

VLIV VĚKOVÉ STRATIFIKACE NA ODHADY V KOHORTOVÝCH STUDIÍCH

Ladislav TOMÁŠEK¹

SÚRO, Praha

Abstract: In the absence of a control group, two methods are usually used for estimating the parameters of relative risk models in cohort studies. The broadly used internal approach consists in age-stratified estimates of baseline mortality, whereas the external approach makes use of appropriate external mortality data. The quality of the two approaches have been studied by simulation. The results have demonstrated that the internal estimates are potentially biased due to age stratification, particularly for models that include temporal modifying factors and in cohorts with a small proportion of low exposed individuals.

Резюме: В случае отсутствия контрольной группы в когортных изучениях для оценки параметров в моделях относительного риска применяют, как правило, два метода. Общеприменяемый внутренний подход исходит из оценки расслоенной по возрасту смертности гипотетической не экспонированной группы, в то время как внешний подход пользуется данными о смертности общей популяции. Качество обоих подходов изучали при помощи симуляций. Результаты показали, что внутренние оценки могут быть смещенными вследствие расслоения по возрасту, главным образом в моделях, включающих модифицирующие во времени факторы, и в когортах с низкой долей слабо экспонированных индивидов.

Úvod

Epidemiologické studie vlivu dlouhodobých expozic a zejména metody kvantifikace míry zdravotního rizika jsou založeny na prospektivních (kohortových) šetřeních. V nich se obecně sledují dvě populace - exponovaná a neexponovaná. Často však bývá obtížné a nákladné pořídit neexponovanou populaci, která by až na expozici byla srovnatelná s populací exponovanou. Proto se většina studií omezuje na analýzu rizika ve vztahu k různým úrovním expozic pouze u exponované skupiny. Hypotetická kontrolní neexponovaná skupina se pak konstruuje na základě projekce směrem k nulovým hodnotám expozic.

¹Projekt byl řešen s finanční podporou IGA MZ ČR reg.č.18591-3.

Při hodnocení profesionálního rizika se většinou uvažuje nezávisle proměnná vyjádřená jako kumulativní expozice. Ta však koreluje s délkou expozice, časový faktor vystupuje zákonitě při samotném šetření rizika v dané populaci, dále se projevuje prostřednictvím věku u většiny chorob, zejména rakovin, a také hraje určitou roli v procesu odeznívání rizika. Navíc je známo, že věk v době expozice (zejména radiační) má vliv na velikost kancerogenního rizika. Vzhledem k tomu, že čas je takto korelován s řadou proměnných, bývá často dosti obtížné správně interpretovat výsledky sledování.

Metody odhadu rizika

Odhady parametrů v modelech rizika vycházejí z dat uspořádaných ve tvaru tabulek, tříděných vzhledem ke kumulativní expozici, věku, případně dalším uvažovaným faktorům. V nejjednodušším uspořádání se předpokládá, že v každém poli třídění má počet (O) pozorovaných případů Poissonovo rozdělení, jehož parametr λ závisí na 'velikosti' sledovaného pole vyjádřené veličinou person-years (PY), což je součin délky sledování a počtu osob vstupujících do příslušné kategorie, dále na relativním výskytu (r) případů v obecné populaci bez expozice a na velikosti expozice (W). Parametr předpokládaného rozdělení lze tedy psát ve tvaru

$$\lambda = r(k) PY (1 + bW), \quad k = 1, \dots, K. \quad (1)$$

V tomto modelu se metodou maximální věrohodnosti odhadují parametry r a b , tj. předpokládaná úmrtnost neexponované kohorty (background rate) a koeficient b relativního rizika. K odhadu úmrtnosti neexponované kohorty se uvažuje třídění dle věku, resp. kalendářního období. Parametry $r(k)$ v každé z K věkových kategorií se odhadují odděleně, zatímco parametr b se uvažuje společný všem polím věkového třídění.

Alternativní postup odhadu spočívá ve využití informací o výskytu uvažované choroby v obecné populaci dané země či oblasti, které bývají většinou k dispozici, zejména pokud jde o úmrtnost. Protože obecná a projektovaná neexponovaná hypotetická populace nemusí vykazovat stejné zákonitosti (pokud jde o úmrtnost), zavádí se do modelu korekční multiplikativní člen (c) a model má tvar

$$\lambda = c E (1 + bW), \quad (2)$$

přičemž veličina E označuje tzv. očekávaný počet výskytu případů v obecné populaci odpovídající věku, pohlaví, kalendářnímu období, resp. oblasti. V modelu (2) se odhadují parametry c a b .

Postup (1) bývá označován jako interní přístup, postup (2) jako externí přístup (striktně externí přístup neuvažuje faktor c).

Modifikující faktory

Výše uvažované modely předpokládají, že riziko, zvýšené v důsledku expozice, bude stejné po celé období života od ukončení expozice (model konstantního rizika). V praxi se však ukazuje (zejména u dlouhodobých studií), že předpoklad konstantního rizika neplatí. Faktor časového odstupu od expozice se do modelu zavádí tak, že se místo jedné souhrnné veličiny kumulované expozice uvažuje několik veličin kumulované expozice pocházejících z různých časových období (time since exposure windows - TSE). Modely (1) a (2) pak přejdou do tvaru

$$\lambda = r(k) PY (1 + \sum b_j W_j), \quad (3)$$

$$\lambda = c E (1 + \sum b_j W_j), \quad (4)$$

kde W_j označují frakce expozic v časových obdobích a b_j jsou pak časově specifické odhady koeficientů relativního rizika. Tyto modely se označují zkratkou TSE.

Podobným způsobem se modeluje vliv věku při expozici. Expozice v časových obdobích se dále dělí podle toho, v jaké věkové kategorii docházelo k expozici. Modely (3) a (4) pak mají tvar:

$$\lambda = r(k) PY (1 + \sum \sum a_i b_j W_{ij}), \quad (5)$$

$$\lambda = c E (1 + \sum \sum a_i b_j W_{ij}), \quad (6)$$

přičemž se fixuje jedna z odhadovaných konstant, např. $b_1 = 1$. Veličiny W_{ij} označují kumulovanou expozici odpovídající i -té věkové kategorii a j -tému časovému období. Parametry a_i mají význam věkově specifických odhadů koeficientů relativního rizika vzhledem k expozicím v prvním časovém období a parametry b_j charakterizují relativní změnu koeficientů relativního rizika s časem. Tyto modely se označují zkratkou TSE+AE (time since exposure + age at exposure).

Simulační studie

Rozdílné výsledky mezi odhady uskutečněnými interním a externím postupem (Tab.1) ve studii rizika rakoviny plic vzhledem k pozdním účinkům expozice radonu v uranových dolech (kohortová studie Jáchymov - Ševc et al, 1991) vedly k otázce, který z postupů je správnější. Proto byla realizována tato simulační studie.

Table 1: Studie Jáchymov - odhady parametrů v modelu TSE+AE

	Interní			Externí			Populační	
	Odhad	CI		Odhad	CI		úmrtnost	Věk
r(1)	.072	.010	.541	1.71	1.23	2.37	.012	(15-29)
r(2)	.283	.128	.627				.046	(30-39)
r(3)	.761	.372	1.56				.336	(40-49)
r(4)	1.46	.709	3.00				1.57	(50-59)
r(5)	2.24	1.06	4.76				3.67	(60-69)
r(6)	1.72	.76	3.91				5.39	(70-84)
c								
q	.023	.009	.058	.086	.054	.137		
a_1	1.00			1.00			Věk při expozici	
a_2	1.25	.74	1.76	.64	.37	.91	15-29	
a_3	1.96	1.25	2.68	.25	.16	.33	30-39	
b_1	1.00			1.00			40-59	
b_2	.99	.72	1.27	.26	.18	.35	Doba od expozice	
b_3	.83	.58	1.09	.05	.02	.08	5-14	
							15-24	
							25-34	

CI označuje 95% interval spolehlivosti, odvozený za předpokladu asymptotické normality odhadů

V modelech (5) a (6) použitých v Tab.1 je fixován parametr $a_1 = 1$ a zaveden koeficient relativního rizika (q), který odpovídá expozici před 5 až 14 lety a věku při expozici do 30 let. Uvedené odhady, třebaže pořízené na základě stejného věkového a expozičního třídění, poskytují zcela rozdílné závěry. V případě interního přístupu je faktor doby od expozice nevýznamný a věk při expozici zvyšuje koeficienty rizika, zatímco externí přístup vede k závěru, že s dobou od expozice a vyšším věkem při expozici se koeficienty rizika snižují. Stojí za povšimnutí, že věkově specifické odhady úmrtnosti neexponované kohorty (na 1000 mužů a 1 rok) v modelu (5) jsou výrazně vyšší v mladším věku než v obecné populaci, zatímco ve starších věkových skupinách je tomu naopak.

Simulační studie pro tři základní popsané modely spočívaly v sestavení tří hypotetických kohort o rozsahu 16 000 osob exponovaných po dobu 5 až 15 let v různém věku. Distribuce věku nástupu byla obdobná, jako tomu bylo u sledované kohorty jáchymovských horníků (tj. 20-40 let). Expozi-

ce se pohybovaly v podobném pásmu jako ve skutečných dolech. U každé simulované osoby bylo v jednotlivých letech od vstupu do hypotetické kohorty simulováno úmrtí na rakovinu plic, resp. na ostatní příčiny (Tab.2), přičemž úmrtnost pro rakovinu plic brala v úvahu relativní zvýšení rizika v důsledku předchozí individuální expozice.

Table 2: Simulované úmrtnosti (na 1000 mužů)

Věk	Rakovina plic	Ostatní příčiny
-29	0.01	1.5
39-39	0.05	2.5
40-49	0.50	3.5
50-59	2	15
60-69	4	35
70-	5	150

U každé osoby a v každém roce sledování tedy mohly nastat tři výsledky: 1-úmrtí na rakovinu plic, 2-úmrtí z jiné příčiny, 3-pokračování ve studii.

Kromě výše uvedených tří simulovaných kohort pro jednotlivé modely rizika (konstantní, TSE, TSE+AE) byla sestavena tzv. kontrolní kohorta o stejném rozsahu, avšak s nulovou expozicí. Sumární charakteristiky těchto kohort jsou uvedeny v tabulce 3 .

Table 3: Základní charakteristiky simulovaných kohort

Model	PY	O	E	O/E	W
konst.riziko	589 348	4 949	787.8	6.28	185
TSE	592 443	3 757	820.7	4.58	185
TSE+AE	590 589	3 563	824.5	4.32	157
kontrolní kohorta	616 012	998	982.0	1.02	0

O = pozorované počty případů,

E = počty případů očekávané na základě předpokládané úmrtnosti,

W = průměrné hodnoty kumulované expozice v kohortě.

Výsledky a závěry

Na základě odhadů v simulovaných kohortách lze porovnat interní a externí přístup. Toto hodnocení je založeno na statistice chí-kvadrát, která porovnává maximálně věrohodné odhady s hodnotami skutečných parametrů. Zatímco v případě modelu konstantního rizika (Tab.4) vedou oba přístupy v podstatě ke shodným odhadům koeficientů rizika, je situace ve složitějších modelech (Tab.5 a 6) odlišná.

Table 4: Model konstantního rizika - simulovaná kohorta

	Skutečný parametr	Interní			Externí		
		Odhad	CI		Odhad	CI	
r(1)		.011	.002	.081			
r(2)		.055	.028	.105			
r(3)		.435	.236	.802			
r(4)		1.66	.888	3.10			
r(5)		3.45	1.85	6.44			
r(6)		4.14	2.22	7.75			
c					1.11	.65	1.90
b	<i>.020</i>	.024	.007	.042	.019	.007	.031
Test χ^2		0.31	<i>p</i> = 0.578		0.04	<i>p</i> = 0.842	

Table 5: Model TSE - simulovaná kohorta

	Skutečný parametr	Interní			Externí		
		Odhad	CI		Odhad	CI	
r(1)		.024	.007	.078			
r(2)		.057	.032	.102			
r(3)		.504	.290	.875			
r(4)		1.67	.954	2.92			
r(5)		2.97	1.69	5.22			
r(6)		3.80	2.14	6.75			
c					1.09	.67	1.77
<i>b</i> ₁	<i>.050</i>	.051	.020	.081	.051	.024	.078
<i>b</i> ₂	<i>.025</i>	.028	.010	.047	.023	.010	.035
<i>b</i> ₃	<i>.005</i>	.008	.001	.014	.005	.000	.009
Test χ^2		13.09	<i>p</i> = 0.004		3.94	<i>p</i> = 0.268	

V obou modelech s modifikujícími faktory (Tab.5 a 6) jsou odhady založené na interním přístupu signifikantně zkreslené. Společným rysem tohoto zkreslení je redukce, resp. eliminace (model 5) časových závislostí a obecně zkreslené odhady úmrtností bez expozice v jednotlivých věkových skupinách, kdy v nižších věkových skupinách jsou tyto odhady nadhodnoceny a ve vyšších věkových skupinách jsou podhodnoceny. Z tohoto důvodu

Table 6: Model TSE+AE - simulovaná kohorta

	Skutečný parametr	Interní			Externí		
		Odhad	CI		Odhad	CI	
r(1)		.040	.015	.108			
r(2)		.138	.101	.187			
r(3)		1.01	.819	1.26			
r(4)		2.66	2.20	3.21			
r(5)		3.77	3.20	4.44			
r(6)		4.80	4.17	5.53			
c					1.04	.92	1.18
q	<i>.100</i>	.031	.023	.042	.114	.097	.134
a_1	<i>1.00</i>	1.00			1.00		
a_2	<i>.70</i>	1.01	.85	1.17	.64	.55	.72
a_3	<i>.40</i>	.91	.74	1.07	.38	.32	.43
b_1	<i>1.00</i>	1.00			1.00		
b_2	<i>.30</i>	.46	.39	.53	.28	.24	.31
b_3	<i>.10</i>	.22	.18	.27	.09	.08	.10
Test χ^2		63.35	$p < 0.0001$		4.55	$p = 0.473$	

byly v kohortách časově modifikovaných modelů odhadovány parametry také za situace, kdy k původní kohortě byla připojena kontrolní neexponovaná kohorta. Ukázalo se, že přidání téměř tisíce neexponovaných případů sice zmenšilo výše popsání zkreslení (Tab.7 a 8), avšak rozdíly mezi skutečnými hodnotami parametrů a interními odhady zůstaly statisticky významné.

Table 7: Model TSE - odhady s využitím kontrolní kohorty

	Skutečný parametr	Interní			Externí		
		Odhad	CI		Odhad	CI	
r(1)		.025	.009	.066			
r(2)		.062	.047	.083			
r(3)		.570	.512	.634			
r(4)		1.95	1.80	2.12			
r(5)		3.59	3.33	3.86			
r(6)		4.78	4.35	5.24			
c					1.02	.96	1.08
b_1	<i>.050</i>	.044	.038	.049	.055	.049	.060
b_2	<i>.025</i>	.024	.021	.026	.024	.022	.027
b_3	<i>.005</i>	.006	.005	.006	.005	.004	.006
Test χ^2		9.59	$p = 0.022$		4.15	$p = 0.246$	

Table 8: Model TSE+AE - odhady s využitím kontrolní kohorty

	Skutečný parametr	Interní			Externí		
		Odhad	CI		Odhad	CI	
r(1)		.028	.011	.066			
r(2)		.082	.065	.104			
r(3)		.667	.601	.739			
r(4)		2.03	1.87	2.20			
r(5)		3.42	3.18	3.69			
r(6)		4.85	4.44	5.30			
c					1.02	.97	1.08
q	<i>.100</i>	.057	.049	.067	.116	.103	.131
a ₁	<i>1.00</i>	1.00			1.00		
a ₂	<i>.70</i>	.86	.74	.98	.64	.56	.72
a ₃	<i>.40</i>	.66	.56	.77	.38	.32	.43
b ₁	<i>1.00</i>	1.00			1.00		
b ₂	<i>.30</i>	.40	.34	.45	.28	.24	.31
b ₃	<i>.10</i>	.17	.14	.20	.09	.08	.10
Test χ^2		37.11	$p < 0.0001$		5.64	$p = 0.343$	

Simulační studie ukázala, že odhady založené na interním přístupu, které jsou běžně užívané v modelech relativního rizika (Lubin et al, 1994), mohou být zkreslené vlivem věkové stratifikace, a to zejména v modelech, které obsahují věkové nebo jiné časové faktory a v situacích, kdy je v kohortě relativně nízký podíl neexponovaných či málo exponovaných osob.

Literatura

Ševc J, Tomášek L, Plaček V. Riziko zhoubného novotvaru plic při inhalaci dceřiných produktů radonu. Čs.hygiena 36:3-13, 1991.

Lubin JH, Boice JD, Hornung RW, Edling C, Howe GR, Kunz E, Kusiak RA, Morrison HI, Radford EP, Samet JM, Woodward A, Xiang ZS, Pierce DA. Radon and lung cancer risk: A joint analysis of 11 underground miners studies. US Dept of Health and Human Services, NIH Publ No 94-3644, 1994.