth et al., 2017, s. 20).

Jako alternativa se nabízelo využít vztahu popsaneho vzorcem (7), který využívá pro odhad středního stavu průměr z dvou sousedních koncových stavů, protože v roce 1920 (a pro následující roky také) jsou publikovány střední stavy obyvatel v potřebném členění (dle pohlaví a jednotek věku). Upravený vzorec (7) by pak vypadal následovně:

$${}_{1920}P_x^z = \frac{{}_{1920}S_x^{z-1,z} + {}_{1921}S_x^{z,z+1}}{2}. \quad (45)$$

Střední stav populace v roce 1920 však vykazuje výraznou nekonzistenci se středními stavy následujících let. V Tab. 1 je naznačený vývoj středních stavů v čase a také difference mezi jednotlivými středními stavy ${}_tS_x - {}_{t+1}S_{x+1}$. Mezi lety 1920 a 1921 jsou difference velmi vysoké. Střední stav žen v nejnižší věkové skupině v roce 1920 je 114 295 osob. Stejná skupina o jeden rok později již vykazuje pouze 90 104 žen, což je úbytek o 24 191 osob. Stejně věkové skupiny mezi lety 1921 a 1922 mají diferenci pouze 5 262, ta je rozhodně blíže realitě ve smyslu úrovně úmrtnosti, která v této věkové kategorii platila.

Podobné problémy lze vidět i u věkových kategorií vyšších. Někde dokonce dochází k nárůstu, který pak musí být způsoben migrací, ale v tomto případě jsou přírůstky na velmi vysoké úrovni. Pro věkovou skupinu v dokončeném věku 3 roky se mezi roky 1920 a 1921 (skupina v dokončeném věku 4 roky v roce 1921) zvýší počet o 2 821 osob, pro skupinu o rok starší je nárůst dokonce 10 315 osob. O jeden kalendářní rok později jsou pro stejné věky difference o několik řádů nižší.

Také počet osob v dokončeném věku 0 let v roce 1920 neodpovídá jiným podpůrným údajům, jako je třeba počet narozených osob v odpovídajícím období. Střední stav nejnižší věkové skupiny v roce t závisí na počtu narozených osob ve 2. polovině roku $t - 1$ a 1. polovině roku t . Pro druhou polovinu

roku 1919 platí, že se živě narodilo 58 161 děvčat (SÚS, 1929a) a v 1. pololetí roku 1920 60 792 (SÚS, 1929a), což je celkově 118 953 živě narozených dívek. Z těchto pak v polovině roku 1920 žilo 114 295, tedy více než 96 %. O rok později jsou počty následující: počet živě narozených děvčat ve 2. pololetí roku 1920 byl 57 461 (SÚS, 1929a) a v 1. pololetí následujícího roku 62 685 (SÚS, 1929b), což dohromady činí 120 146. Střední stav nejnižší věkové kategorie byl v roce 1921 109 372 dívek, což tvořilo přibližně 91 % ze skupiny narozených. Pro rok 1922 pak byl poměr dokonce pouze 86 %, celkově se narodilo za 2. pololetí roku 1921 a 1. pololetí roku 1922 živě 125 572 dívek (SÚS, 1929b).

Tabulka 1: Střední stavy žen v roce 1920, 1921 a 1922 a meziroční změny mezi lety 1920–1921 a 1921–1922

Dokončený věk	1920	1921	${}_{1920}S_x - {}_{1921}S_{x+1}$	1922	${}_{1921}S_x - {}_{1922}S_{x+1}$
0	114 295	109 372	.	108 379	.
1	76 316	90 104	24 191	104 111	5 262
2	46 527	60 747	15 569	89 074	1 029
3	47 005	46 607	-80	60 638	109
4	52 835	49 827	-2 821	46 527	80
5	73 479	63 149	-10 315	49 783	44

Zdroj: Mazouch, Hulíková Tesárková (2018), vlastní zpracování

Bez ohledu na úroveň kojenecké úmrtnosti, která v té době platila, lze vidět, že střední stav žen v dokončeném věku 0 let je v roce 1920 nadhodnocen. Ke stejným závěrům dojdeme i v případě analýzy mužů. Opět je zde značně nekonzistentní střední stav roku 1920, kde jsou dokončené věky 0 a 1 nadhodnoceny, a pak další věkové skupiny nekorespondují s následujícími hodnotami středních stavů z následujících let.

Pro odhad středního stavu 1920, publikovaného Státním úřadem statistickým, neexistuje bohužel žádný metodický popis, jak byl stav dopočítán. Pokud se vycházelo opět z prvního známého údaje o stavu a struktuře populace (15. února 1921), je možné, že vzdálenost okamžiku, ke kterému odhadujeme střední stav od okamžiku sčítání, je již příliš velká, a proto vyústí ve velmi nepřesné odhady. Toto je však pouze domněnka, kterou nemůžeme ověřit. Z důvodů nekonzistence publikovaného středního stavu v roce 1920 byl pro další práci odhadnut střední stav dle vztahu (7), který využívá odhadu koncového stavu 1920 (přepočítaného ze sčítání lidu 15. 2. 1921) a koncového stavu 1921 (který je, dle Mazouch, Hulíková Tesárková, 2018, s. 39, založen na průměru středních stavů 1921 a 1922).

5.4 Rozdělení zemřelých do DES a HES ve věkové skupině 70 a více let v roce 1913

Jak již bylo uvedeno v kapitole 4.1, v roce 1918 vychází konečně publikace „*Bewegung der Bevölkerung der im Reichsrathe vertretenen Königreiche und Länder*“ s daty za rok 1913. Její struktura v některých částech nekoresponduje s předchozími publikacemi za předešlé roky. V oblasti počtu událostí je to v bodě, kdy záznamy o zemřelých ve věku 70 a více let v pramenném díle chybí. Na jiném místě však můžeme nalézt třídění do agregovaných pětiletých skupin (pro každé pohlaví zvlášť) s posledním otevřeným intervalem 95+ let, které byly rozděleny do jednoletých věkových skupin (s poslední věkovou kategorií 99+) dle metodiky popsané v kapitole 3.3.

Pro rozdělení zemřelých do jednotlivých DES a HES bylo využito vztahů z kap. 3.2 výše, konkrétně vzorců (20)–(24). Data tak svojí strukturou respektují rozdělení předchozích i následujících let a lze tak považovat v tomto ohledu časovou řadu za koherentní.

5.5 Odhad počtu zemřelých nezaznamenaných v civilních matrikách v období 1. světové války

Pro korektní odhad úrovně úmrtnosti v období 1. světové války je nutné civilní záznamy doplnit o odhad záznamů nevidovaných v civilních matrikách, zejména o válečné oběti. Prvním krokem je odhad celkového počtu záznamů, které v civilních matrikách chybí, a následně jejich rozdělení v čase a mezi jednotlivé věkové skupiny v daném kalendářním roce.

V publikaci Mazoucha a Hulíkové Tesárkové (2018, s. 28) se pracuje s počtem nezaznamenaných událostí na úrovni 200 000 zemřelých mužů. Ti jsou následně rozděleni mezi jednotlivé věkové skupiny dle podílů odhadnutých Ulanisem (převzato z článku Fialové, 2014) a v čase byly počty rozděleny téměř rovnoměrně s respektem ke kratšímu období bojů v roce 1914 a 1918 (Mazouch, Hulíková Tesárková, 2018, s. 28). Tento přístup byl vzhledem k využití dat pro kohortní analýzu přijatelný. Pokud však budou data analyzována také transversálně, je nutné předpoklady o celkové úrovni a rozdělení dat do jednotlivých věkových skupin a kalendářních let revidovat. Transverzální přístup je mnohem citlivější na meziroční změny, které se v případě kohortního přístupu v rámci generace kompenzují a případné nepřesnosti nejsou ve výsledných hodnotách nijak zásadní.

Základní otázka válečných ztrát je zajímavá také z jiného pohledu – jak to, že Rakousko-Uhersko, které disponovalo velmi dobrou statistikou na vysoké úrovni (což zmiňuje například Srb a Kučera, 1959), nedokázalo evidovat tak důležité údaje, jako byl počet válečných ztrát? Odpověď lze nalézt v popisu procesů záznamů válečných ztrát, které diskutuje Vojenský historický ústav v Praze (VHÚ, 2017): „*Do války v roce 1914 vstupovala rakousko-uherská armáda vybavena poměrně propracovaným systémem evidence válečných ztrát. ... Teorie však narazila na praxi úvodních tažení a záhy se ukázalo, jak*

rakousko-uherská vojenská správa podcenila nárůst této agendy.“ Údaje nebyly plánovány pro využití pouze vojenské správy, ale měly být také zveřejňovány tiskem (Seznam ztrát vydávaný od č. 1 (12. 8. 1914) do č. 709 (14. 2. 1919), čítá celkem 39 144 stran) z důvodu informování obyvatelstva a měl z nich čerpat také systém sociální péče pro pozůstalé (VHÚ, 2017).

Důležité je připomenout metodologický nesoulad mezi evidencí válečných ztrát a demografií. Proti demografickým událostem, které zde sledujeme (počet zemřelých), jsou mezi válečné ztráty počítány i další události, jako ranění, nemocní, nezvěstní či zajatí (VHÚ, 2017). Tohoto terminologického nesouladu si všímá i Fialová (2014), která uvádí i další možné členění například na bojové a nebojové ztráty. Evidence válečných ztrát v období 1. světové války se realizovala formou vyplňování seznamů ztrát (s uvedením podrobných informací o vojáku z jeho legitimačního lístku) za jednotlivé pododdíly armády. Seznamy následně centrálně zpracovávala *Skupina pro seznam ztrát při 10. oddělení ministerstva války*, která z nich přepisem vyhotovovala lístkový katalog, jenž je nyní (i spolu s jinými datovými zdroji) postupně digitalizován Vojenským historickým ústavem v Praze (VHÚ, 2017).

Situace s evidencí válečných ztrát je ještě složitější v tom, že voják mohl být během války evidován v těchto počtech i několikrát (vícekrát zraněn či zajat, z nezvěstného se mohl stát například zajatým, atp.). Lístky také obsahovaly chyby či neúplné informace, které se pochopitelně snažili úředníci doplňovat nebo opravovat, ale ne vždy to bylo možné (VHÚ, 2017). Pro demografii jsou důležité záznamy o zemřelých, tedy padlých v boji, či zemřelých na následky zranění nebo zemřelí v zajetí. Tyto počty nejde jednoduše zjistit a zejména absence informací o osudech zajatých či nezvěstných ukazuje, že využití tohoto zdroje je značně omezené.

Přesto na sklonku války začal se Seznamy ztrát pracovat Wilhelm Winkler, který se pokusil analyzovat část záznamů (které pak zobecňoval na celou armádu). V případě pohřešovaných použil Winkler konstantu 2/3, a ty považoval za zemřelé (Winkler, 1919). Odhady Winklera, které lze považovat za první odhady válečných ztrát, sklidily tvrdou kritiku díky použité metodice. A. Boháč (1920) uvádí ve své recenzi Winklerova díla, že „*Winkler se první pokusil o tento úkol, avšak jeho pokus selhal, naprosto*“ (Boháč, 1920).

V roce 1923 publikoval F. J. Netušil (1923) odhady dopadů války v Čechách a na Moravě (bez Slezska). Odhadl počet padlých „v poli“ na přibližně 165 000 osob. Vedle těchto přímých ztrát odhadoval také demografické ztráty vzniklé díky vyšší úmrtnosti (zhoršením životních podmínek) a reprodukční ztráty na úrovni 678 000 (Jindra, 2014, převzato z Netušil, 1923), což však není předmětem tohoto textu. Podobně pojal své odhady také Srb (2004, s. 206), kdy počítá se ztrátami přímo spojenými s válkou na úrovni 300 000 osob a další reprodukční ztráty 610 000. Srb tedy uvažuje počet padlých téměř dvojnásobný oproti Netušilovi (1923).

Jak však popisují Mazouch a Hulíková Tesárková (2018, s. 28), pokud předpokládáme přibližně stejné migrační chování mužů a žen, ukazovala by válečná ztráta 300 tisíc mužů na situaci, kdy ženy mají v období 1910–1920 saldo způsobené migrací (a případnými neevidovanými událostmi) na úrovni asi 90 tisíc¹⁵ osob a muži za stejné období naopak mírný přebytek přibližně 10 tisíc osob. Odhad Netušila (1923) by naopak vedl k deficitu na straně mužů na úrovni 120 tisíc. Fialová (2014), která pracuje s odhadem Srba (2004), validuje odhad padlých mužů prostřednictvím počtu vdov, které po padlých ženatých mužích zůstaly. Tuto úvahu dále rozvíjí Mazouch a Hulíková Tesárková (2018, s. 28), a pokud by platil předpoklad 300 tisíc padlých mužů, měl by být počet vdov v roce 1920 o 50 % vyšší, než tomu ve skutečnosti bylo.

Skutečný počet zemřelých vojáků, který není evidován v rámci civilních záznamů, nikdy nebude znám, a tak lze počty pouze odhadovat. Z hlediska koherence s počty vdov a migračním chováním žen, je odhadován počet těchto událostí na úrovni 200 tisíc mužů, tak jak to bylo využito již v publikaci Mazoucha a Hulíkové Tesárkové (2018, s. 28).

Z hlediska odhadu celkového počtu zemřelých vojáků je nutné přistupovat k uvedeným datovým zdrojům evidujícím padlé obezřetně. Pro další parametry rozdělení událostí jsou však data více než vhodná, což využil například Urlanis (1963), který odhadl podíl jednotlivých kohort na skupině padlých vojáků. Graficky je rozdělení publikováno v knize Mazoucha a Hulíkové Tesárkové (2018, s. 28).

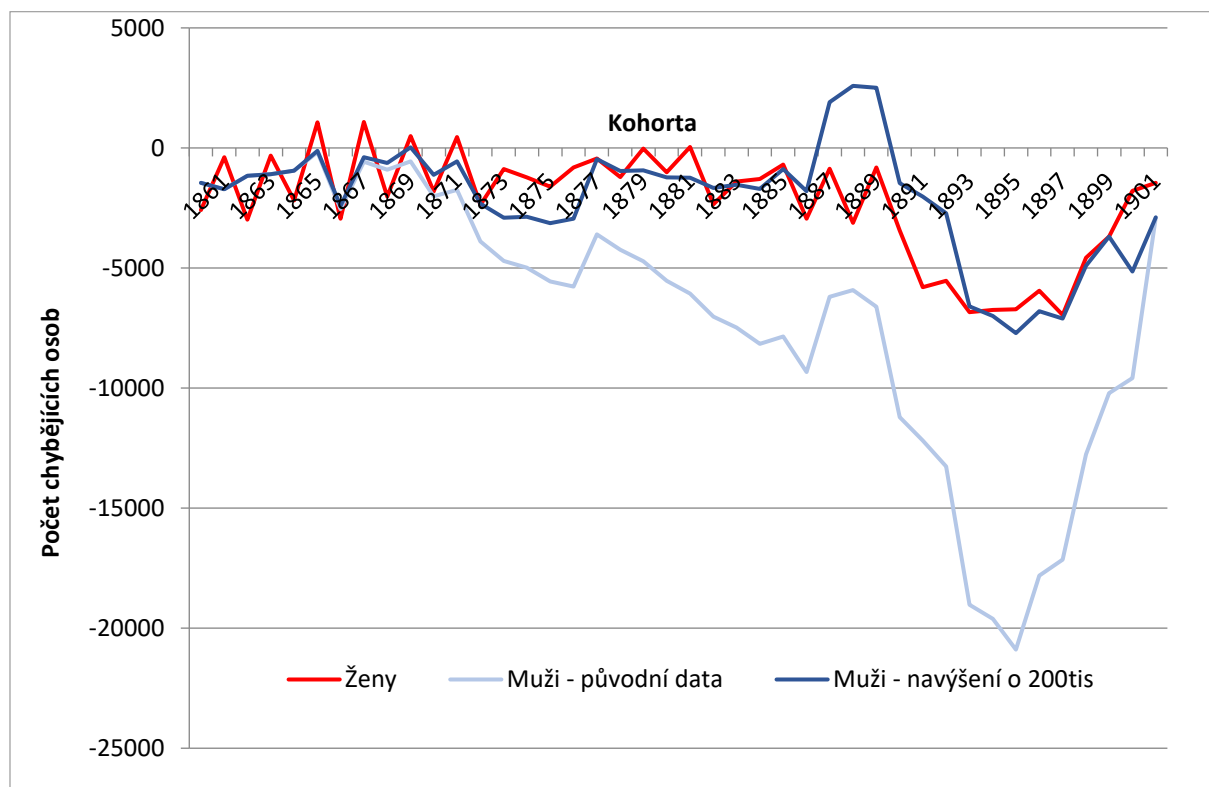
Na Obr. 5 je znázorněn deficit či přebytek jednotlivých kohort, zjištěný v roce 1920, proti evidovaným událostem v období 1910–1920. Je zde patrné, že ženy mají deficit zejména v kohortách 1891–1898, kdy byly deficity více než 5 tisíc osob. To koresponduje se závěry Fialové (1989), která analyzovala migraci v období 1870–1910, jež byla nejvíce deficitní právě v mladých věkových skupinách, často po dokončení odborného vzdělání. Mladí lidé měli problém najít zaměstnání, a tak odcházeli do zahraničí a většina emigrantů opouštěla zemi jako svobodní a sňatky uzavírali až později (Fialová, 1989).

Pro muže by měl tento předpoklad platit také, ale při pohledu na Obr. 5 můžeme vidět, že nerevidované počty událostí (nezahrnutí zemřelých ve válce) způsobují výrazné deficity. Největší deficit dosahují kohorty 1880–1900. Deficit činí v těchto případech více než 5 tisíc mužů, u kohort 1890–1900 dokonce více než 10 tisíc a pro kohorty 1893–1897 dokonce více než 15 tisíc. Nejvyššího salda dosahuje ročník mužů narozených v roce 1895. Počet těchto mužů byl 31. 12. 1910 105 658 a k 31. 12. 1920 pouze 78 410. Rozdíl tedy činil 27 248 osob. V období 1911–1920 zemřelo z této kohorty, dle záznamů v civilních matrikách, 6 357 osob a zbývá tedy deficit 20 891 osob, které tvoří z části počet zemřelých neevidovaných v civilních matrikách a dále saldo migrace z daného období pro kohortu 1895. Tato

¹⁵ K mírnému rozdílu proti publikaci Mazoucha a Hulíkové Tesárkové (2018, s. 28) došlo díky zpřesnění odhadu koncového stavu 1920.

kohorta je jednou z nejvíce zasažených kohort i z hlediska počtu padlých, jak rozdělení odhadoval Urlanis (1969), který předpokládal podíl kohorty 1895 na celkovém počtu padlých více než 6 % (Mazouch, Hulíková Tesárková, 2018, s. 28).

Obrázek 5: Deficit/přebytek počtu osob mezi lety 1910 a 1920, kohorty 1860-1901



Zdroj: vlastní výpočty autora

Při zahrnutí odhadovaného počtu nezaznamenaných událostí na úrovni 200 tisíc mužů, dojde k výrazné korekci salda, jak je patrné i z Obr. 5. Křivka kohort mužů pak přibližně kopíruje průběh křivky kohort žen, pouze s výjimkou ročníků 1887, 1888 a 1889, které vykazují migrační přebytek (tento efekt je vidět již na křivce s nerevidovanými počty událostí). Důvodem je skutečnost, že právě tyto ročníky byly v období před začátkem 1. světové války ve věku povinné vojenské služby. Ta byla v Rakousku-Uhersku stanovena na 21 let (pokud nebyl muž odveden, šel k odvodu ještě ve věku 22 a případně i 23 let) a část odvedených mužů pak absolvovala 3letou prezenční službu (zkrácenou od roku 1912 na 2 roky), ostatní prošli jen kratším výcvikem (Sigumbelli1914, 2020a). Muži v prezenční vojenské službě neměli možnost emigrovat a po vypuknutí 1. světové války byli rovnou nasazeni do bojů.

Posledním parametrem při rozdělení neevidovaných zemřelých v civilních matrikách, který bylo nutné upravit, byla distribuce mezi jednotlivé kalendářní roky 1. světové války. V publikaci Mazoucha a

Hulíkové Tesárkové (2018, s. 28) bylo využito rovnoměrného rozdělení s podíly 0,4-1-1-1-0,9¹⁶ pro jednotlivé roky 1914 až 1918 a podobný přístup zvolil i Netušil (1923). Rovnoměrné rozdělení však nerespektuje vývoj války a válečných ztrát, které byly (s výše uvedenými limitacemi) evidovány.

Tabulka 2: Rozdělení počtů neevidovaných zemřelých mužů během 1. světové války

Kohorta	Kalendářní rok				
	1914	1915	1916	1917	1918
1900	0	0	0	0	4 435
1899	0	0	0	3 725	2 794
1898	0	0	2 456	3 089	2 317
1897	0	1 312	2 726	3 429	2 572
1896	0	3 675	2 292	2 882	2 162
1895	3 298	3 298	2 057	2 587	1 940
1894	3 157	3 157	1 968	2 476	1 857
1893	3 110	3 110	1 939	2 439	1 829
1892	2 639	2 639	1 645	2 069	1 552
1891	2 544	2 544	1 586	1 996	1 497
1890	2 441	2 441	1 522	1 914	1 436
1889	2 282	2 282	1 423	1 790	1 343
1888	2 130	2 130	1 328	1 670	1 253
1887	2 026	2 026	1 263	1 589	1 192
1886	1 885	1 885	1 175	1 478	1 109
1885	1 745	1 745	1 088	1 369	1 027
1884	1 617	1 617	1 008	1 268	951
1883	1 489	1 489	928	1 168	876
1882	1 340	1 340	836	1 051	788
1881	1 210	1 210	754	949	712
1880	1 080	1 080	673	847	635
1879	950	950	592	745	559
1878	820	820	511	643	482
1877	784	784	489	615	461
1876	707	707	441	554	416
1875	607	607	378	476	357
1874	530	530	330	415	312
1873	452	452	282	355	266
1872	396	396	247	310	233
1871	300	300	187	235	176
1870	223	223	139	175	131
1869	147	147	92	115	86
1868	71	71	44	55	42
1867	47	47	29	37	28
1866	33	33	21	26	19
1865	14	14	9	11	8

Zdroj: Vlastní výpočty autora

¹⁶ Rok 1914 a 1918 zde mají menší podíl, protože válka netrvala po celou dobu roku.

Jak již bylo zmíněno výše, zemřelí nemuseli přímo padnout na bitevním poli, či krátce po bitvě, mohlo dojít ke zranění a voják mohl zemřít až za určitou dobu ve zdravotnickém zařízení. Úplnou neznámou jsou osudy osob nezvěstných či části zajatých, kteří se nikdy ze zajetí nevrátili. Informace o okamžiku jejich úmrtí, pokud zemřeli, chybí zcela. Určitým vodítkem pro rozdělení během trvání války může být popis průběhu bojů v jednotlivých letech, který je uveden například na stránkách spolku *Signumbelli*1914 (2020b).

Jindra (2014) považuje počáteční roky války 1914 a 1915 za velmi náročné z hlediska obnovy vojska, protože ztráty byly velmi vysoké. Počet padlých v Rakousko-Uherské armádě se odhaduje za tyto dva roky na 370 tisíc vojáků a dalších téměř 1 milion zraněných (hrubý odhad podílu Českých zemí tvořil na těchto počtech přibližně 20 %), k tomu je nutné počítat s nemalými počty nemocných a zajatých či nezvěstných (*Signumbelli*1914, 2020b). Následující roky již nebyly tak významné počtem padlých, ale i v roce 1916 zůstal vysoký počet zraněných a v roce 1917 výrazně rostl také počet nemocných (*Signumbelli*1914, 2020b). Na základě těchto dat pak bylo odhadnuto rozdělení zemřelých mezi kalendářní roky 1. světové války v následujícím poměru: 1914 20 %, 1915 23 %, 1916 16 %, 1917 22 % a rok 1918 19 %. Rozdělení 200 000 neevidovaných zemřelých mužů mezi jednotlivé kohorty a roky války je uvedeno v Tab. 2.

5.6 Odhad rozdělení migrace a neregistrovaných událostí v období 1911–1920

V případě odhadu intercenzálních koncových stavů je dle metodiky popsané v kap. 3.5 výše uveden předpoklad o rovnoměrném rozdělení Δ^z (diference vzniklá díky migraci, neregistrovaných událostech a případnými chybami při sčítání lidu pro kohortu z) ve vzorcích (31) a (35). Již v předchozí podkapitole 5.5 byl tento předpoklad popřen, když byly počty zemřelých během 1. světové války, nezaznamenání v civilních matrikách, rozdělení do jednotlivých věkových skupin jinak než rovnoměrně, čímž se částečně korigoval výrazný rozdíl mezi počtem osob zjištěných při sčítání v roce 1921 a odhadovaným počtem osob na základě bilanční metody.

I po této korekci však stále přetrvávají rozdíly mezi zjištěným a odhadovaným počtem (korekce v předchozí podkapitole 5.5 se týkala pouze mužů), a proto je nutné tuto nesrovnalost vyřešit, aby byla data srovnatelná s výstupy oficiální statistiky. Vzniklý rozdíl je v tomto případě tvořen zejména migrací, která nebyla evidovaná a v předchozích obdobích bylo rozdělení takového rozdílu odhadováno jako rovnoměrné po celou dobu intercenzálního období (viz kap. 3.5). Pro období 1. světové války je tento předpoklad problematický, protože za standardní roky lze z hlediska migrace považovat pouze období 1911–1914 (s tím, že rok 1914 je do tohoto období možné započítat jen jeho první polovinou, před vypuknutím války), u kterých není důvod předpokládat změnu trendu proti předchozímu období.

Během války byla přeshraniční migrace spíše výjimečná a k další vlně migrace docházelo až po skončení války.

Pro úvodní část období, trvající 3,5 roku, byla uvažována intenzita migrace na úrovni předcházejícího období (1901–1910) a pro tento účel byl upraven vzorec (31). Jádrem vzorce, tedy odčítání elementárních souborů zemřelých od posledního známého koncového stavu, zůstává nezměněno. Pouze velikost migrace a neevidovaných událostí je odhadnuta na základě koeficientu, který se vypočítá jako průměr z posledních tří let před sčítáním lidu (1908–1910) pro každý věk zvlášť a vyjádří se následně jako počet osob. Odhad pak pro kohortu z vypadá následovně:

$${}_{1911}P_{x+1}^z = {}_{1910}C_x^z - {}_{1911}D_x^z - {}_{1911}D_{x+1}^z + \frac{\frac{{}_{1900-1910}\Delta^{z+3}}{{}_{1908}P_x^{z+3}} + \frac{{}_{1900-1910}\Delta^{z+2}}{{}_{1909}P_x^{z+2}} + \frac{{}_{1900-1910}\Delta^{z+1}}{{}_{1910}P_x^{z+1}}}{3} \cdot {}_{1910}P_x^z, \quad (46)$$

$${}_{1912}P_{x+2}^z = {}_{1911}P_{x+1}^z - {}_{1912}D_{x+1}^z - {}_{1912}D_{x+2}^z + \frac{\frac{{}_{1900-1910}\Delta^{z+4}}{{}_{1908}P_x^{z+4}} + \frac{{}_{1900-1910}\Delta^{z+3}}{{}_{1909}P_x^{z+3}} + \frac{{}_{1900-1910}\Delta^{z+2}}{{}_{1910}P_x^{z+2}}}{3} \cdot {}_{1911}P_{x+1}^z, \quad (47)$$

$${}_{1913}P_{x+3}^z = {}_{1912}P_{x+2}^z - {}_{1913}D_{x+2}^z - {}_{1913}D_{x+3}^z + \frac{\frac{{}_{1900-1910}\Delta^{z+5}}{{}_{1908}P_x^{z+5}} + \frac{{}_{1900-1910}\Delta^{z+4}}{{}_{1909}P_x^{z+4}} + \frac{{}_{1900-1910}\Delta^{z+3}}{{}_{1910}P_x^{z+3}}}{3} \cdot {}_{1912}P_{x+2}^z, \quad (48)$$

$${}_{1914}P_{x+4}^z = {}_{1913}P_{x+3}^z - {}_{1914}D_{x+3}^z - {}_{1914}D_{x+4}^z + \frac{\frac{{}_{1900-1910}\Delta^{z+6}}{{}_{1908}P_x^{z+6}} + \frac{{}_{1900-1910}\Delta^{z+5}}{{}_{1909}P_x^{z+5}} + \frac{{}_{1900-1910}\Delta^{z+4}}{{}_{1910}P_x^{z+4}}}{3} \cdot {}_{1913}P_{x+3}^z \cdot 0,5, \quad (49)$$

kde ${}_{1900-1910}\Delta^z$ je difference způsobená migrací a chybami v evidenci počtu událostí pro kohortu z, zjištěná mezi sčítáními 1900 a 1910.

Tímto způsobem byla rozdělena pouze část celkové difference mezi sčítáními 1910 a 1920 ${}_{1910-1920}\Delta^z$ (v tomto případě je to přepočítaný koncový stav roku 1920 ze sčítání lidu s rozhodným okamžikem 15. 2. 1921). Zbývající část difference ${}_{1918-1920}\Delta^z$ (připadající na poslední tři roky sledovaného období) je odhadnuta na základě následujícího vztahu:

$$\begin{aligned}
{}_{1918-1920}\Delta^z = {}_{1910-1920}\Delta^z & - \left(\frac{\frac{1900-1910\Delta^{z+3}}{1908P_x^{z+3}} + \frac{1900-1910\Delta^{z+2}}{1909P_x^{z+2}} + \frac{1900-1910\Delta^{z+1}}{1910P_x^{z+1}}}{3} \cdot {}_{1910}P_x^z \right. \\
& + \frac{\frac{1900-1910\Delta^{z+4}}{1908P_x^{z+4}} + \frac{1900-1910\Delta^{z+3}}{1909P_x^{z+3}} + \frac{1900-1910\Delta^{z+2}}{1910P_x^{z+2}}}{3} \cdot {}_{1911}P_{x+1}^z \\
& + \frac{\frac{1900-1910\Delta^{z+5}}{1908P_x^{z+5}} + \frac{1900-1910\Delta^{z+4}}{1909P_x^{z+4}} + \frac{1900-1910\Delta^{z+3}}{1910P_x^{z+3}}}{3} \cdot {}_{1912}P_{x+2}^z \\
& \left. + \frac{\frac{1900-1910\Delta^{z+6}}{1908P_x^{z+6}} + \frac{1900-1910\Delta^{z+5}}{1909P_x^{z+5}} + \frac{1900-1910\Delta^{z+4}}{1910P_x^{z+4}}}{3} \cdot {}_{1913}P_{x+3}^z \cdot 0,5 \right). \quad (50)
\end{aligned}$$

${}_{1918-1920}\Delta^z$ byla rozpočítána mezi roky 1918, 1919 dle předpokladu rovnoměrného rozdělení:

$${}_{1918}P_{x+1}^z = {}_{1917}P_x^z - {}_{1917}D_x^z - {}_{1917}D_{x+1}^z + \frac{1}{13} \cdot {}_{1918-1920}\Delta^z, \quad (51)$$

$${}_{1919}P_{x+2}^z = {}_{1918}P_{x+1}^z - {}_{1918}D_{x+1}^z - {}_{1918}D_{x+2}^z + \frac{7}{13} \cdot {}_{1918-1920}\Delta^z. \quad (52)$$

Odhadovaný koncový stav roku 1920 je v souladu s předchozími vztahy – zejména vzorcem (32) – roven stavu zjištěnému při druhém sčítání lidu (v tomto případě přepočítaným koncovým stavem roku 1920 ze sčítání lidu s rozhodným okamžikem 15. 2. 1921) a tím pádem na rok 1920 připadá zbytek nezjištěné migrace a nevidovaných událostí (konkrétně $6/13$ z ${}_{1918-1920}\Delta^z$).

Pro podporu odhadu rozdělení migrace v období 1911–1920 neexistuje příliš mnoho pramenů, o které by se bylo možné opřít. Důvodem je již několikrát zmíněná 1. světová válka, pro jejíž období je horší kvalita i jiných ukazatelů (viz text o kvalitě rekonstruovaných dat v publikaci ČSÚ, 2005, s. 1), a pak fakt, že údaje z období těsně před válkou byly ještě zpracovávány rakousko-uherskou statistikou, ale opět v období trvání války, protože zpracovaný rok 1912 vyšel v roce 1915 a rok 1913 vyšel až v roce 1918 (Mazouch, Hulíková Tesárková, 2018, s. 75). Poválečné období je spojeno s rychlým vznikem Státního úřadu statistického, který však také nebyl snadný (viz poznámku v publikaci ČSÚ, 2015, s. 24) a první kompaktní pohled na dopady 1. světové války (a pohybu obyvatelstva krátce po ní) dává až sčítání lidu v roce 1921.

Hodnocení migrace v období 1. světové války a krátce po ní se ve svém díle „1000 let obyvatelstva českých zemí“ vyhýbá také Srb (Srb, 2005, s. 210). Popisuje pouze odhady migrace do roku 1914, které

ukazují, že předpoklad o zachování trendu v letech do vypuknutí 1. světové války, jsou správné. Následující období, které Srb popisuje, je až od roku 1920 dále (Srb, 2005, s. 211). Velmi opatrně a spíše obecně popisuje migraci ve své stati o obyvatelstvu Českých zemí za 1. světové války i Fialová (2014) a v již několikrát citované publikaci ČSÚ (2005) se analýza migrace nevyskytuje vůbec.

Celkové hodnoty $_{1910-1920}\Delta$ dosahují přibližně úrovně 1,3 % celkové populace v roce 1920, což je velmi nízká úroveň podílu migrace a nevidovaných událostí (která se navíc rozdělí do více let). Faktem však zůstává, že nejsou rovnoměrně zasaženy všechny generace. Nejvyšší rozdíl je u generace mužů narozených v roce 1895 (generace, která měla také nejvyšší počet zemřelých v boji), kde je difference na úrovni 10 % a u žen generace 1892 (rozdíl přibližně 8 %). Je zde však nutné pamatovat na to, že vzniklý rozdíl je distribuován do více kalendářních let a jeho dopad se tak rozloží. Dopad výše popsaného rozdělení ve vzorcích (46)–(52) vede k situaci, kdy pro muže většina $_{1910-1920}\Delta$ připadá na období 1911–1914 (přibližně 60 %) a pro ženy je tento podíl nižší, pouze přibližně 40 %. Zbývající části jsou pak rozděleny mezi roky 1918–1920.

5.7 Rekonstrukce koncových stavů v období 1911–1919

Po rekonstrukci koncového stavu roku 1920, který byl odhadnut z výsledků sčítání lidu 15. 2. 1921, vznikly podmínky, kdy bylo možné aplikovat postup popsaný vzorci (29)–(35) z kapitoly 3.5, protože sčítání lidu 1910 proběhlo k rozhodnému okamžiku 31. 12. Nejprve však bylo nutné doplnit a upravit případné údaje o událostech:

1. Rozdělit zemřelé do DES a HES ve věkové skupině 70 a více let v roce 1913.
2. Odhadnout počet zemřelých nezaznamenaných v civilních matrikách v období 1. světové války.
3. Rozdělit zemřelé v otevřených věkových intervalech 90+ let v období 1914–1918.
4. Odhadnout rozdělení migrace a neregistrovaných událostí v období 1911–1920.

Až po všech těchto krocích bylo možné rekonstruovat koncové stavy v období 1911–1919. Je jasné, že toto období vyžadovalo, v porovnání s předchozími obdobími, nejvíce zásahů, protože data nebyla v optimálním stavu.

6 Doplněk – data po roce 1920

Zrekonstruovaná data v období 1870–1919 plynule navazují (s drobnou výjimkou středního stavu roku 1920) na již publikovaná data od roku 1920 do současnosti. Vzniká tak přibližně 150 let dlouhá časová řada, během níž lze analyzovat jak vývoj transversálních ukazatelů, tak vzniká prostor pro doposud velmi okrajově zpracovanou kohortní analýzu. Pro doplnění popisu datových zdrojů je vhodné krátce zmínit i datové zdroje po roce 1920 (podrobněji se datovým zdrojům věnuje publikace Mazoucha a Hulíkové Tesárkové, 2018). Součástí publikovaných datových sad je řada pravděpodobností úmrtí v kohortním uspořádání, kde bylo nutné vyřešit i období kolem 2. světové války.

6.1 Počet událostí

Od roku 1919 (od vzniku Státního úřadu statistického) začíná nepřerušovaná časová řada publikací „*Pohyb obyvatelstva*“, která obsahuje podrobné údaje o jednotlivých demografických procesech v příslušném roce. V období 1919–1937 lze označit časovou řadu publikací za naprosto srovnatelnou, jak z hlediska územního pokrytí (tehdejší území se shoduje s územím současné České republiky), tak i z hlediska publikovaných ukazatelů. Následující období 1938–1944 je poznamenáno Mnichovskou dohodou a publikace obsahují data pouze za území tehdejšího protektorátu Čechy a Morava a navíc bez německých státních příslušníků (Mazouch, Hulíková Tesárková, 2018, s. 25). Od roku 1945 (který se vyznačuje velmi problematickým zpracováním dat za některé části území) již opět publikace obsahují celé území shodující se se současnou Českou republikou. Od roku 2006 (data za rok 2005) se publikace jmenuje „*Demografická ročenka České republiky*“ (Mazouch, Hulíková Tesárková, 2018, s. 23–26).

V celém období jsou události tříděny do elementárních souborů a od roku 1957 již není žádná osoba evidována v kategorii „věk neznámý“ (Mazouch, Hulíková Tesárková, 2018, s. 23–26). V období 2. světové války lze opět předpokládat zvýšený počet událostí, které nejsou evidovány v civilních matrikách, jako například zemřelí v Terezínském ghettu nebo oběti válečných zločinů (Mazouch, Hulíková Tesárková, 2018, s. 25).

6.2 Stav populace

Také pro stav populace existuje nepřetržitá časová řada, která, podobně jako časová řada událostí, obsahuje určitá období nekonzistentní se zbytkem řady. V pramenných dílech „*Pohyb obyvatelstva*“ vychází informace o stavu populace (s výjimkou výsledků sčítání lidu, které jsou také informací o stavu a struktuře populace) až od roku 1982 (Mazouch, Hulíková Tesárková, 2018, s. 36). Do té doby je zdrojem dat sada poskytnutá pracovníky Českého statistického úřadu, jež je součástí elektronických příloh publikace Mazoucha a Hulíkové Tesárkové (2018).

Střední stavy populace jsou k dispozici již od roku 1920 za každý kalendářní rok, opět s nekonzistencí v období 1938–1944, kdy jsou data pouze za území Protektorátu Čechy a Morava. Od roku 1945 se opět údaje územně shodují se současným územím České republiky a od tohoto roku jsou k dispozici, vedle středních stavů, také koncové stavy obyvatelstva pro každý kalendářní rok (Mazouch, Hulíková Tesárková, 2018, s. 36–39).

6.3 Využití dat pro transversální a kohortní analýzu

Publikované časové řady událostí a středních stavů od roku 1920 umožňují konstruovat transversální úmrtnostní tabulky pro celé sledované období. S určitou mírou opatrnosti při interpretaci výsledků i pro 2. světovou válku, kde je počet zaznamenaných událostí spíše podhodnocen a stav populace je odhadován zpětnou projekcí ze stavu roku 1945 bez migrace (Mazouch, Hulíková Tesárková, 2018, s. 38). Úmrtnostní tabulky pro každý kalendářní rok od roku 1920 jsou zkonstruovány Českým statistickým úřadem a publikovány na webových stránkách, včetně metodiky použité při konstrukci úmrtnostních tabulek (ČSÚ, 2020).

Proti tomu využití stejných časových řad událostí a stavů (v podobě středních stavů) je problematické pro kohortní analýzu. Nejspíše nekonzistence dat z hlediska území v období 2. světové války způsobila, že není zrekonstruována časová řada koncových stavů populace pro období před rokem 1945. Před rokem 1938 by bylo možné využít jednoduchého vztahu, popsaného vzorcem (45), kterého využívá i Mazouch a Hulíková Tesárková (2018, s. 40) a v letech 2. světové války využít stejného postupu jako v případě středních stavů. Vystává však problém s odhadem základního ukazatele pro konstrukci úmrtnostní tabulky, což je pravděpodobnost úmrtí (nebo míra úmrtnosti).

Vzorec (10) ukazuje, jak se pro pravděpodobnost úmrtí v 1. hlavním souboru událostí (kohortní perspektiva) kombinují počty zemřelých ze dvou kalendářních let, které se následně vztahují k počtu osob dožívajících se přesného věku (či ke koncovému počtu osob v daném věkovém intervalu v případě míry úmrtnosti). To v případě nekonzistentních let může způsobit nadhodnocení nebo podhodnocení pravděpodobnosti úmrtí (či míry úmrtnosti). Z tohoto důvodu a také z důvodu nespolehlivosti dat z období 2. světové války bylo využito jiné metody pro odhad vývoje úmrtnosti během 2. světové války.

6.4 Odhad úrovně úmrtnosti během 2. světové války

Byť období 2. světové války trvalo 6 let, v oficiální statistice je vliv patrný již od roku 1938 (viz předmluvu k Pohybu obyvatelstva 1938 (SÚS, 1944)) do roku 1945. Dopad na kohortní analýzu lze vidět na Obr. 8 v publikaci Mazoucha a Hulíkové Tesárkové (2018, s. 48). Je zde patrné, že byť jsou data oficiální statistiky ovlivněna až v roce 1938, nelze pro vzorec (10) použít již data dřívější. Pro období před rokem 1945 nebyly publikovány koncové stavy obyvatelstva, jak již bylo uvedeno dříve, a koncové stavy byly dopočítány na základě vzorce (45) ze středních stavů obyvatelstva. Pro rok 1938 byl však

střední stav dopočítán zpětnou projekcí z roku 1945, a proto nenavazoval na rok předcházející (Mazouch, Hulíková Tesárková, 2018, s. 38). Aplikace vzorce (45) pro výpočet koncového stavu roku 1937 nebyla možná, a tak první koncový stav, který je koherentní s předchozí časovou řadou, je pro rok 1936 (spočten jako průměr středních stavů roku 1936 a 1937). Pro rok 1945 již koncový stav publikován byl, ale pro výpočet pravděpodobnosti úmrtí dle vzorce (10) by došlo opět ke kombinaci nesrovnatelných období, protože by se sčítal DES 1945, do kterého ještě stále zasahovala nemalou částí válka, a HES 1946. Proto pro období po skončení 2. světové války byla použita první data o událostech až za dvouletí 1946 (DES) a 1947 (HES), která byla vztažena ke koncovému stavu 1946 (viz obr. 8 v publikaci Mazoucha a Hulíkové Tesárkové, 2018, s. 48). V každé z kohort tak vznikla mezera o délce 9 let věku, kterou bylo nutné nějakým způsobem doplnit.

Interpolace použitá v knize Mazoucha a Hulíkové Tesárkové (2018, s. 48–52) je založena na dvou základních krocích. V prvním kroku byl interpolován vývoj úmrtnosti pro každou generaci ve formě hladké exponenciální funkce, která propojovala poslední známé hodnoty pravděpodobnosti úmrtí před válkou a po jejím skončení. Pro eliminaci výkyvů pravděpodobnosti úmrtí díky různým vlivům byly poslední (a po válce první) známé hodnoty pravděpodobností úmrtí spočteny jako vážený průměr posledních (resp. prvních) tří hodnot (podrobněji viz Mazouch, Hulíková Tesárková, 2018, s. 49). Tento postup simuloval situaci, kdy by v období války nebyl žádný růst úmrtnosti a ta se vyvíjela pouze přirozeně.

Druhým krokem bylo zohlednění války a jejího vlivu na úroveň úmrtnosti. To proběhlo prostřednictvím koeficientů navyšujících „přirozenou“ úroveň úmrtnosti. Jednotlivé koeficienty byly odhadnuty na základě výsledků analýz předchozích autorů a zohledňovaly jak nárůst úmrtnosti z důvodu zhoršení životních podmínek během války (zhoršení přístupu k potravinám a zdravotní péči, úmrtí na vnější příčiny), tak i události přímo spojené s válečným konfliktem (vraždy, neevidované události v civilních matrikách). U první skupiny událostí je výraznější nárůst patrný až na konci války, v letech 1944 a 1945, druhá skupina je hůře evidovaná, a tak jsou koeficienty odhadnuty jako souhrnné a jejich výše je uvedena v publikaci Mazoucha a Hulíkové Tesárkové (2018, s. 51), podobně jako podrobnější popis postupu odhadů koeficientů.

Vzhledem k tomu, že všechny koeficienty dosahují hodnoty větší než jedna (dochází k nárůstu úrovně úmrtnosti), je otázkou dopad na střední délku života jednotlivých kohort, které prožily 2. světovou válku. Kvantifikaci těchto dopadů lze nalézt v publikaci Mazoucha a Hulíkové Tesárkové (2019), která využívá vztahů tabulkových funkcí a odhaduje dopad dočasného zvýšení úmrtnosti na střední délku života. Obdobný přístup lze nalézt také u dalších autorů. Vaupel (1986) zkoušel odhadnout dopad procentní redukce v počtu zemřelých na střední délku života v různých věkových kategoriích (se

zjištěním, že největší dopad je v období kolem normální délky života – modu počtu zemřelých). Pollard (1988) zkoumal efekt změny úrovně úmrtnosti v konkrétní věkové skupině na střední délku života a velmi podobně Wryca a Baudisch (2012) podrobně popsali celkový vliv síly závislosti mezi změnou úmrtnosti a změnou střední délky života při narození.

Formálně lze relativní změnu úmrtnosti vyjádřit jako (Mazouch, Hulíková Tesárková, 2019):

$$A = \frac{p_x^*}{p_x} \Rightarrow p_x^* = A \times p_x, \quad (53)$$

kde p_x je pravděpodobnost přežití z přesného věku x do přesného věku $x + 1$, a A je relativní změna v pravděpodobnosti přežití, tedy p_x^* je změněná (navýšená nebo snižená) pravděpodobnost přežití a p_x je původní (nezměněná) hodnota.

Další podmínky pro aplikaci postupu založeného na využití tabulkových funkcí jsou následující (Mazouch, Hulíková Tesárková, 2019):

1. $l_0 = 1$, což je kořen úmrtnostní tabulky
2. $a > 0$, kde a reprezentuje věk (nebo věkovou skupinu) zasaženou změnou úmrtnosti (vzorec nelze aplikovat na situaci, kdy je změnou úmrtnosti zasažena nejmladší věková skupina v dokončeném věku 0),
3. distribuce událostí ve věkovém intervalu je rovnoměrná.

V případě 2. světové války je pro každou kohortu, která 2. světovou válku přežila, nutné rekonstruovat pravděpodobnost úmrtí pro 9 jednoletých věkových intervalů. Rozdíl mezi původní (nezměněnou) střední délkou života narozeného a střední délkou života, zahrnující navýšení úrovně úmrtnosti ($e_0 - e_0^*$), pak lze vyjádřit jako (upravený vztah dle Mazouch, Hulíková Tesárková, 2019):

$$\begin{aligned} e_0 - e_0^* &= l_{a+9} \times (1 - A_a \times A_{a+1} \dots A_{a+7} \times A_{a+8}) \times e_{a+9} \\ &\quad + (p_a \times (1 - A_a) + p_a \times p_{a+1} \times (1 - A_a \times A_{a+1}) + \dots \\ &\quad + p_a \times p_{a+1} \dots p_{a+7} \times (1 - A_a \times A_{a+1} \dots A_{a+7}) \\ &\quad + p_a \times p_{a+1} \dots p_{a+7} \times p_{a+8} (1 - A_a \times A_{a+1} \dots A_{a+7} \times A_{a+8}) \times 0,5) \times l_a \\ &= (e_{a+9} + 0,5) \times l_{a+9} \times \left(1 - \prod_{i=a}^{a+8} A_i\right) \\ &\quad + \sum_{i=a+1}^{a+8} \left(l_i \times \left(1 - \prod_{s=a+1}^i A_{s-1}\right) \right), \end{aligned} \quad (54)$$

kde a je první věk, který je zasažen zvýšením úrovně úmrtnosti, p_x je pravděpodobnost přežití z přesného věku x do přesného věku $x + 1$, a A_x je relativní změna v pravděpodobnosti přežití pro věkovou skupinu x .

Závěrečný tvar vztahu popsaného vzorcem (54) je složen ze dvou sčítanců. První z nich reprezentuje snížení počtu osob, které přežily období vyšší úrovně úmrtnosti a tím snížily celkový počet let prožitých danou kohortou v následujících letech a druhá část sčítance ukazuje dopad na počet prožitých let přímo v období se zvýšenou úrovní úmrtnosti. Arriaga (1984) označuje tyto efekty jako nepřímý a přímý efekt a využívá to k vyjádření rozdílu mezi úrovní úmrtnosti dvou populací (Arriaga, 1984). V publikaci Mazoucha a Hulíkové Tesárkové (2019) lze nalézt i další možnosti, které uvažují dočasnou změnu úmrtnosti například jen v jednom nebo dvou letech.

V případě vyjádření dopadu navýšení úmrtnosti v období 2. světové války, je vidět, že rozložení navýšení do celého věkového spektra vedlo k dopadům, které se v kohortní střední délce života narozeného pohybují na velmi nízkých úrovních, v případě mužů přibližně $-0,5$ roku a u žen dokonce pouze $-0,3$ roku života (Mazouch, Hulíková Tesárková, 2019).

7 Struktura publikovaných dat

Datové sady, které vznikly na základě postupů popsanych v předchozích kapitolách, jsou publikovány na webu mortality.vse.cz. Součástí webu je také publikace Mazoucha a Hulíkové Tesárkové (2018), která obsahuje jako elektronické přílohy původní data pocházející z pramenných děl, případně dalších zdrojů, které jsou podkladem pro vznik rekonstruovaných řad, vymezených v kap. 2. Následující podkapitoly obsahují stručný popis datových sad, včetně vymezení zahrnutých kalendářních let a věkových skupin, případně další úpravy, které ještě bylo nutné realizovat, aby data bylo možné publikovat v jednotném formátu.

Na zmíněném webu mortality.vse.cz jsou publikována nejen nově rekonstruovaná data, ale také návazné již publikované datové řady (začínající nejčastěji v roce 1920), čímž vzniká datový podklad, obsahující materiál pro analýzy od roku 1870 do současnosti. Publikované základní datové sady jsou následující:

1. Zemřelí v členění na elementární soubory DES a HES od roku 1870
2. Koncové stavy obyvatel v období 1870–1937 a 1945 do současnosti
3. Střední stavy obyvatel od roku 1870
4. Pravděpodobnosti úmrtí pro kohorty narozené v letech 1820–1922

7.1 Zemřelí v členění na elementární soubory DES a HES od roku 1870

Počty událostí jsou tříděny do elementárních souborů pro všechny kalendářní roky od roku 1870. Proti zdrojovým datům jsou na DES a HES rozděleny i události v letech 1870–1894 a pro období 1870–1918 jsou mezi ostatní věkové skupiny rozpočítány i počty událostí v neznámých věcích. Otevřené věkové intervaly byly sjednoceny na 99+ nebo 100+ let. Od roku 1919 jsou publikovaná data z původních pramenných děl (viz výčet Mazoucha a Hulíkové Tesárkové, 2018) bez úprav.

7.2 Koncové stavy obyvatel v období 1870–1937 a 1945 do současnosti

Stavy obyvatel k 31. 12. jednotlivých kalendářních let lze rozdělit do 4 období. V období 1870–1920 bylo využito jednotlivých sčítání lidu a koncové stavy pro intercenzální období byly dopočteny dle metodiky v kapitole 3, případně v kapitole 5. Pro období 1921–1936 bylo využito informace o středních stavech a vzorce (45) a koncové stavy byly vypočteny jako průměry ze dvou sousedních středních stavů (po rozdělení neznámých věků do ostatních věkových skupin). Pro období 1937–1944 nejsou koncové stavy obyvatel rekonstruovány a od roku 1945 do současnosti jsou data k dispozici z původních pramenných děl bez dalších úprav. Údaje za všechny sledované roky končí otevřeným věkovým intervalem 99+ nebo 100+ let.

7.3 Střední stavy obyvatelstva od roku 1870

Stavy obyvatel k 30. 6. za jednotlivé kalendářní roky jsou pro období 1870–1920 spočteny z dopočtených koncových stavů, jako průměry sousedních hodnot dle vzorce (7). Od roku 1921 jsou data k dispozici z původních pramenných děl bez dalších úprav. Všechny roky končí otevřeným věkovým intervalem 99+ nebo 100+ let. Pro období 1938–1944 jsou data spočtena zpětnou projekcí od roku 1945 (viz poznámku v příslušném datovém zdroji).

7.4 Pravděpodobnosti úmrtí pro kohorty narozené 1820–1922

Pramenná díla neobsahují ukazatel pravděpodobnosti úmrtí, který musel být pro jednotlivé kohorty dopočten na základě rekonstruovaných počtů událostí a koncových stavů. Třídění je zde z důvodu kohortní analýzy vždy za kombinaci jednoleté věkové kategorie, jednoleté kohorty a dvouleté období kalendářních let (viz Obr. 1 a vyznačení I. hlavních souborů událostí).

Pro jednotlivé věkové kategorie, s výjimkou nejnižšího dokončeného věku 0 let, bylo využito vztahu popsaného Wilmothem et al. (2017, s. 40):

$$q_x^z = \frac{{}_tD_x^z + {}_{t+1}D_x^z}{{}_tP_x^z + {}_{t+1}D_x^z}, \quad (55)$$

kde ${}_tD_x^z$ je počet zemřelých ve věku x , z kohorty z v roce t , ${}_{t+1}D_x^z$ je počet zemřelých ve věku x , z kohorty z v roce $t + 1$ a ${}_tP_x^z$ je koncový stav populace ve věku x , z kohorty z v roce t .

Nejnižší věková kategorie pak využívá analogického vztahu

$$q_0^z = \frac{{}_tD_0^z + {}_{t+1}D_0^z}{{}_tN^z}, \quad (56)$$

kde ${}_tD_0^z$ je počet zemřelých ve věku do 1 roku z generace z v roce t , ${}_{t+1}D_0^z$ je počet zemřelých ve věku do 1 roku z generace z v roce $t + 1$ a ${}_tN^z$ je počet narozených v roce t .

Pro nejvyšší věkové kategorie bylo z důvodu kvality datových zdrojů, kterou zmiňují také Mazouch a Hulíková Tesárková (2018, s. 45 a 46), využito jednoduchého nahrazení empirických hodnot interpolovanými hodnotami. Ty pro věky nad 95 let uvažují exponenciální růst až do věku 102, pro který je pravděpodobnost úmrtí uvažována rovna jedné (Mazouch, Hulíková Tesárková, 2018, s. 46). Výjimkou jsou kohorty, které dosáhly nejvyšších věků v období 2. světové války (během které tyto kohorty téměř vymřely). Zde není možné aplikovat postup popsaný v kap. 6.4, protože pro interpolaci nelze využít úroveň úmrtnosti po skončení války. Proto je zde využito posledních známých hodnot a analogicky využita interpolace do věku 102 let (kde je uvažována pravděpodobnost úmrtí rovna jedné), pouze počáteční věk pro interpolaci není 95 let, ale pro kohortu 1842 věk 94 a postupně až po kohortu 1848 věk 88 let.

Datová sada obsahuje údaje o pravděpodobnostech úmrtí pro kohorty od roku narození 1820, které však nemají kompletní data pro celé věkové spektrum (data byla rekonstruována od kalendářního roku 1870) a první generace začíná věkem 50 let (a stejně jako ostatní kohorty končí věkem 102). Následující generace mají vždy o jeden rok věku více a první kompletní kohortou jsou narození v roce 1870. Poslední kompletní kohortou jsou narození v roce 1922. Tento údaj však není konečný, protože s publikací nových dat bude každý rok docházet k prodloužení časové řady o další kohortu.

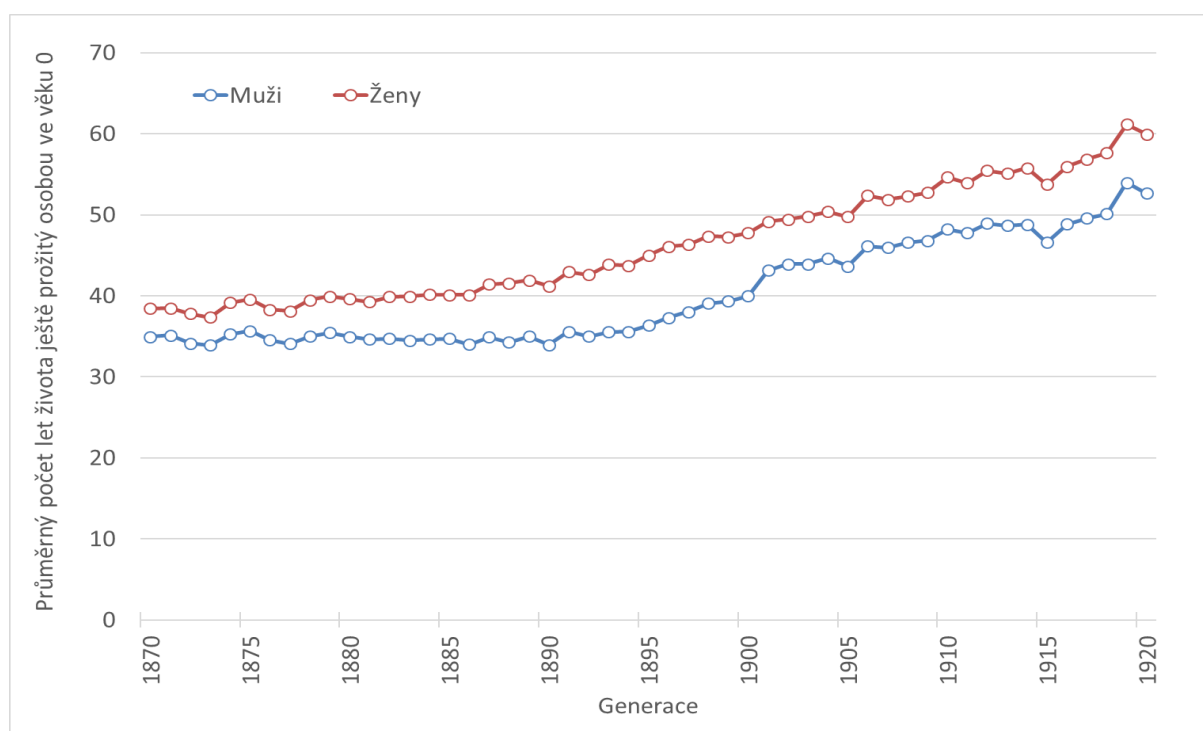
8 Využití rekonstruovaných dat

Využití zrekonstruovaných datových sad je z hlediska demografické analýzy velmi široké. Cílem této kapitoly je alespoň částečně ilustrovat základní výsledky, které již byly publikovány, či jsou ve fázi přípravy k publikaci a také porovnat výsledky s dalšími zdroji, aby se ověřila koherence dat. Základní výstupy v podobě kohortních úmrtnostních tabulek byly již součástí publikace Mazoucha a Hulíkové Tesárkové (2018), ale transversální úmrtnostní tabulky pro intercenzální období v letech 1870–1919 doposud nikde publikovány nebyly¹⁷.

8.1 Základní výsledky kohortní analýzy¹⁸

V rámci stručného popisu výsledků je na místě představit alespoň základní vývoj hodnot nejběžnějších ukazatelů. Mezi ně patří především průměrná délka života osob živě narozených v jednotlivých generacích (neboli střední/průměrný počet let života prožitý právě narozenou osobou ve studovaných generacích). Její vývoj je zobrazen v rámci Obr. 6.

Obrázek 6: Průměrný počet let života prožitý osobou v přesném věku 0 (při narození), generace 1870–1920, muži, ženy



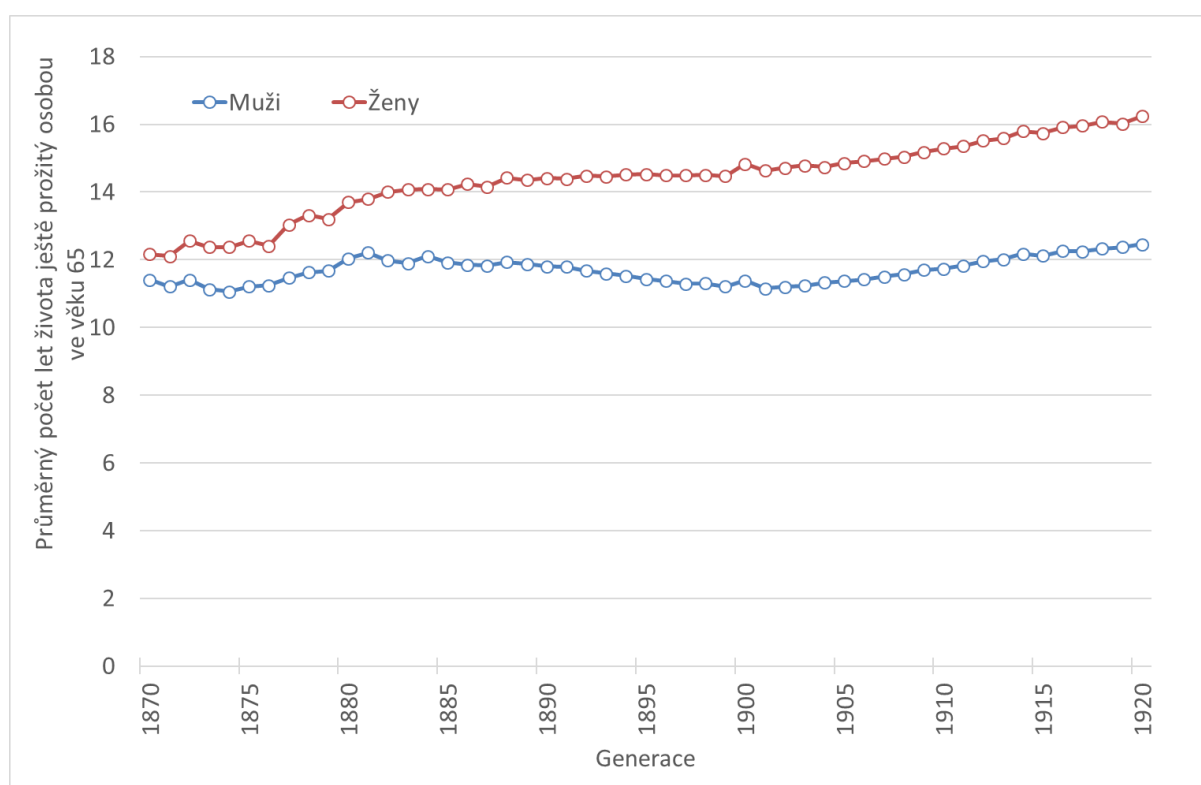
Zdroj: Mazouch, Hulíková Tesárková (2018, s. 59)

¹⁷ Na začátku roku 2021 je plánovaná publikace transversálních tabulek za období 1870–1910, autorského kolektivu Hulíková Tesárková, Mazouch a Fialová. Tabulky jsou založeny na rekonstruovaných datech popsanych v předchozích kapitolách.

¹⁸ Podkapitola je převzata z publikace Mazoucha a Hulíkové Tesárkové (2018)

Je zřejmé, že hodnoty průměrné délky života pro živě narozené osoby ve studovaných generacích byly od 70. let 19. století až do 90. let 19. století v podstatě stabilní, především v případě mužů. Pro muže se hodnota průměrné délky života pohybovala na úrovni okolo 35 let, pro ženy velmi mírně narůstala až k hodnotám lehce nad 40 let. Pro ženy narozené po roce 1890 a pro muže narozené po roce cca 1895 se však průměrná délka života začíná viditelně prodlužovat. Dodejme, že tyto a následující mladší generace prožily mladý věk v úmrtnostně relativně příznivém období po 1. světové válce a do vyššího věku se dostávaly v dalším příznivém období po 2. světové válce. Nárůst hodnot studovaného ukazatele je patrný i pro následující generace, ačkoli ty byly ve věku mladé dospělosti zasaženy 2. světovou válkou. Pro osoby narozené okolo roku 1920 je již předpokládána průměrná délka života nad 50 lety v případě mužů, a dokonce až nad 60 lety v případě žen. Lze předpokládat, že nárůst střední délky života pokračoval i pro následující generace, které již nebyly předmětem studia tohoto projektu.

Obrázek 7: Průměrný počet let života ještě prožitý osobou v přesném věku 65 let, generace 1870–1920, muži, ženy



Zdroj: Mazouch, Hulíková Tesárková (2018, s. 60)

Zajímavý je pohled na vývoj ukazatele střední (průměrné) délky života pro osobu v přesném věku 65 let (Obr. 7). V případě žen je viditelný nárůst pro všechny zkoumané generace. Zatímco 65letá žena narozená v roce 1870 měla před sebou ještě v průměru okolo 12 let života, stejně stará žena narozená v roce 1920 měla před sebou ještě v průměru více než 16 let k prožití. Nárůst hodnot tohoto ukazatele

byl nejrychlejší především pro ženy narozené v letech 1870 až 1880. Tyto ženy do věku 65 let dospěly v letech 1935–1945, tedy v době 2. světové války. Jak bylo uvedeno výše, dopad 2. světové války na civilní obyvatelstvo a na ženy nebyl radikálního charakteru v porovnání s mnoha jinými evropskými státy, po skončení války pak navíc přišlo období rychlého zlepšování úmrtnosti, což lze považovat za klíčový faktor rychlého růstu střední délky života v přesném věku 65 let těchto generací – právě jejich život ve věku 65 a více let se prolíná s obdobím zmíněného rychlého zlepšení úrovně úmrtnosti.

Naopak nárůst střední délky života 65letých žen byl relativně jen velmi mírný pro generace narozených v letech 1885–1905, ale ani pro další generace nenabral vysokého tempa. Ženy narozené v těchto generacích dospěly do věku 65 let přibližně v 50. až 70. letech 20. století, tedy v období stagnace nebo jen velmi pozvolných změn v úrovni úmrtnosti. Opět i v tomto případě lze očekávat, že mladší generace, narozené po roce 1920, zaznamenají rychlejší růst hodnot průměrného počtu let prožitých ještě osobou 65letou, neboť osoby narozené po roce 1920 dospěly do věku 65 a více let v období konce 80. let 20. století a později, tedy v období rychlého obecného poklesu úmrtnosti.

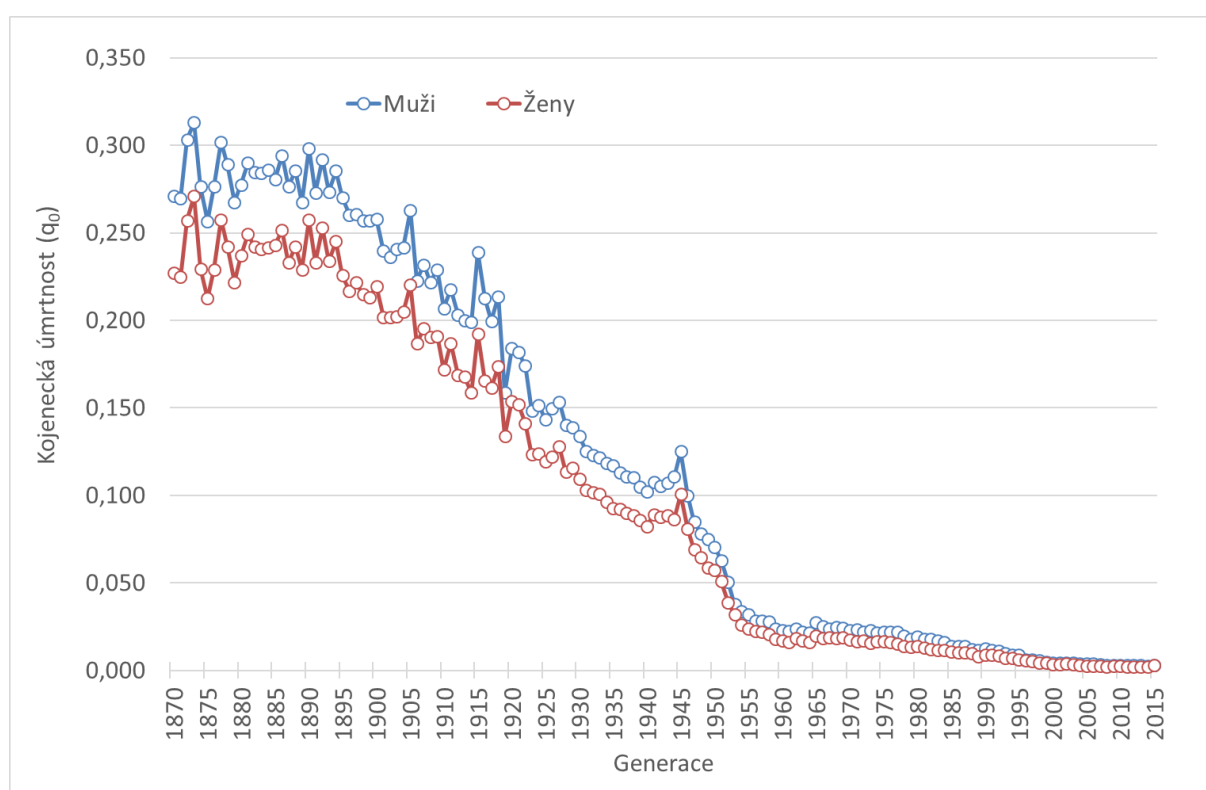
V případě mužů je vývoj v zásadě podobný. Rychlejší růst hodnot střední délky života osoby 65leté byl zaznamenán, obdobně jako u žen, u generací, které do věku 65 let přešly během 2. světové války a na jejím konci. Tyto generace tak mohly těžit z pozitivního vývoje úmrtnosti v poválečných letech. Zajímavý je však vývoj studovaného ukazatele pro generace mužů narozených především ve druhé polovině 80. a v 90. letech 19. století. V rámci těchto generací je totiž zaznamenán dokonce pokles, tedy spíše postupné zhoršování úmrtnostních poměrů ve vyšším věku, projevující se také v průměru kratší délkou života po 65. narozeninách. Jedná se o generace, které dospěly do věku 65 let v 50. a 60. letech 20. století, zbytek života tedy prožily převážně v období do začátku 80. let 20. století, kdy se teprve začaly projevovat výraznější tendence ke zlepšování úmrtnostních poměrů české populace. Naproti tomu 60. a 70. léta 20. století jsou typická spíše stagnací v celkovém vývoji úmrtnosti, na průřezových ukazatelích lze dokonce zaznamenat zhoršení právě v případě mužů (například Rychtaříková, 1987).

Ani nárůst hodnot mladších generací není rychlý, což odpovídá spíše stagnaci vývoje úmrtnosti ve 3. čtvrtině 20. století. Lze však, stejně jako u žen, předpokládat, že studovaný ukazatel nabere trend k pozitivnímu vývoji především u mužů narozených po roce 1920, neboť až ti se dožili 65. narozenin v období rychlého zlepšování úmrtnosti po roce 1990.

Pro doplnění prezentovaného vývoje kohortně počítaných ukazatelů lze uvést ještě vývoj dalšího z tradičních ukazatelů charakterizujících demografické a úmrtnostní poměry. Jedná se o kojeneckou úmrtnost, tedy pravděpodobnost úmrtí v prvním roce života (Obr. 8). Na vývoji hodnot tohoto ukazatele je patrné především enormní snížení jeho hodnot v jednotlivých generacích. V tomto případě

není nutné ukončit vynášené hodnoty u generace 1920, ale lze zobrazit hodnoty odpovídající také dalším generacím, neboť pro výpočet kvocientu kojenecké úmrtnosti lze využít i nejmladší generace, které přežily na konec prvního roku života. Z grafu (Obr. 8) je patrný pokles především po roce 1890. Ten byl narušen oběma světovými válkami. Ve druhé polovině 20. století se úroveň kojenecké úmrtnosti dostala na tak nízké hodnoty, že další pokles byl již velmi pozvolný. Zatímco v současné době se pravděpodobnost úmrtí v prvním roce života pohybuje na úrovni lehce nad 0,2 %, pro generace na konci 19. století se jednalo o hodnoty okolo 25 % (v případě dívek) a až 30 % (v případě mužů).

Obrázek 8: Vývoj kojenecké úmrtnosti (pravděpodobnost úmrtí v prvním roce života), generace 1870–1920, muži, ženy



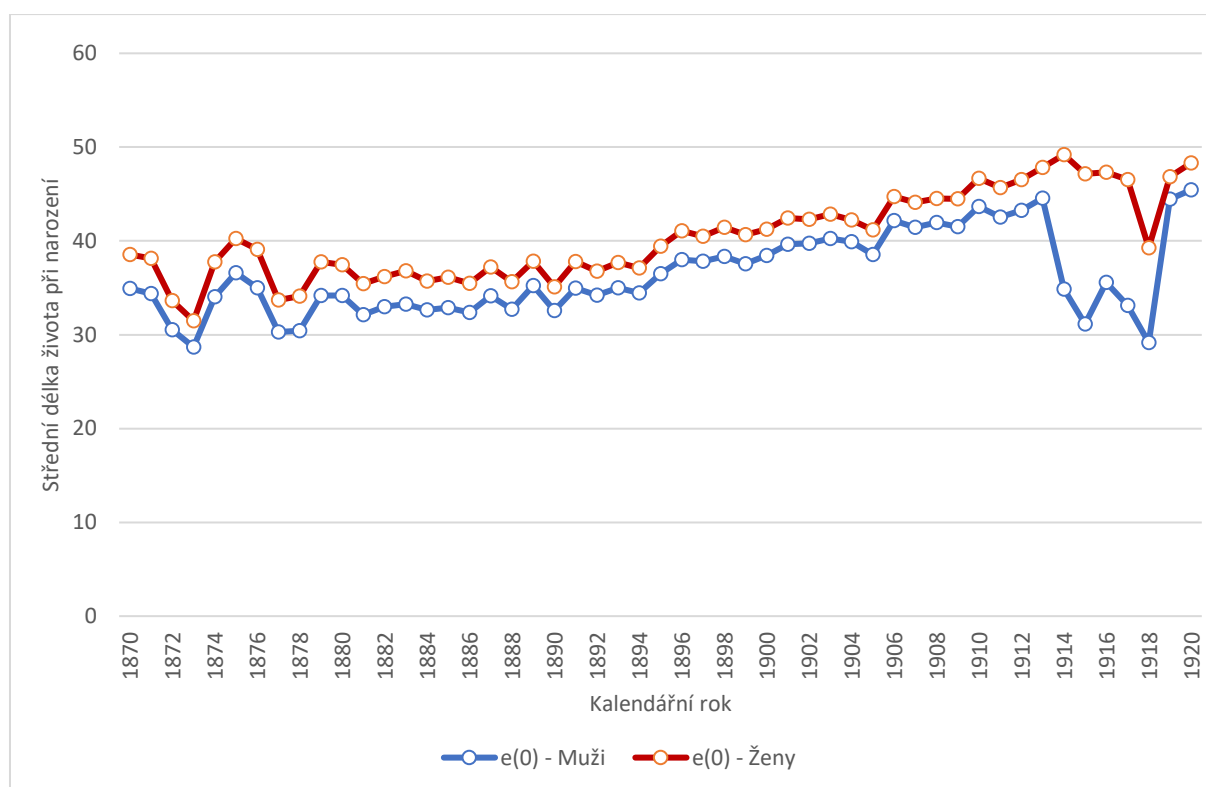
Zdroj: Mazouch, Hulíková Tesárková (2018, s. 61)

8.2 Základní výsledky transverzální analýzy

Výsledky vývoje kojenecké úmrtnosti, prezentované na Obr. 8, jsou určitým můstkem k výsledkům transverzální analýzy, protože podobně jako můžeme vidět silný vliv kojenecké úmrtnosti na kohortní ukazatele, lze tento vliv pozorovat také u transverzálních ukazatelů. Jsou zde patrné výkyvy zejména v prvních letech, kdy se v datech vyskytuje vysoká volatilita. Výrazné propady střední délky života lze pozorovat v období 1872–1873, kdy České země zasáhl jeden z posledních výskytů epidemie cholery a také dopad nekvalitní vakcíny proti neštovicím, dále pak další propad v letech 1877–1878, způsobený zejména epidemií záškrtu (Dokoupil, Nesládková, 1987). Další výraznější propady jsou pak patrné

v období 1. světové války, které lze přisoudit jak zhoršeným životním podmínkám, tak u mužů válečným ztrátám. Na konci analyzovaného období, zejména v roce 1918, můžeme vidět výrazný dopad španělské chřipky (Salfellner, 2020), který u žen způsobil propad střední délky života při narození o více než 7 let. Pro téměř celé období však lze vidět rostoucí trend střední délky života při narození, která se zvýšila pro obě pohlaví o přibližně 10 let.

Obrázek 9: Střední délka života při narození v období 1870–1920, muži, ženy



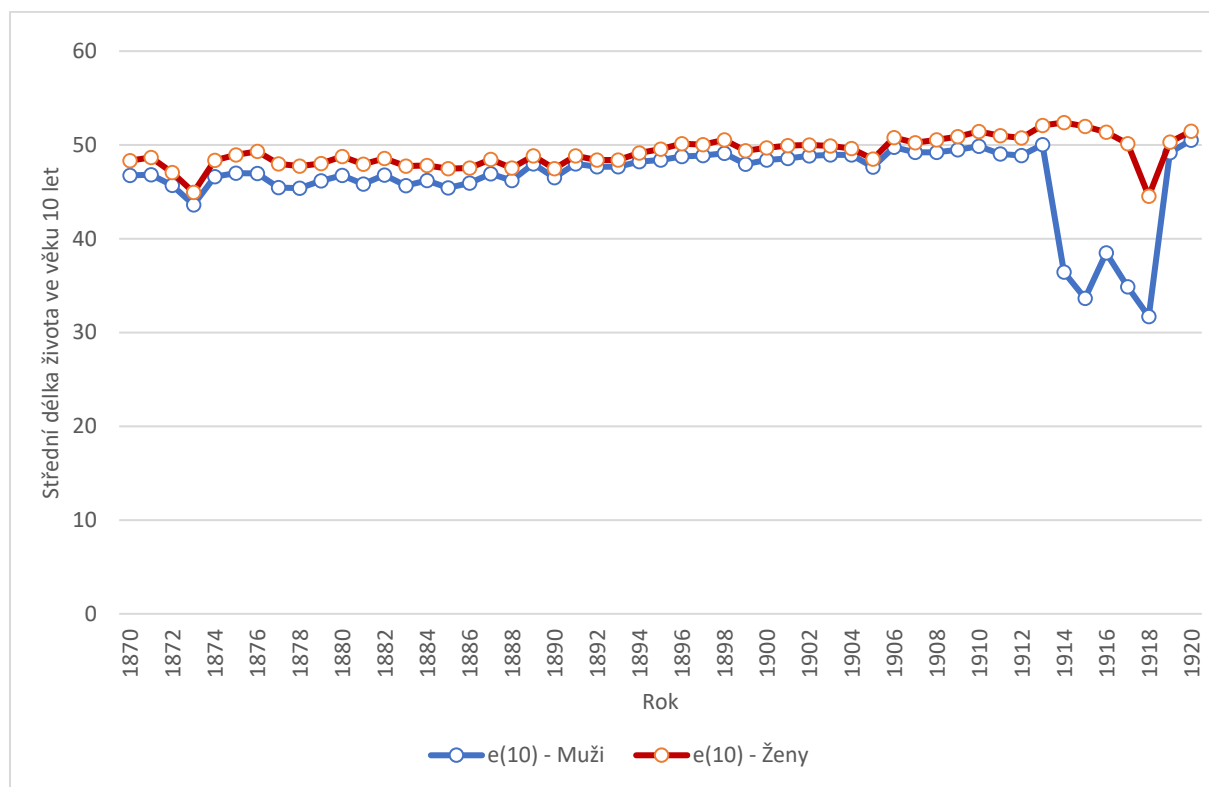
Zdroj: Vlastní zpracování

Pro odstranění vlivu úmrtnosti v nejnižších věcích je na Obr. 10 a 11 znázorněn vývoj střední délky života ve věku 10 a 65 let. U střední délky života ve věku 10 let je patrná výrazně větší stabilita, některé roky, které se u střední délky života při narození jeví jako určité lokální extrémy, zde nejsou tak výrazné, což je způsobeno tím, že dočasný výkyv v úrovni úmrtnosti zasáhl více věkovou skupinu do 10 let. Také je zde patrný *paradox střední délky života* – střední délka života ve věku 10 let je vyšší než střední délka života při narození, což je způsobeno zejména vysokou kojeneckou úmrtností a úmrtností v nejmladších věcích.

Opět je zde patrný výrazný propad střední délky života pro muže v období 1. světové války, kdy hodnota klesla o více než 15 let. Celkový vývoj se opět vyznačuje rostoucím trendem, který však již není tak výrazný. Celkový nárůst je za sledované období přibližně 10 %, proti střední délce života při narození, která vzrostla o přibližně 20 %. Zajímavé je porovnání mezi střední délkou života mužů a žen,

kteřá bývá pro ženy většinou vyšší. Zde se v období konce 19. století střední délka života mužů ve věku 10 let výrazně přibližuje střední délce života žen, ale stále zůstává pod touto úrovní.

Obrázek 10: Střední délka života ve věku 10 let v období 1870–1920, muži, ženy

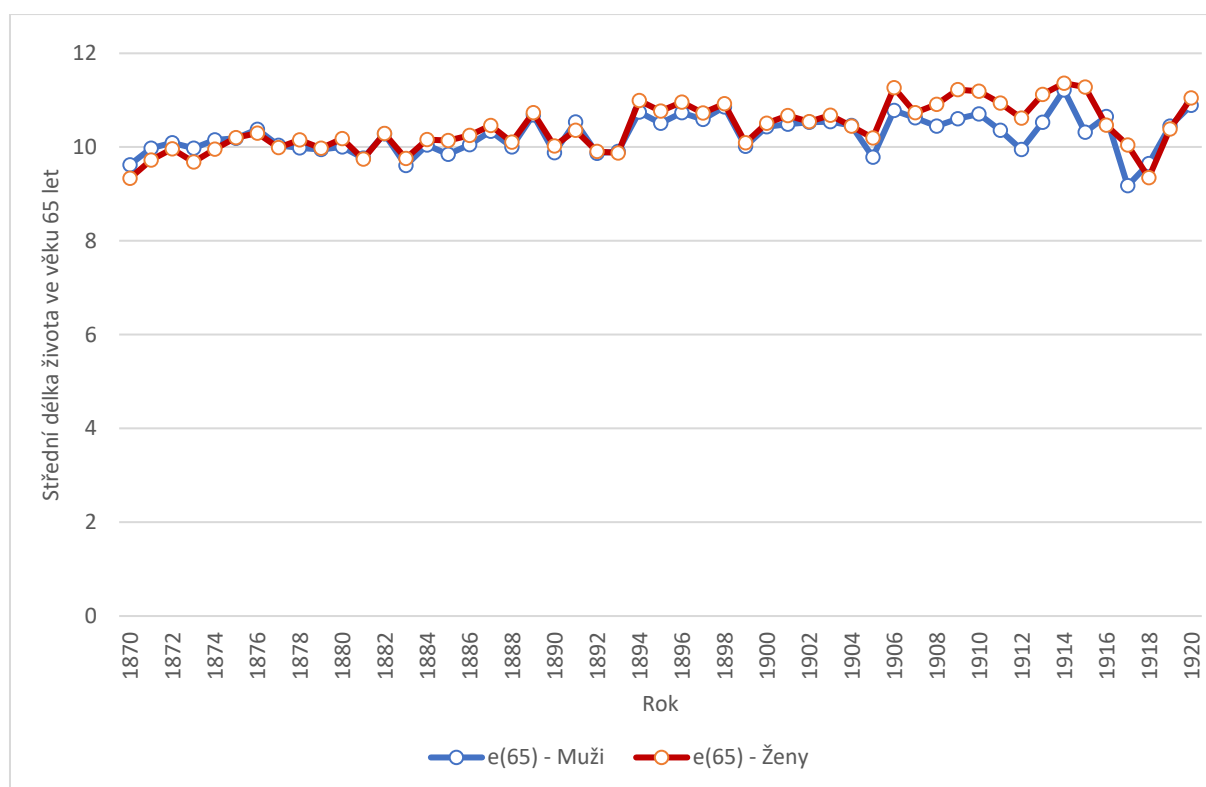


Zdroj: Vlastní zpracování

Podobný trend jako u střední délky života ve věku 10 let můžeme pozorovat také u střední délky života ve věku 65 let. Opět zde dochází k nárůstu přibližně na úrovni 10 % za celé sledované období, což je přibližně 1 rok. Vývoj však nebyl monotónní, protože střední délka života ve věku 65 let se začala zvyšovat až v posledních pěti letech 19. století, do té doby se držela na přibližně konstantní úrovni.

Zajímavým zjištěním je, že v případě střední délky života ve věku 65 let byla hodnota pro muže a ženy po většinu sledovaného období srovnatelná a lišila se jen velmi málo (na začátku období byla dokonce delší pro muže než pro ženy). Určitou nepatrnou výhodu začaly ženy získávat až kolem roku 1905. Při srovnání mužů a žen v případě střední délky života při narození, kde byly rozdíly na úrovni přibližně 4 let (s výjimkou období 1. světové války), je zřejmé, že rozdíl v úrovni úmrtnosti mezi muži a ženami byl zejména na začátku života. Podrobnější analýza, například formou dekompozice, by mohla ukázat, které věkové skupiny přispívají k tomuto rozdílu nejvíce, to však přesahuje cíl této práce.

Obrázek 11: Střední délka života ve věku 65 let v období 1870–1920, muži, ženy



Zdroj: Vlastní zpracování

8.3 Diskuse výsledků¹⁹

Již v úvodních částech bylo zmíněno, že v České republice nebyla před publikací výsledků projektu „Generační úmrtnostní tabulky České republiky: data, biometrické funkce a trendy“ k dispozici ucelená studie týkající se kohortní analýzy úmrtnosti, která by zahrnovala kompletní data celé kohorty. Výjimkou jsou fragmenty kohortních úmrtnostních tabulek v článku Růžičky (1959) a v knize Pavlíka et al. (1984), kde jsou v případě článku Růžičky obsaženy výsledky zkrácené úmrtnostní tabulky za kohorty 1875 a 1900 (pro muže a ženy zvlášť) včetně základních funkcí – tabulkový počet dožívajících se, tabulkový počet zemřelých, pravděpodobnost úmrtí a přežití pro věkové skupiny a střední délka života (Růžička, 1959). V případě knihy kolektivu vedeného Pavlíkem je k dispozici kompletní úmrtnostní tabulka po jednotkách věku pro generaci žen narozených v roce 1875, a pak odhady střední délky života pro právě narozeného z generace 1875, 1900, 1910, 1930, 1950, 1960 a 1971 s rozdělením na muže a ženy (Pavlík et al., 1984, s. 180 a 188).

Obě analýzy jsou velmi kvalitní a vzhledem k částečné (či v případě Růžičky úplné) absenci výpočetní techniky jde o velmi cenné zdroje. Oba autoři pracovali s dostupnými daty z období, ve kterých byly zkonstruovány průřezové úmrtnostní tabulky, a v období mezi nimi hodnoty interpolovali (Růžička

¹⁹ Podkapitola je převzata z publikace Mazoucha a Hulíkové Tesárkové (2018)

zmiňuje konkrétně lineární interpolaci). Protože tabulky vznikaly v době, kdy některé generace ještě nemohly být kompletně vymřelé, pro nejvyšší věky byly hodnoty odhadnuty z posledních známých hodnot průřezových tabulek s předpokladem, že úmrtnost v nejvyšších věcích se v čase příliš nesnižuje.

Přestože mohou tyto předpoklady výsledky mírně zkreslovat, byly srovnány s výsledky prezentovanými výše. V Tab. 3 jsou srovnány výsledky kohorty narozené v roce 1875. V případě žen bylo možné srovnat střední délky života nejen pro novorozence, ale i pro další věkové skupiny, pro muže byly hodnoty omezeny v případě Pavlíka et al. (1984) pouze na základní ukazatel střední délky života novorozence.

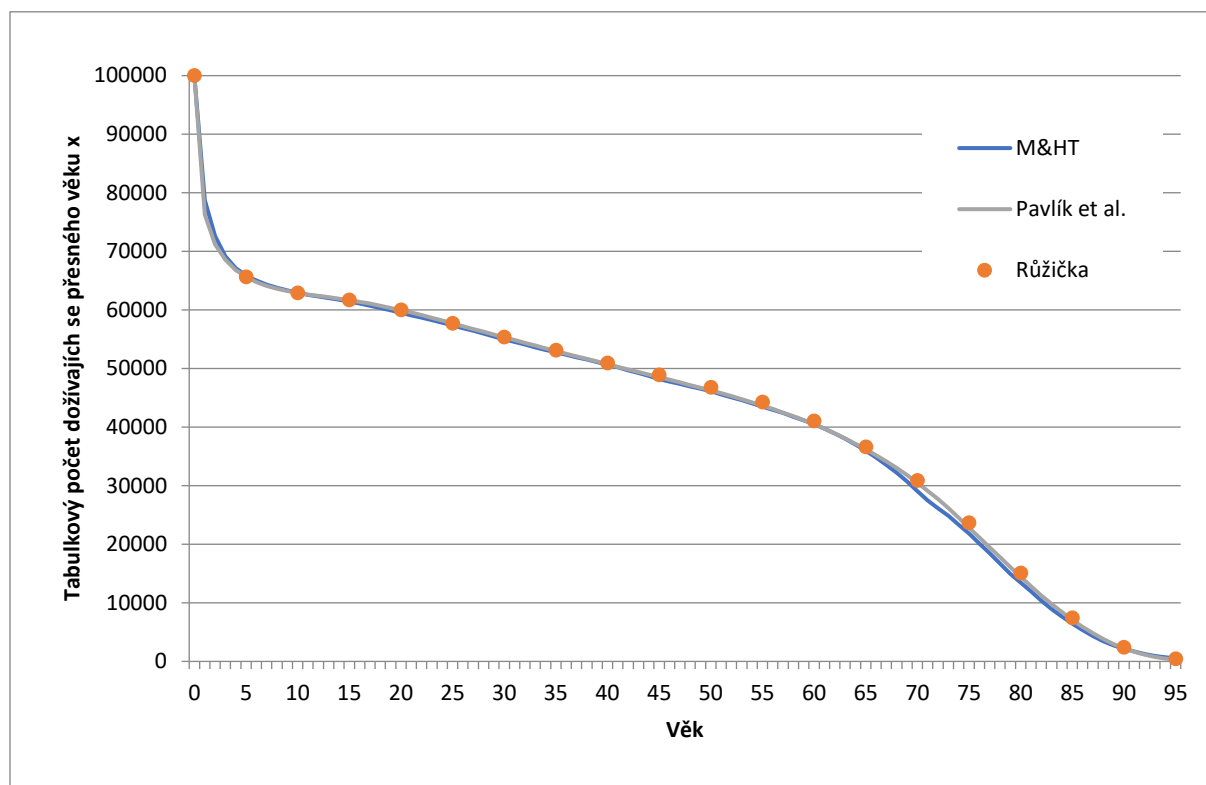
Z výsledků je patrné, že v případě střední délky života novorozené holčičky jsou hodnoty nejvyšší v případě odhadu Růžičky, následně Pavlíka et al., a nejnižší jsou v případě výsledků Mazoucha a Hulíkové Tesárkové. Rozdíly však nejsou nijak dramatické a činí pouze 0,5 roku. Při srovnání výsledků za další věky lze pozorovat, že hodnoty Růžičky i Pavlíka et al. jsou mírně vyšší až do věku 80 let, kdy se hodnoty téměř rovnají. To může být způsobeno vlivem 2. světové války (generace 1875 byla zasažena touto událostí ve věku přibližně 62–70 let), protože v případě interpolace Růžička ani Pavlík et al. nezahrnovali do odhadu jakýkoliv efekt války nebo jiného možného dočasného zvýšení/snížení úmrtnosti a uvažovali pouze lineární interpolaci mezi dvěma hodnotami (Růžička, 1959 a Pavlík et al., 1984). V případě mužů je situace u střední délky života novorozence obdobná, pouze difference je zde mírně vyšší, a to 0,6 roku, detailněji však lze srovnat pouze data Růžičky a Mazoucha a Hulíkové Tesárkové.

Tabulka 3: Srovnání střední délky života ve vybraných věcích pro kohortu mužů a žen narozených v roce 1870 dle výsledků Růžičky (1959), Pavlíka et al. (1984) a Mazoucha a Hulíkové Tesárkové (2018)

Věk	Ženy			Muži		
	Růžička	Pavlík et al.	M&HT	Růžička	Pavlík et al.	M&HT
0	40,2	39,9	39,7	36,3	36,0	35,7
10	52,9	52,6	52,1	50,7	-	49,7
20	45,2	45,0	44,8	42,7	-	41,8
30	38,6	38,3	38,0	35,9	-	34,8
40	31,5	31,3	30,9	28,7	-	27,6
50	23,9	23,9	23,5	21,6	-	21,2
60	16,5	16,5	16,0	14,9	-	14,2
70	10,1	10,1	10,0	9,2	-	9,0
80	5,4	5,6	5,6	4,9	-	5,1
90	3,0	2,6	3,2	2,9	-	2,8

Zdroj: Růžička (1959), Pavlík et al. (1984) a Mazouch, Hulíková Tesárková (2018)

Obrázek 12: Srovnání vývoje tabulkového počtu dožívajících se přesného věku x pro ženy narozené v roce 1875 dle výsledků Růžičky (1959), Pavlíka et al. (1984) a Mazoucha a Hulíkové Tesárkové (2018)



Zdroj: Růžička (1959), Pavlík et al. (1984) a Mazouch, Hulíková Tesárková (2018)

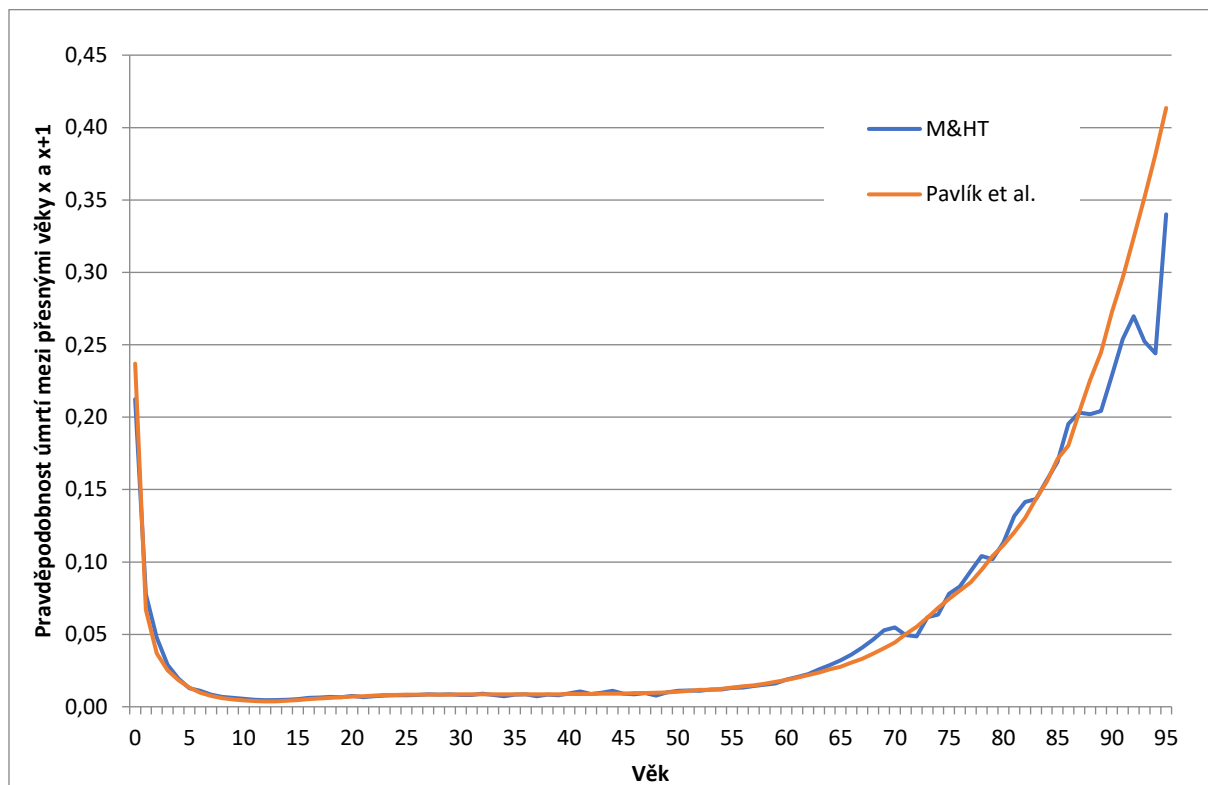
Díky publikování kompletní úmrtnostní tabulky za kohortu žen narozených 1875 v případě Pavlíka et al. (1986) a zkrácené úmrtnostní tabulky za stejnou skupinu v případě Růžičky (1959) lze srovnat i další tabulkové funkce. Obr. 12 obsahuje srovnání tabulkového počtu dožívajících se přesného věku x pro všechny tři studie. Obrázek demonstruje velmi dobrou shodu výsledků všech tří studií. Mírné odklonění lze vidět v případě výsledků současné studie, které se vyznačuje rychlejším poklesem počtu dožívajících se přibližně kolem věku 65 let, což se shoduje s obdobím, kdy tato kohorta prožívala první roky 2. světové války.

Stejný závěr lze udělat také při pohledu na Obr. 13, kde jsou znázorněny průběhy pravděpodobností úmrtí mezi přesnými věky x a $x+1$ pro kohortu žen narozenou v roce 1875. Zde lze porovnat bohužel pouze výsledky Pavlíka et al. a Mazoucha a Hulíkové Tesárkové (2018), protože zkrácené úmrtnostní tabulky Růžičky jednoduchou komparaci s úplnými tabulkami neumožňují.

I z tohoto výstupu lze vidět zvýšenou intenzitu úmrtnosti v již zmiňovaném věkovém období kolem 65.–71. roku věku, což je výše uvedené období 2. světové války. Z výsledků Pavlíka et al. je také patrné, že data byla v případě ukazatelů úmrtnosti vyrovnána a celá křivka je pak hladká s čistě exponenciálním růstem v nejvyšších věcích. Tato skutečnost může také vést k určitému zkreslení, ale protože počet

osob dožívajících se tohoto věku je již velmi malý, nemá tato skutečnost téměř žádný vliv na celkové hodnoty střední délky života, jak bylo patrné již z Tab. 3.

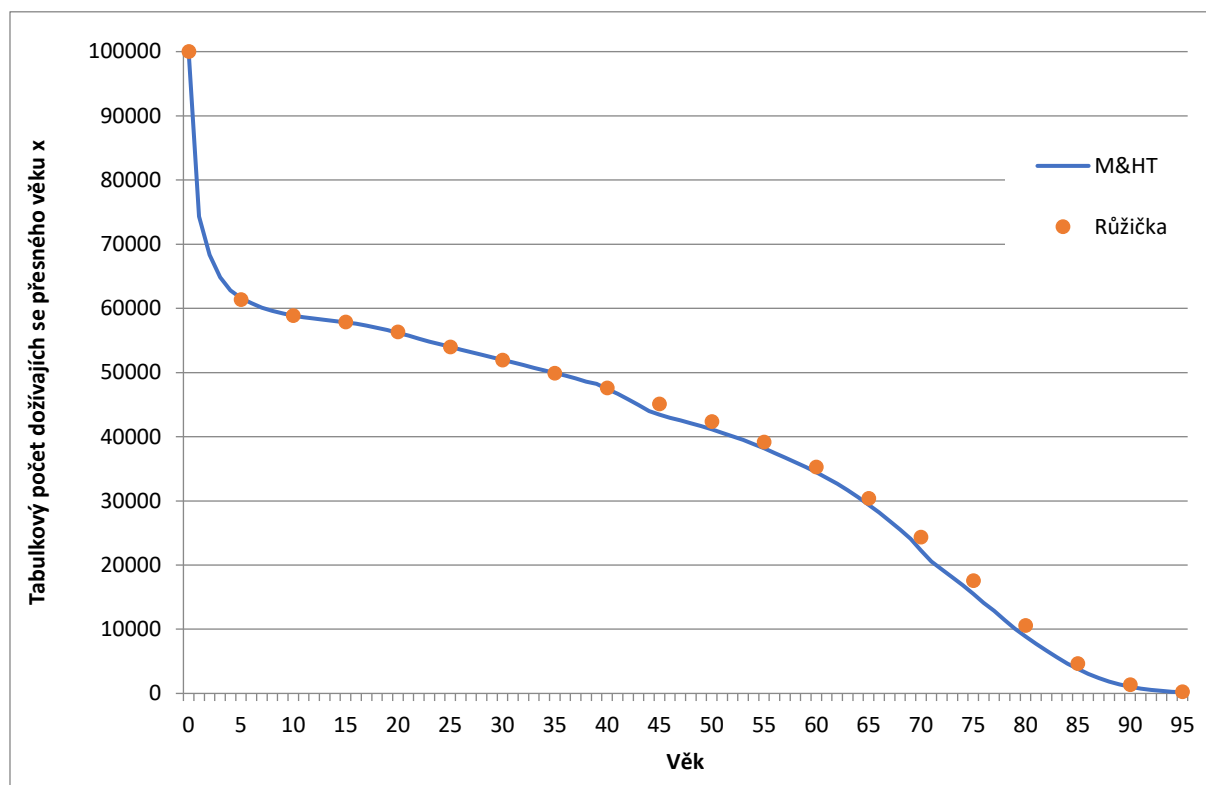
Obrázek 13: Srovnání vývoje pravděpodobnosti úmrtí mezi přesnými věky x a $x+1$ pro ženy narozené v roce 1875 dle výsledků Pavlíka et al. (1984) a Mazoucha a Hulíkové Tesárkové (2018)



Zdroj: Pavlík et al. (1984) a Mazouch, Hulíková Tesárková (2018)

Srovnání výsledků pro kohortu mužů narozených v roce 1875 z hlediska jiných ukazatelů než pouze střední délky života je možné jen pro výsledky Růžičky a a Mazoucha a Hulíkové Tesárkové (2018) a pouze za ukazatel tabulkového počtu dožívajících se přesného věku x . Z Obr. 14 je zřetelně vidět, že výsledky jsou srovnatelné, až na období obou světových válek, kdy počet dožívajících se přesného věku x klesá v případě výsledků Mazoucha a Hulíkové Tesárkové rychleji než v případě výsledků článku Růžičky. Na Obr. 14 je to patrné ve věku kolem 40–45 let, a pak identicky jako pro ženy kolem věku 65–71 let s dopadem i na následující věky.

Obrázek 14: Srovnání vývoje tabulkového počtu dožívajících se přesného věku x pro muže narozené v roce 1875 dle výsledků Růžičky (1959) a Mazoucha a Hulíkové Tesárkové (2018)



Zdroj: Růžička (1959) a Mazouch, Hulíková Tesárková (2018)

Následující srovnatelná data jsou pro kohortu narozenou v roce 1900. Zde již bohužel nejsou k dispozici podrobná data Pavlíka et al., kde byly publikovány pouze základní hodnoty střední délky života novorozence pro muže a ženy. Střední délka života žen je v tomto případě téměř shodná, s výjimkou rozdílu 0,5 roku pro věkovou kategorii 10 let a pak pro nejvyšší věkové kategorie, kdy hodnoty Mazoucha a Hulíkové Tesárkové jsou vyšší než v případě výsledků Růžičky (je nutné mít na paměti, že v době publikace článku Růžičky bylo této kohortě pouhých 59 let a nejvyšší věky byly odhadovány na základě průřezových dat, blíže viz Růžička, 1959). Pavlík odhaduje střední délku novorozené holčičky na 49 let, ale bez podrobnějších dat nelze zjistit, kterými věky jsou nejspíše rozdíly způsobeny.

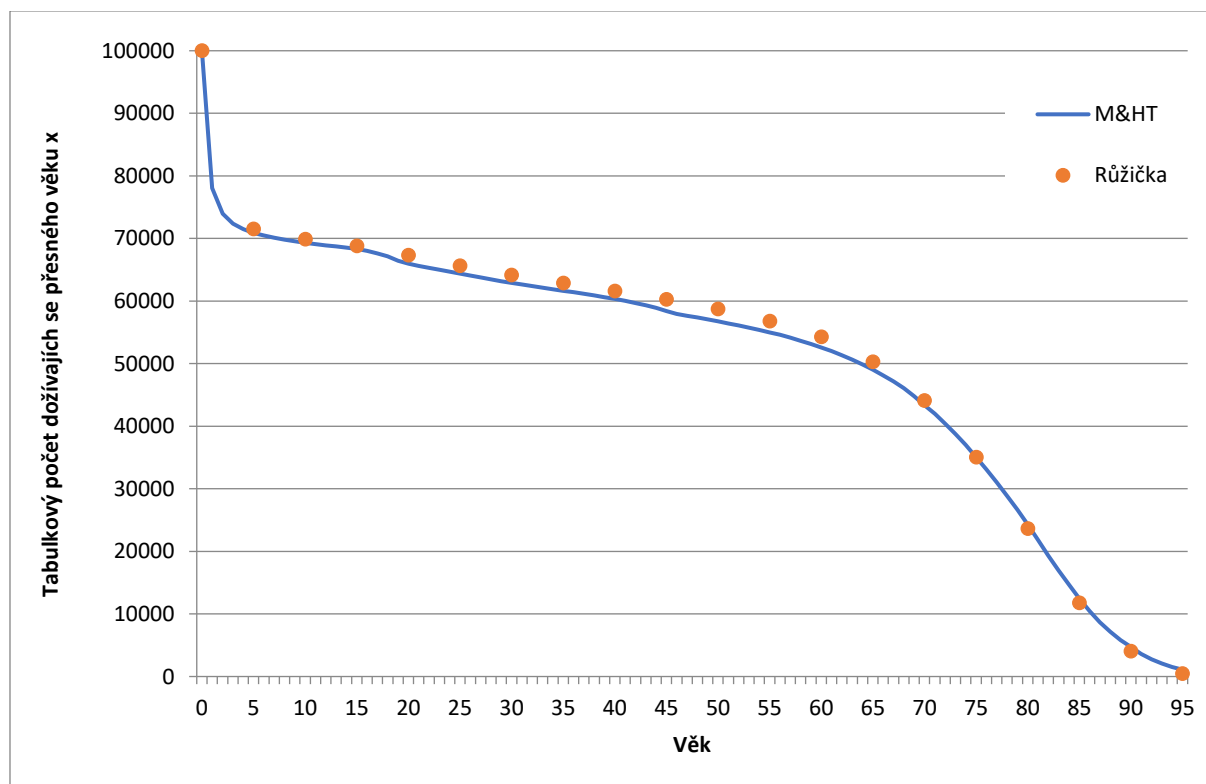
Úplně jiná je situace v případě mužů, protože zde vykazují výsledky Mazoucha a Hulíkové Tesárkové (2018) výrazně odlišné hodnoty od studií Růžičky a Pavlíka et al., a to o 3,7 resp. o 4,1 roku. Tento rozdíl je způsoben opět válkou, zejména 1. světovou, protože tato generace byla ještě generací, která byla postižena výraznými přímými válečnými ztrátami. Ve vyšších věkových kategoriích se odhady přibližují, ale kvůli následující válce a pak také nejspíše díky tomu, že nejvyšší věky byly Růžičkou odhadnuty pouze hrubě, je rozdíl v celém věkovém spektru.

Tabulka 4: Srovnání střední délky života v jednotlivých věkových skupinách pro kohortu mužů a žen narozených v roce 1900 dle výsledků Růžičky (1959), Pavlíka et al. (1984) a Mazoucha a Hulíkové Tesárkové (2018)

Věk	Ženy			Muži		
	Růžička	Pavlík et al.	M&HT	Růžička	Pavlík et al.	M&HT
0	47,8	49,0	47,8	43,6	44,0	39,9
10	59,0	-	58,5	55,4	-	50,0
20	51,0	-	51,2	47,0	-	45,4
30	43,3	-	43,4	39,2	-	37,6
40	34,9	-	35,0	30,9	-	29,3
50	26,4	-	27,0	22,8	-	21,9
60	18,1	-	18,7	15,5	-	14,5
70	11,0	-	11,4	9,8	-	8,9
80	5,9	-	6,1	5,8	-	5,2
90	3,1	-	3,2	3,2	-	3,1

Zdroj: Růžička (1959), Pavlík et al. (1984) a Mazouch, Hulíková Tesárková (2018)

Obrázek 15: Srovnání vývoje tabulkového počtu dožívajících se přesného věku x pro ženy narozené v roce 1900 dle výsledků Růžičky (1959) a Mazoucha a Hulíkové Tesárkové (2018)

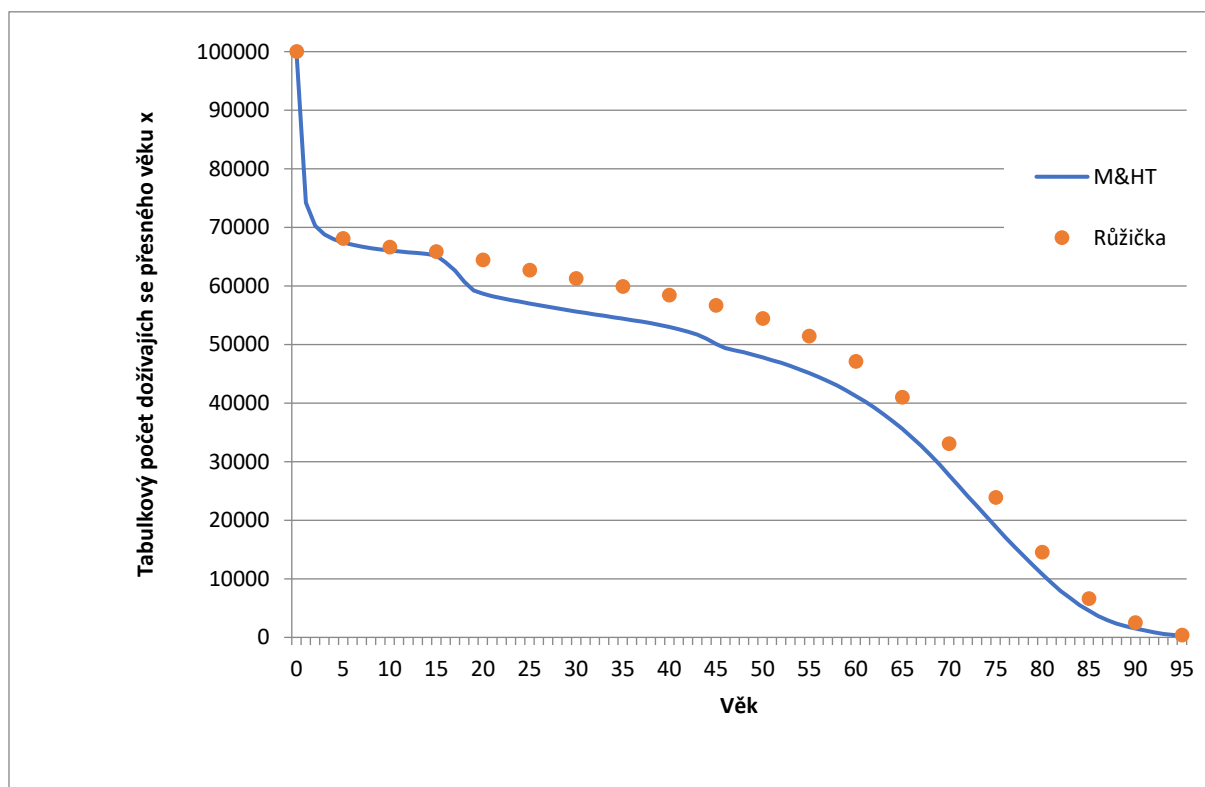


Zdroj: Růžička (1959) a Mazouch, Hulíková Tesárková (2018)

Výsledky pro ženy, prezentované v Tab. 4, jsou zřetelné i z Obr. 15, kde se srovnává ukazatel tabulkového počtu dožívajících se přesného věku x a mírná odchylka je patrná opět v období, kdy tato kohorta prožívala 2. světovou válku, což bylo přibližně ve věku 38–45 let s mírným dopadem na následující věky.

Již zmíněný odlišný vývoj u mužů narozených v roce 1900 je zřetelný také z Obr. 16, kde je srovnán tabulkový počet mužů dožívajících se přesného věku x . Zřejmý je efekt 1. světové války, která tuto populaci zasáhla ve věku 15–19 let s výraznějším dopadem až ve věku kolem 18 let, protože až v tomto věku mohli být muži povoláni na frontu. Druhý viditelný propad u této kohorty nastává přibližně ve věku 43 let, což je opět již zmiňovaný efekt 2. světové války. Ani jeden z těchto efektů neměl ve své studii Růžička zahrnutý, což vede následně k rozdílným hodnotám.

Obrázek 16: Srovnání vývoje tabulkového počtu dožívajících se přesného věku x pro muže narozené v roce 1900 dle výsledků Růžičky (1959) a Mazoucha a Hulíkové Tesárkové (2018)



Zdroj: Růžička (1959) a Mazouch, Hulíková Tesárková (2018)

Dalšími srovnatelnými daty jsou již pouze základní hodnoty střední délky života novorozence pro generaci narozenou v roce 1910, jejichž odhady provedl Pavlík et al. Pro ženy je odhadována hodnota 54 let a pro muže 49 let. Mazouch a Hulíková Tesárková (2018) odhaduje hodnoty pro stejnou kohortu 54,7 let pro ženy a 48,2 let pro muže. Rozdíly zde odpovídají přibližně diferencím, které jsme viděli u kohort 1875 (pro obě pohlaví) a pro ženy narozené 1900. Hlubší analýza rozdílů zde není možná, ale

opět lze uvažovat, že část půjde na vrub Pavlíkem et al. neuvažovaného dopadu 2. světové války a také odhadu nejvyšších věků, kterých v době publikace výsledků Pavlíkem tato generace ještě nedosáhla.

Celkové srovnání ukazuje, že výsledky můžeme považovat za koherentní. Výjimkou jsou pouze kohorty, které byly výrazněji zasaženy některou z válek, což nebylo v odhadech Růžičky a Pavlíka et al. zahrnuto. Nesporné je, že předchozí autoři ukázali, že i s jednoduchými předpoklady lze vytvořit dostatečně spolehlivý odhad, který byl ověřen na základě důslednější analýzy zdrojových dat v této knize.

Pro srovnání výsledků transversální analýzy je výrazně méně prostoru. Jak již bylo zmíněno v kapitole 1, pro absenci dat v intercenzálních obdobích nejsou pro České země zkonstruovány žádné úmrtnostní tabulky s výjimkou zmíněných let kolem realizovaných sčítání lidu. Ty však nedává příliš smysl srovnávat, protože jak tabulky například Srba a Kučery (1959), tak výsledky transversální analýzy uvedené v této kapitole jsou opřeny o stejný datový zdroj – počty zemřelých v jednotlivých letech a stav obyvatelstva v okamžiku sčítání, tedy rozdíl může být způsoben pouze různým přístupem například k vyrovnání měr úmrtnosti v nejvyšších věcích nebo jiným metodickým aspektem.

Mnohem důležitější je však validovat výsledky transversálních úmrtnostních tabulek právě v období mezi sčítáními, kdy došlo k modelování stavu populace (a před rokem 1895 také počtu zemřelých). Při srovnání ukazatelů nově konstruovaných úmrtnostních tabulek a předchozích studií, které sice využívaly hrubých měr (či jiných ukazatelů), můžeme vidět, že v obdobích, které autoři jako Dokoupil a Nesládková (1987), Kárníková (1965), Srb (2004) nebo Salfellner (2020) zmiňují jako období s výjimečnými úmrtnostními poměry, vykazují i výsledky transversálních úmrtnostních tabulek koherentní hodnoty. Tato skutečnost umožňuje považovat podkladová data za věrohodná a vhodná pro další podrobnější analýzy, které předchozí studie doplní o možnost kvantifikovat změny v úmrtnosti (a nejen v ní) propracovanějšími ukazateli, než jsou například pouze hrubé míry.

Závěr

Podobně jako jiné vědní disciplíny je i demografie závislá na datových zdrojích. Bez nich by celá věda byla založena pouze na teoretickém vymezení vztahů, které by postrádaly potřebnou číselnou demonstraci. Demografická statistika má na území Českých zemí velmi dlouhou tradici, sahající až do 18. století, což je srovnatelné s jinými rozvinutými zeměmi. Jiné země však díky rozvoji metod rekonstrukce demografických dat dokázaly i na tato historická data aplikovat moderní metody demografické analýzy a vytěžit tak informace o demografickém vývoji téměř kompletně, tedy z pohledu dnešních limitů demografické analýzy.

Dobrá praxe některých zemí podnítila motivaci pokusit se rekonstruovat datové podklady, původně za účelem konstrukce kohortních úmrtnostních tabulek za území Českých zemí. Ta ukázala, že výsledky jsou velmi kvalitní a nabízejí širokou paletu dalších možných návazných či hlubších demografických analýz. Přesto se jednalo pouze o rekonstrukci dat pro kohortní analýzu, která má svá určitá specifika a v rámci výzkumného projektu byl kladen důraz spíše na publikaci výsledných úmrtnostních tabulek; rekonstruovaná data byla sice popsána, ale publikována již nebyla.

Nadšení z kvalitních výstupů kohortní analýzy mne vedlo k úvahám, proč neposunout stav poznání o demografickém vývoji v období před vznikem Československa na vyšší úroveň, odpovídající ostatním rozvinutým státům. K tomuto kroku však bylo třeba zrekonstruovat datové zdroje pečlivěji (zejména popisem aplikovaných metod) a šířeji (upravit je i pro jiné metody než pouze kohortní analýzu), aby na jejich základě bylo možné následně aplikovat nástroje demografické analýzy.

Práci na rekonstrukci datových zdrojů, byť jsou to datové zdroje demografické a mají sloužit zejména k demografické analýze, považuji spíše za statistickou a vytyčil jsem její cíl jednoduše – podrobně popsat původní datové zdroje a veškeré kroky, které vedly ke vzniku rekonstruovaných datových sad vhodných pro demografickou analýzu období před vznikem Československé republiky.

Výsledkem práce je publikace unikátních datových souborů za období 1870–1920, které obsahují koherentní údaje pro demografickou analýzu, a to vše ve formátu vhodném k okamžitému využití bez nutnosti výraznějších úprav a také volně přístupné v elektronické formě. Pevně věřím, že toto obohacení demografické statistiky ve formě prodloužení délky časových řad o 50 % bude s využitím moderních metod demografické analýzy krokem, který umožní dále rozvíjet současný stav poznání.

Seznam použité literatury

ARRIAGA, E.E., 1984. *Demography*. 21: 83. Springer-Verlag. Dostupné také z <https://doi.org/10.2307/2061029>

BOHÁČ, A., 1920. Recenze: Dr. Wilhelm Winkler, Die Totenverluste der öst.-ung. Monarchie nach Nationalitäten. – Die Altersgliederung der Toten. – Ausblicke in die Zukunft. *Československý statistický věstník 1920*. Ročník 1, Praha: Státní úřad statistický, 1920. 336 s.

ČSÚ, 2005. *Přirozená měna obyvatelstva v zemích Koruny české v letech 1. světové války*. Praha: Český statistický úřad, 2005. Dostupné také z <https://www.czso.cz/csu/czso/prirozena-mena-obyvatelstva-v-zemich-koruny-ceske-v-letech-1-svetove-valky-1914-az-1918-n-y1naizbqp2>

ČSÚ, 2015. *Historie státní statistické služby 1919–2014*. Praha: Český statistický úřad, 2015. Dostupné také z <https://www.czso.cz/csu/czso/historie-statni-statisticke-sluzby-1919-2014-n9e3pk3ay4>

ČSÚ, 2020. *Úmrtnostní tabulky*. Praha: Český statistický úřad, 2020. [cit. 15. 6. 2020] Dostupné také z https://www.czso.cz/csu/czso/umrtnostni_tabulky

DOKOUPIL, L. a L. NESLÁDKOVÁ, 1987. Charakteristické rysy vývoje úmrtnosti obyvatelstva českých zemí v 19. století. *Historická demografie* 12, 1987, s. 193–206. Dostupné také z http://www.eu.avcr.cz/opencms/export/sites/eu/.content/files/historicka-demografie/HD_12_1987.pdf

FIALOVÁ, L., 1989. Age structure of the balance of migration in the Czech lands in 1870–1910. *Historická demografie* 13, 1989, s. 229–241. Dostupné také z http://www.eu.avcr.cz/opencms/export/sites/eu/.content/files/historicka-demografie/HD_13_1989.pdf

FIALOVÁ, L., 2014. První světová válka a obyvatelstvo českých zemí. *Historická sociologie*. Praha: Univerzita Karlova, Nakladatelství Karolinum, 2014, č. 2/2014. ISSN 1804-0616. Dostupné také z <https://historicalsociology.cuni.cz/HS-23-version1-7fialova.pdf>

FISCHER, J., JEŘÁBKOVÁ, V., PETKOVÁ, L., PTÁČKOVÁ, V. a P. ŠVARCOVÁ, 2019. *Základní metody statistického srovnávání* [online]. 1. vyd. Praha : Oeconomica, Nakladatelství VŠE, 2019. 88 s. ISBN 978-80-245-2342-2. Dostupné také z: https://oeconomica.vse.cz/wp-content/uploads/page/17038/EKNIHA_Skripta-4ES402_v4.0.pdf

JINDRA, Z., 2014. Kardinální problém Rakousko-Uherska za Velké války 1914–1918: Lidský potenciál na frontě a v zázemí. *Historická sociologie*. Praha: Univerzita Karlova, Nakladatelství Karolinum, 2014, č.

2/2014. ISSN 1804-0616. Dostupné také z <https://historicalsociology.cuni.cz/HS-23-version1-6jindra.pdf>

KALIBOVÁ, K., PAVLÍK, Z. a A. VODÁKOVÁ, 2009. *Demografie (nejen) pro demografy*. Praha: Sociologické nakladatelství, 2009, 241 s. ISBN 978-80-7419-012-4.

KÁRNÍKOVÁ, L., 1965. *Vývoj obyvatelstva v českých zemích 1754-1914*. Praha: Nakladatelství Československé akademie věd, 1965.

KORČÁK, J., 1946. Skryté oslabení československé populace. *Statistický zpravodaj 1946*. Praha: Státní úřad statistický, 1946. Roč. IX, č. 1.

KOSCHIN, F., 2000. *Aktuárská demografie*. 2.vyd. Praha: VŠE, 2000. 123 s.

KOSCHIN, F., 2005. *Demografie poprvé*. Vyd. 2. přepracované. Praha: Oeconomica, 2005. 122 s. ISBN 8024508591.

MATESOVÁ, J., 1983. Základní trendy v dlouhodobém vývoji řádu vymírání české populace. *Demografie* 25, 1983, s. 309–321. Dostupné také z <https://www.czso.cz/csu/czso/demografie-revue-pro-vyzkum-populacniho-vyvoje-1959-az-2010-n-b5146vsfjk>

MAZOUCH, P. a K. HULÍKOVÁ TESÁRKOVÁ, 2018. *Kohortní úmrtnostní tabulky v ČR: Metodické aspekty zpracování* [online]. 1. vyd. Praha: Oeconomica, 2018. 90 s. ISBN 978-80-245-2248-7. [cit. 15. 6. 2020] Dostupné z: <https://oeconomica.vse.cz/vydavani-publikaci/informace-pro-ctenare/seznam-vsech-vydanych-titulu/publikace-2018-az-2002/mazouch-petr/>

MAZOUCH, P. a K. HULÍKOVÁ TESÁRKOVÁ, 2019. Short period (or cohort) mortality shifts and estimation of their effect on life expectancy: formal expression using the biometric functions with practical application. In: *New Generations in Demography: New Challenging Adventures in the Population Science* [online]. Praha: Oeconomica, 2019, s. 117–131. 145 s. ISBN 978-80-245-2302-6. [cit. 15. 6. 2020] Dostupné z: https://oeconomica.vse.cz/wp-content/uploads/page/15835/EBOOK_New-Generations-in-Demography.pdf

MUSIL, J., 1959. Rozbor dynamiky některých ukazatelů úmrtnostních tabulek za léta 1875–1955. *Demografie* 1, 1959, s. 77–87. Dostupné také z <https://www.czso.cz/csu/czso/demografie-revue-pro-vyzkum-populacniho-vyvoje-1959-az-2010-n-b5146vsfjk>

NETUŠIL, F., J., 1923. Vývoj obyvatelstva v Čechách a na Moravě v letech válečných. *Antropologie. Časopis věnovaný fyzické antropologii, nauce o plemenech, demografii*. Roč. I. Praha: Anthropologický ústav Karlovy University, 1923.

PAVLÍK, Z., RYCHTAŘÍKOVÁ J. a A. ŠUBRTOVÁ, 1986. *Základy demografie*. Praha: Academia Praha, 1986, 736 s.

POLLARD, J., H., 1988. On the decomposition of changes in expectation of life and differentials in life expectancy. *Demography* 25(2): 265–276. doi:10.2307/2061293.

PRESSAT, R., 1972. *Demographic analysis: methods, results, applications*. Londýn: Edward Arnold, 1972.

ROUBÍČEK, V., 1997. *Úvod do demografie*. 1. vyd. Praha: Codex Bohemia, 1997. 349 s. ISBN 80-85963-43-4.

RŮŽIČKA, L., 1959a. Generační úmrtnost v českých zemích. *Demografický sborník 1959*. Praha: Státní úřad statistický, 1959, 187 s.

RŮŽIČKA, L., 1959b. K vývoji úmrtnosti v českých zemích. *Statistika a demografie 1*, Praha 1959, s. 181–211.

RYCHTAŘÍKOVÁ, J. 1987. Vývoj úmrtnosti v ČSR podle pohlaví a věku v období 1950–1984. *Demografie*. 1987, 29(3), s. 193–207. Dostupné také z <https://www.czso.cz/csu/czso/demografie-revue-pro-vyzkum-populacniho-vyvoje-1959-az-2010-n-b5146vsfjk>

RYCHTAŘÍKOVÁ, J., ŘEHÁK, J., CASELLI, G., MESLÉ, F. a J. VALLIN, 1994: *Analysis of mortality in the Czech Republic: causal models of mortality changes in generations and the international comparative analysis*, 67s. Final report.

Salfellner, H., 2020. Odhad počtů obětí pandemie španělské chřipky v Českých zemích v letech 1918–1920. *Demografie*. 2020, 62(3), s. 182–196. Dostupné také z <https://www.czso.cz/csu/czso/demografie-revue-pro-vyzkum-populacniho-vyvoje-c-32020>

SIGNUMBELLI1914, 2020a. *Povinná vojenská služba v Rakousku-Uhersku*. [cit. 15. 6. 2020] Dostupné z https://www.velkavalka.info/wiki/doku.php?id=mirova_sluzba

SIGNUMBELLI1914, 2020b. *Vojenské ztráty armád v 1. světové válce*. [cit. 15. 6. 2020] Dostupné z <https://www.velkavalka.info/wiki/doku.php?id=ztraty>

SRB, V., 2004. *1000 let obyvatelstva českých zemí*. 1. vyd. Praha: Univerzita Karlova, Nakladatelství Karolinum, 2004. 275 s. ISBN 80-246-0712-3

SRB, V. a M. KUČERA, 1959. Vývoj obyvatelstva českých zemí v XIX. století. In: *Statistika a demografie 1*, s. 109–156, Praha 1959.

STŘÍTESKÝ, J., K., 1976. *Zdravotní a populační vývoj českých zemí*, Praha 1976.

SÚS, 1929a. *Pohyb obyvatelstva v Československé republice v letech 1919–1920*. Praha. Státní úřad statistický, 1929. Dostupné také z

https://www.czso.cz/csu/czso/casova_rada_demografie_1929_1919

SÚS. 1929b. *Pohyb obyvatelstva v Československé republice v letech 1921–1922*. Praha. Státní úřad statistický, 1929. Dostupné také z

https://www.czso.cz/csu/czso/casova_rada_demografie_1929_1919

SÚS. 1935: *Mimořádné zprávy Státního úřadu statistického*. Praha, roč. 4., č. 17–23.

SÚS. 1944. *Pohyb obyvatelstva v Československé republice v letech 1938–1940*. Praha. Státní úřad statistický, 1944. Dostupné také z

https://www.czso.cz/csu/czso/casova_rada_demografie_1949_1930

TALACKO, J. 1941. Dynamická pozorování ve statistice úmrtnosti. In: *Statistický obzor (věstník)*, sv. 43, Praha: Ústřední statistický úřad, 1941, 66 s.

URLANIS, B., C., 1963. *Války a evropské obyvatelstvo. Lidské ztráty ozbrojených sil evropských zemí ve válkách XVII.–XX. století*. Praha: Naše vojsko.

VARGA, M., HUDEČEK, M., RUCKÁ, G. a H. SÝKOROVÁ, 2018. *Srdečně zdravím a vřele líbám. Rodáci a občané okresu Hodonín ve velké válce*. Hodonín: Masarykovo muzeum, 2018, 322 s. ISBN 978-80-87375-11-2.

VAUPEL, J., W., 1986. How change in age-specific mortality affects life expectancy. *Population Studies* 40(1): 147–157. doi:10.1080/0032472031000141896.

VÁVRA, Z., 1960. Změny ve specifické úmrtnosti obyvatelstva českých zemí za léta 1870–1944. *Demografie* 2, 1960, s. 37–49. Dostupné také z <https://www.czso.cz/csu/czso/demografie-revue-pro-vyzkum-populacniho-vyvoje-1959-az-2010-n-b5146vsfjk>

VHU, 2017. *Seznamy ztrát rakousko-uherské armády z první světové války – online*. Vojenský historický ústav Praha. cit. 25. 6. 2020] Dostupné z <http://www.vhu.cz/kde-patrat-po-svych-predcich-padlych-zemrelych-nezvestnych-ranenych-ci-zajatych-v-prvni-svetove-valce/> [cit. 15. 6. 2020]

WILMOTH, J., R., ANDREEV, K., JDANOV D., GLEI, D., A. a T. RIFFE, 2017. *Methods Protocol for the Human Mortality Database*. [online]. Verze metodiky z 27. 11. 2017. 2017. [cit. 15.6.2020]. Dostupné z www: <http://www.mortality.org/Public/Docs/MethodsProtocol.pdf>

WINKLER, W., 1919. *Die Totenverluste der öst.-ung. Monarchie nach Nationalitäten. – Die Altersgliederung der Toten. – Ausblicke in die Zukunft.* Vídeň, 1919.

WRYCZA, T. a A. BAUDISCH, 2012. How life expectancy varies with perturbations in age-specific mortality. *Demographic Research* 27(13): 365-376. doi: 0.4054/DemRes.2012.27.13