

VYUŽITÍ NÁRODNÍHO ONKOLOGICKÉHO REGISTRU PRO MODELOVÁNÍ Vlivu SCREENINGOVÝCH PROGRAMŮ V CÍLOVÉ POPULACI: AGE-PERIOD-COHORT MODEL

DATA OF CZECH NATIONAL CANCER REGISTRY IN MODELLING OF EFFECT OF SCREENING PROGRAMS IN TARGET POPULATION: AGE-PERIOD-COHORT MODELS

GELNAROVÁ E.¹, NEUVIRTOVÁ L.¹, SVOBODNÍK A.¹, KOMOLÍKOVÁ L.¹, DANEŠ J.², SKOVAJSOVÁ M.³, BARTOŇKOVÁ H.⁴, MUŽÍK J.¹, KOPTÍKOVÁ J.¹, DUŠEK L.¹

¹ LÉKAŘSKÁ FAKULTA MASARYKOVY UNIVERZITY, BRNO

² VŠEOBECNÁ FAKULTNÍ NEMOCNICE V PRAZE, RADIOLOGICKÁ KLINIKA

³ DTC PRAHA, MAMMACENTRUM

⁴ MASARYKŮV ONKOLOGICKÝ ÚSTAV BRNO

Souhrn

Článek analyzuje význam dat Národního onkologického registru ČR z hlediska možnosti jejich využití pro modelování a predikci vývoje incidence a mortality karcinomu prsu. Bližší pozornost je věnována problematice tzv. Age-Period-Cohort modelů (AP, AC, APC modely). Kromě modelového odhadu incidence a mortality jsou významným výstupem těchto modelů také predikce vývoje, což má značný administrativní i vědecký význam. Na základě AP modelu pro mortalitu karcinomu prsu v české populaci byly predikovány mortality pro období 1997–2001. Srovnání hodnot předikovaných AP, AC nebo APC modely se skutečně pozorovanými incidencemi a mortalitami prokázalo významnost vlivu působících faktorů, které ovlivňují věkově specifickou incidenci a mortalitu. Práce je úvodním metodickým příspěvkem do hodnocení programu screeningu karcinomu prsu v České republice.

Klíčová slova: mamografický screening, age-period-cohort modely.

Summary

The paper presents analysis of data of Czech National Cancer Registry from the viewpoint of modelling of incidence and mortality related to breast carcinoma. Special attention is paid to Age-Period-Cohort models (AP, AC, APC models). Both model estimates and predictions produced by these algorithms are of administrative and scientific importance. AP models were sufficiently applied to predict breast carcinoma mortality for the period 1997 – 2001. Comparative evaluation of model predictions with real data indicated significant impact of many factors that can interact with age-specific incidence and mortality in this disease. Paper is pilot methodical contribution to the evaluation of Czech national breast cancer screening program.

Key words: screening, mammography, age-period-cohort models.

ÚVOD

Instituce Národního onkologického registru (NOR) shromažďuje údaje o výskytu nádorových onemocnění v České republice od roku 1976. Kromě evidence základních dat o typu a lokalizaci nádorového onemocnění jsou v databázi NOR shromažďovány také údaje o datu úmrtí onkologických pacientů. Tyto informace jsou od určitého období validovány prostřednictvím údajů z registru zemřelých. Tyto údaje jsou velmi cenné nejen z hlediska deskripce časových trendů v incidenci a mortalitě nádorových onemocnění v ČR a pro srovnání statistických údajů se zahraničím, ale umožňují rovněž hodnotit vliv národních programů primární a sekundární prevence v cílových populacích. V následujícím textu analyzujeme význam dat NOR z hlediska možnosti jejich využití pro modelování a predikci vývoje incidence a mortality karcinomu prsu. Bližší pozornost je věnována problematice tzv. Age-Period-Cohort modelů.

TEORIE

Specifické problémy modelování dat incidence a mortality

Incidence je definována jako počet nových případů (např. onemocnění) zjištěných ve sledovaném období v hodnocené populaci (nejčastěji na 100 000 obyvatel dané populace). Incidenci lze vypočítat podle následujícího vzorce:

$$incidence = \frac{\text{počet nových případů ve sledovaném období}}{\text{počet člověkoroků v riziku}} \times 100\,000$$

Pokud je sledované období dlouhé pouze jeden rok, pak ve jmenovateli zlomku dostáváme pouze počet osob v sledované populaci. Pokud je sledované období delší, pak jmenovatel zlomku tvoří počet **člověkoroků** v riziku. Počet člo-

věkoroků dostaneme sečtením délek sledování všech osob v riziku (např. sledovaná populace měla dva subjekty A a B; sledované období bylo 3 roky, osoba A byla sledována celé 3 roky, osoba B byla sledována 2 roky ze tří sledovaných, celkový počet člověkoroků je tedy $3+2=5$). Počet člověkoroků je tedy vážený součet osob v rizikové, sledované populaci, kde váhy jsou délky sledování dané osoby. Sledujeme-li počet úmrtí v populaci za dané období, hovoříme o tzv. **mortalitě**.

$$\text{mortalita} = \frac{\text{počet zemřelých osob ve sledovaném období}}{\text{počet člověkoroků v riziku}} \times 100\,000$$

Pokud sledovaná populace značí jen osoby v definovaném věkovém intervalu, pak mluvíme o **věkově specifické incidenci (věkově specifické mortalitě)**. Pokud do studie zahrnujeme osoby z celé populace, mluvíme o **hrubé incidenci (mortalitě)** a pokud počítáme jen s určitou skupinou osob z blíže definované oblasti (zaměstnanci podniku, obyvatelé určitého města, atd.), mluvíme o **kohortové incidenci (mortalitě)**.

Incidenci (mortalitu) vyjadřujeme buď jako **absolutní** počet případů, nebo jako **relativní incidenci (mortalitu)**. **Relativní incidence (mortalita)** je číslo, které nabývá hodnoty od 0 do 1 a vypočte se jako prostý podíl počtu nových případů (úmrtí) a počet člověkoroků v riziku:

$$\text{relativní incidence mortalita} = \frac{\text{počet nových případů (úmrtí) ve sledovaném období}}{\text{počet člověkoroků v riziku}} \times 100\,000$$

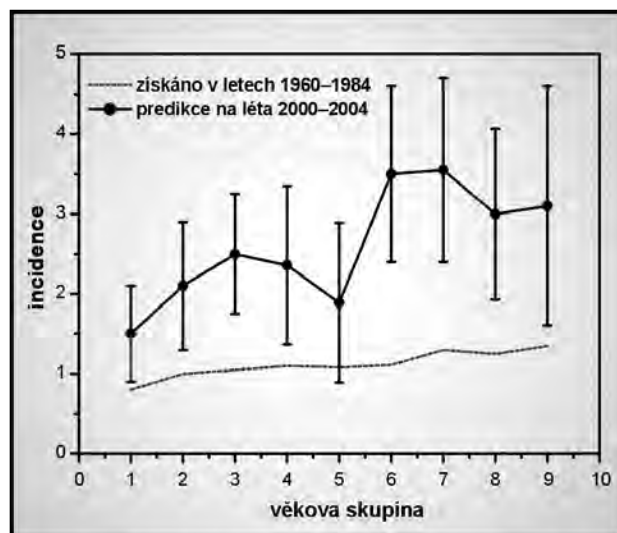
Absolutní incidence a mortalitu (věkově specifickou, hrubou, kohortovou) lze modelovat stejnými nástroji, neboť se z matematického hlediska jedná o veličiny stejného charakteru – vždy se jedná o počty nových případů (nemoci, úmrtí), které jsou vztaženy na 100 000 obyvatel.

Časový vývoj věkově specifických incidencí a mortalit v závislosti na datu diagnózy lze korektně modelovat několika odlišnými způsoby, například prostřednictvím několika regresních modelů (pro každou věkovou skupinu zvlášť), prostřednictvím časových řad (Tiwari, 2004), nebo pomocí tzv. Age-Period-Cohort (APC) modelů. Modelování s využitím několika oddělených regresních modelů není nejvhodnější metodou, protože může vést (např. Dyba, 1997) k významným zkreslením. Vlivem existujících fluktuací v datech mohou být výsledné predikce pro některou kategorii podhodnocené a neodpovídat očekávaným hodnotám.

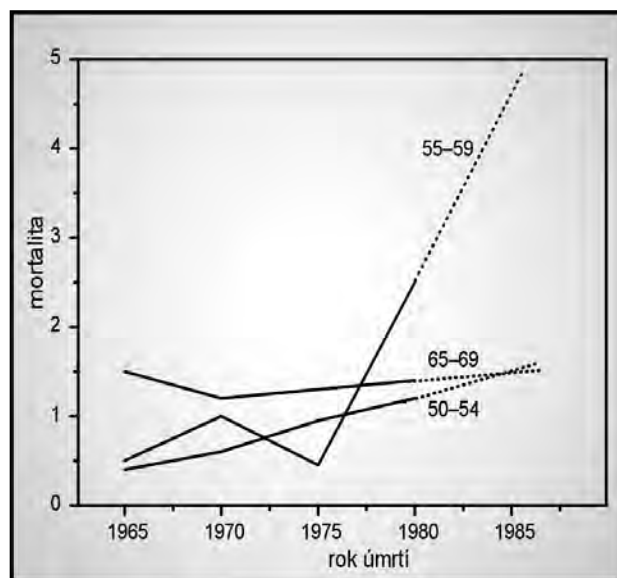
Dyba ve své práci analyzoval onkologická data z regionu Stockholm-Gotland ve Švédsku. Byly srovnávány průměrné věkově specifické incidence maligního melanomu žen v letech 1960–1984 a incidence predikované pro období 2000–2004. Predikce incidencí pro 5. věkovou skupinu byly významně nižší než pro 2., 3. a 4. věkovou skupinu, ačkoliv je známo, že starší věkové skupiny mají vyšší incidence než mladší (Obrázek č.1). Obdobný případ, kdy místo regresních modelů byla pro predikci budoucího vývoje použita prostá extrapolace (protažení trendu posledních dvou pozorování) ukazuje Bray, 2000 – vývoj mortality rakoviny hrtanu v závislosti na datu diagnózy pro vybrané

věkové skupiny maďarských žen (Obrázek č. 2). Predikce extrapolací byla pro skupinu žen ve věku od 55 do 59 let zcela zavádějící.

APC modely vycházejí oproti výše uvedeným z předpokladu, že „věkový profil“, (proporcionální vztah mezi věkově specifickými incidencemi) se v čase nemění. Data nejsou modelována odděleně pro různé věkové skupiny (jako v případě některých regresních modelů), ale jsou vyhodnocována jako jeden celek. Proto jsou APC modely vhodným nástrojem pro modelování časového vývoje věkově specifických incidencí a mortalit nádorových onemocnění.



Obrázek 1: Ilustrační obrázek, upraveno podle Dyba, 1997. T.Dyba: Melanom, ženy, onkologický region Stockholm-Gotland. Přerušovaná čára – průměrné, věkově specifické incidence v letech 1960–1984. Plná čára – predikce incidencí pro periodu 2000–2004.



Obrázek 2: Ilustrační obrázek, upraveno podle Bray, 2000. I.Bray: Rakovina hrtanu, ženy, Maďarsko. Plná čára – pozorované mortality. Přerušovaná čára – predikce mortalit na základě extrapolace dvou posledních pozorování.

Tabulka 1 – Lexisův diagram: Jednotlivé kohorty tvoří diagonály Lexisova diagramu. Tučně je zvýrazněna kohorta označená 1957.

Věková kategorie (průměrný věk) – AGE	Kategorie dle roku diagnózy (střední datum diagnózy) – PERIOD				
	1977–1981 (1979)	1982–1986 (1984)	1987–1991 (1989)	1992–1986 (1994)	1987–2001 (1999)
20 – 24 (22)	1957	1962	1967	1972	1977
25 – 29 (27)	1952	1957	1962	1967	1972
30 – 34 (32)	1947	1952	1957	1962	1967
35 – 39 (37)	1942	1947	1952	1957	1962
40 – 44 (42)	1937	1942	1947	1952	1957
45 – 49 (47)	1932	1937	1942	1947	1952
50 – 54 (52)	1927	1932	1937	1942	1947
55 – 59 (57)	1922	1927	1932	1937	1942
60 – 64 (62)	1917	1922	1927	1932	1937
65 – 69 (67)	1912	1917	1922	1927	1932
70 – 74 (72)	1907	1912	1917	1922	1927
75 – 79 (77)	1902	1907	1912	1917	1922
80 – 84 (82)	1897	1902	1907	1912	1917
85 – 89 (87)	1892	1897	1902	1907	1912

Age-period-cohort modely

Vývoj incidencí a mortalit je ovlivněn mnoha rizikovými faktory, jako např. životním stylem, stavem životního prostředí, dědičnými faktory apod. Některé působící faktory nelze uspokojivě kvantifikovat (např. množství tuku v potravě), jiné nejsou zatím validované, nebo je jejich význam v etiologii onemocnění diskutabilní (např. vliv elektromagnetického záření z mobilních telefonů). Proto v APC modelech uvažujeme čas jako univerzální faktor, který zastupuje vícero působících faktorů.

V APC modelech je modelována incidence (mortalita) v závislosti na třech časových faktorech: **age** (věk při diagnóze, úmrtí), **perioda** (rok diagnózy, úmrtí) a **cohort** (rok narození). Existuje jednoznačný vztah mezi rokem narození, věkem při diagnóze a rokem diagnózy:

$$\text{cohort} = \text{period} - \text{age}$$

Nejčastěji se však data zkoumají v agregované podobě, například za pětiletá období. V takovém případě hovoříme sledování tzv. souhrnná incidence. Vzájemný vztah mezi věkem při diagnóze (nebo úmrtí), rokem diagnózy (úmrtí) a rokem narození již nelze vyjádřit vzorcem, ale pomocí tzv. **Lexisova diagramu**, který je znázorněn v Tabulce 1. Řádky Lexisova diagramu jsou tvořeny jednotlivými věkovými kategoriemi (age), sloupce jsou tvořeny kategoriemi dle data diagnózy (period). Každé věkové kohortě odpovídá jedna diagonála tabulky.

Věková kohorta označená **1942** zahrnuje skupinu lidí, jejichž datum narození mohlo být od 1. 1. 1937 do 31. 12. 1947. Věková kohorta označená **1947** zahrnuje skupinu lidí, jejichž datum narození mohlo být od 1. 1. 1942 do 31. 12. 1952. Vidíme, že jednotlivé kohorty tvoří skupiny lidí, jejichž data narození se vzájemně překrývají – jsou to uměle vytvořené skupiny. Je-li znám věk a rok diagnózy (věková kategorie a perioda), je kohorta jednoznačně určena.

APC modely slouží k modelování závislosti incidence na faktorech věku, periody a kohorty. Data nejsou modelována odděleně pro různé věkové skupiny, ale jsou vyhodnocována jako jeden celek. Předpokládáme, že známe incidence pro jednotlivé věkové kategorie (A věkových

kategorií), periody (P period) a kohorty (C kohort) tak, jak je uvedeno v Lexisově diagramu. Pro Tabulku 1 platí, že $A=14$, $P=5$ a $C=18$. Faktory věku, periody a kohorty jsou čísla a mají v APC modelech následující interpretace a značení:

- I. **Vliv věku** (α_i) reprezentuje různá rizika spojená s různými věkovými skupinami – starší věkové skupiny jsou více ohroženy nádorovými onemocněními.
- II. **Vliv periody** (β_j) reprezentuje změny v trendu incidence způsobené vlivy, které působily na všechny věkové skupiny a kohorty současně. Mezi takovéto vlivy lze pro ilustraci zařadit například ozáření obyvatel Hirošimy po svržení atomové pumy – inkubační doby případů leukemie způsobené ozářením jsou popsány v Cobb et al, 1959.
- III. **Vliv kohorty** (γ_k) je spojen s dlouhodobými zvyky a návyky – například obvyklý počet dětí, životospráva nebo užívání antikoncepce. Například v současné době se v ČR výrazně mění věk při porodu prvního dítěte a celkový počet dětí. Mezi další dlouhodobé vlivy, které jsou obsaženy v efektu kohorty, patří celkové dlouhodobé znečištění ovzduší emisemi.

Postup při tvorbě APC modelů je následující. Nejdříve je zvolena jedna referenční perioda nebo kohorta (obvykle prostřední). Pro tuto referenční periodu se odhadnou pro každou věkovou kategorii efekty věků, α_i , $i=1, \dots, A$. Incidence pro jednotlivé věkové kategorie v referenční periodě, kohortě získáme exponováním efektů věků, $\exp(\alpha_i)$. Incidence pro ostatní referenční periody jsou odvozeny z incidencí v referenční periodě. Incidence v i -té věkové skupině a j -té periodě je rovna součinu $\exp(\alpha_i) \times \exp(\beta_j) \times \exp(\gamma_k)$, kde k značí příslušnou kohortu (kterou lze určit jednoznačně pomocí Lexisova diagramu).

Označme incidence v i -té věkové skupině a j -té periodě R_{ij} . Necht jsou data agregována do A věkových kategorií, P skupin dle roku diagnózy (period) a C kohort. Formálně lze APC model zapsat následujícím způsobem

$$R_{ij} = \exp(\alpha_i) \times \exp(\beta_j) \times \exp(\gamma_k), \quad i=1, \dots, A, \quad j=1, \dots, P, \quad k=1, \dots, C, \text{ obvyklá forma zápisu je}$$

$$\ln(R_{ij}) = \alpha_i + \beta_j + \gamma_k, \quad i=1, \dots, A, \quad j=1, \dots, P, \quad k=1, \dots, C.$$

V případě, že vliv kohorty není považován za významný, lze data modelovat pomocí jednoduššího Age-Period (AP) modelu. Také v tomto případě je nejdříve zvolena jedna referenční perioda. V této referenční periodě se odhadnou pro každou věkovou kategorii efekty věků, $\alpha_i, i=1, \dots, A$. Incidence pro jednotlivé věkové kategorie v referenční periodě získáme exponováním efektů věků, $\exp(\alpha_i)$. Incidence pro ostatní referenční periody jsou odvozeny z incidence v referenční periodě. Incidence v i -té věkové skupině a j -periodě je rovna pouze součinu $\exp(\alpha_i) \times \exp(\beta_j)$. Efekt kohorty není zahrnut. Pro referenční periodu je $\exp(\beta_{\text{ref}}) = 1$. Pro ostatní periody číslo $\exp(\beta_j)$ udává poměr mezi incidencemi v referenční periodě a zkoumané j -té periodě. Číslo $\exp(\beta_j)$ je stejné pro všech A zkoumaných věkových kategorií.

Formálně lze AP model zapsat následujícím způsobem

$$R_{ij} = \exp(\alpha_i) \times \exp(\beta_j), i=1, \dots, A, j=1, \dots, P,$$

obvyklá forma zápisu je

$$\ln(R_{ij}) = \alpha_i + \beta_j, i=1, \dots, A, j=1, \dots, P.$$

Analogicky je možné definovat Age-Cohort (AC) model, ve kterém není obsažen efekt periody. Konkrétní aplikaci APC, AP a AC modelů na modelování a predikci incidence a mortality v případě rakoviny prsu uvádíme v kapitole o praktické aplikaci této práce.

Úvod do problematiky AP, AC a APC modelů a řešení technických problémů spojených s odhady parametrů v APC modelech lze nalézt v řadě publikací (Clayton, Schifflers, 1987a, Clayton, Schifflers, 1987b, Holford, 2005, Carstensen et al., 2005, Robertson et al., 1998, Robertson et al., 1999). Existují volně přístupné programy a procedury pro odhady parametrů APC modelů (např. BAMP, Schmid, 2004, Carstensen et al., 2005). Model APC s efektem screeningu je popsán v Rostgaard et al., 2001.

Predikce incidence a mortality karcinomu prsu

Kromě vlastního modelování vývoje jsou zajímavé také predikce vývoje do budoucna na základě historicky naměřených dat (Møller, 2003). Predikce vývoje incidence má význam

- I. **Administrativní** – predikce by měly být co nejpřesnější, aby bylo možné odhadnout množství zdrojů potřebných pro léčbu a rehabilitaci.
- II. **Vědecký** – význam mají, kromě správných predikcí, i predikce, které selžou. Rozdíl mezi naměřenými daty a predikovanými hodnotami poukazuje na existenci rizikových faktorů, které nebyly vzaty v úvahu nebo na úspěšnost intervenčních programů, např. kampaně proti kouření nebo screening (Dyba, 2000).

Při predikci incidence do budoucna se předpokládá, že se působící faktory nezmění. Proto i proporce ve věkové struktuře incidence zůstanou stejné, tj. tvar „věkového profilu“ se nezmění. Dále se předpokládá, že celkový trend se nezmění a nárůst (nebo úbytek) incidence bude také stejný rychlý, jako byl ve sledovaném období.

Při predikování hodnot incidence pomocí APC modelu lze tedy předpokládat, že i v predikovaných budoucích periodách bude možné incidence v různých věkových skupinách odhadnout na základě odhadu incidence pro různé věkové kategorie v referenční periodě a kohortě $\exp(\alpha_i)$. Tyto hodnoty se opět vynásobí hodnotou $\exp(\beta_{j\text{pred}}) \times \exp(\gamma_{k\text{pred}})$, kde $j\text{pred}$ a $k\text{pred}$ značí příslušnou periodu a kohortu, jejichž incidence chceme predikovat.

Hodnotu β_{pred} lze predikovat například pomocí lineárního regresního modelu na několika posledních hodnotách β_j (Osmond, 1985) nebo lineární extrapolací posledních dvou hodnot β_j . Podobně lze predikovat i hodnoty $\gamma_{k\text{pred}}$. Kromě bodových odhadů incidence, lze spočítat i intervaly spolehlivosti pro tyto odhady (Naser B. Elkun, 2005). Jestliže skutečně pozorované hodnoty incidence budou ležet mimo tento interval, lze usoudit, že data generující mechanismus se liší a data byla ovlivněna novými faktory (např. efekt screeningu).

Faktory ovlivňující kvalitu modelu

Pro kvalitu modelu a predikci jsou důležité nejen výběr vhodného modelu a metody odhadu parametrů. Nejsou-li data, která slouží k tvorbě modelu kvalitní, pak také odhady parametrů modelu jsou nepřesné a zavádějící. Data mohou znehodnotit zásahy administrativního charakteru, jako např.

- I. změny v systému registrace dat,
- II. změny v klasifikaci (diagnóza C34, NOR, změna počtu stadií v roce 1982).

Charakter dat může být ovlivněn také

- III. změnami v diagnostice,
- IV. změnami v životním stylu, režimu.

Reálná data jsou výsledkem působení dvou vlivů (faktorů). První lze nazvat „systematický“ vliv (např. systematická změna v závislosti na čase pozorování). Druhý faktor lze nazvat „stochastický“, náhodný, (vliv působení náhody, náhodné fluktuace v datech). Agregace dat, použitá ve výše uvedeném případě je vhodná pro „vyhlazení“ náhodných fluktuací v datech. Nevýhodou je modelování v příliš dlouhých časových úsecích. Problémem při odhadech parametrů APC modelů je nestejný počet pozorování v jednotlivých kohortách. V Tabulce 1 je vidět, že pro kohortu 1957 je k dispozici pět pozorování, pro kohorty 1892 a 1977 pouze jedno. Proto jsou odhady γ_k v krajních kohortách méně přesné. Intervaly spolehlivosti pro odhady γ_k jsou velmi široké.

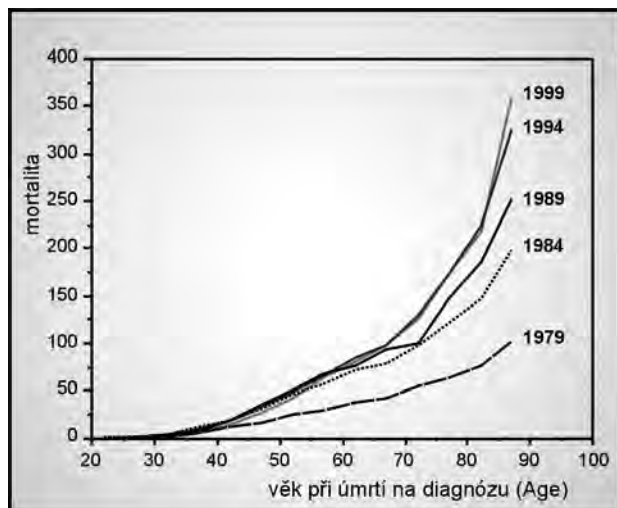
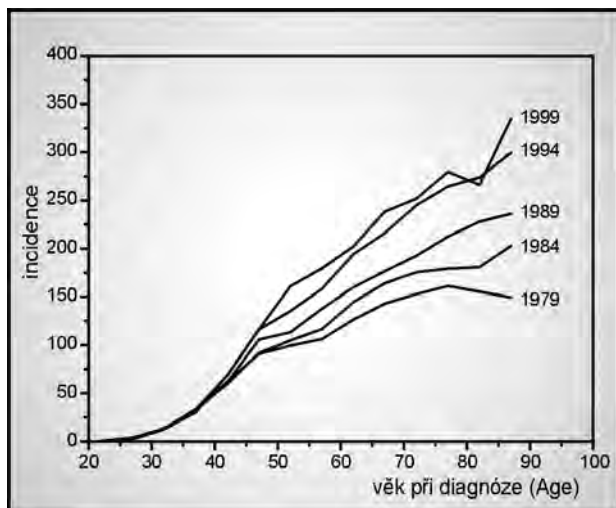
PRAKTICKÁ APLIKACE

V následující sekci budou analyzována data NOR ČR týkající se incidence a mortality rakoviny prsu. Jedná se o data z období 1977–2001, která jsou rozdělena podle věku do 14 věkových kategorií a 5 pětiletých period a 18 kohort. Struktura datového souboru je stejná jako v Tabulce 1. Ačkoli rakovina prsu postihuje i muže, jsou zde analyzována pouze data žen.

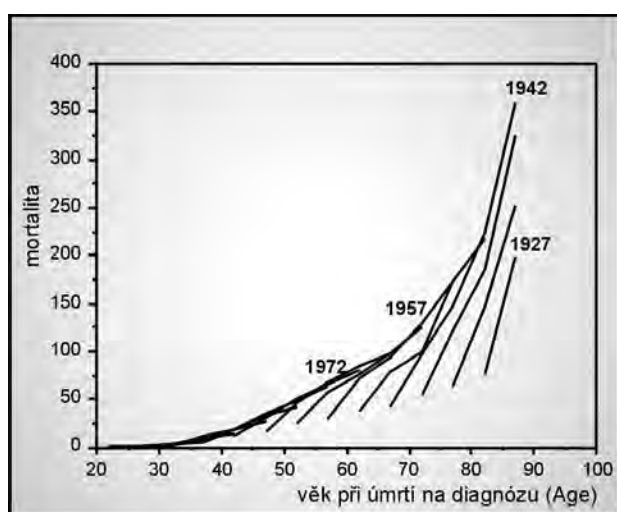
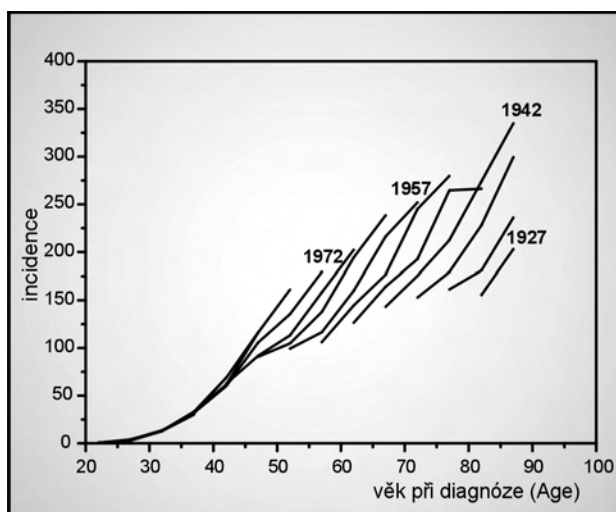
Vizualizace dat

Pro prvotní orientaci a náhled na vývoj výskytu onkologických onemocnění jsou dostatečné různé způsoby vizualizace incidence a mortality, v různých kategoriích dle pohlaví, věku nebo dalších demografických charakteristik. Na Obrázcích 3–6 jsou popisné grafy věkově specifických incidence (vlevo) v závislosti na periodě a kohortě a incidence pro různé periody a kohorty vykreslené v závislosti na věkové kategorii. Analogické grafy jsou uvedeny i pro mortalitu (vpravo).

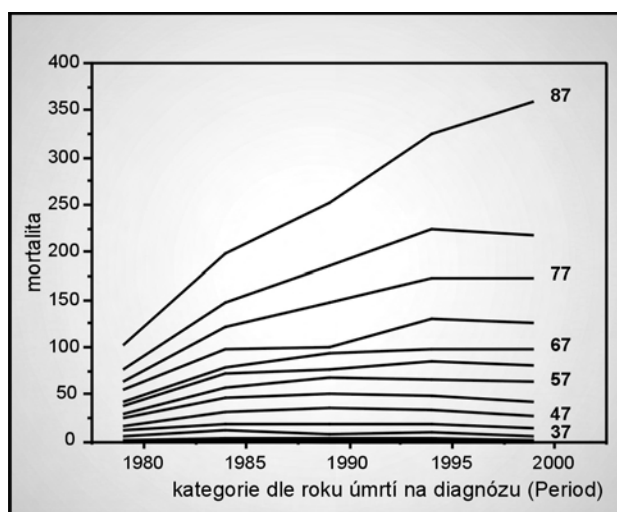
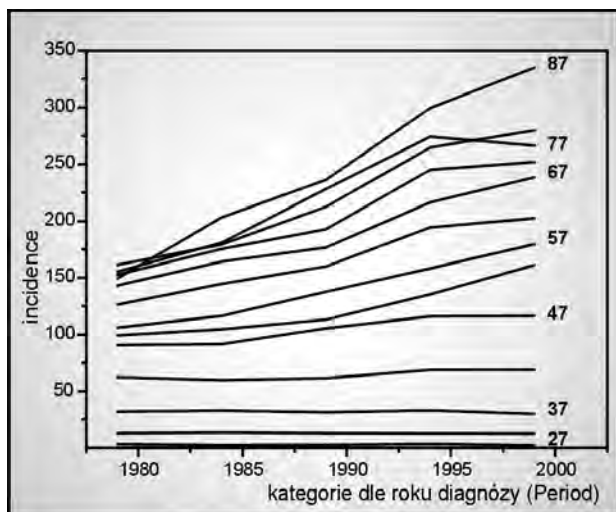
Na Obrázku 3 jsou zakresleny věkově specifické incidence pro různé periody. V obou případech (pro incidence i mortalitu) platí, že starší věkové skupiny jsou významně více



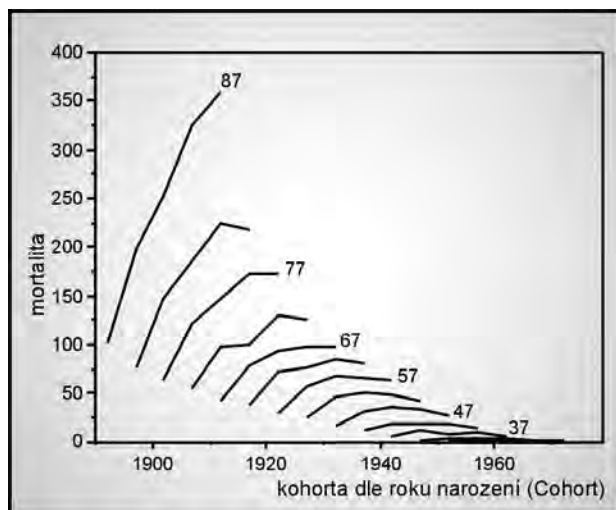
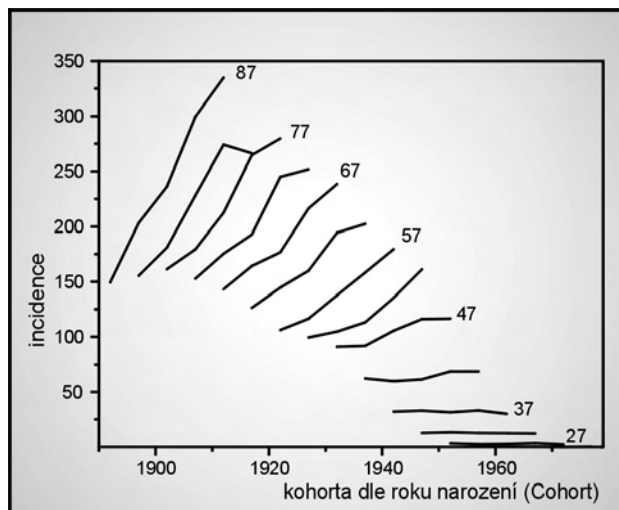
Obrázek 3.: Věkově specifické incidence (3a) a mortality (3b) pro diagnózu C50, kategorizace dle periody. Popisné grafy.



Obrázek 4.: Věkově specifické incidence (4a) a mortality (4b) pro diagnózu C50, kategorizace dle kohorty. Popisné grafy.



Obrázek 5.: Závislost věkově specifické incidence (5a) a mortality (5b) na datu diagnózy (Period). Popisné grafy.



Obrázek 6.: Závislost věkově specifické incidence (6a) a mortality (6b) na datu narození (Cohort). Popisné grafy.

ohroženy vznikem onemocnění. Na Obrázku 4 jsou opět zakresleny věkově specifické incidence, nejsou ale spojeny hodnoty příslušející jedné periodě, ale hodnoty příslušející jedné kohortě. Tyto grafy přibližují strukturu datového souboru. Pro starší kohorty (např. 1897) jsou k dispozici pouze data pro starší věkové skupiny (75–89 let). Pro mladší kohorty jsou k dispozici data pouze pro mladší věkové skupiny.

Na Obrázku 5 jsou zakresleny věkově specifické incidence v závislosti na datu diagnózy. V obou případech nastává nárůst ve starších věkových skupinách a pokles v mladších věkových skupinách. Nárůst a pokles je intenzivnější v případě mortality. Na Obrázku 6 jsou zakresleny věkově specifické incidence v závislosti na kohortě. Pro starší věkové kategorie platí: čím později se pacientka narodila, tím větší má šanci, že v určitém věku onemocní nebo zemře na rakovinu prsu. Například incidence rakoviny prsu ve věkové kategorii 72 let je u žen, které se narodily dříve (např. před rokem 1920) nižší než u žen, které se narodily později (např. po roce 1920). Pro mladší věkové kategorie platí v případě mortality opačný trend. Je patrný pokles mortality v nižších věkových kategoriích. Jako příklad uvedme mortalitu ve věkové kategorii 47 let. Ženy této věkové kategorie, které se narodily po roce 1947, mají menší mortalitu než ženy narozené v roce 1942 a jejich mortalita je srovnatelná s mortalitou před rokem 1937. Z Obrázku 6 je opět patrná struktura datového souboru.

Modelování incidence a mortality karcinomu prsu

Byly odhadnuty parametry AP, AC a APC modelů pro incidence a mortalitu. Na základě odhadu parametrů těchto modelů byly odhadnuty incidence a mortality ($\exp(\alpha_i)$) a $\exp(\beta_j)$ a $\exp(\gamma_k)$. Tyto odhady jsou uvedeny v Tabulce 2.

Odhady incidencí z AP modelu jsou zobrazeny na Obrázku 7. Odhady incidencí $\exp(\alpha_i)$ pro sledované věkové kategorie v referenční periodě 1987–1991 (střed referenční periody je 1989) jsou na obrázku č. 7 vlevo. Na Obrázku 7 vpravo jsou vykresleny odhady $\exp(\beta_j)$, které udávají poměr mezi incidencemi v referenční periodě a zkoumané j -té periodě. Pokud bychom chtěli získat podobný „věkový profil“ věko-

vě specifických incidencí jako máme pro referenční periodu, například pro periodu 1982–1986 (1985), pak stačí všechny odhady incidencí referenční periody vynásobit číslem $\exp(\beta_j) = 0.89$. Tlustou čarou jsou vykresleny bodové odhady incidencí a poměry mezi incidencemi v referenční periodě a zkoumané j -té periodě. Tenkou čarou jsou vykresleny dolní a horní hranice 95 % konfidenčního intervalu. Odhady incidencí z AC modelu jsou zobrazeny na Obrázku 8. Konfidenční intervaly pro $\exp(\gamma_k)$ pro krajní kohorty jsou velice široké. Šířka konfidenčního intervalu v krajních kohortách je způsobena příliš malým počtem pozorování v těchto kohortách (viz. Tabulka 1). Odhady $\exp(\gamma_k)$ nemají monotónní průběh, nejdříve rostou, což lze interpretovat jako soustavný nárůst incidencí. V roce 1942 ale dochází k poklesu odhadů $\exp(\gamma_k)$.

Na Obrázku 9 jsou odhady incidencí získané z APC modelu. Pro odhad parametrů APC modelu byla zvolena sekvenční a Holfordova metoda (Carstensen, Keiding 2005), (Holford, 1983).

Na Obrázku 10 jsou zobrazeny odhady mortality $\exp(\alpha_i)$ a poměrů mezi mortalitami v referenční periodě a zkoumané j -té periodě, $\exp(\beta_j)$ AP modelu. Odhady mortality z AC modelu jsou zobrazeny na Obrázku 11. Konfidenční intervaly pro $\exp(\gamma_k)$ pro krajní kohorty jsou opět velice široké. Odhady $\exp(\gamma_k)$ nemají monotónní průběh, nejdříve rostou, což lze interpretovat jako soustavný nárůst mortality ve starších věkových kategoriích. Od roku 1942 ale dochází k poklesu odhadů $\exp(\gamma_k)$. Pro ženy, které se narodily po roce 1942, mají věkově specifické mortality klesající tendenci. Tuto tendenci zatím lze pozorovat pouze u nižších věkových kategorií. Data pro starší věkové kategorie zatím nejsou k dispozici.

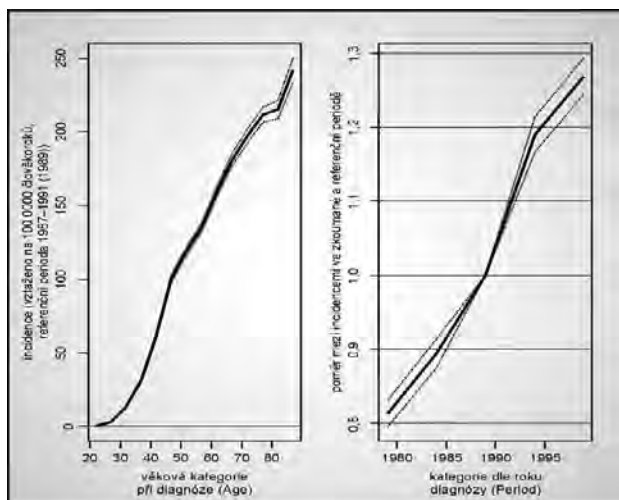
Na Obrázku 12 jsou odhady mortality získané z APC modelu. Pro odhad parametrů APC modelu byla zvolena sekvenční a Holfordova metoda (Carstensen, Keiding 2005), (Holford, 1983).

Orientační predikce mortality

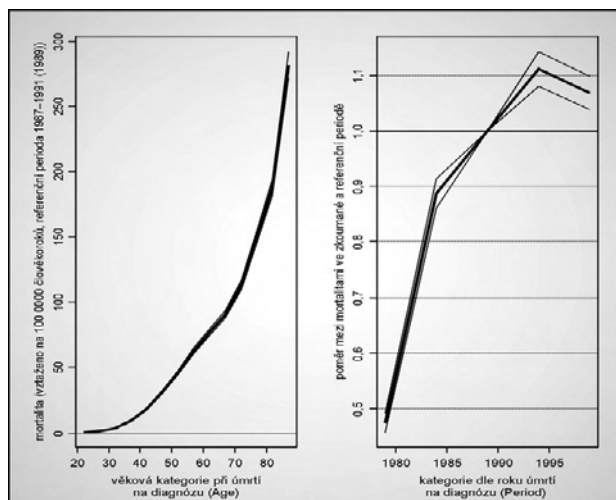
Na základě AP modelu pro mortalitu byly predikovány mortality pro periodu 1997–2001 (1999). Srovnání predikova-

Tabulka 2: Odhady incidencí a mortalit $\exp(\alpha_i)$ a dále $\exp(\beta_j)$ a $\exp(\gamma_k)$ pro AP, AC a APC modely.

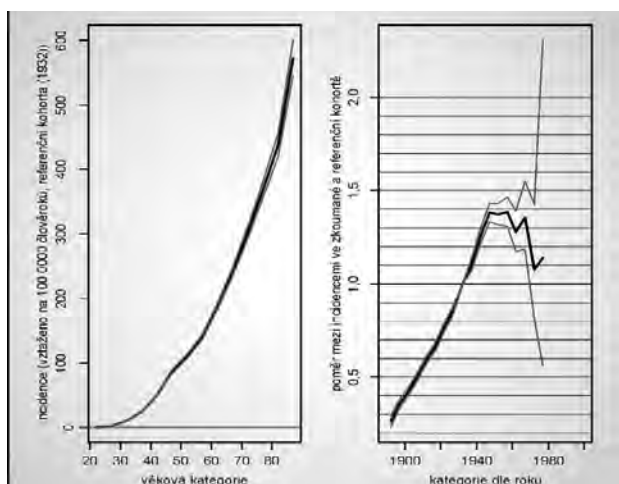
	Incidence			Mortalita		
	AP	AC	APC (sekvenční metoda)	AP	AC	APC
$\exp(\alpha_1)$	0,48	0,41	0,41	0,11	0,14	0,14
$\exp(\alpha_2)$	3,16	2,52	2,52	0,71	0,72	0,72
$\exp(\alpha_3)$	12,93	9,69	9,69	2,96	2,64	2,64
$\exp(\alpha_4)$	31,22	23,94	23,94	8,96	7,61	7,61
$\exp(\alpha_5)$	61,53	49,15	49,15	17,74	14,73	14,73
$\exp(\alpha_6)$	99,94	85,17	85,17	31,35	26,10	26,10
$\exp(\alpha_7)$	119,47	110,47	110,47	45,93	39,20	39,20
$\exp(\alpha_8)$	135,17	140,18	140,18	62,53	58,12	58,12
$\exp(\alpha_9)$	160,42	188,61	188,61	77,12	82,90	82,90
$\exp(\alpha_{10})$	181,47	241,57	241,57	90,64	109,64	109,64
$\exp(\alpha_{11})$	197,34	301,12	301,12	113,40	161,60	161,60
$\exp(\alpha_{12})$	212,00	369,67	369,67	149,49	257,61	257,61
$\exp(\alpha_{13})$	215,24	437,56	437,56	187,64	410,93	410,93
$\exp(\alpha_{14})$	242,32	572,65	572,65	281,86	751,44	751,44
$\exp(\beta_1)$	0,81	-	1,03	0,47	-	0,76
$\exp(\beta_2)$	0,89	-	0,98	0,89	-	1,16
$\exp(\beta_3)$	1,00	-	0,97	1,00	-	1,10
$\exp(\beta_4)$	1,19	-	1,03	1,11	-	1,6
$\exp(\beta_5)$	1,27	-	1,00	1,07	-	0,91
$\exp(\gamma_1)$	-	0,26	0,26	-	0,14	0,14
$\exp(\gamma_2)$	-	0,35	0,35	-	0,22	0,22
$\exp(\gamma_3)$	-	0,42	0,42	-	0,31	0,31
$\exp(\gamma_4)$	-	0,51	0,51	-	0,42	0,42
$\exp(\gamma_5)$	-	0,59	0,59	-	0,52	0,52
$\exp(\gamma_6)$	-	0,66	0,66	-	0,60	0,60
$\exp(\gamma_7)$	-	0,77	0,77	-	0,75	0,75
$\exp(\gamma_8)$	-	0,86	0,86	-	0,85	0,85
$\exp(\gamma_9)$	-	1,00	1,00	-	1	1
$\exp(\gamma_{10})$	-	1,10	1,10	-	1,09	1,09
$\exp(\gamma_{11})$	-	1,25	1,25	-	1,17	1,17
$\exp(\gamma_{12})$	-	1,38	1,38	-	1,21	1,21
$\exp(\gamma_{13})$	-	1,37	1,37	-	1,10	1,10
$\exp(\gamma_{14})$	-	1,38	1,38	-	1,05	1,05
$\exp(\gamma_{15})$	-	1,28	1,28	-	0,79	0,79
$\exp(\gamma_{16})$	-	1,36	1,36	-	0,76	0,76
$\exp(\gamma_{17})$	-	1,08	1,08	-	0,64	0,64
$\exp(\gamma_{18})$	-	1,14	1,14	-	0,34	0,34



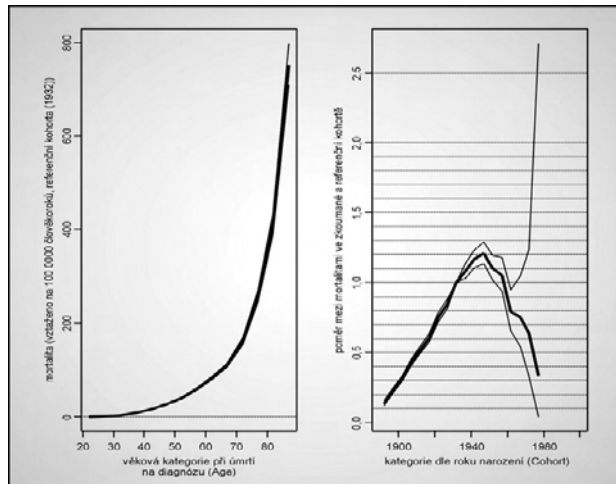
Obrázek 7.: Age-period model (incidence). Odhad incidencí v referenční periodě $\exp(\alpha_i)$ (vlevo) a poměrů mezi incidencemi v referenční periodě a zkoumané j -té periodě, $\exp(\beta_j)$ (vpravo).



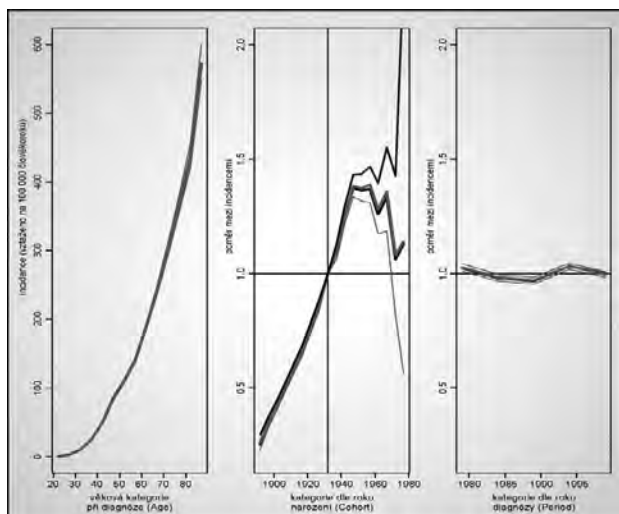
Obrázek 10.: Age-period model (mortalita). Odhad mortalit v referenční periodě $\exp(\alpha_i)$ (vlevo) a poměrů mezi mortalitami v referenční periodě a zkoumané j -té periodě, $\exp(\beta_j)$ (vpravo).



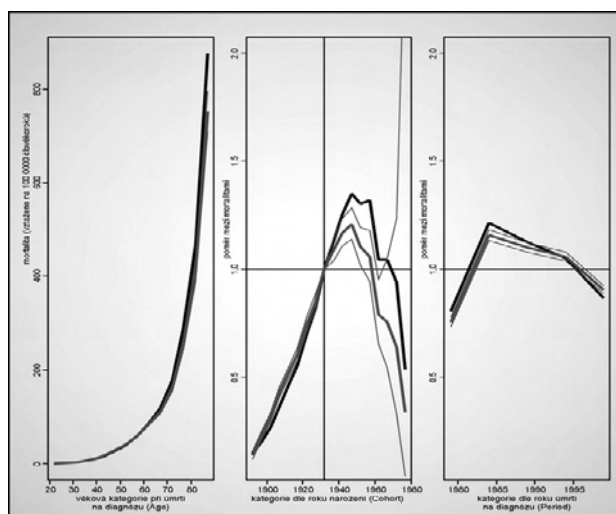
Obrázek 8.: Age-cohort model (incidence). Odhad incidencí v referenční cohortě $\exp(\alpha_i)$ (vlevo) a poměrů mezi incidencemi v referenční cohortě a zkoumané k -té cohortě, $\exp(\gamma_k)$ (vpravo).



Obrázek 11.: Age-cohort model (mortalita). Odhad mortalit v referenční cohortě $\exp(\alpha_i)$ (vlevo) a poměrů mezi mortalitami v referenční cohortě a zkoumané k -té cohortě, $\exp(\gamma_k)$ (vpravo).



Obrázek 9.: Age-period-cohort model (incidence). Černá čára – Holfordova metoda, šedá čára – sekvenční metoda.



Obrázek 12.: Age-period-cohort model (mortalita). Černá čára – Holfordova metoda, šedá čára – sekvenční metoda.

Tabulka 3: Skutečně pozorované mortality pro periodu 1997–2001 (1999) a predikce mortalit na základě AP modelu pro čtyři periody období 1977–1996.

Věková kategorie (průměrný věk) – AGE	Skutečně pozorované mortality	Mortality predikované na základě AP modelu
20 – 24 (22)	0,05	0,16
25 – 29 (27)	0,4	1,01
30 – 34 (32)	1,7	4,14
35 – 39 (37)	5,36	12,35
40 – 44 (42)	13,67	23,70
45 – 49 (47)	27,17	41,40
50 – 54 (52)	42,34	59,90
55 – 59 (57)	62,82	78,80
60 – 64 (62)	80,09	95,99
65 – 69 (67)	98,71	111,29
70 – 74 (72)	126,79	137,84
75 – 79 (77)	173,04	179,11
80 – 84 (82)	218,17	226,76
85 – 89 (87)	359,52	319,01

ných hodnot se skutečnými mortalitami je uvedeno v Tabulce 3. AP model byl vytvořen na základě pozorování čtyř period období 1977–1996. Predikované hodnoty mortalit se liší od skutečně naměřených mortalit. Skutečná mortalita v mladších věkových kategoriích byla nižší než byla mortalita predikovaná AP modelem. Nejstarší věková kategorie měla vyšší mortalitu než byla mortalita predikovaná AP modelem.

Data byla analyzována prostřednictvím nástroje pro statistické výpočty R 2.0.1 (library Epi s využitím procedur a materiálů Carstensen, 2005).

ZÁVĚR

Modelování incidencí a mortalit nádorových onemocnění pomocí AP, AC nebo APC modelů přináší ucelený pohled na data a pochopení jejich vývoje nejen ve sledovaném období, ale i do budoucna. Odhady parametrů modelů lze velmi dobře interpretovat a přehledně znázornit v grafech. Srovnání hodnot predikovanými AP, AC nebo APC modely se skutečně pozorovanými incidencemi a mortalitami může poukázat na významnost vlivu působících faktorů, které ovlivňují věkově specifickou incidenci a mortalitu.

LITERATURA

BAMP („Bayesian Age-Period-Cohort Modelling and Prediction“), www.stat.uni-muenchen.de/NSchmidt/bamp .

Bray I. (2000) Projecting Cancer Rates Using Age-Period-Cohort Models: A Bayesian Approach, poster, Annual General Meeting of International Biometric Society and 195th Ordinary Meeting of the British Region, 23rd November 2000.

Clayton, D., Schifflers, E. (1987): Models for temporal variation in cancer rates. I. Age-period and age-cohort models. *Statistics in medicine*, **6**, pp.449–467

Clayton, D., Schifflers, E. (1987): Models for temporal variation in cancer rates. II. Age-period-cohort models. *Statistics in medicine*, **6**, pp.469–481

Carstensen, B., Keiding, N. (2005) Age-Period-Cohort models: Statistical inference in the Lexis diagram, www.biostat.ku.dk/~bxc/APC .

Cobb, S., Miller, N., and Wald, N. (1959) On the estimation of the incubation period in malignant disease. *J.Chron.Dis.* 9:385–393.

Dyba, T., Hakulinen, T. and Päävrinta, L. (1997) A simple non-linear model in incidence prediction. *Stats. in Med.*, **16**, pp.2297–2309.

Dyba, T. (2000) *Precision of cancer incidence predictions based on the poisson distributed observations*, PhD. Thesis, Helsinki

Holford, T.R. (1983) The estimation of age, period, cohort effects for vital rates. *Biometrics*, **39**, pp. 311–324

Holford, T. R. (2005) in *Encyclopedia of biostatistics* (eds. Armitage P., Colton T.), Wiley

Moller, H, Fekjaer, H., Hakulinen, T., Sigvaldason, H., Storm, HH., Talbäck a Haldorsen, T. (2003) Prediction of cancer incidence in the Nordic Countries: empirical comparison of different approaches. *Statistics in medicine*, **22**, p. 2751–2766

Naser B. Elkun (2005): Predicting confidence intervals for the Age-period-cohort model. *Journal of Data Science*, **3** (2005), p. 403–414

Osmond, C. (1985) Using age, period and cohort models to estimate future mortality rates. *Int. J. Epidemiol.*, **14**, 124–129.

R Development Core Team (2004). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.

Robertson, Ch., Gandini, S., Boyle, P. (1999) Age-Period-Cohort Models: A Comparative Study of Available Methodologies. *J Clin Epidemiol*, vol. **52**, No. 6, pp. 569–583.

Robertson, Ch., Boyle, P. (1998) Age-Period-Cohort Analysis of Chronic Disease Rates. I.: Modelling Approach, *Stats. in Med.*, **17**, pp.1305–1323.

Rostgaard, K. et Al. (2001) Age-period-cohort modelling of breast cancer incidence in Nordic countries. *Statistics in medicine*, **20**, pp 47–61

Schmid, V. (2004) Bayesianische Raum-Zeit-Modellierung in der Epidemiologie <http://edoc.ub.uni-muenchen.de/archive/00003000/>

Tiwari, R.C. and col. (2004) A new Method of Predicting US and State-Level Cancer Mortality Counts for the Current Calendar Year, *CA Cancer J Clin*, **54**, pp. 30 – 40.