



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

# Informační brožura k edukačnímu semináři

## Analýza klinických a epidemiologických dat: od přípravy dat až po vizualizaci výsledků

11.–12. září 2018  
Rakvice



**Editoři: Tomáš Pavlík, Jakub Gregor**



**MUNI  
MED** Institut  
biostatistiky  
a analýz

Ústav zdravotnických informací a statistiky České republiky  
Institut biostatistiky a analýz Lékařské fakulty Masarykovy univerzity  
Společné pracoviště



**Ústav zdravotnických informací a statistiky České republiky**  
**Institut biostatistiky a analýz Lékařské fakulty Masarykovy univerzity**  
Společné pracoviště

## **Analýza klinických a epidemiologických dat: od přípravy dat až po vizualizaci výsledků**

**11.–12. září 2018**  
**Rakvice**

**Editoři:**  
**Tomáš Pavlík**  
**Jakub Gregor**

Konání semináře a vydání této publikace byly podpořeny Evropským sociálním fondem, Operační program Zaměstnanost, projekt „Centrum pro rozvoj technologické platformy registrů Národního zdravotnického informačního systému, modernizace vytěžování jejich obsahu a rozšíření jejich informační kapacity: Rozvoj technologické platformy NZIS“ – CZ.03.4.74/0.0/0.0/15\_019/0002748.



Evropská unie  
Evropský sociální fond  
Operační program Zaměstnanost

Všechna práva vyhrazena. Žádná část této elektronické publikace nesmí být reprodukována nebo šířena v papírové, elektronické či jiné podobě bez předchozího písemného souhlasu vykonavatele majetkových práv k dílu, kterého je možné kontaktovat na adrese: Nakladatelství Masarykovy univerzity Munipress, Rybkova 19, 602 00 Brno.

© 2018 Masarykova univerzita

ISBN 978-80-210-9031-6

ISBN 978-80-210-9030-9 (brož. vaz.)

## Obsah

Úvod.....	4
Národní onkologický registr (NOR) .....	5
Základní epidemiologické charakteristiky kolorektálního karcinomu v ČR.....	11
Aplikace joinpoint regrese v epidemiologii .....	18
Principy klasifikace a nákladového ocenění hospitalizovaných pacientů .....	28
Využití regresních stromů při kultivaci DRG klasifikačního systému.....	36
Program screeningu karcinomu tlustého střeva a konečníku a hodnocení jeho výkonnosti...	41
Interaktivní online reporting v praxi .....	53

# Úvod

Vážení čtenáři, milí účastníci semináře,

po roce se vám opět dostává do ruky sborník příspěvků edukačního semináře, tentokrát na téma *Analýza klinických a epidemiologických dat: od přípravy dat až po vizualizaci výsledků*. Letošní ročník tematicky spojuje dva předchozí ročníky, které se zabývaly strukturou a obsahem dat Národního zdravotnického informačního systému a následně analýzou, zpracováním a vizualizací dat. Analýza zdravotnických dat je oblast neustále aktuální a její význam lze téměř donekonečna dokumentovat na příkladech aplikací matematických a statistických modelů v různých oblastech medicíny, epidemiologie a biologie. Analytické výstupy v podobě souhrnných reportů, ročenek, prezentací a publikací hrají klíčovou roli od plánování klinických studií pro vzácná onemocnění až po rozhodovací procesy na celostátní úrovni, které mohou ovlivnit životy milionů lidí. Co je v současné době možná ale ještě důležitější, je správná vizualizace a prezentace výsledků jakožto nezbytný předpoklad pro jejich korektní pochopení a interpretaci. S ohledem na dnešní uspěchanou dobu je navíc stále častěji žádoucí sdělení nejpodstatnějších výstupů a informací na omezeném prostoru a v omezeném čase, kdy bez adekvátních prezentačních dovedností a nástrojů pro interpretaci získaných výsledků cílové skupině roste riziko nepochopení nebo dokonce nesprávného výkladu toho, co chce analytik svými výsledky sdělit.

Letošní seminář proto bude problematiku obou předchozích ročníků propojovat a bude zaměřen prakticky tak, aby si studenti jednotlivé metodické přednášky vyzkoušeli na simulovaných datech nejen po stránce analytické, ale i po stránce prezentační a interpretační. Hlavní pozornost je stejně jako loni soustředěna na týmovou spolupráci. Jednotlivé úvodní příspěvky a jim odpovídající workshopy reflektují hlavní metodické oblasti, které Ústav zdravotnických informací a statistiky ČR řeší v rámci hodnocení onkologických onemocnění: epidemiologické modelování trendů, identifikace nákladných pacientů a hodnocení dostupnosti screeningového programu. Vyvrcholením programu semináře je pak prezentace výsledků jednotlivých týmů s využitím volně dostupných vizualizačních nástrojů.

Za celý organizační tým edukačního semináře se na setkání v Rakvicích těší

RNDr. Jiří Jarkovský, Ph.D.

RNDr. Martin Komenda, Ph.D.

RNDr. Ondřej Májek, Ph.D.

RNDr. Tomáš Pavlík, Ph.D.

odborní garanti semináře, ÚZIS ČR & IBA LF MU

# Národní onkologický registr (NOR)

Jan Mužík

*Ústav zdravotnických informací a statistiky ČR, Praha  
Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
e-mail: muzik@iba.muni.cz*

## Abstrakt

Národní onkologický registr (NOR) je nejdůležitějším zdrojem dat o epidemiologii novotvarů v České republice. Údaje jsou sledovány na populační úrovni kontinuálně od roku 1976 doposud, do roku 2016 včetně obsahuje NOR více než 2,4 milionu záznamů novotvarů. Data NOR jsou využívána nejen pro hodnocení epidemiologické zátěže (incidence, mortalita, prevalence), ale spolu s dalšími zdroji dat (údaje o zemřelých, údaje o hospitalizacích a léčbě) v rámci Národního zdravotnického informačního systému (NZIS), slouží také k hodnocení výsledků léčebné péče o pacienty s novotvary na populační úrovni a pro plánování léčebné péče o pacienty s novotvary pomocí predikčních modelů epidemiologie a léčebné zátěže.

## Klíčová slova

novotvary, registry, epidemiologie

## 1. Co je to Národní onkologický registr

### 1.1. Účel registru a legislativní ukotvení

Národní onkologický registr (NOR) je součástí Národního zdravotnického informačního systému (NZIS), jeho účelem je registrace onkologických onemocnění a periodické sledování jejich dalšího vývoje, tj. shromažďování dat, jejich verifikace, ukládání, ochrana a zpracování. NOR poskytuje souhrnné údaje pro statistické přehledy jak na národní, tak i mezinárodní úrovni, dále pro epidemiologické studie a zdravotnický výzkum. NOR je celoplošným populačním registrem, který navazuje na sledování novotvarů v populaci ČR zavedené v 50. letech 20. století a jako populační registr záznamů jednotlivých novotvarů je provozován ÚZIS ČR od roku 1976. Do roku 2016 včetně obsahuje NOR více než 2,4 milionu záznamů novotvarů.

Údaje NOR slouží také k podpoře včasné diagnostiky a léčby novotvarů a přednádorových stavů, ke sledování trendů jejich výskytu, příčinných faktorů a společenských důsledků. Na populační úrovni jsou hodnoceny také výsledky léčby novotvarů formou analýzy přežití.

Souhrnná data jsou podkladem pro tvorbu, realizaci a vyhodnocování preventivních zdravotnických programů a pro odhady potřebných finančních nákladů na zabezpečení komplexní onkologické péče.

Anonymní individuální data mohou být poskytnuta pro epidemiologické studie a zdravotnický výzkum v souladu s platnými zákony a se souhlasem správce NOR (ÚZIS ČR) a Rady NOR (odborný poradní orgán správce NOR).

NOR je členem IACR (Mezinárodní asociace onkologických registrů) v Lyonu, spolupracuje s Evropskou sítí onkologických registrů (ENCR) a udržuje kontakt s registry v zahraničí. NOR je nedílnou součástí komplexní onkologické péče v ČR.

Předávání údajů do NOR se řídí platnou legislativou, zejména těmito právními předpisy:

- Zákon č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování (zákon o zdravotních službách) ve znění pozdějších předpisů
- Vyhláška č. 373/2016 Sb., o předávání údajů do Národního zdravotnického informačního systému, včetně příloh
- Zákon č. 101/2000 Sb., o ochraně osobních údajů a o změně některých zákonů, ve znění pozdější pozdějších předpisů
- Nařízení Evropského parlamentu a Rady (EU) 2016/679 ze dne 27. dubna 2016 o ochraně fyzických osob v souvislosti se zpracováním osobních údajů a o volném pohybu těchto údajů a o zrušení směrnice 95/46/ES (obecné nařízení o ochraně osobních údajů)

Předávání údajů do NOR a jeho organizace jsou blíže specifikovány v metodice NZIS pro NOR, jejíž aktuální verze je k dispozici na stránkách ÚZIS ČR ([www.uzis.cz](http://www.uzis.cz)) v sekci věnované národnímu onkologickému registru.

## **1.2. Klasifikační systémy používané v NOR**

Při předávání údajů do databáze NOR je nutno vycházet z klasifikací platných v České republice:

### **Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů – 10. revize**

Zkráceně označovaná jako MKN-10, aktualizovaná verze platná od 1. 1. 2013, viz předpis č. 426/2012 Sb. – sdělení Českého statistického úřadu o aktualizaci Mezinárodní statistické klasifikace nemocí a přidružených zdravotních problémů (MKN-10).

Aktuální verze MKN-10 je k dispozici na adrese <http://www.uzis.cz/katalog/klasifikace/mkn>.

Během vývoje byly v NOR postupně používány MNK 8. revize (do roku 1978), 9. revize (do roku 1993) a nyní 10. revize (od roku 1994 do současnosti).

### **Mezinárodní klasifikace nemocí pro onkologii – 3. vydání**

Zkráceně označovaná jako MKN-O-3, vydal ÚZIS ČR v roce 2004. MKN-O-3 je v ČR zavedena od 1. 1. 2005, viz Sdělení ČSÚ č. 49 ze dne 17. 1. 2005, částka 11/2005 Sb.

MKN-O-3 představuje dvojí klasifikaci s kódovacími systémy jednak pro topografii, jednak pro morfologii. Topografický kód popisuje místo původu nádoru a používá stejné třímístné a čtyřmístné kategorie (kódy). To umožňuje vyšší specifitu pro kódování nemaligních nádorů, než je možné v MKN-10. Morfologický kód popisuje buněčný typ tvořící nádor, biologickou aktivitu nádoru (popisuje charakteristiky nádoru jako takového) a stupeň jeho diferenciace (grade) od okolní tkáně.

Elektronická verze této publikace je k dispozici na internetové adrese

<http://www.uzis.cz/katalog/klasifikace/mkn-mezinarodni-klasifikace-nemoci-pro-onkologii>.

Během vývoje byly v NOR postupně používány MKN-O 1. vydání (od roku 1983 do roku 1993), 2. vydání (od roku 1994 do roku 2004) a nyní 3. vydání (od roku 2005 do současnosti).

### **TNM klasifikace zhoubných novotvarů**

TNM je systém pro popis anatomického rozsahu novotvaru založený na určení tří složek: T – rozsah primárního nádoru, N – nepřítomnost či přítomnost a rozsah metastáz v regionálních lymfatických uzlinách, M – nepřítomnost či přítomnost vzdálených metastáz. Na základě T, N a M



určeného klinickým vyšetřením a/nebo pooperačním/patologickým vyšetřením je odvozeno stadium onemocnění.

Elektronická verze této publikace je k dispozici na internetové adrese <http://www.uzis.cz/katalog/klasifikace/tnm-klasifikace-zhoubnych-novotvaru>.

Během vývoje byly v NOR postupně používány TNM klasifikace 2. revize (do roku 1981), 3. revize (1982–1994), 4. revize (1995–2000), 5. revize (2001–2004), 6. revize (2005–2010) a nyní 7. revize (od roku 2011 do současnosti). TNM klasifikace 8. revize bude v datech NOR v platnosti od 1. 1. 2019.

### **1.3. Jak vznikají záznamy novotvarů v NOR**

Povinnost nahlásit novotvar do NOR má každý poskytovatel zdravotních služeb, který stanoví diagnózu, a každý poskytovatel zdravotních služeb, který odpovídá za léčení a dispenzarizaci onemocnění vybraných diagnóz ze skupiny novotvarů (bližší specifikace viz metodika NOR na stránkách registru):

- zhoubné novotvary (C00–C97),
- novotvary in situ (D00–D09),
- novotvary nejistého nebo neznámého chování (D37–D48).

Údaje jsou předávány v rozsahu podle formuláře „Incidence a léčba zhoubného novotvaru“ (viz obrázek 1), který je třeba vyplnit a předat na příslušné územní (okresní) pracoviště NOR podle místa trvalého bydliště pacienta, a to ve dvou termínech od data stanovení diagnózy: do jednoho měsíce 1. část formuláře o diagnóze a jejím stanovení, do osmi měsíců 2. část formuláře o léčbě novotvaru.

Územní pracoviště NOR následně zajistí kontrolu, validaci a kompletaci údajů uvedených na zaslaných hlášeních spolu s dalšími informacemi dostupnými ve zdravotnické dokumentaci a zapíše novotvar do centrální databáze pomocí aplikace NOR v prostředí registrů rezortu ministerstva zdravotnictví.

Poskytovatel zdravotních služeb, který poskytuje pacientovi s novotvarem dispenzární péči, má za povinnost vyplnit a zaslat příslušnému územnímu pracovišti NOR „Kontrolní hlášení novotvaru“, ve kterém se sleduje vývoj nádorového onemocnění (remise, relaps, atd.) a následovně prováděná protinádorová léčba od uzavření hlášení, nebo od posledního kontrolního hlášení. Kontrolní hlášení se požaduje ve stanovených intervalech jeden rok, dva, tři, čtyři, pět, sedm, deset let a dále vždy po pěti letech a při úmrtí.

## **2. Využití dat Národního onkologického registru**

### **2.1. Hodnocení epidemiologie novotvarů**

Nejběžnějším typem výstupů, které využívají data NOR, jsou analýzy zaměřené na epidemiologii definovaných nádorových onemocnění. Epidemiologie je standardně popisována pomocí incidence, mortality a prevalence, a to v trendech či za definované období, za celou populaci nebo za jednotlivé územní celky v rámci ČR (kraje, okresy atd.). Dále jsou uváděny věkové charakteristiky pacientů (věková struktura, věkově specifická incidence, mortalita či prevalence) a údaje o stadiu, tedy pokročilosti onemocnění při diagnóze (za období nebo v trendu). Na významu v posledních letech nabývá také informace o pořadí novotvaru zjištěného u pacienta – v populaci pozorujeme rostoucí podíl novotvarů, které byly diagnostikovány u osob, u nichž se již jiné nádorové onemocnění dříve vyskytlo.

### Incidence a léčba zhoubného novotvaru

Vypište trojčíslovinou sítě odřezovanou část.  
\*) lze zaškrtnout více možností

Pořadí ZN  Evid. č. ZN

Razítko zdravot. pracoviště, které diagnostikuje novotvar

Kodné číslo <input type="text"/>		Příjmení a jméno <input type="text"/>		M <input type="text"/> Z <input type="text"/>	
Trvalé bydliště: Obec, okres, PSC <input type="text"/>				Datum narození <input type="text"/>	
(do registru se přenáší jen obec)					
Sociální postavení: 1 <input type="checkbox"/> zaměstnanec 2 <input type="checkbox"/> samost. pracující 3 <input type="checkbox"/> nepracující 4 <input type="checkbox"/> nezaměstnaný 5 <input type="checkbox"/> bezdomovec 9 <input type="checkbox"/> neznámo					
Stav (žijící) 1 <input type="checkbox"/> v rodině 2 <input type="checkbox"/> osamocené 9 <input type="checkbox"/> neznámo		Země původu (slovně) <input type="text"/>			
Zjištěno při: 1 <input type="checkbox"/> skrining 3 <input type="checkbox"/> klin. manif. 5 <input type="checkbox"/> jiný způsob		Datum 1. návštěvy lékaře <input type="text"/>		Datum stanovení diagnózy <input type="text"/>	
2 <input type="checkbox"/> prev. prohl. 4 <input type="checkbox"/> pitva 9 <input type="checkbox"/> neznámo					
Kouření: 1 <input type="checkbox"/> ano 3 <input type="checkbox"/> bývalý kuřák 2 <input type="checkbox"/> ne 9 <input type="checkbox"/> neznámo		Laterality: 3 <input type="checkbox"/> oboustranně 1 <input type="checkbox"/> vpravo 4 <input type="checkbox"/> obojí 2 <input type="checkbox"/> vlevo 9 <input type="checkbox"/> neznámo		K dg. vedla tato vyšetř. *) 02 <input type="checkbox"/> operace 16 <input type="checkbox"/> endoskopie 00 <input type="checkbox"/> klin. jasné 04 <input type="checkbox"/> cytologie 32 <input type="checkbox"/> lab. vys., markery 01 <input type="checkbox"/> histologie 08 <input type="checkbox"/> RTG, zobr. met. 64 <input type="checkbox"/> pitva 99 <input type="checkbox"/> DCO	
Diagnóza (slovně) <input type="text"/>					
Morfologie hist. <input type="checkbox"/> cyt. <input type="checkbox"/>					
TNM T <input type="text"/> mm <input type="text"/> N <input type="text"/> M <input type="text"/> pTNM y <input type="text"/> pT <input type="text"/> mm <input type="text"/> pN <input type="text"/> pN <input type="text"/> pN <input type="text"/> pN <input type="text"/> pN <input type="text"/> pM <input type="text"/> JTC <input type="text"/>					
Lokalizace metastáz *) 0001 <input type="checkbox"/> plicé 0004 <input type="checkbox"/> kost 0016 <input type="checkbox"/> játra 0064 <input type="checkbox"/> mozek 0256 <input type="checkbox"/> uzliny 1024 <input type="checkbox"/> jiný orgán 0002 <input type="checkbox"/> kostní dřeň 0008 <input type="checkbox"/> pleura 0032 <input type="checkbox"/> peritoneum 0128 <input type="checkbox"/> nadledviny 0512 <input type="checkbox"/> kůže 0000 <input type="checkbox"/> neznámo					
Riziková kategorie (trojčíslo) <input type="text"/> Sérové nádorové markery (jen u varlat) <input type="checkbox"/> Onemocnění					
Klinické stadium (0, I, II, III, IV), 6 metastázy u nezn. prim., 7 neuvádí se, 9 neznámo 1 <input type="checkbox"/> lokalizované 2 <input type="checkbox"/> pokročilé 3 <input type="checkbox"/> neznámo					
Praktický lékař, u kterého je pacient registrován <input type="text"/>					
Lečebný postup navržený pracovištěm zodpovědným za léčbu: Datum převzetí do péče <input type="text"/>					
Razítko zdravot. pracoviště zodpovědného za léčbu <input type="text"/>					
Datum zahájení <input type="text"/>					
Jméno a podpis hlavního lékaře <input type="text"/>					
Operace: operační výkon (slovně) <input type="text"/>					
zdr. zařízení a oddělení <input type="text"/>					
Datum operace <input type="text"/>					
Druh operace 1 <input type="checkbox"/> extirpace nádoru 3 <input type="checkbox"/> radik. operace s odst. regionálních uzlin 7 <input type="checkbox"/> pokus o nebo odlehčovací operace 0 <input type="checkbox"/> neoperován 2 <input type="checkbox"/> odstranění orgánu s nádorem 4 <input type="checkbox"/> radik. odst. uzlin bez zásahu na prim. nádor 9 <input type="checkbox"/> neznámo					
Odstranění nádoru 0 <input type="checkbox"/> R0 žádný reziduální nádor 2 <input type="checkbox"/> R2 makroskop. rezid. nádor 9 <input type="checkbox"/> neznámo 1 <input type="checkbox"/> R1 mikroskop. rezid. nádor 3 <input type="checkbox"/> nádor ponechán Endoskopická operace 1 <input type="checkbox"/> ano 3 <input type="checkbox"/> neznámo 2 <input type="checkbox"/> ne					
Radioterapie (slovně) <input type="text"/>					
zdr. zařízení a oddělení <input type="text"/>					
Datum zahájení radioterapie <input type="text"/>					
Druh záření *) 01 <input type="checkbox"/> RTG terapie 04 <input type="checkbox"/> terapie kopenskulárním zářením 16 <input type="checkbox"/> brachyterapie 00 <input type="checkbox"/> neozařován 02 <input type="checkbox"/> teleterapie Co 60, Cs 137, Gama nůž 08 <input type="checkbox"/> lineární urychlovač 32 <input type="checkbox"/> terapie otevřeným zářením					
Forma léčby 1 <input type="checkbox"/> předoperační 3 <input type="checkbox"/> před i pooperační 9 <input type="checkbox"/> neznámo 2 <input type="checkbox"/> pooperační 4 <input type="checkbox"/> samostatná (bez operace) Konkomitance s chemoterapií 1 <input type="checkbox"/> ano 2 <input type="checkbox"/> ne					
Chemoterapie (slovně) <input type="text"/>					
zdr. zařízení a oddělení <input type="text"/>					
Datum zahájení chemoterapie <input type="text"/>					
Forma léčby 1 <input type="checkbox"/> předoperační 3 <input type="checkbox"/> před i pooperační 9 <input type="checkbox"/> neznámo 2 <input type="checkbox"/> pooperační 4 <input type="checkbox"/> samostatná (bez operace) 0 <input type="checkbox"/> nepodána Typ léčby *) 01 <input type="checkbox"/> systémová 08 <input type="checkbox"/> lokální 02 <input type="checkbox"/> regionální					
Hormonální léčba (slovně) <input type="text"/>					
zdr. zařízení a oddělení <input type="text"/>					
Datum zahájení hormonální léčby <input type="text"/>					
Druh léčby *) 1 <input type="checkbox"/> chirurgická 2 <input type="checkbox"/> medikamentózní 4 <input type="checkbox"/> radiální 0 <input type="checkbox"/> neprovedena					
Jiná léčba (slovně) <input type="text"/>					
zdr. zařízení a oddělení <input type="text"/>					
Datum zahájení jiné léčby <input type="text"/>					
Druh léčby *) 001 <input type="checkbox"/> imunoterapie 004 <input type="checkbox"/> kryoterapie 016 <input type="checkbox"/> fotodynamická 064 <input type="checkbox"/> jiná léčba 000 <input type="checkbox"/> neprovedena 002 <input type="checkbox"/> hypertermie 008 <input type="checkbox"/> transplantace kostní dřeně 032 <input type="checkbox"/> regulační terapie 128 <input type="checkbox"/> cílená (biologická)					
Protinádorová léčba nepodána pro 1 <input type="checkbox"/> stav nevyžaduje 2 <input type="checkbox"/> místní rozsah 3 <input type="checkbox"/> generalizaci 4 <input type="checkbox"/> kontréind. nes. s nád. 5 <input type="checkbox"/> odmítnuti 6 <input type="checkbox"/> umrti 9 <input type="checkbox"/> neznámo					
Příčina smrti z LPZ (kód MKN-10) Bezprostřední (Ia) <input type="text"/> Základní (Id) <input type="text"/>					
Oprava Bezprostřední (Ia) <input type="text"/> Základní (Id) <input type="text"/>					
Datum úmrti <input type="text"/>					
Pitva 1 <input type="checkbox"/> ano 2 <input type="checkbox"/> ne Předchozí novotvary: <input type="text"/>					
Počet novotvarů u tétož nemocného <input type="text"/>					
Zařízení dispensární péče <input type="text"/>					
IČ <input type="text"/> PČZ <input type="text"/> PČDP <input type="text"/> odd. <input type="text"/>					

verze 2014/06  
Národní onkologický register, ÚZIS ČR, http://www.onkologickyregister.cz

Obrázek 1. Formulář „Incidence a léčba zhoubného novotvaru“ pro předávání údajů do Národního onkologického registru.

Základní epidemiologické výstupy z dat NOR jsou nejsnadněji dostupné na portálu epidemiologie nádorů [www.svod.cz](http://www.svod.cz), další údaje jsou k dispozici v publikační řadě Novotvary vydávané ÚZIS ČR, na portálu Regionální zpravodajství NZIS (<http://reporting.uzis.cz/>) a na Portálu zdravotnických ukazatelů (<https://pzu.uzis.cz/>). Epidemiologické charakteristiky jsou také standardní součástí řady publikačních výstupů zaměřených na konkrétní onemocnění.

## **2.2. Hodnocení výsledků léčby na populační úrovni – přežití**

Data NOR poskytují kromě dlouhodobých epidemiologických trendů také údaje o dlouhodobém sledování pacientů a umožňují hodnocení výsledků léčby na populační úrovni formou celkového (pozorovaného) nebo relativního přežití. Relativní přežití je kalkulováno jako poměr celkového přežití (představuje celkovou pozorovanou mortalitu v souboru pacientů s nádorem) a tzv. očekávaného přežití (představuje mortalitu v obecné populaci, která odpovídá sledované skupině pacientů věkem a pohlavím), jinak řečeno, relativní přežití vyjadřuje celkové přežití korigované na mortalitu spojenou s dalšími chorobami, na něž může pacient zemřít. Vzhledem k lišící se věkové struktuře pacientů v jednotlivých časových úsecích bývá relativní přežití dále věkově standardizováno, tzn. výsledné přežití je získáno jako vážený průměr hodnot relativního přežití v definovaných věkových kategoriích.

Výsledky hodnocení přežití jsou k dispozici v publikačních výstupech zaměřených na jednotlivá onemocnění a na portálu Regionální zpravodajství NZIS (<http://reporting.uzis.cz/>).

## **2.3. Predikce vývoje epidemiologie zhoubných novotvarů**

Díky dlouhodobým trendům dostupným v datech NOR je možné kromě popisné epidemiologické statistiky provádět také pokročilejší analytické zpracování v podobě odhadů budoucího vývoje incidence a prevalence pacientů s novotvary v následujících letech. Metodika predikce incidence vychází z epidemiologického trendu za známé období a koriguje jej s ohledem na pravděpodobné demografické změny populace. Prospektivní odhad prevalence pak kombinuje odhady počtu nově diagnostikovaných pacientů v jednotlivých letech s pravděpodobností jejich přežití do predikovaného kalendářního roku (tedy kombinuje odhady incidence s odhady 1letého až x-letého přežití dle roku diagnózy daného pacienta).

Tyto výsledky jsou následně využívány v plánování onkologické péče a představují jednu z nejdůležitějších údajů pro zásadní rozhodnutí v této oblasti.

Výsledky predikcí jsou standardně součástí publikace Modrá kniha České onkologické společnosti, která definuje doporučení léčby onkologických pacientů v ČR (dostupné na [www.linkos.cz](http://www.linkos.cz)).

## **3. Další vývoj Národního onkologického registru**

S ohledem na rozvoj a rozšiřováním datových zdrojů dostupných v rámci NZIS a s ohledem na změny v organizaci léčebné péče o onkologické pacienty, které od založení NOR proběhly, se stává současný způsob sběru dat NOR déle neudržitelným. V současné době probíhá transformační proces, jehož cílem je snížení rozsahu údajů předávaných poskytovateli na nezbytný základ zaměřený na diagnózu novotvaru a rozšíření využití již existujících zdrojů dat dostupných v NZIS. Zásadním krokem je také přechod na elektronické předávání údajů přímo z informačních systémů poskytovatelů.

Všechny výše popsané transformační kroky však musí proběhnout tak, aby nenarušily dlouhodobou kontinuitu sběru dat NOR a neovlivnili úplnost a kvalitu záznamů. Data

Národního onkologického registru jsou totiž nenahraditelným zdrojem informací a plným právem je lze označit za „rodinné stříbro“ české onkologie.

#### 4. Literatura

Ederer, F., Axtell, L. M., Cutler, S. J.: The relative survival rate: A statistical methodology. National Cancer Institute Monograph 1961; 6: 101–121.

Brenner, H., and Hakulinen, T.: On crude and age-adjusted relative survival rates. Journal of Clinical Epidemiology 2003; 56: 1185–1191.

Internetové zdroje:

URL 1: Ústav zdravotnických informací a statistiky ČR: Národní onkologický registr (NOR), dostupné z <http://www.uzis.cz/registry-nzis/nor> [citováno dne 17. 8. 2018]

URL 2: Dušek, L., Mužík, J., Kubásek, M., Koptíková, J., Žaloudík, J., Vyzula, R.: Epidemiologie zhoubných nádorů v České republice [online]. Masarykova univerzita, Brno 2005, dostupné z <http://www.svod.cz>. Verze 7.0 [2007], ISSN 1802 – 8861. [citováno dne: 29. 7. 2018]

URL 3: Český statistický úřad, dostupné z <https://www.czso.cz/> [citováno dne: 29. 7. 2018]

# Základní epidemiologické charakteristiky kolorektálního karcinomu v ČR

Denisa Krejčí

*Ústav zdravotnických informací a statistiky ČR, Praha  
Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
e-mail: denisa.krejci@iba.muni.cz*

## Abstrakt

Prezentovaná data dokládají vysokou zátěž české populace kolorektálním karcinomem, ročně je v ČR nově diagnostikováno téměř 8 000 pacientů s tímto zhoubným novotvarem a přibližně 3 500 pacientů na tento nádor umírá. Prevalence, tedy počet žijících osob, které mají nebo měly diagnostikovaný tento typ nádoru, dosáhla v roce 2016 hodnoty 57 730 osob ve srovnání s rokem 2006 tak vzrostla o 43 %. V ČR je nadto skoro polovina nových pacientů s kolorektálním karcinomem diagnostikována v klinickém stadiu III nebo vyšším, což významně zhoršuje prognózu pacientů a dosažitelné výsledky léčby. Již z těchto základních epidemiologických hodnot je jasné, že kolorektální karcinom představuje závažný zdravotní problém v české populaci.

## Klíčová slova

Kolorektální karcinom, incidence, mortalita, prevalence

## 1. Dostupná data

Hlavním zdrojem dat o epidemiologii zhoubných nádorů je Národní onkologický registr ČR (NOR). Dnes je NOR nedílnou součástí komplexní onkologické péče a při 100% pokrytí české populace obsahuje za období 1977–2016 více než 2,4 milionu záznamů. Registrace novotvarů je legislativně zakotvena a je povinná (URL 1). Široké veřejnosti jsou tato data přístupná na webovém portálu [www.svod.cz](http://www.svod.cz) (URL 2).

Jako součást monitoringu české populace zpracovává Český statistický úřad (URL 3) data o demografické struktuře obyvatelstva ČR. Data postihují demografické charakteristiky populace, zejména celkový počet obyvatel, věkovou strukturu, očekávanou délku života, atd. Zdrojem údajů o populační mortalitě zhoubných novotvarů v ČR je databáze příčin úmrtí, kterou rovněž zpracovává ČSÚ dle mezinárodních metodik na základě údajů z listu o prohlídce mrtvého (ICD-10 klasifikace).

## 2. Základní epidemiologické ukazatele

Mezi základní epidemiologické ukazatele patří incidence, mortalita a prevalence.

### 2.1. Incidence

Incidence je definována jako počet nově diagnostikovaných onemocnění ve sledované populaci v určitém časovém intervalu. Časový interval je nejčastěji volen jako jeden rok. V epidemiologii novotvarů je incidence často vztažena na 100 000 obyvatel dané populace. V tomto případě pak hovoříme o tzv. **hrubé incidenci** (*crude incidence*), která je definována jako podíl počtu nově diagnostikovaných onemocnění v dané populaci v daném období a počtu

osob v dané populaci v daném období (podle použití se osobami rozumí celá populace, pouze muži nebo pouze ženy). Hrubá incidence tedy zohledňuje rozsah populace, nezohledňuje však její věkovou strukturu. V případě, kdy sledovaná populace představuje osoby v určitém věkovém intervalu, hovoříme o tzv. **věkově specifické incidenci** (*age-specific incidence*), která je definována obdobně jako hrubá incidence, ovšem pro vymezenou věkovou skupinu.

## 2.2. Mortalita

Mortalita neboli úmrtnost je definována jako počet zaznamenaných úmrtí na onemocnění ve sledované populaci v určitém časovém intervalu. Časový interval je opět nejčastěji volen jako jeden rok. Analogicky jako v případě incidence vztahujeme mortalitu na 100 000 obyvatel a hovoříme o tzv. **hrubé mortalitě** (*crude mortality*) definované jako podíl počtu zaznamenaných úmrtí v dané populaci v daném období a počtu osob v dané populaci v daném období. Stejně jako u incidence hovoříme také zde o **věkově specifické mortalitě** (*age-specific mortality*), tj. mortalitě vztažené k určitému věkovému intervalu.

## 2.3. Prevalence

Prevalence je definována jako počet žijících osob, které mají v dané době určité onemocnění či prodělaly toto onemocnění v minulosti. Tedy například do prevalence zhoubných novotvarů (ZN) kolorekta v roce 2016 se budou počítat všechny žijící osoby, kterým byl v roce 2016 diagnostikován ZN kolorekta, ale také všechny žijící osoby, kterým byl ZN kolorekta diagnostikován kdykoli v minulosti.

U prevalence rozlišujeme dva základní typy – prevalenci okamžikovou a intervalovou.

**Okamžiková prevalence** (*point prevalence*) se vztahuje ke konkrétnímu okamžiku v čase a udává, kolik se v daném časovém okamžiku vyskytuje v populaci osob s uvažovanou diagnózou. Okamžiková prevalence je definována jako podíl počtu všech nemocných osob v dané populaci v daném okamžiku a počtu osob v dané populaci v témže okamžiku (nejčastěji k 31.12), nejčastěji vztažená na 100 000 obyvatel.

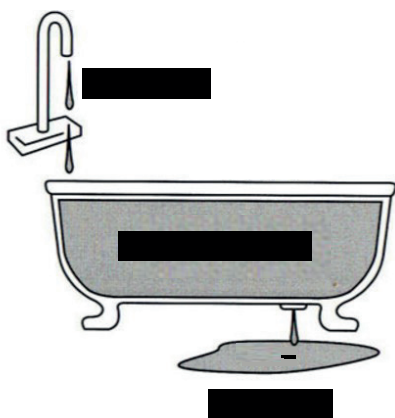
**Intervalová prevalence** (*period prevalence*) vyjadřuje podíl počtu žijících pacientů v daném časovém intervalu a počtu středního stavu obyvatelstva ve sledované populaci pro stejný časový interval, opět nejčastěji vztažená na 100 000 obyvatel.

Pro nádorová onemocnění, která jsou dobře léčitelná a doba přežití pacientů je tedy dlouhá, platí, že prevalence v určitém roce bude mnohonásobně vyšší než incidence pro daný rok. Naopak pro nádorová onemocnění, u kterých je dožití pacientů krátké, budou hodnoty prevalence a incidence srovnatelné (Bencko et al., 2003). Vzájemný vztah základních epidemiologických ukazatelů – incidence, mortality a prevalence je ilustrován na obrázku 1.

## 3. Věková standardizace epidemiologických ukazatelů

Jedním z častých úkolů v epidemiologii zhoubných nádorů je porovnání epidemiologických ukazatelů v různých populacích. V případě že populace nebudou mít stejné věkové složení, bude využití hrubé incidence či mortality značně zkreslené. Tomu lze předejít tak, že ukazatele standardizujeme, tj. převedeme na společný základ a místo hrubých ukazatelů se pak porovnávají ukazatele standardizované.

Pro výpočet věkově standardizovaných ukazatelů existuje mnoho různých přístupů. Obecně lze tyto přístupy rozdělit do dvou hlavních kategorií – přímá standardizace (*direct standardization*) a nepřímá standardizace (*indirect standardization*) (Bencko et al., 2003).



**Obrázek 1.** Vzájemný vztah základních epidemiologických ukazatelů (upraveno z: <https://www.pinterest.co.uk/pin/570057265307885435/>)

### 3.1. Přímá standardizace

Principem přímé standardizace je přepočítání epidemiologických ukazatelů na teoretický počet, který by příslušel dané populaci, kdyby mělo obyvatelstvo této populace stejné věkové složení jako zvolená standardní populace (Klener, 2002). Výsledkem je teoretický ukazatel **ASR** (*age-standardized rate*), který vyjadřuje, jaký by byl počet onemocnění na 100 000 osob, kdyby zkoumaná populace měla stejnou věkovou strukturu jako populace standardní.

Pro volbu standardní populace nejsou žádná obecně uznávaná pravidla a její volba závisí především na cílech naší analýzy. Budeme-li například porovnávat incidenci v jednotlivých krajích České republiky, můžeme jako standardní populaci zvolit populaci celé České republiky (Bencko et al., 2003). V případě mezinárodního srovnání se jako standardní populace nejčastěji volí světová či evropská standardní populace. Standardní světová populace byla stanovena v roce 1960 (Segi, 1960), evropská pak v roce 1976 (Waterhouse, 1976). Světová zdravotnická organizace v roce 2001 zareagovala na kontinuální stárnutí populace a vydala novou světovou populaci (Ahmad et al., 2001). Ze stejného důvodu vydal statistický úřad Evropské unie (Eurostat) v roce 2013 také novou evropskou populaci (Pace et al., 2013).

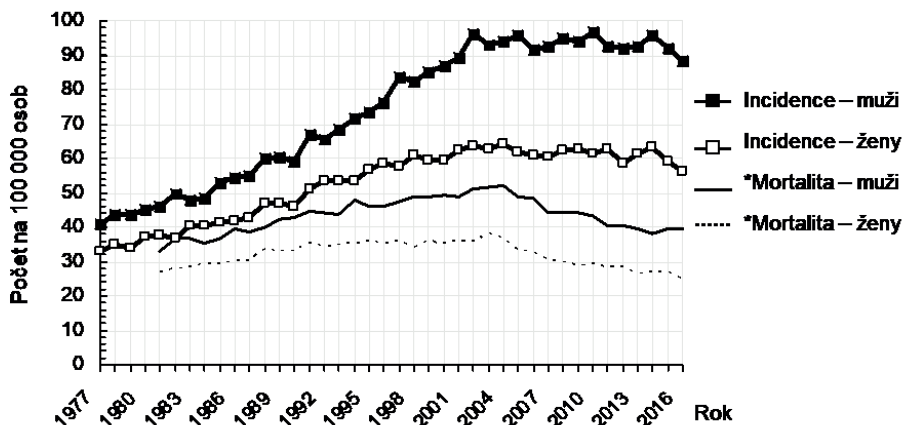
## 4. Epidemiologie kolorektálního karcinomu v ČR

### 4.1. Incidence a mortalita

Kolorektální karcinom je jednou z nejčastějších onkologických diagnóz a ve všech vyspělých státech jeho incidence setrvale narůstá. Česká republika navíc bohužel obsazuje přední místa v mezinárodním srovnání. Podle nejnovějších dat GLOBOCAN z roku 2012 (URL 4) stojí česká populace mužů v evropském přehledu na 3. místě za Slovenskem a Maďarskem, u žen potom obsazujeme 10. nejvyšší pozici v Evropě. Populační zátěž je skutečně vysoká, ročně je v ČR nově diagnostikováno kolem 8 000 pacientů s tímto karcinomem a téměř 3 500 pacientů na něj zemře. U mužů je incidence i mortalita kolorektálního karcinomu vyšší než u žen. Kromě absolutních počtů nově diagnostikovaných případů a úmrtí za 1 rok lze incidenci (a mortalitu) kolorektálního karcinomu vyjádřit formou přepočtu na 100 000 osob v populaci (obrázek 2) či standardizovat na určitý věkový standard (obrázek 3); nejčastěji je používán světový věkový standard ASR-W nebo evropský věkový standard ASR-E. Tyto přepočty

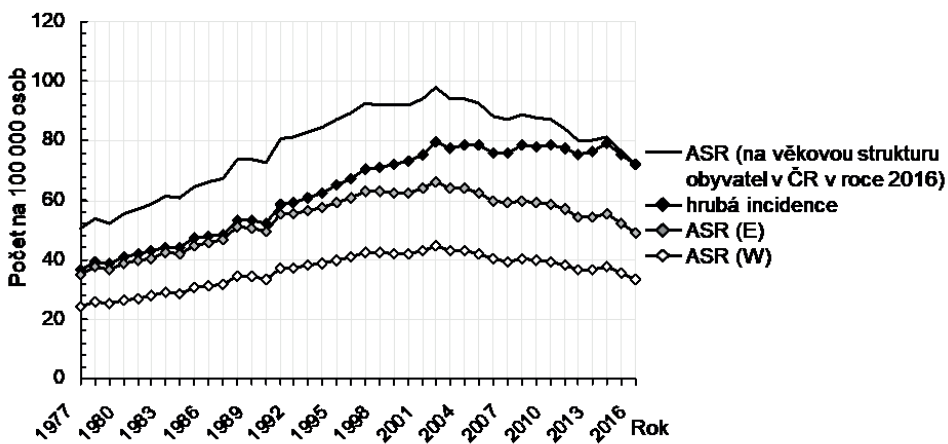


umožňují srovnávat incidenci (i mortalitu) kolorektálního karcinomu v České republice s ostatními zeměmi.



Zdroj: Národní onkologický registr ČR, \*Český statistický úřad

Obrázek 2. Hrubá incidence a mortalita ZN kolorekta v ČR.



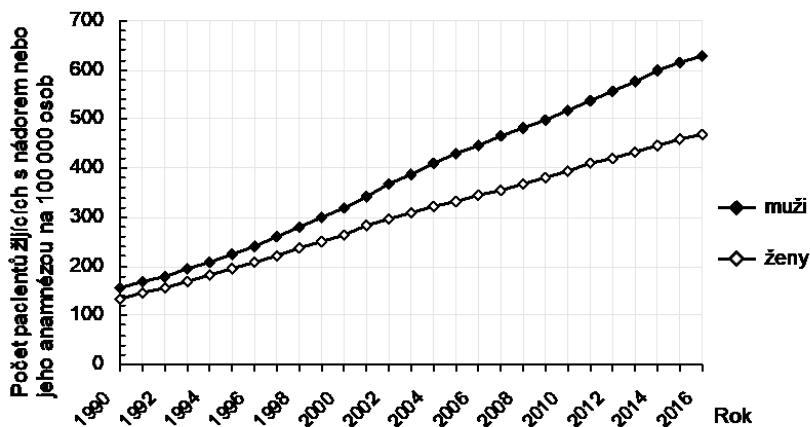
Zdroj: Národní onkologický registr ČR

Obrázek 3. Věkově standardizovaná incidence ZN kolorekta v ČR

#### 4.2. Prevalence

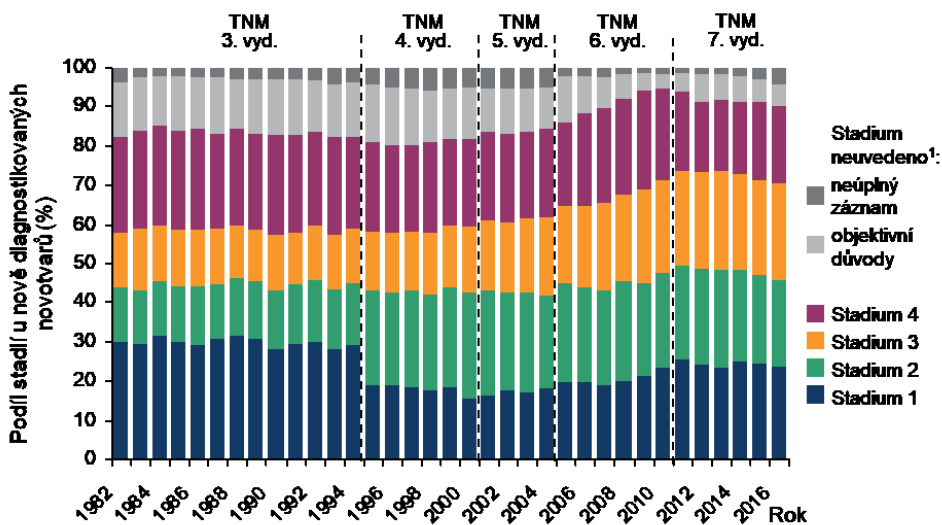
V posledních letech lze pozorovat náznaky zlepšení epidemiologické situace, především stabilizací či dokonce poklesu mortality (viz výše), což ovšem vede k nárůstu prevalence, tedy počtu žijících osob, které mají nebo měly diagnostikovaný kolorektální karcinom (obrázek 4). Prevalence dosáhla v roce 2016 hodnoty 57 730 osob a ve srovnání s rokem 2006 tak vzrostla téměř o 43 %.





Zdroj: Národní onkologický registr ČR

Obrázek 4. Trendy prevalence ZN kolorekta v ČR



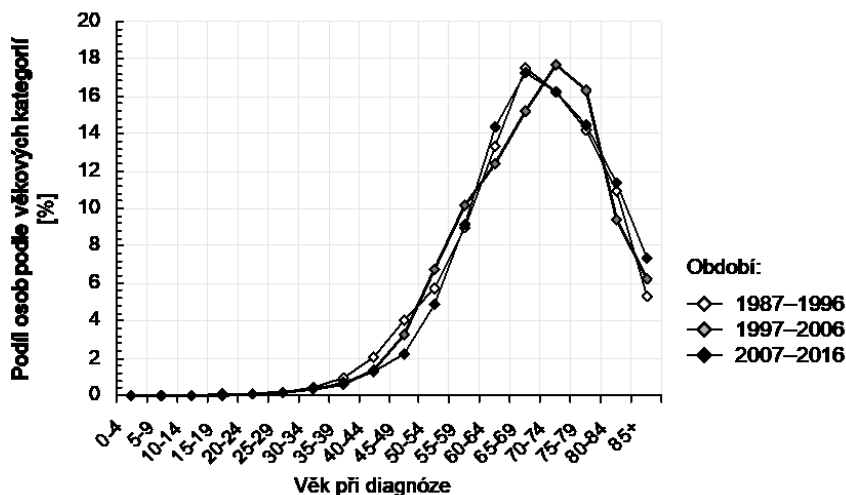
TNM: klasifikace zhoubných nádorů v různých vydáních

<sup>1</sup>Objektivními důvody pro neuvvedení stadia onemocnění jsou nálezy DCO nebo při pitvě, velmi časná úmrtí pacienta, nemocní neléčení pro kontraindikace onkologické léčby, odmítnutí léčby pacientem. Není-li neuvvedení stadia vysvětleno, je záznam považován za chybně neúplný (stadium neznámo bez důvodu).

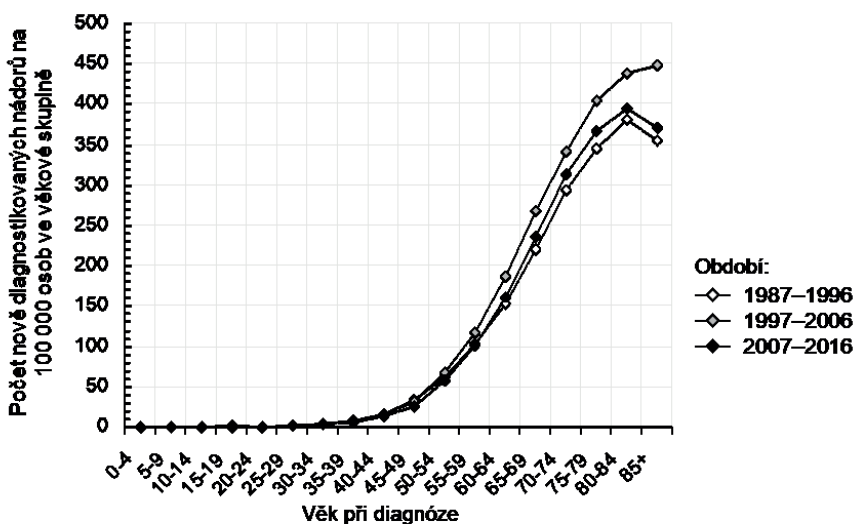
Zdroj: Národní onkologický registr ČR

Obrázek 5. Záchyt klinických stadií u kolorektálního karcinomu v ČR

### 6A. Věková struktura pacientů



### 6B. Věkově specifická incidence



Zdroj: Národní onkologický registr ČR

Obrázek 6. Věk pacientů s kolorektálním karcinomem

### 4.3. Klinická stadia

Je všeobecně známo, že záchyt onkologického onemocnění v méně pokročilém klinickém stadiu (anebo nejlépe ve fázi prekancerózy, ještě před vznikem vlastního zhoubného nádoru) výrazně zvyšuje naději na dobrý výsledek léčby a na dlouhodobé přežití. Dostupná populační data o kolorektálním karcinomu bohužel ukazují, že v ČR je i nadále velmi vysoký podíl nových pacientů s kolorektálním karcinomem diagnostikován v klinickém stadiu III nebo vyšším (obrázek 5). To samozřejmě významně zhoršuje dosažitelné výsledky léčby a také

zvyšuje náklady na léčbu, která je nejdražší právě u pokročilých stadií nemoci. Situace se navíc v čase nijak zvlášť nelepší.

#### 4.4. Věková struktura pacientů

Kolorektální karcinom ve velké míře postihuje pacienty v produktivním věku. Typický věk českého pacienta s kolorektálním karcinomem sice leží v intervalu 60–79 let, ale téměř 19 % všech nemocných je mladších než 60 let (obrázek 6A). Prudký nárůst incidence tohoto onemocnění pak začíná již po 45. roce života. Profil věkově specifické incidence kolorektálního karcinomu (obrázek 6B) ukazuje, jak závažným problémem je kolorektální karcinom v nejvyšších věkových kategoriích (URL 5).

### 5. Závěr

Tato kapitola dokládá, že i přes vysokou incidenci kolorektálního karcinomu se daří dlouhodobě stabilizovat mortalitu, což následně vede k významnému nárůstu celkové prevalence pacientů.

### 6. Literatura

- Ahmad, O. B., Boschi-Pinto, C., Lopez, A. D., et al. 2001. Age standardization of rates: A new WHO standard. GPE Discussion Paper Series. 31, World Health Organization.
- Bencko, V., Zvárová, J., MALÝ, M. 2003. Statistické metody v epidemiologii. Praha: Karolinum. svazek 1. ISBN 978-80-246-0763-4.
- Klener, P. 2002. *Klinická onkologie*. Praha: Karolinum. ISBN 978-80-246-0468-8.
- Pace, M., Cayotte, E., Agafitei, L., et al. 2013. Revision of the European Standard Population: report of Eurostat's task force: 2013. ISBN 978-92-79-31094-2.
- Segi, M. 1960. Cancer mortality for selected sites in 24 countries. Tokoku, University School of Medicine, Japan.
- Waterhouse, J., Muir, C., Correa, P., Powell, J. 1976. Cancer incidence in five continents. Lyon: IARC. 3: 456.
- URL 1: Ústav zdravotnických informací a statistiky ČR: Národní onkologický registr (NOR), dostupné z <http://www.uzis.cz/registry-nzis/nor> [citováno dne 29. 7. 2018]
- URL 2: Dušek, L., Mužík, J., Kubásek, M., Koptíková, J., Žaloudík, J., Vyzula, R.: Epidemiologie zhoubných nádorů v České republice [online]. Masarykova univerzita, Brno 2005, dostupné z <http://www.svod.cz>. Verze 7.0 [2007], ISSN 1802 – 8861. [citováno dne: 29. 7. 2018]
- URL 3: Český statistický úřad, dostupné z <https://www.czso.cz/> [citováno dne: 29. 7. 2018]
- URL 4: Ferlay, J., Soerjomataram, I., Ervik, M., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F.: GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [online]. International Agency for Research on Cancer, Lyon (France) 2013, dostupné z <https://globocan.iarc.fr> [citováno dne 29. 7. 2018]
- URL 5: Dušek, L., Zavoral, M., Májek, O., Suchánek, Š., Mužík, J., Pavlík, T., Šnajdrová, L., Gregor, J. Kolorektum.cz – Program kolorektálního screeningu v České republice [online]. Masarykova univerzita, Brno, 2018. dostupné z <http://www.kolorektum.cz>. ISSN 1804-0888. Verze 1.6f. [citováno dne: 29. 7. 2018]

# Aplikace joinpoint regrese v epidemiologii

Lucie Pechalová

*Ústav zdravotnických informací a statistiky ČR, Praha  
Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
e-mail: pehalova@iba.muni.cz*

## Abstrakt

Príspevek se zabývá metodou joinpoint regrese, která je představena jak po stránce teoretické, tak po stránce praktické. V teoretické části je uvedena základní definice modelu a je vysvětlen rozdíl mezi roční procentuální změnou a průměrnou roční procentuální změnou. Dále je objasněna problematika výběru optimálního joinpoint modelu a jsou shrnuty základy dvouvýběrového porovnávání. V praktické části je pak detailně popsána práce s Joinpoint regression programem.

## Klíčová slova

Joinpoint regrese, časové trendy v epidemiologii, roční procentuální změna, průměrná roční procentuální změna, dvouvýběrové porovnávání, Joinpoint regression program

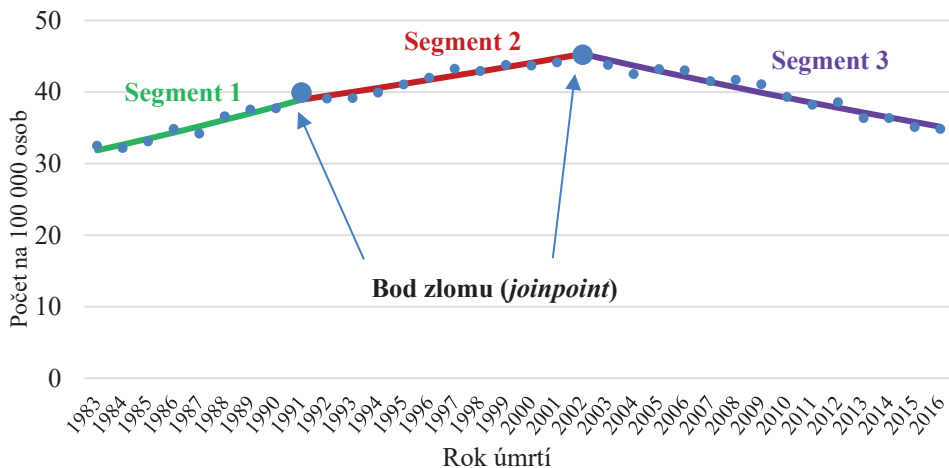
## 1. Úvod

Joinpoint regrese je ústřední metodou pro analýzu časových trendů v epidemiologii nádorových onemocnění. Hlavní výhodou joinpoint regrese oproti ostatním metodám je přímočará interpretace parametrů, která nám umožňuje jednoduchý popis změn ve sledovaném časovém trendu. Konkrétním příkladem využití joinpoint regrese může být hodnocení změn v trendu novotvarů prostaty po zavedení PSA testu (*prostate specific antigen*) či hodnocení trendu po zavedení určitého screeningového programu (Kim et al., 2000).

Joinpoint regrese je doporučovanou statistickou metodou pro hodnocení časových trendů na webu *National Cancer Institute of USA* a je proto možné dohledat stovky medicínských článků s onkologickou tematikou, které využívají právě tento model.

## 2. Definice modelu

Joinpoint regrese představuje regresní model, ve kterém je vysvětlována proměnná (incidence či mortalita) popsána po částech lineární spojitou funkcí. Tato funkce mění svoji směrnici v několika obecně neznámých bodech zlomu. Body zlomu se někdy nazývají jako tzv. **joinpoints**, z čehož vyplývá název joinpoint regrese. Jednotlivé části lineární funkce se označují jako **segmenty**. Ukázkový výstup joinpoint regrese pro mortalitu na zhoubné novotvary tlustého střeva a konečníku (C18–C20) ilustruje obrázek 1.



**Obrázek 1.** Joinpoint regrese pro mortalitu zhoubných novotvarů tlustého střeva a konečnicku

Z matematického úhlu pohledu je model joinpoint regrese popisující závislost proměnné  $Y$  (incidence/mortalita) na veličině  $x$  (čas) definován následovně (Kim et al., 2000):

$$E(Y) = \mu(x) = \beta_0 + \beta_1 x + \delta_1(x - \tau_1)^+ + \delta_2(x - \tau_2)^+ + \dots + \delta_k(x - \tau_k)^+,$$

kde:

$\beta_0, \beta_1, \delta_1, \delta_2, \dots, \delta_k$  jsou odhadované regresní koeficienty,

$\tau_1, \tau_2, \dots, \tau_k$  jsou neznámé body zlomu,

$(x - \tau_i)^+$  značí kladnou část přímky  $(x - \tau_i)$ ,  $i = 1, 2, \dots, k$ .

Z hlediska analýzy trendu je pro nás zajímavé kvantifikovat rychlost nárůstu či poklesu v jednotlivých segmentech. Pro kvantifikaci rychlosti nárůstu či poklesu je vhodné využít joinpoint regresi ve formě tzv. **log lineárního modelu**. Využití log lineárního modelu znamená, že pozorované hodnoty incidence či mortality jsou nejprve zlogaritmovány a následně jsou zlogaritmované hodnoty proloženy po částech lineární spojitou funkcí. Při využití log-lineárního modelu jsou tedy hodnoty incidence (mortality) popsány pomocí segmentů, které v tomto případě již nejsou částmi lineárních přímek, nýbrž částmi exponenciál.

### 3. Roční procentuální změna

V případě využití log-lineárního modelu můžeme rychlost nárůstu či poklesu v jednotlivých segmentech kvantifikovat pomocí tzv. **roční procentuální změny** (*annual percent change – APC*). APC nám udává konstantní procentuální změnu incidence (mortality) oproti předchozímu roku. Bude-li např. hodnota roční procentuální změny v segmentu popisující období 1990–2005 rovna 5 %, z interpretačního hlediska to znamená, že v období 1990–2005 dochází k nárůstu incidence (mortality) průměrně o 5 % ročně.

V případě kvantifikace nárůstu či poklesu pomocí roční procentuální změny je možné srovnávat pozorovaný trend napříč různými věkovými skupinami nebo odlišnými diagnózami,

neboť roční procentuální změna není závislá na absolutních počtech. Tato možnost srovnání mezi různými skupinami je hlavním důvodem, proč je joinpoint regrese nejčastěji využívána ve formě log-lineárního modelu.

Bodový odhad roční procentuální změny pro  $i$ -tý segment je odvozen z parametru  $\gamma_i$  reprezentující sklon přímky  $i$ -tého segmentu (Kleinbaum a Kupper, 1988):

$$APC = (e^{\gamma_i} - 1) * 100.$$

Bodový odhad je vhodné doplnit příslušným **intervalem spolehlivosti** (IS). Interval spolehlivosti nám jednak poskytuje informaci o přesnosti daného bodového odhadu APC a také na základě hodnot intervalu spolehlivosti můžeme zhodnotit statistickou významnost pro příslušnou roční procentuální změnu. V případě, kdy interval spolehlivosti zahrnuje hodnotu 0, je pozorovaná roční procentuální změna statisticky nevýznamná. V opačném případě je hodnota APC statisticky významná a můžeme říci, že v daném segmentu dochází ke statisticky významnému nárůstu či poklesu.

95% interval spolehlivosti pro roční procentuální změnu v  $i$ -tém segmentu je popsán následujícím vztahem (Kleinbaum a Kupper, 1988):

$$IS_{0,95} = (e^{(\gamma_i \pm t_{0,975}(N-2) * SE(\gamma_i))} - 1) * 100,$$

kde:

$\gamma_i$  je parametr reprezentující sklon  $i$ -tého segmentu,

$N$  je celkový počet pozorování,

$SE(\gamma_i)$  je standardní chyba parametru  $\gamma_i$ ,

$t_{0,975}(n - 2)$  značí hodnotu kvantilu Studentova rozdělení s  $n - 2$  stupni volnosti.

#### 4. Průměrná roční procentuální změna

V předchozí kapitole jsme si ukázali, jak určit roční procentuální změnu pro jednotlivé segmenty joinpoint modelu. Někdy však může být užitečné sumarizovat trend pro dané fixní období (např. posledních 10 let či celé analyzované období). V tomto případě je možné využít tzv. **průměrnou roční procentuální změnu** (*average annual percent change – AAPC*). Průměrná roční procentuální změna umožňuje popis trendu v předem stanoveném fixním intervalu pomocí jednoho čísla.

Výpočet AAPC je založen na váženém průměru ročních procentuálních změn v jednotlivých segmentech, přičemž váhy jsou rovny délce příslušných segmentů. Matematické vyjádření průměrné roční procentuální změny je dáno vztahem (Clegg et al., 2009):

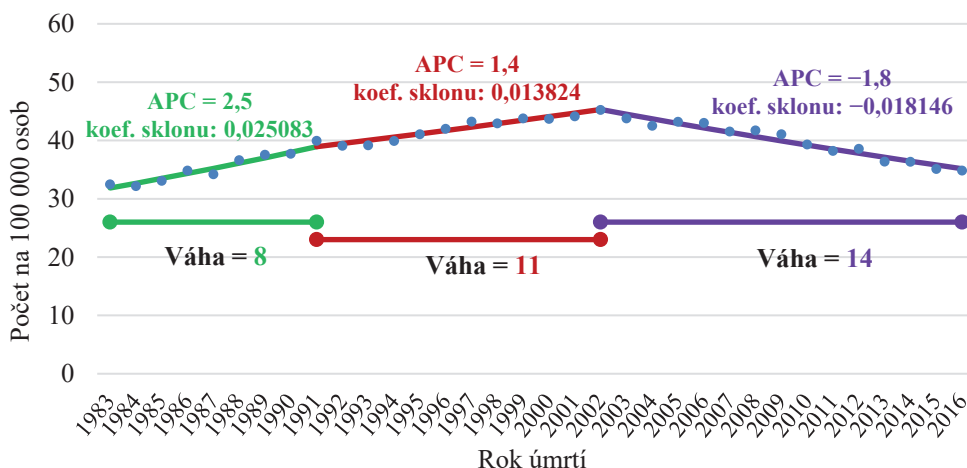
$$AAPC = \left\{ \exp \left( \frac{\sum w_i \gamma_i}{\sum w_i} \right) - 1 \right\} * 100,$$

kde:

$\gamma_i$  je parametr reprezentující sklon přímky  $i$ -tého segmentu,

$w_i$  je délka  $i$ -tého segmentu.

Ilustrace výpočtu průměrné roční procentuální změny pro mortalitu na zhoubné novotvary tlustého střeva a konečnicku je znázorněna na obrázku 2. Z tohoto obrázku si můžeme všimnout, že ve chvíli, kdy je průměrná roční procentuální změna počítána pro období, do kterého zasahuje pouze jeden segment joinpoint modelu, je AAPC rovno APC tohoto segmentu.



$$\text{AAPC pro } 1983\text{--}2016 = \{ \exp[(8 * 0,025083 + 11 * 0,013824 + 14 * -0,018146) / (8 + 11 + 14)] - 1 \} * 100 = 0,3$$

$$\text{AAPC pro } 2000\text{--}2016 = \{ \exp[(2 * 0,013824 + 14 * -0,018146) / (2 + 14)] - 1 \} * 100 = -1,4$$

$$\text{AAPC pro } 2007\text{--}2016 = \{ \exp[(10 * -0,018146) / 10] - 1 \} * 100 = -1,8$$

**Obrázek 1.** Joinpoint regrese pro mortalitu zhoubných novotvarů tlustého střeva a konečnicku

## 5. Poloha bodů zlomu

Po teoretické stránce je nalezení optimálního joinpoint modelu poměrně komplikovanou záležitostí. Při hledání optimálního modelu se musíme totiž vypořádat se hned dvěma úkoly. Prvním úkolem je určení optimálního počtu bodů zlomu a druhý úkol spočívá ve stanovení polohy těchto bodů zlomu. Počtu bodů zlomu bude věnována kapitola 6. V této chvíli předpokládejme, že počet bodů zlomu je daný, a zaměříme se na odhad jejich optimální polohy. Pro řešení tohoto problému lze využít obecně dva různé přístupy. Jedná se o metodu prohledávání mřížky nebo tzv. Hudsonovu metodu.

Metoda **prohledávání mřížky** (*grid search*) předpokládá, že se bod zlomu bude nacházet v  $x$ -ové souřadnici některé z pozorovaných hodnot (Lerman, 1980). Tato metoda tak vytvoří „mřížku“ možných poloh bodů zlomu a mezi těmito modely vybere takový, pro který je suma reziduálních čtverců minimální. Z omezující podmínky, kterou klademe na polohu bodu

zlomu, vyplývá, že pomocí metody prohledávání mřížky nemusíme vždy dostat zcela optimální řešení. Lepšího výsledku můžeme dosáhnout „zjemněním mřížky“, ve které budeme optimální model hledat. „Zjemnění mřížky“ znamená, že mezi každý bod x-ové souřadnice pozorovaných hodnot přidáme několik dalších bodů, ve kterých se může bod zlomu nacházet.

Druhým možným přístupem pro nalezení polohy bodů zlomu je **Hudsonova metoda** (*Hudson's method*) (Hudson, 1966). Tato metoda prochází prostor polohy bodů zlomu spojitě, a tak vede k nalezení skutečně optimálního modelu (Yu et al., 2007). Daní za nalezení optimálního modelu je mnohonásobně vyšší výpočetní náročnost.

Pro hledání optimálního modelu v analýze trendu onkologických dat je ve většině případů naprosto dostatečné využití metody prohledávání mřížky. Někdy je dokonce žádoucí, aby byly body zlomu určeny jako hodnoty z pozorovaných dat. Z interpretačního hlediska je totiž vhodnější, aby byl bod zlomu určen jako např. rok 2000 a ne jako rok 2000,37.

## 6. Počet bodů zlomu

Druhým cílem joinpoint regrese je určení optimálního počtu bodů zlomu. V ideálním případě by počet bodů zlomu měl být určen na základě apriorních znalostí problému. Tato situace však v praxi nastane pouze málokdy. Častěji se setkáme s přístupem, že je vytvořeno více modelů s různým počtem bodů zlomu a následně se optimální model vybere na základě určitého statistického přístupu.

Prvním možným přístupem je využití některého z informačních kritérií, přičemž nejdoporučovanějším je **Bayessovo informační kritérium** (BIC) (Kim et al., 2009). Jako optimální je zvolen model s tolika body zlomu, pro který je hodnota BIC nejnižší.

Definice Bayessova informačního kritéria pro joinpoint model s  $k$  body zlomu je následující:

$$BIC_k = \ln\left(\frac{SSE_k}{N}\right) + \frac{2 * (k + 1)}{N} + \ln(N),$$

kde:

$SSE_k$  značí sumu reziduálních čtverců joinpoint modelu s  $k$  body zlomu,  $N$  je celkový počet pozorování.

Jiným možným přístupem pro určení optimálního počtu bodů zlomu je **metoda permutačního testování** navržená v Kim et al. (2000). V rámci této metody je využita sekvence permutačních testů založených na F-statistice. Každý permutační test provádí testování nulové hypotézy  $H_0$ : počet bodů zlomu je roven  $k_a$  proti alternativní hypotéze  $H_1$ : počet bodů zlomu je roven  $k_b$ , kde  $k_a < k_b$ . Testování začíná s nastavením  $k_a = 0$  a  $k_b =$  maximální počet bodů zlomu. Jinými slovy, testujeme hypotézu, zda model s maximálním počtem bodů zlomu popisuje data statisticky významně lépe než model nulový (s nula body zlomu).

V případě, že je nulová hypotéza zamítnuta, hodnota  $k_a$  se zvýší o jedna. V opačném případě je hodnota  $k_b$  snížena o jedna. Procedura pokračuje až do doby, než  $k_a = k_b$  a optimální počet bodů zlomu je roven  $k = k_a = k_b$ .



Maximální možný počet bodů zlomu je volen buď na základě znalosti dané problematiky nebo podle délky analyzovaného období. Doporučovaný maximální počet bodů zlomu v závislosti na délce analyzovaného období je uveden v tabulce 1.

**Tabulka 1.** Doporučovaný maximální počet bodů zlomu v joinpoint modelu

Délka analyzovaného období	Doporučovaný maximální počet bodů zlomu
0–6 let	0
7–11 let	1
12–16 let	2
17–21 let	3
22–26 let	4
27 let a více	5

V publikaci Kim et al. (2009) je diskutována rozdílnost optimálního modelu při využití BIC a permutačního testování. Obecně lze říci, že Bayessovo informační kritérium lépe odhalí menší změny v trendu, zatímco permutační testování je více konzervativní a na lehké změny v trendu není příliš citlivé. V praxi to znamená, že BIC reaguje i na velmi malé změny ve sledovaném trendu a výsledný model tak často obsahuje velký počet bodů zlomu, což z interpretačního hlediska nemusí být příliš vhodné. Pro analýzu trendu onkologických onemocnění je z tohoto důvodu doporučováno využívat spíše permutační testování.

## 7. Dvouvýběrové porovnávání

Cílem dvouvýběrového porovnávání (*pairwise comparison*) je statistické srovnání joinpoint modelu pro dvě skupiny dat. Může nás například zajímat, zda se od sebe statisticky významně liší výsledek joinpoint regrese pro muže a ženy. Jinými slovy, pomocí dvouvýběrového porovnávání můžeme odpovědět na otázku, zda je pozorovaný trend v jedné skupině statisticky významně odlišný od pozorovaného trendu v jiné skupině.

Pro párové porovnávání můžeme využít dva různé testy – test paralelismu nebo test identity (Kim et al., 2004). **Test paralelismu** (*test of parallelism*) testuje hypotézu, že dvě regresní přímky popsané pomocí joinpoint modelu jsou paralelní. Jinak řečeno, testuje, zda je pozorovaný trend v různých skupinách stejný, s tím, že nebere v potaz absolutní hodnoty incidence/mortality (dovoluje rozdílný intercept). **Test identity** (*test of coincidence*) pak testuje hypotézu, zda dvě regresní přímky popsané joinpoint modelem jsou naprosto identické. Nepřipouští tedy ani rozdílný intercept.

## 8. Software pro joinpoint regresi

Pro modelování joinpoint regrese v epidemiologii je v současné době nejvhodnějším softwarem **Joinpoint regression program (JRP)**, který je k dispozici zdarma po zaregistrování na webových stránkách <https://surveillance.cancer.gov/joinpoint/>. Tento software byl vyvinut speciálně pro potřeby analýzy trendu onkologických onemocnění v čase.

V rámci tohoto softwaru je pro výběr optimálního modelu možné využít metodu prohledávání mřížky i Hudsonovu metodu. Výběr optimálního počtu bodů zlomu je možné volit na základě Bayesova informačního kritéria, modifikovaného Bayesova informačního kritéria nebo permutačního testování. Pro každý segment může být pomocí softwaru stanovena roční procentuální změna nebo průměrná roční procentuální změna za libovolné období. JRP umožňuje také provádět dvouvýběrové porovnávání pomocí testu paralelismu či identity.

Pro lepší představu práce s Joinpoint regression programem jsou popsány základní kroky analýzy pomocí dialogových oken softwaru JRP verze 4.6.0.0. Joinpoint regression program vyžaduje specifický formát vstupních dat. Musí se jednat o textový soubor, kde v prvním sloupci jsou hodnoty nezávisle proměnné (rok diagnózy/úmrtí), ve druhém sloupci hodnoty závisle proměnné (např. absolutní počty incidence/mortality, hrubá incidence/mortalita nebo věkově standardizovaná incidence/mortalita) a v posledním sloupci standardní chyba pro závisle proměnnou. Vstupní data musejí být seřazena vzestupně podle hodnot nezávisle proměnné a musí obsahovat desetinnou tečku nikoli desetinnou čárku.

Nastavení joinpoint analýzy se specifikuje pomocí 3 záložek – *Input file*, *Method and Parameters* a *Advanced analysis tool*. Záložka **Input file** (viz obrázek 2) slouží ke specifikaci vstupních dat a výběru klasického či log-lineárního modelu.

Možnost výpočtu věkové standardizace a standardní chyby pomocí softwaru – volba *Calculated From Data File*

Specifikace typu závisle proměnné

Specifikace závisle proměnné a standardní chyby

Specifikace klasického či log-lineárního modelu

Ukáзка vstupních dat

Specifikace nezávisle proměnné

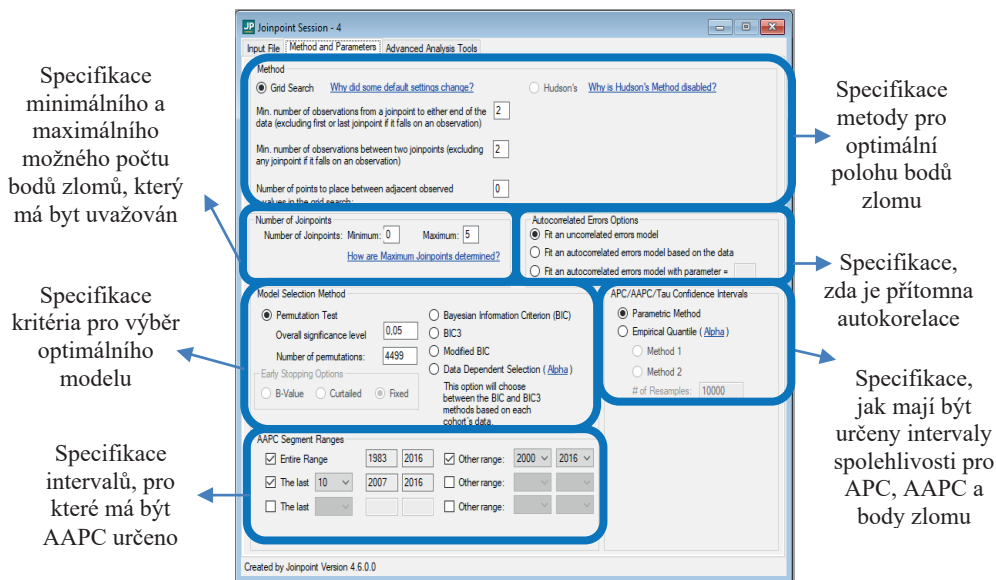
Specifikace typu standardní chyby

Výběr proměnné pro stratifikaci

ROK	Hrubá_mortalita	SE
1983	32.47946115	0.560824898
1984	32.16597799	0.557912707
1985	33.08403451	0.565642363
1986	34.83138593	0.580281364
1987	34.18742633	0.574761267

Obrázek 2. Joinpoint regression program – záložka *Input file*

**Záložka *Method and Parameters*** slouží ke specifikaci, jakým způsobem má být joinpoint regrese provedena (viz **obrázek 4**). V rámci této záložky jsou vybrána kritéria pro volbu optimálního modelu. Dále je možné v této záložce vybrat libovolné intervaly pro výpočet průměrné roční procentuální změny – AAPC.

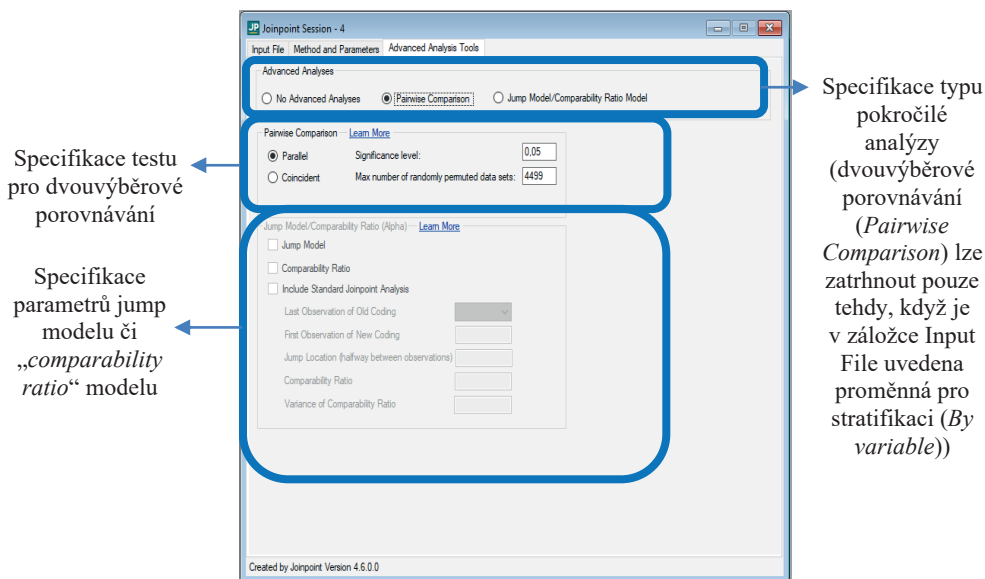


**Obrázek 4.** Joinpoint regression program – záložka *Method and Parameters*

**Záložka *Advanced analysis tool*** umožňuje provedení některé z pokročilejších analýz (viz **obrázek 3**). Pomocí této záložky můžeme provést dvouvýběrové porovnávání, případně vytvořit tzv. jump model či „*comparability ratio*“ model.

*Jump* modely a *comparability ratio* modely jsou modifikací klasických joinpoint modelů. Tyto modely umožňují provedení joinpoint regrese pro data, ve kterých je přítomný určitý „skok“ v pozorovaných hodnotách incidence či mortality, přičemž tento „skok“ není způsoben změnou v trendu. Takovýto „skok“ v datech může být přítomen například z důvodu změny v kódování nebo jiným administrativním zásahem. Typickým příkladem může být změna klasifikace jednotlivých klinických stádií nádorových onemocnění při přechodu na novější vydání TNM klasifikace (klasifikace pro hodnocení zhoubných novotvarů).

*Jump* modely jsou využívány ve chvíli, kdy od určitého roku začne být používáno pouze nové kódování a staré kódování již dále používáno není. Naproti tomu *comparability ratio* modely slouží k modelování dat, ve kterých sice došlo ke změně v kódování, ale staré a nové kódování je pak používáno simultánně (Anderson et al., 2001).



**Obrázek 3.** Joinpoint regression program – záložka *Advanced analysis tool*

Po výpočtu analýzy se v Joinpoint regression programu vytvoří nové okno s výsledky analýzy rozdělenými do pěti záložek – *Graph*, *Data*, *Model Estimates*, *Trends* a *Model Selection*.

**Záložka *Graph*** slouží ke grafickému znázornění získaných výsledků. V rámci této záložky si můžeme zobrazit modely s různým počtem bodů zlomu a graficky tak posoudit rozdílnosti mezi modelem s optimálním počtem bodů zlomu a jinými modely.

V **záložce *Data*** jsou zobrazeny pozorované a namodelované hodnoty incidence či mortality. Tyto hodnoty můžeme využít k vykreslení vlastního grafu pro joinpoint regresi.

**Záložka *Model Estimates*** nám poskytuje informace o odhadnutých parametrech modelu. V této záložce najdeme základní statistiky modelu (např. reziduální sumu čtverců, počet stupňů volnosti atd.). Dále zde také nalezneme odhadnuté body zlomy s příslušnými intervaly spolehlivosti nebo intercepty a sklony jednotlivých segmentů.

V **záložce *Trends*** jsou shrnuty informace týkající se analýzy trendu. Nalezneme zde roční procentuální změny s odpovídajícími intervaly spolehlivosti pro jednotlivé segmenty a průměrné roční procentuální změny pro námi specifikovaná období. V případě, že bychom chtěli získat průměrnou roční procentuální změnu pro jiné období, než které jsme specifikovali v počátečním nastavení analýzy, můžeme k tomu využít volby *Output* v horní liště. Pomocí *Specify AAPC ranges* můžeme specifikovat interval, pro který má být AAPC dodatečně určena.

Poslední **záložka *Model Selection*** nám poskytuje informace o postupu při výběru optimálního modelu. V případě využití permutačního testování je v této záložce zobrazen hierarchický postup optimálního modelu. Při využití Bayesova informačního kritéria jsou v této záložce hodnoty BIC pro modely s různými počty bodů zlomu.

V případě, kdy budeme provádět dvouvýběrové porovnávání, získáme ve výstupu ještě další šestou **záložku *Comparison***, ve které je zobrazen výsledek dvouvýběrového porovnávání pro příslušné skupiny.

Získané výstupy si můžeme přizpůsobit prostřednictvím volby *Output* v horní liště. Pomocí *Options* můžeme například volit počet desetinných míst, který má být zobrazen pro dané parametry modelu. *Options* nám také může posloužit pro úpravu popisků či stylu grafu, který je zobrazen v záložce *Graph*. Získané grafy poté můžeme exportovat. Případně si můžeme uložit celou sekci s analýzou nebo okno s výsledky.

## 9. Literatura

- Anderson, R. N., Miniño A. M., Hoyert D. L., Rosenberg H. M. 2001. Comparability of cause of death between ICD-9 and ICD-10: preliminary estimates. *National Vital Statistics Reports*, 49: 1–32.
- Clegg L. X., Hankey B. F., Tiwari R., Feuer E. J., Edwards B. K. 2009. Estimating average annual per cent change in trend analysis. *Statistics in Medicine*, 28: 3670–3682.
- Hudson D. J. 1966. Fitting Segmented Curves Whose Join Points Have to be Estimated. *Journal of the American Statistical Association*, 61: 1097–1129.
- Kim H. J., Yu B., Feuer E. J. 2009. Selecting the number of change-points in segmented line regression. *Statistica Sinica*, 19: 597–609.
- Kim, H. J., Fay M. P., Yu B., Barrett M. J., Feuer E. J. 2004. Comparability of segmented line regression models. *Biometrics*, 60: 1005–1014.
- Kim, H. J., Fay M. P., Feuer E. J., Midthune D. N. 2000. Permutation tests for joinpoint regression with applications to cancer rates. *Statistics in Medicine*, 19: 335–351
- Kleinbaum, D. G., Kupper L. L. *Applied regression analysis and other multivariable methods*. North Scituate: Duxbury Press. 1988. 556 p. ISBN 978-0-87872-139-9.
- Lerman, P. M. 1980. Fitting Segmented Regression Models by Grid Search. *Journal of the Royal Statistical Society*, 29: 77–84.
- Yu, B., Barrett M. J., Kim H. J., Feuer E. J., 2007. Estimating joinpoints in continuous time scale for multiple change-point models. *Computational Statistics & Data Analysis*, 51: 2420–2427.

# Principy klasifikace a nákladového ocenění hospitalizovaných pacientů

Tomáš Pavlík<sup>1,2</sup>, Michal Uher<sup>1,2</sup>, Zbyněk Bortlíček<sup>1,2</sup>, Petr, Klika<sup>1,2</sup>, Markéta Bartůňková<sup>1</sup>

<sup>1</sup> Ústav zdravotnických informací a statistiky ČR, Praha

<sup>2</sup> Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
e-mail: Tomas.Pavlik@uzis.cz

## Abstrakt

Data sbíraná o hospitalizovaných pacientech a informace z nich získané jsou důležitým nástrojem pro hodnocení procesů ve zdravotnictví, návrh koncepce a řízení zdravotní politiky státu. Informace o pacientovi a jemu poskytnuté zdravotní péči je potřeba popsat pomocí kódů, které většinou vycházejí z mezinárodních klasifikačních systémů. Tyto informace dále využívají tzv. *pacientské klasifikační systémy*, které umožňují třídění pacientů do konečného počtu navzájem disjunktních skupin na základě jejich známých charakteristik. Nejznámějšími patientskými klasifikačními systémy jsou klasifikační systémy odvozené z amerického systému DRG. Použití DRG systému v akutní lůžkové péči v ČR je spojováno zejména s přerozdělováním finančních prostředků směrem od plátců zdravotní péče k poskytovatelům. Cílem tohoto příspěvku je představit základní principy klasifikace a kalkulace celkových nákladů na hospitalizaci pacienta dle mezinárodně platné metodiky s využitím standardizovaných datových zdrojů.

## Klíčová slova

Hospitalizace, akutní lůžková péče, patientské klasifikační systémy, DRG klasifikace

## 1. Hospitalizace pacienta

Hospitalizace pacienta je definována dle §3 zákona č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování jako doba zpravidla delší než 24 hodin, po kterou je pacientovi přijatému na lůžko ve zdravotnickém zařízení poskytovatele lůžkové péče poskytována lůžková péče. Aby byla definice úplná, je třeba vysvětlit i pojem lůžková péče, což je dle §9 zákona 372/2011 Sb. zdravotní péče, kterou nelze poskytnout ambulantně a pro její poskytnutí je nezbytná hospitalizace pacienta. Lůžková péče musí být navíc poskytována v rámci nepřetržitého provozu zdravotnického zařízení.

Data sbíraná o hospitalizovaných pacientech a informace z nich získané jsou důležitým nástrojem pro hodnocení procesů ve zdravotnictví, návrh koncepce a řízení zdravotní politiky státu, které jsou nezbytné pro optimální nastavení sítě lůžkových zdravotnických zařízení a dostupnosti péče. Hlavními databázemi obsahujícími informace o hospitalizovaných pacientech jsou Národní registr hospitalizovaných (NRHOSP) a Národní registr hrazených zdravotních služeb (NRHVS), které jsou zdrojem údajů předávaných do databáze Světové zdravotnické organizace (WHO) a dalším mezinárodním organizacím podle smluvních závazků Ústavu zdravotnických informací a statistiky ČR a Ministerstva zdravotnictví ČR.

## 2. Údaje o hospitalizaci pacienta

Informace o pacientovi a jemu poskytnuté zdravotní péči je potřeba popsat pomocí kódů, které většinou vycházejí z mezinárodních klasifikačních systémů, tzv. *primárních klasifikací*. Těmito primárními systémy jsou nejčastěji systémy pro kódování diagnóz, léčivých přípravků a výkonů:

1. Pro *kódování diagnóz* se v České republice používá Mezinárodní statistická klasifikace nemocí a přidružených zdravotních problémů, konkrétně její 10. revize (MKN-10), která je publikována Světovou zdravotnickou organizací (WHO, 1992). Hlavním cílem použití MKN-10 je zajistit, aby ve světě probíhalo kódování nemocí, úrazů a příčin smrti obdobně a mohlo docházet k mezinárodnímu srovnávání.
2. Pro *kódování zvlášť účtovaného zdravotnického materiálu (ZUM) a zvlášť účtovaných léčivých přípravků (ZULP)* využívá systém CZ-DRG číselník ZP, který vydává Všeobecná zdravotní pojišťovna ČR (<https://www.vzp.cz/poskytovatele/ciselniky>), a číselníky HVLP a IVLP, které vydává Státní ústav pro kontrolu léčiv.
3. Pro *kódování zdravotnických výkonů* se v České republice používá tzv. Seznam zdravotních výkonů s bodovými hodnotami. Jak napovídá název, nejedná se o klasifikační systém v pravém slova smyslu, ale pouze o seznam výkonů prováděných ve zdravotnických zařízeních. Seznam výkonů je vydáván vyhláškou Ministerstva zdravotnictví ČR a je rovněž pravidelně aktualizován (<https://szv.mzcr.cz/>). Pro kódování výkonů, které zatím neexistují v Seznamu zdravotních výkonů, jsou využívány signální kódy (tzv. DRG markery), které mají z důvodu snadné aplikovatelnosti stejnou strukturu kódu jako standardní zdravotní výkony.

Údaje o hospitalizovaných pacientech v rámci databází NRHOSP i NRHZS, které jsou obsahově podobné, ale slouží jiným účelům, lze ve stručnosti rozdělit do následujících oblastí:

*Souhrnná informace o hospitalizaci pacienta* obsahuje základní přehled o hospitalizaci v jednom zdravotnickém zařízení ukončené ve sledovaném období, za které jsou data sbírána:

- Datum zahájení a ukončení
- Délka hospitalizace
- Kód doporučení hospitalizace
- Hlavní diagnóza případu
- Kritické výkony
- Odbornost přijímajícího pracoviště
- Pohlaví a věk pacienta (roky/dny)
- Porodní hmotnost
- Kód ukončení léčení
- Vedlejší diagnózy
- DRG dle systému IR-DRG (viz níže)
- Odbornost propouštějícího pracoviště

*Souhrnná informace o hospitalizaci pacienta na jednom oddělení* obsahuje především tyto klíčové informace:

- Datum zahájení a ukončení pobytu na oddělení
- Identifikace lůžkového pracoviště (odbornost)
- Kód doporučení hospitalizace a ukončení léčení na oddělení
- Základní diagnóza dokladu 02
- Vedlejší diagnózy příslušné pobytu na daném oddělení

*Informace o vykázaných výkonech* – obsahuje přehled všech výkonů příslušejících jednotlivým hospitalizacím:

- Identifikace a odbornost provádějícího pracoviště
- Datum provedení výkonu
- Kód výkonu
- Počet provedení výkonu



*Informace o zvlášť účtovaných položkách* – obsahuje přehled všech zvlášť účtovaných položek, kterými jsou materiálové položky (např. endoprotetický materiál, kardiostimulátory, operační nástroje, apod.) a léčivé přípravky (např. specializovaná antibiotika, protinádorové přípravky, apod.):

- Datum spotřeby/aplikace
- Typ a kód zvlášť účtovaného zdravotnického prostředku
- Množství/počet
- Částka/cena v Kč

*Informace o intenzitě ošetrovatelské péče, kterou daný hospitalizovaný pacient vyžadoval* – pro každý hospitalizační případ je pro jednotlivé kategorie uveden počet dní, v rámci nichž pacient splňoval kritéria pro zařazení do dané kategorie.

*Informace o pobytu pacienta na operačním sále* – obsahuje přehled operačních protokolů k hospitalizačním případům:

- Délka operace
- Délka anestezie
- Celkový čas využití operačního sálu
- Operační diagnóza
- Počty jedinců dle profesí přítomných během operace na sále (operatér/asistent atd.)

### 3. Pacientské klasifikační systémy jako sekundární klasifikace

Hospitalizace hrazené z veřejného zdravotního pojištění, kterých je v ČR přibližně dva miliony ročně, představují z datového pohledu velmi heterogenní množinu objektů, „ideálních“ pro následnou analýzu a klasifikaci do omezeného množství skupin, v nichž by byly jednotlivé hospitalizace vzájemně klinicky podobné. K tomuto účelu slouží tzv. *pacientské klasifikační systémy*, což jsou takové systémy, které umožňují třídění pacientů do konečného počtu navzájem disjunktních skupin na základě jejich známých charakteristik (parametrů, atributů). Mluvíme-li o akutní lůžkové péči, pak cílem je klasifikace hospitalizovaných pacientů, případně tzv. hospitalizačních případů, což jsou sloučené nebo naopak rozdělené hospitalizace dle definovaných klinických pravidel. Hlavní použití klasifikačních systémů hospitalizačních případů je v oblasti úhrad akutní lůžkové péče, monitoringu a srovnávání poskytnuté zdravotní péče a jednotlivých poskytovatelů zdravotních služeb (Busse a kol., 2011).

Obecně lze klasifikační systémy hospitalizačních případů rozdělit na dvě skupiny:

1. **klasifikační systémy odvozené z amerického systému DRG** (z anglického *Diagnosis Related Groups*, tedy skupiny případů o příbuzné diagnóze), které jsou v celosvětovém měřítku nejpoužívanější;
2. **klasifikační systémy vyvinuté nezávisle na systému DRG**. Do této skupiny patří systémy používané ve Velké Británii, Polsku, Rakousku a Nizozemsku.

Pacientské klasifikační systémy, a zejména systémy DRG, jsou ve skutečnosti sekundárními klasifikačními systémy, neboť pro svá pravidla využívají klasifikace primární. Charakteristiky, které jsou standardně využívány pro klasifikaci případů, jsou hlavní diagnóza pacienta, pro kterou byl léčen, další diagnózy popisující jeho klinický stav, dále věk a pohlaví pacienta a samozřejmě závažné (většinou operační nebo endoskopické) výkony, které pacient v průběhu hospitalizace podstoupil. Na klasifikaci může mít také vliv způsob přijetí



a propuštění pacienta ze zdravotnického zařízení, případně u novorozenců také porodní hmotnost.

V ČR je již od roku 2007 využívána varianta amerického systému IR-DRG (z anglického *International Refined Diagnosis Related Groups*), kterou vyvinula společnost 3M (Koženy a kol., 2010). Stejně jako ostatní DRG systémy, i česká modifikace IR-DRG využívá tři úrovně klasifikace:

- první úroveň je tzv. **MDC skupina** (z anglického *Major Diagnostic Category*), která představuje úroveň systému založenou na podobnosti případů z hlediska orgánové soustavy odpovídající danému onemocnění;
- druhou úroveň je tzv. **DRG báze**, která zohledňuje formu péče o pacienta v rámci dané MDC. Rozlišujeme primárně tzv. chirurgické DRG báze (např. s operačním výkonem) a nechirurgické DRG báze (např. konzervativní farmakologická léčba);
- třetí úroveň je samotné DRG, někdy také tzv. **DRG skupina**, která má za cíl zohlednit i celkový stav konkrétního léčeného pacienta skrze jeho další přidružená onemocnění a komplikace léčby poskytované za hospitalizace.

I když jsou DRG systémy používány pro úhradu a monitoring akutní lůžkové péče v mnoha zemích, úspěšnost jejich použití je různá, což lze částečně dát do souvislosti s aktivním zapojením legislativních autorit a finančním zázemím v jednotlivých zemích (Busse a kol., 2011). Použití a kvalita DRG systémů se však napříč zeměmi liší také z toho důvodu, že každý ze systémů poskytování a úhrady zdravotní péče je specifický. V důsledku toho se nelze divit, že za celou dobu vývoje systémů pro klasifikaci hospitalizačních případů akutní lůžkové péče nebyl sestaven obecně aplikovatelný návod na to, jak nejlépe takový systém definovat, sestavit, implementovat a kultivovat.

## 4. Základy nákladového oceňování hospitalizačních případů

Použití DRG systému v akutní lůžkové péči v ČR je spojováno zejména s přerozdělováním finančních prostředků směrem od plátců zdravotní péče k poskytovatelům. Proměnnou, kterou klasifikátorem vysvětlujeme, jsou celkové náklady na hospitalizační případ, které by měly být vyčísleny v korunách a nikoliv v jiných pseudonákladových hodnotách (např. v bodech uvedených v Seznamu zdravotních výkonů s bodovými hodnotami jakožto pomocný indikátor). Naším cílem je optimalizovat schopnost klasifikačního systému tyto náklady s určitou přesností vysvětlit a následně predikovat.

Vedle adekvátní ekonomické predikční síly DRG systému je důležitý ještě další aspekt, a tím je klinická srozumitelnost struktury klasifikačního systému, respektive jeho jednotek. Klinicky srozumitelná definice a obsah jednotlivých tříd jsou klíčové zejména pro uživatele systému na straně poskytovatelů. Je třeba si uvědomit, že právě lékaři a kodéři jsou těmi, kteří v rámci DRG systému produkují primární data a jsou tak zodpovědní za jejich přesnost a správnost. Nesrozumitelnost a nejednoznačnost systému zbytečně zvyšuje chybovost, ať již úmyslnou či neúmyslnou, která se následně projeví ve zvýšené variabilitě neboli nehomogenitě systému.

### 4.1. Východiska pro nákladové oceňování

Východiskem pro oceňování hospitalizačních případů je vždy kalkulace nákladů poskytovatelů akutní lůžkové péče (PALP). Pro účely hodnocení nákladovosti musí být vybudován ekonomický model, jehož základnu tvoří informace z finančního účetnictví, údaje o personálním zabezpečení a informace o produkci zdravotních služeb PALP. Jeho základní funkcionalitou musí být sofistikovaný kalkulační model, na jehož základě budou

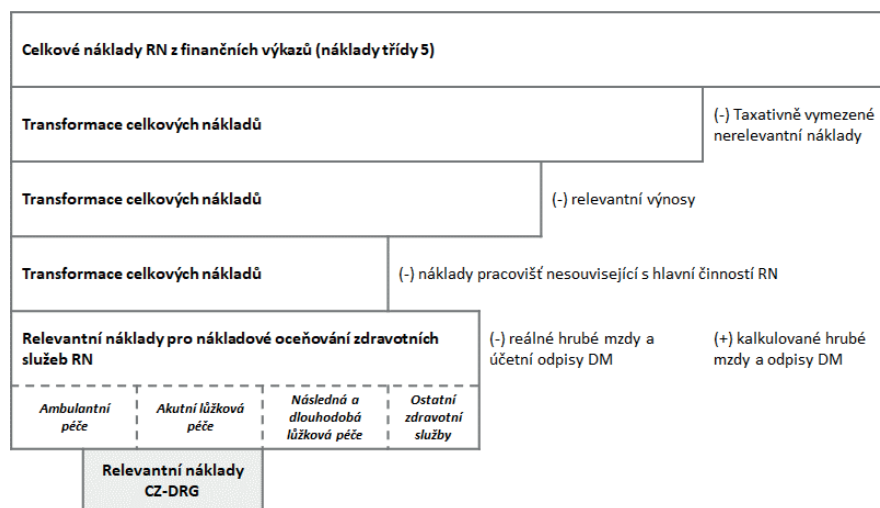
z verifikovaných dat PALP stanoveny náklady zdravotních služeb, poskytované hospitalizovaným pacientům.

V rámci mezinárodních modelů kalkulace nákladů v DRG systémech jsou využívány 4 modely oceňování hospitalizací (Busse a kol., 2011), a to:

- Metoda *Top-down gross-costing* oceňuje na základě hrubých nákladových komponent a průměrných hodnot skupiny případů.
- Metoda *Bottom-up gross-costing* oceňuje na základě hrubých nákladových komponent a individuálních případů.
- Metoda *Top-down micro-costing* oceňuje na základě detailních nákladových komponent a průměrných hodnot skupiny případů.
- Metoda *Bottom-up micro-costing* oceňuje na základě detailních nákladových komponent a individuálních případů.

Po předchozích zkušenostech se jako nejefektivnější metodikou nákladového oceňování HP ukázal postup *bottom-up micro-costing* využívaný např. ve Švédsku, Holandsku, Německu a Finsku (Tan a kol., 2009). Metodika musí současně definovat formální a obsahovou strukturu dat (aby se z uvedených dat daly náklady počítat standardizovaným postupem) a kalkulační postupy, na jejichž základě jsou vyhodnocovány náklady související s poskytováním akutní lůžkové péče v ČR.

Klíčovou komponentou oceňování je také **vymezení uznatelných nákladů** souvisejících s poskytováním akutní lůžkové péče, kdy je pro stanovení relevantních nákladů zapotřebí celkové náklady nemocnice očistit o náklady, které nesouvisí s hlavní činností PALP, resp. s hospitalizačními případy, které tvoří základní klasifikační a oceňovanou jednotku systému. Transformace celkových nákladů nemocnice je zobrazena na obrázku 1.



**Obrázek 1.** Transformace celkových nákladů nemocnice pro nákladové oceňování

## 4.2. Kalkulační model

Nezbytnou součástí objektivního hodnocení nákladů na hospitalizaci pacienta by měl tvořit transparentní *kalkulační model* na jehož základě dojde k vyčíslení relevantních normovaných nákladů PALP pro oceňování. Výstupem kalkulačního modelu by měl být strukturovaný

záznam o spotřebovaných nákladech, dle jejich druhu, ve vazbě na zdravotní služby, které byly poskytnuty jednotlivým hospitalizovaným pacientům.

Celkové náklady (CN)  $i$ -tého hospitalizačního případu jsou dány součtem jeho nepřímých nákladů (NN) a přímých nákladů (PN):

- **Nepřímé náklady** jsou náklady za zdravotní služby, které nelze jednoznačně přiřadit ke konkrétnímu hospitalizovanému pacientu (případu). Nepřímé náklady jsou vztaženy k tzv. *kalkulačním jednicím* (kalkulační jednice je nákladový objekt nebo také objekt představující jednotku výkonu, na kterou jsou zjišťovány náklady) zdravotních služeb, přes které budou k případu hospitalizace přiřazeny náklady, které nelze jednoznačně přiřadit ke konkrétnímu pacientovi. Příkladem kalkulační jednice minuta využití operačního sálu, přes kterou jsou oceňovány náklady na operační výkony pacientů.
- **Přímé náklady** jsou přímo přiřaditelné k jednotlivým případům hospitalizace na základě jejich evidence v produkčních datech PALP. Přímé náklady hospitalizačních případů tedy nejsou stanoveny na úrovni rozpoustění nákladů přes alokační klíče na případ, ale jsou evidovány v nemocničních informačních systémech pod identifikačním číslem konkrétního pacienta, lze je tak jednoduše vyčíslit. Příkladem přímého nákladu je podání konkrétního léčivého přípravku pacientovi, který je evidován na jeho rodné číslo (tzv. zvlášť účtovaný léčivý přípravek, ZULP).

**Nepřímé náklady** případu lze rozdělit na náklady na standardní pobytové služby ( $NN^{SPS}$ ), intenzivní pobytové služby ( $NN^{IPS}$ ), intramurální péči ( $NN^{IP}$ ) a operační služby ( $NN^{OS}$ ) dle množství spotřebovaných jednotek.

**Přímé náklady** tvoří náklady na zvlášť účtovaný materiál ( $PN^{ZUM}$ ), zvlášť účtované léčivé přípravky ( $PN^{ZULP}$ ), interní materiálové položky vztažené k rodnému číslu ( $PN^{IMP}$ ), extramurální péči ( $PN^{EP}$ ) a léčivé přípravky nebo speciální zdravotnický materiál zahrnutý přímo v hodnotě výkonu vyjádřený v Kč ( $PN^{PMAT}$ ).

Celkem tedy dostáváme celkové náklady  $i$ -tého hospitalizovaného pacienta v podobě:

$$CN_i = PN_i^{ZUM} + PN_i^{ZULP} + PN_i^{IMP} + PN_i^{EP} + PN_i^{PMAT} + NN_i^{SPS} + NN_i^{IPS} + NN_i^{IP} + NN_i^{OS} \quad (1)$$

### 4.3. Klíče pro rozpoustění nepřímých nákladů na pacienta

Pro každý typ pracoviště v nemocnici musí být v souvislosti s tím, co je zde prováděno za péči nebo služby, navržena výkonová jednotka, která nejlépe vyjadřuje spotřebu nákladů a která je závislá na množství poskytnutých dílčích služeb. Do ocenění výkonových jednotek všech přímých pracovišť vstupují pouze relevantní nepřímé náklady (pouze ty související s akutní lůžkovou péčí). Pro úplnost je třeba dodat, že do výpočtů výkonových jednotek jsou samozřejmě zahrnuty pouze relevantní náklady celého sledovaného období, tj. kalendářního roku.

Nepřímé náklady dle kapitoly 4.2 dělíme podle poskytovaných zdravotních služeb jednotlivých pracovišť na:

- **nepřímé náklady za standardní pobytové služby** – zde je kalkulační jednicí pro nákladové ocenění ošetrovací den na standardním oddělení,
- **nepřímé náklady za intenzivní pobytové služby** – zde je kalkulační jednicí pro nákladové ocenění ošetrovací den na oddělení intenzivní péče,
- **nepřímé náklady za intramurální péči** – zde je kalkulační jednicí pro nákladové ocenění jeden bod jako měřítko poskytnuté zdravotní péče dle Seznamu zdravotních výkonů,

- **nepřímé náklady za operační služby** – zde je kalkulační jendicí pro nákladové ocenění minuta využití operačního sálu.

Výpočet nákladové ceny standardního ošetrovacího dne na daném pracovišti ( $OD_{SPS}$ ) lze vyjádřit jako podíl nepřímých nákladů daného pracoviště standardní péče ( $NN_{SPS}$ , ekonomicky vyjádřeno jako zůstatky nákladových druhů evidovaných na pracovišti po odečtení přímých nákladů vztažených přímo k pacientovi) a celkového počtu ošetrovacích dnů evidovaných na daném pracovišti (do počtu ošetrovacích dnů se nepočítá propustka na žádost pacienta, propustka z léčebných důvodů, doprovod pacienta do 6 let věku a doprovod pacienta nad 6 let věku):

$$OD_{SPS} = \frac{NN_{SPS}}{\text{počet ošetrovacích dnů na standardním oddělení}} \quad (2)$$

Výpočet nákladové ceny ošetrovacího dne intenzivní péče ( $OD_{IPS}$ ) pak lze zjednodušeně vyjádřit jako podíl nepřímých nákladů daného pracoviště intenzivní péče ( $NN_{IPS}$ , ekonomicky vyjádřeno jako zůstatky nákladových druhů evidovaných na pracovišti po odečtení přímých nákladů vztažených přímo k pacientovi) a celkového počtu ošetrovacích dnů intenzivní péče evidovaných na daném pracovišti (do počtu ošetrovacích dnů se nepočítá propustka na žádost pacienta, propustka z léčebných důvodů, doprovod pacienta do 6 let věku a doprovod pacienta nad 6 let věku):

$$OD_{IPS} = \frac{NN_{IPS}}{\text{počet ošetrovacích dnů intenzivní péče}} \quad (3)$$

Vzorec (3) je pro účely tohoto článku zjednodušený, v praxi se většinou využívá vážená varianta výpočtu dle jednotlivých typů poskytované intenzivní péče.

Výpočet nákladové ceny zdravotního výkonu ( $BOD_{IP}$ ) je kalkulován jako podíl nepřímých relevantních nákladů ambulantního pracoviště a celkového souhrnu bodů zdravotních výkonů dle Seznamu zdravotních výkonů s bodovými hodnotami evidovaných na ambulantním pracovišti:

$$BOD_{IP} = \frac{NN_{ambulance}}{\text{počet bodů}} \quad (4)$$

A konečně, výpočet nákladové ceny minuty operace ( $MIN_{OS}$ ) je kalkulován jako podíl nepřímých relevantních nákladů pracoviště operačního sálu a celkového souhrnu minut doby trvání operací evidovaných na daném pracovišti (operačním sálu):

$$MIN_{OS} = \frac{NN_{OS}}{\text{počet minut na sále}} \quad (5)$$

Stanovíme-li hodnoty kalkulačních jednic dle vzorců (2)–(5), můžeme pak kalkulovat celkové nepřímé náklady  $i$ -tého hospitalizovaného pacienta následujícím postupem:

$$NN_i = NN_i^{SPS} + NN_i^{IPS} + NN_i^{IP} + NN_i^{OS}, \text{ kde} \quad (6)$$

$NN_i^{SPS} = OD_{SPS} * \text{počet spotřebovaných jednotek pacientem,}$

$NN_i^{IPS} = OD_{IPS} * \text{počet spotřebovaných jednotek pacientem,}$

$NN_i^{IP} = BOD_{IP} * \text{počet spotřebovaných jednotek pacientem,}$

$NN_i^{OS} = MIN_{OS} * \text{počet spotřebovaných jednotek pacientem.}$

## 5. Závěr

Proces klasifikace a nákladového ocenění hospitalizovaných pacientů je netriviální postup vycházející z heterogenních a často obtížně dostupných datových zdrojů. V ČR těmito datovými podklady disponuje Ústav zdravotnických informací a statistiky ČR, který je současně garantem vývoje klasifikačního systému DRG pro akutní lůžkovou péči v ČR. Výstupy výše zmíněného procesu jsou nezbytné pro hodnocení procesů v akutní lůžkové péči, např. hodnocení její dostupnosti a nákladové efektivity a zejména v rámci prospektivního procesu kultivace klasifikačního systému DRG. Kvalita výstupů však souvisí nejen s kvalitou samotného procesu, ale také s kvalitou a validitou vstupních dat. Tento článek měl za cíl seznámit čtenáře se základními principy a strukturou vstupů procesu klasifikace a nákladového ocenění hospitalizovaných pacientů a doložit potřebu sběru a validace potřebných údajů v rámci Národního zdravotnického informačního systému.

## 6. Literatura

- Busse, R., Geissler, A., Quentin, W., Wiley, M. Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals. Open University Press, 2011. 458 p. ISBN 978-0-33-524557-4. [http://www.euro.who.int/data/assets/pdf\\_file/0004/162265/e96538.pdf](http://www.euro.who.int/data/assets/pdf_file/0004/162265/e96538.pdf).
- International statistical classification of diseases and related health problems, 10th revision (ICD-10). Geneva: World Health Organization; 1992.
- Kožený, P., Němec, J., Kárníková, J., Lomíček, M. Klasifikační systém DRG. Praha: Grada Publishing, 2010. 208 p. ISBN 978-80-247-2701-1.
- Tan, S. S., Rutten, F. F., van Ineveld, B. M., Redekop, W. K., Hakkaart-van Roijen, L. 2009. Comparing methodologies for the cost estimation of hospital services. European Journal of Health Economics, 10: 39–45.

# Využití regresních stromů při kultivaci DRG klasifikačního systému

Michal Uher<sup>1,2</sup>, Monika Mazalová<sup>1</sup>

<sup>1</sup> Ústav zdravotnických informací a statistiky ČR, Praha

<sup>2</sup> Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
e-mail: uher@iba.muni.cz

## Abstrakt

DRG (Diagnosis Related Groups) systémy představují klasifikační systémy pacientů celosvětově využívané jako nástroj pro úhradu a monitoring hospitalizační péče (Busse a kol., 2011). Na základě DRG systémů se úhrada hospitalizačního případu odvíjí od jeho klasifikace do finální jednotky systému, do tzv. DRG skupiny. Pro vytvoření těchto skupin mohou být využity regresní stromy, které jsou ve svém principu jednoduchými matematickými modely popisujícími krokovým způsobem závislost vysvětlované proměnné na proměnných vysvětlujících.

## Klíčová slova

DRG klasifikační systémy pacientů, regresní stromy, CART, CHAID, skupinové modely, náhodné lesy, bootstrap, bumping

## 1. Úvod

DRG systémy, v českém překladu systémy skupin vztahených k diagnóze, jsou v obecnosti založeny na víceúrovňové klasifikaci pacientů do nákladově homogenních skupin na základě klinických a demografických faktorů. Prvním požadavkem těchto systémů je, aby jednotlivé úrovně až do finálních DRG skupin byly klinicky interpretovatelné. Tento požadavek je základním předpokladem přijetí systému v kruhu zdravotnických pracovníků, avšak sám o sobě nezaručí vytvoření nákladově homogenních skupin pacientů. Nejpodobnější náklady na hospitalizaci v rámci DRG skupin představují požadavek druhý a jsou podstatou kvalitní predikce nákladovosti jednotlivých hospitalizací. Jako nástroj pro návrh nákladově homogenních skupin s těmito dvěma specifiky bude představena metoda regresních stromů.

## 2. Metody regresních stromů

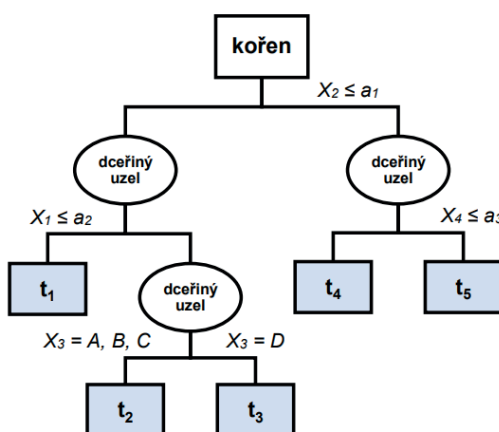
Regresní stromy řadíme mezi neparametrické metody strojového učení, které jsou oblíbené v mnoha oborech mimo jiné díky své intuitivní interpretaci, relativně snadné implementaci a „shovívatosti“ vůči charakteru vstupních dat.

Analogicky ke klasickému stromu i tyto jsou složeny z kořene, větví a listů. Kořen v našem případě obsahuje všechna pozorování sestávající ze sady vysvětlujících proměnných (prediktorů) a jedné spojité vysvětlované proměnné, která je (alespoň to předpokládáme) závislá na těchto prediktorech. Regresní stromy tuto závislost modelují pomocí rekurzivního dělení souboru. Smyslem dělení je, s velkou měrou zjednodušení, nejlépe rozdělit soubor případů do disjunktních skupin. V každém kroku dělení jsou vytvořeny (nejčastěji) dvě skupiny objektů, které se mezi sebou v hodnotách závisle proměnné (tedy například nákladech) nejvíce liší a zároveň v rámci skupiny jsou případy v těchto hodnotách

nejpodobnější (Komprdová, 2012). Tímto způsobem se kořen rozrůstá do dceřiných uzlů až do úrovně listů (terminálních uzlů).

Mimo stromy regresní, existují i stromy klasifikační, které se vyznačují kategoriální závisle proměnnou. Oba tyto typy souhrnně označujeme jako stromy rozhodovací.

Jednoduché schéma zobrazené na obrázku 1 demonstruje binární rozhodovací strom. Prediktory jsou označeny písmenem  $X$  se spodním číselným indexem. U každého větvení je uvedeno rozhodovací pravidlo. Toto pravidlo má buď tvar nerovnosti (kde  $a_1, a_2, a_3$  představují konkrétní číselnou hodnotu prediktoru, dle které se soubor dělí), nebo množiny kategorií kategoriálního prediktoru (v tomto případě prediktor  $X_3$  s kategoriemi  $A, B, C$  a  $D$ ). Terminální uzly stromu jsou znázorněny světle modrými obdélníky a písmenem  $t$  se spodním číselným indexem.



**Obrázek 1.** Obecné schéma rozhodovacího stromu (vytvoreno dle Komprdová, 2012)

Predikce závisle proměnné pro budoucí případy je nejčastěji dána průměrem závisle proměnné v terminálním uzlu, do kterého byl nový případ zařazen. Modelem regresního stromu nemusíme pouze predikovat závisle proměnnou, ale i sledovat vztah prediktorů a vysvětlované proměnné či identifikovat specifické skupiny objektů (Breiman a kol., 1993).

Matematicky si můžeme regresní strom  $T$  vyjádřit jako funkci vysvětlujících proměnných  $\mathbf{x}$  zapsanou předpisem

$$T(\mathbf{x}) = \sum_{m=1}^M \beta_m \cdot I(\mathbf{x} \in t_m),$$

kde  $t_m$  je list (region) stromu  $T$ ,  $\beta_m$  označuje funkci (např. průměr závisle proměnné) v regionu  $t_m$ ,  $M$  označuje počet regionů, tedy počet terminálních uzlů, a  $I(\mathbf{x} \in t_m)$  indikátorovou funkci, která nabývá hodnoty 1, pokud  $\mathbf{x} \in t_m$ , a hodnoty 0 v opačném případě (Hastie a kol., 2009). Případ, jehož indikátorová funkce je rovna 1 v určitém terminálním uzlu, je do tohoto uzlu klasifikován a je mu přiřazena predikce odpovídající průměru závisle proměnné v tomto uzlu.

V rámci tohoto příspěvku si představíme dvě metody regresních stromů, a to metodu CART (Classification and Regression Trees) a metodu CHAID (Chi-square Automatic Interaction Detection).



## 2.1. Classification and Regression Trees (CART)

Mezi nejznámější algoritmy tvorby jak regresních, tak klasifikačních stromů patří metoda CART představená Leo Breimanem a jeho vědeckým týmem v roce 1984.

Růst stromu je založen na rekurzivním binárním dělení souboru s cílem maximálně snížit nepřesnost stromu. Nepřesnost stromu  $T$  určenou průměrnou kvadratickou chybou značíme  $R(T)$  a vypočítáme ji pomocí vzorce

$$R(T) = \frac{1}{N} \sum_{t \in \tilde{T}} \sum_{\mathbf{x}_n \in t} (y_n - \bar{y}(t))^2,$$

kde, pro  $n = 1, \dots, N$  označuje  $\mathbf{x}_n$  vektor hodnot prediktorů a  $y_n$  hodnotu závisle proměnné  $n$ -tého pozorování,  $t$  značí terminální uzel,  $\tilde{T}$  množinu všech terminálních uzlů a  $\bar{y}(t)$  predikovanou hodnotu závisle proměnné v terminálním uzlu (Breiman a kol., 1993).

Dělení souboru pokračuje do doby uplatnění některého z předem nastavených pravidel, kterým je nejčastěji maximální počet pozorování v terminálním uzlu.

Tímto postupem naroste složitý strom, který bude mít malou chybu na datech, na kterých byl vytvořen, tj. na tzv. trénovacím souboru. Jelikož by jeho predikční schopnost na nezávislém (testovacím) souboru byla určitě nižší, je strom zpětně prořezáván. Prořezáváním rozumíme odstraňování „nejslabších“ větví, které nejméně snižují nepřesnost stromu.

Takto je vytvořena posloupnost různě velkých stromů, ze kterých vybereme finální. Pro výběr optimálně složitého stromu je využit nezávislý odhad nepřesnosti modelu na základě metody krosvalidace nebo testovacího souboru (Breiman a kol., 1993).

## 2.2. Chi-square Automatic Interaction Detection (CHAID)

CHAID je metoda tvorby rozhodovacích stromů představená Gordonem Kassem v roce 1980. Primární význam měla tato metoda pro řešení problematiky klasifikace kategoriálních dat, avšak s jistými modifikacemi může být uplatněna i v případě spojité závisle proměnné (Kass, 1980).

V kontrastu s metodou CART, CHAID vytváří nebinární stromy, kdy se každý neterminální uzel může rozdělit i více jak na dva uzly dceřiné. Při tomto postupu se soubor dělí rychleji a v porovnání se stromy binárními vznikají rozrostlé stromy s méně patry. CHAID je tudíž vhodný spíše pro větší datové soubory (Komprdová, 2012). Zásadní charakteristikou metody CHAID a jejím určitým omezením je, že pracuje pouze s kategoriálními prediktory. Spojité prediktory je před samotným procesem tvorby stromu potřeba rozdělit do ordinálních kategorií, v ideálním případě se stejným počtem pozorování (Nisbet a kol., 2009).

Celý postup tvorby stromu se skládá ze tří kroků, které se cyklicky opakují. Prvním z nich je slučování kategorií prediktorů, které jsou v hodnotách závisle proměnné nejpodobnější. Dále dochází k dělení uzlu na základě nově vzniklé sady kategorií prediktoru. Nakonec jsou testována pravidla pro zastavení růstu stromu (velikost skupin, míra snížení nepřesnosti, počet pater stromu atd.). Slučení kategorií prediktoru a nejlepší dělení uzlu je určeno pomocí analýzy rozptylu (Nisbet a kol., 2009).

## 3. Skupinové modely

Regresní stromy trpí velkou nestabilitou. K řešení tohoto problému se nabízí využití skupinového modelu založeného na kombinaci několika různých slabých modelů, které jsou



charakterizovány malým zkreslením a velkým rozptylem. Kombinací více regresních stromů je možné snížit rozptyl a vytvořit predikčně silný model (Komprdová, 2012).

Skupinové modely jsou založeny na bootstrapových výběrech, což jsou náhodné výběry s opakováním. Z tohoto popisu vyplývá, že se některá pozorování mohou do jednoho bootstrapového výběru dostat i vícekrát, jiná naopak vůbec, a tudíž nejsou jednotlivé výběry nezávislé (Breiman, 1996).

Ve stručnosti bude představena metoda náhodného lesa a metoda bumping (Bootstrap Umbrella of Model Parameters).

### 3.1. Náhodné lesy

Intuitivně seskupením mnoha stromů vzniká les. Každý ze skupiny neprořezaných stromů tohoto tzv. náhodného lesa je vytvořen na vlastním bootstrapovém výběru z trénovacího souboru. Pro snížení korelace mezi jednotlivými stromy se navíc využívá náhodného výběru určitého počtu vysvětlujících proměnných pro dělení uzlu. Výsledná predikce je dána průměrem predikce přes všechny stromy lesa (Breiman, 2001).

Tato metoda nám může v porovnání se samotným prořezaným stromem zvýšit predikční schopnost, avšak její výstup není již tak dobře čitelný.

### 3.2. Bootstrap Umbrella of Model Parameters (Bumping)

Výhody samotného stromu a lesa kombinuje specifická skupinová metoda bumping. Tato metoda vytváří jednotlivé stromy na bootstrapových výběrech z trénovacího souboru. V tomto kroku již nedochází ke zprůměrování predikce, ale dle požadavků na přesnost predikce, složitost modelu a smysluplnost jednotlivých dělení je z této sady stromů vybrán finální strom (Tibshirani a Knight, 1999).

## 4. Závěr

Proces kultivace DRG systému je klíčový pro jeho správnou funkčnost, což je nezbytný předpoklad korektní kontroly výdajů a spravedlivého finančního ocenění nemocnic. Regresní stromy mohou sloužit pro nalezení významných klasifikačních faktorů a s tím související vytvoření hierarchické struktury klasifikace hospitalizačních případů.

Síla regresních stromů spočívá v jejich jednoduchosti a současné schopnosti zachytit kumulativní účinek faktorů ovlivňujících nákladnost hospitalizace pacienta.

## 5. Literatura

Breiman, L., Friedman, J., Olshen, R., Stone, C. 1993. Classification and regression trees. Chapman & Hall, Boca Raton, Fla., 358 pp. ISBN 04-120-4841-8.

Breiman, L. 1996. Bagging predictors. Machine Learning [online]. 24(2), pp. 123– 140. [cit. 16. 2. 2018]. DOI: 10.1023/A:1018054314350. ISSN 08856125. Dostupné z: <http://link.springer.com/10.1023/A:1018054314350>.

Breiman, L. 2001. Random Forests. Machine Learning [online]. 45(1), pp. 5–32. [cit. 16. 2. 2018]. DOI: 10.1023/A:1010933404324. ISSN 08856125. Dostupné z: <http://link.springer.com/10.1023/A:1010933404324>.

- Busse, R., Geissler, A., Quentin, W., Wiley, M. 2011. Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals [online]. Open University Press, Maidenhead, England, 458 pp. [cit. 1. 2. 2018]. ISBN 978-0-33- 524557-4. Dostupné z: [http://www.euro.who.int/\\_data/assets/pdf\\_file/0004/162265/e96538.pdf](http://www.euro.who.int/_data/assets/pdf_file/0004/162265/e96538.pdf).
- Hastie, T., Tibshirani, R., Friedman, J. 2009. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. Springer, New York, NY, 745 pp. ISBN 978-0-387-84857-0.
- Kass, G. V. 1980: An Exploratory Technique for Investigating Large Quantities of Categorical Data. *Applied Statistics* [online]. 29(2), pp. 119–127. [cit. 6. 2. 2018]. DOI: 10.2307/2986296. ISSN 00359254. Dostupné z: <http://www.jstor.org/stable/10.2307/2986296?origin=crossref>.
- Komprdová, K. 2012. Rozhodovací stromy a lesy. Akademické nakladatelství CERM, Brno, 98 pp. ISBN 978-80-7204-785-7.
- Nisbet, R., Elder, J. F., Miner, G., 2009. Handbook of statistical analysis and data mining applications [online]. Elsevier, Amsterdam, 824 pp. [cit. 6. 2. 2018]. ISBN 978- 0-12-374765-5. Dostupné z: <https://doc.lagout.org/Others/Data%20Mining/Handbook%20of%20Statistical%20Analysis%20%26%20Data%20Mining%20Applications%20%5BNisbet%2C%20Elder%20%26%20Miner%202009-06-05%5D.pdf>
- Tibshirani, R., Knight, K. 1999. Model Search by Bootstrap "Bumping". *Journal of Computational and Graphical Statistics* [online]. 8(4), pp. 671–686. [cit. 16. 2. 2018]. DOI: 10.2307/1390820. ISSN 10618600. Dostupné z: <http://www.jstor.org/stable/1390820?origin=crossref>

# Program screeningu karcinomu tlustého střeva a konečníku a hodnocení jeho výkonnosti

Ondřej Ngo<sup>1,2</sup>, Ondřej Májek<sup>1,2</sup>

<sup>1</sup> *Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno*

<sup>2</sup> *Ústav zdravotnických informací a statistiky ČR, Praha*

## Abstrakt

Programy screeningu nádorových onemocnění představují nepostradatelnou součást prevence těchto onemocnění. Dle evropských doporučení by měly být zaváděné programy plně organizované a populační. Současně je potřeba pro zajištění očekávané účinnosti a zachování příznivého poměru mezi přínosy a riziky screeningového programu zavést a dodržovat komplexní standardy kvality péče. Tyto aktivity jsou jedním z předmětů nově vznikajícího Národního screeningového centra (NSC), které se také současně snaží inovovat zavedené onkologické screeningové programy v České republice, tedy screening karcinomu prsu, děložního hrdla a karcinomu tlustého střeva a konečníku (kolorektálního karcinomu). V případě screeningu kolorektálního karcinomu je snahou NSC zavést centralizované vyhodnocování screeningového testu a současně inovovat zavedený systém adresného zvaní pro zajištění odpovídající senzitivity programu a podpoření vyšší aktivní participace osob z cílové populace na screeningu. Cílem tohoto příspěvku je popsat obecné principy screeningových programů ve vztahu ke kolorektálnímu karcinomu. Dále popsat v současnosti nastavený program screeningu kolorektálního karcinomu v České republice a také popsat nastavený systém kontroly výkonnosti, který zahrnuje monitoring screeningového procesu prostřednictvím klinických a administrativních dat (zejména prostřednictvím tzv. indikátorů výkonnosti) a hodnocení dopadu screeningového programu na populační zátěž těmito onemocněními.

## Klíčová slova

Zhoubné nádory, screeningový program, kolorektální karcinom, indikátory výkonnosti

## 1. Prevence zhoubných nádorů

Národová onemocnění představují velmi závažný zdravotní problém, který se týká řady vyspělých zemí. Tato onemocnění se řadí mezi jednu z nejčastějších příčin úmrtí a počet úmrtí se celosvětově neustále zvyšuje (WHO, 2002; Ferlay et al., 2015). Podstatnou roli pro minimalizaci tohoto negativního dopadu sehrává primární prevence, tedy eliminace rizikových faktorů, které přispívají ke vzniku nádorových onemocnění (např. kouření, obezita, špatné stravování atd.). Současně je ovšem i řada faktorů, které aktivně neovlivníme (např. genetická predispozice onemocnění, stárnutí, faktory vnějšího prostředí atd.) a sehrávají také významnou roli ve vývoji zhoubných nádorových onemocnění. V tomto ohledu jsou klíčovým nástrojem prevence screeningové programy (Abrahámová, 2003).

### 1.1. Onkologické screeningové programy

Onkologické screeningové programy mají za cíl odhalit zhoubné nádorové onemocnění v časných stádiích u osob, které jsou v průměrném riziku onemocnění (např. bez rodinné anamnézy daného nádorového onemocnění) a současně nemají klinické příznaky sledovaného onemocnění. Včasná diagnóza onemocnění v bezpříznakovém stadiu umožňuje detekovat nádor v lokalizovaném a případně i v prekancerózním stadiu. V takovém případě je velmi

často léčba mnohem šetrnější a především úspěšnější. Současně je léčba časných stádií podstatně levnější pro zdravotní systém v porovnání s léčeným pokročilým onemocněním. Nástrojem pro detekci časného stádia onemocnění je screeningový test, pro který se stanovuje odpovídající screeningový interval, tedy interval mezi dvěma screeningovými vyšetřeními (Segnan et al., 2010), který musí být ve vztahu k délce tzv. detekovatelné preklinické fáze onemocnění.

Abychom zachovali odpovídající bilanci mezi přínosy a riziky screeningového programu aplikovaného na širokou populaci, je potřeba zavést vhodné standardy kvality péče a monitorovat jejich naplnění zapojenými zdravotnickými zařízeními. Programy by tedy měly být zaváděny jako organizované se zajištěním odpovídající kvality ve všech bodech screeningového procesu v souladu s evropskými doporučeními (Arbyn et al., 2008; Perry et al., 2006; Segnan et al., 2010). Současně by měla být screeningová politika jako celek součástí zákonů nebo jiného oficiálního legislativního dokumentu (Karsa et al., 2008). Dle IARC Handbook of Cancer Prevention (Vainio et al., 2002) musí v rámci organizovaného screeningového programu existovat:

1. politika specifikující cílovou populaci, screeningovou metodu a interval
2. definovaná cílová populace
3. tým zodpovědný za kontrolu screeningových center
4. struktura pro rozhodování a zodpovědnost za management zdravotní péče
5. systém zajišťující kvalitu programu a sběr příslušných dat
6. monitoring výskytu nádorového onemocnění v cílové populaci

Pro dosažení nejvyšší úrovně organizace programu (tzv. populačního programu) je vyžadováno, aby byly způsobilé osoby, které splňují kritéria pro zařazení do screeningového programu, adresně pozvány ke screeningovému vyšetření ve stanoveném intervalu (Karsa et al., 2008). Úspěch screeningové programu opravdu závisí na pečlivém plánování a organizaci. Neorganizovaný screeningový program (někdy také nazývaný „šedý“ screening) často vede k menší účinnosti a je finančně nákladnější pro zdravotní systém kvůli nadužívání testů a jejich nedostatečné senzitivitě nebo specificitě. Dostupné vědecké důkazy dokládají u organizovaných programů snižování populační zátěže těmito onemocněními a dosažení lepšího pokrytí definované způsobilé populace. Lze tedy konstatovat, že k maximalizaci pozitivního dopadu je potřeba zavádět na populační úroveň pouze organizované screeningové programy (Segnan et al., 2010).

Rada Evropské unie v roce 2003 oficiálně doporučila, na základě pozitivních zkušeností z randomizovaných a observačních studií, provadět screening karcinomu prsu, tlustého střeva a konečníku a děložního hrdla. Principy pro úspěšné zavedení programů publikovala Rada Evropské unie v dokumentu „Council Recommendation of 2 December 2003 on cancer screening“. V současné době jsou v České republice úspěšně implementovány národní screeningové programy pro zmíněná tři onkologická onemocnění. Od roku 2014 byl zaveden systém přímého adresného zvaní osob z cílové populace ke screeningovému vyšetření a programy lze tak považovat za populační. V rámci legislativy screeningových programů vydalo Ministerstvo zdravotnictví ČR vyhlášku, která stanovuje obsah a časové rozmezí preventivních prohlídek. Současně stanovuje ke každému programu jeho kritéria a podmínky Věstník Ministerstva zdravotnictví ČR. Všechny podstatné informace se mohou odborníci a široká veřejnost dozvědět na oficiálních internetových stránkách programů [www.mamo.cz](http://www.mamo.cz), [www.kolorektum.cz](http://www.kolorektum.cz), [www.cervix.cz](http://www.cervix.cz).

## 1.2 Screeningový program karcinomu tlustého střeva a konečníku

Karcinom tlustého střeva a konečníku (kolorektální karcinom) je jedním z nejčastějších zhoubných nádorů (ZN) na světě a v posledních desetiletích dochází ve většině zemí ke zvyšování jeho výskytu. Kolorektálním karcinomem trpí převážně starší osoby a riziko onemocnění se výrazně zvyšuje od přibližně 50 let. Vzhledem k možnosti detekce časných stadií onemocnění (zejména přednádorových kolorektálních neoplazií), existenci účinné léčby v těchto raných stadiích a dostupnosti laciných screeningových testů je toto onemocnění vhodným kandidátem pro zařazení do screeningových programů (Garcia, 2012).

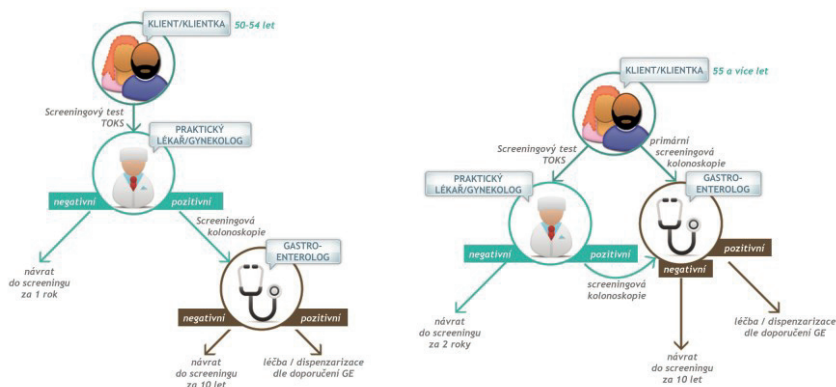
Program screeningu kolorektálního karcinomu prokazatelně snižuje úmrtnost na toto onemocnění. Prostřednictvím metaanalýz rozsáhlých randomizovaných klinických studií byly získány přesvědčivé důkazy o jeho přínosu. Při pravidelném vyšetřování pomocí guajakového testu na okultní krvácení do stolice (TOKS) bylo v americké randomizované studii sledováno snížení mortality na kolorektální karcinom o 33 % (Mandel et al., 1993). Ve studiích provedených v Evropě bylo taktéž dosaženo snížení mortality na kolorektální karcinom při vyšetření guajakovým testem jednou za dva roky o 15 % (Hardcastle et al., 1996), resp. 18 % (Kronborg et al., 1996). Tyto pozitivní výsledky přetrvávaly i po 9 kolech vyšetření a zejména u pacientů, kteří se účastnili pravidelně všech kol, došlo k redukci mortality až o 43 %. Jednoznačně se tak potvrzuje důležitost pravidelné účasti ve screeningu (Kronborg et al., 2004). Pokud zahrneme i nepublikované výsledky ze studie, která proběhla v Göteborgu, tak bylo odhadnuto snížení mortality na 16 % (Hewitson et al., 2008). Prostřednictvím alternativního, a v současné době doporučovaného, imunochemického TOKS lze dle současných poznatků zajistit ještě lepších výsledků než u guajakového, což naznačuje i první publikovaná randomizovaná studie (Zheng et al., 2003). Klíčovým nástrojem ve screeningu kolorektálního karcinomu je i kolonoskopie, kde je dosažitelné snížení incidence odhadováno až na 90 %. Ovšem pro potvrzení uvedené účinnosti v praxi prozatím chybějí přímé důkazy z randomizovaných klinických studií (Ransohoff, 2009). V současné době probíhají dvě velké randomizované studie, jejichž cílem bude potvrzení účinnosti kolonoskopie: jedna ve Španělsku a druhá organizovaná Severoevropskou iniciativou pro kolorektální karcinom (Bretthauer, 2011). V roce 2011 byl také publikován systematický přehled zaměřený na nákladovou efektivitu screeningu kolorektálního karcinomu (Lansdorp-Vogelaar et al., 2011). Studie hodnotila celkem 32 modelů z jednotlivých studií, které proběhly v Evropě, Severní Americe, Asii a Austrálii. Většina z uvedených studií hodnotila náklady na jeden získaný rok života z pohledu plátce zdravotní péče. I přesto, že se studie lišily z pohledu předpokladů a použitých parametrů, což se projevilo ve variabilitě výsledků, tak bylo zjištěno, že je screening kolorektálního karcinomu nákladově efektivní nebo dokonce náklady šetřící ve srovnání se situací, kdy není zaveden.

## 1.3. Epidemiologie kolorektálního karcinomu v ČR a zavedený screening

Česká republika patří v rámci Evropy mezi země nejzatíženější kolorektálním karcinomem. Dle aktuálních údajů z Národního onkologického registru ČR (NOR, dostupné na [www.svod.cz](http://www.svod.cz)) byl v ČR v roce 2015 výskyt tohoto onemocnění 76,4 nových onemocnění na 100 tisíc osob (přibližně 8 tisíc nových onemocnění ročně) a 35,6 úmrtí na 100 tisíc osob (přes 3,5 tisíce úmrtí na kolorektální karcinom ročně).

Pilotní snahy o zavádění časné detekce ZN tlustého střeva a konečníku spadají už do 80. let 20. století a následně pokračovaly i v devadesátých letech. V populaci byl sledován vysoký zájem o kolorektální screening. Samotný organizovaný program byl zahájen v červenci 2000 na základě dohody MZ ČR, plátců zdravotní péče a příslušných odborných společností. Na základě zkušeností byl přijat dvouetapový program: TOKS byl nabízen ve dvouletých screeningových intervalech a v případě positivity byla osoba zaslána na navazující

kolonoskopické vyšetření (Frič et al., 2007). Už od začátku je program sledován Radou pro screening kolorektálního karcinomu České gastroenterologické společnosti ČLS JEP. Sběr agregovaných dat pro monitoring kolonoskopických vyšetření byl realizován ze strany regionálních koordinátorů screeningového programu a sběr výsledků o provedených TOKS byl v gesci Všeobecné zdravotní pojišťovny. Od roku 2006 byl monitoring založen na elektronickém zadávání anonymních individuálních dat o kolonoskopických vyšetřeních do elektronické databáze. Na celý průběh programu dohlíží Komise pro screening kolorektálního karcinomu MZ ČR. Cílovou populací organizovaného programu jsou osoby nad 50 let. Počínaje lednem 2009 byl mírně upraven design screeningového procesu. Pro věkovou skupinu 50–54 let byl nově určen TOKS v jednoletém intervalu a osobám od 55 let byla dána možnost výběru mezi TOKS ve dvouletém intervalu nebo podstoupit jednou za 10 let přímo screeningovou kolonoskopií (schéma screeningového procesu je popsáno na obrázku 1). V prvních letech organizovaného programu byl používán guajakový TOKS, nicméně v posledních letech je nahrazován imunochemickým testem, který je přesnější a jeho použití je pro klienta jednodušší (Zavoral et al., 2014). V roce 2014 bylo v českém organizovaném programu zavedeno klíčové opatření, které mělo směřovat ke zvýšení participace osob z cílové populace na screeningu. Ke screeningovému programu jsou nyní zvány osoby ve věku od 50 do 70 let, které se dlouhodobě neúčastní screeningového vyšetření (Dušek et al., 2014).



**Obrázek 1.** Grafické schéma screeningového procesu v ČR. Vlevo pro osoby ve věku 50–54 let a vpravo pro klienty od 55 let (dostupné z [www.kolorektum.cz](http://www.kolorektum.cz))

#### 1.4. Národní screeningové centrum a plánované inovace kolorektálního screeningu v ČR

Národní screeningové centrum (NSC) je organizační jednotkou Ústavu zdravotnických informací a statistiky ČR (ÚZIS ČR), které má samostatnou koordinační a řídicí strukturu. Hlavním cílem NSC je vybudování metodického a personálního zázemí pro zavádění, realizaci a hodnocení screeningových programů v ČR. Současně je klíčovou aktivitou NSC vybudování systému evaluace screeningových programů, tak aby byl zajištěn jejich maximální pozitivní dopad na zdraví české populace a jejich vysoká nákladová efektivita. Vybudování tohoto zázemí umožňuje ÚZIS ČR realizovat pilotní projekty, které budou ověřovat potenciálně prospěšné screeningové programy vážných onemocnění nebo inovovat v současnosti zavedené programy.

Ve vztahu k zavedenému programu screeningu kolorektálního karcinomu připravuje NSC dva pilotní projekty, které přímo souvisí s jeho optimalizací. První projekt by měl řešit nepříznivou



situaci využívání různorodých imunochemických testů v ČR. Jeho ambicí je ověřit plošné zavedení centralizovaného vyhodnocení jednotného kvantitativního imunochemického testu na okultní krvácení do stolice (qFIT), potenciálně i s možností přímého zaslání odběrových nádobek osobám z cílové populace. Předmětem druhého projektu, který souvisí obecně se všemi aktuálně zavedenými onkologickými screeningovými programy, je zvýšení efektivity zavedeného systému adresného zvaní občanů ke screeningu k zajištění ještě větší účasti osob. Další podrobné informace o NSC a jeho aktivitách je k nalezení na oficiálním portálu [nsc.uzis.cz](http://nsc.uzis.cz) nebo na facebooku.

## **2. Hodnocení výkonnosti screeningových programů**

Základním cílem screeningových programů nádorových onemocnění je snížení populační zátěže těmito onemocněními, tedy zejména snížení mortality nebo v lepším případě incidence sledovaného onemocnění. Tato veličina je směrodatná pro výsledné hodnocení celkové efektivity programu. Ověřování tohoto faktu v klinické praxi je ovšem mnohem náročnější než je tomu v randomizovaných klinických studiích a kýžený efekt populačního screeningového programu navíc můžeme sledovat až za mnoho let. Do hodnocení účinnosti programu se mísí řada dalších faktorů, které ovlivňují změnu v epidemiologických trendech onemocnění – např. úmrtí u pacientů diagnostikovaných před zavedeným programem, postupný růst celopopulačního pokrytí screeningem, kohortový efekt (generačně specifické trendy ve výskytu rizikových faktorů), zavedení modernější léčby nebo uvědomělost cílové populace o časných příznacích nádorových onemocnění (Vainio et al, 2002).

Z výše uvedeného tedy vyplývá nutnost pečlivého monitoringu celého screeningového procesu od úplného začátku, aby bylo možné zajistit jeho nezbytnou výkonnost ve všech screeningových centrech a případně přijmout nápravná opatření tak rychle, jak umožňuje situace ve zdravotním systému. Pro tento účel byla vyvinuta sada indikátorů výkonnosti, která umožňuje průběžně monitorovat celý screeningový proces. Tyto indikátory jsou zejména zaměřeny na organizační aspekty (zda je např. dostatečná a pravidelná účast osob z cílové populace), sledování provedených klinických vyšetření (např. detekce zhoubného onemocnění či prekanceróz dle stanoveného cíle programu) a indikátorů dlouhodobého dopadu ve smyslu sledování trendů ve vývoji epidemiologických charakteristik a intervalových karcinomů. Monitoring vychází zejména ze tří výše zmíněných evropských doporučení (Arbyn et al., 2008; Perry et al., 2006; Segnan et al., 2010) vydaných Evropskou komisí. Tyto dokumenty se staly i zdrojem indikátorů hodnocených v rámci screeningových programů implementovaných v ČR.

### **2.1. Informační podpora screeningových programů v ČR**

Systém informační podpory je nepostradatelnou a vyžadovanou součástí organizovaného screeningového programu. Dle doporučení Rady Evropské unie z roku 2003 (cílový dokument zmíněn výše) je třeba „pravidelně monitorovat screeningový proces a jeho výsledky a s těmito informacemi v krátké době seznámat širokou veřejnost i instituce zapojené do screeningů“ a informační systém by měl být schopný „sbírat, zpracovávat a hodnotit data o všech screeningových vyšetřeních, doplňujících diagnostických vyšetřeních a finálních diagnózách“. V souladu s evropskými doporučeními a mezinárodními standardy byl v České republice zaveden systém informační podpory screeningových programů zhoubných nádorů, který umožňuje monitoring zátěže populace onkologickými onemocněními (Národní onkologický registr ČR, kde jsou evidovány všechny zhoubné novotvary v České republice), monitoring screeningového procesu prostřednictvím klinických (kontinuálně sbíraná data o screeningových vyšetřeních, klinických výsledcích a souvisejících vyšetření z jednotlivých

zdravotních zařízení zapojených do screeningu) a administrativních dat (data zdravotních pojišťoven reprezentující všechny provedené zdravotní výkony související se screeningovým procesem). Současně bylo se zavedením adresného zvaní občanů ke screeningu nastaveno také využívání a hodnocení dat zdravotních pojišťoven pro monitoring projektu. Všechny tři datové zdroje jsou zásadní pro hodnocení programů screeningu nádorových onemocnění a velmi dobře se doplňují pro získání kompletního přehledu o screeningu v ČR. Současně byl přijat legislativní rámec pro komplexnější a integrovanější využití dostupných dat v rámci Národního zdravotnického informačního systému. Samotné komplexní hodnocení programů zajišťuje v současné době Národní screeningové centrum Ústavu zdravotnických informací a statistiky ČR ve spolupráci s Institutem biostatistiky a analýz Lékařské fakulty Masarykovy univerzity (IBA LF MU).

## 2.2. Časné indikátory výkonnosti kolorektálního screeningu

Jak bylo zmíněno v úvodu kapitoly 2, k dosažení podobného účinku jako v řadě klinických studií je potřeba zajistit odpovídající výkonnost ve všech fázích screeningového procesu – od identifikace způsobilé populace ke screeningu a jejich adresného pozvání přes provedení screeningová a související vyšetření až po léčbu a následnou péči. Musí být zachována příznivá bilance přínosů a rizik při zavádění programu do reálné klinické praxe a je tedy potřeba nastavit systém hodnocení měřitelných standardů, resp. indikátorů výkonnosti screeningových programů.

Zásadním dokumentem pro zavedení a vyhodnocování screeningového programu kolorektálního karcinomu představují „*European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis*“ (Segnan et al., 2010). Jak bylo uvedeno v předchozím textu, tento dokument se stal předlohou indikátorů výkonnosti hodnocených v ČR. Uvedené směrnice vznikly po pečlivém prozkoumání vědeckých důkazů souvisejících s účinností kolorektálního screeningu a vzájemném konsenzu renomovaných odborníků v dané oblasti. Tyto směrnice tedy představují moderní doporučení a standardy k zajištění výkonnosti tohoto programu a měly by být nápomocné státům celé EU při implementaci programu se zajištěním vysoké úrovně populačního programu už od jeho samotného počátku. Evropská doporučení dokládají zejména indikátory, které mají doložený epidemiologický význam. Samotná forma výpočtu indikátorů výkonnosti je postavena na podílech a vyjadřují určitou míru jevu, který chceme monitorovat. Následující výpis indikátorů výkonnosti lze členit do dvou skupin dle oblasti hodnocení.

**(1)** První skupinou jsou tzv. **organizační indikátory**, které lze hodnotit při jakémkoliv nastaveném primárním screeningovém testu, a jsou jimi zejména:

*Pokrytí pozváním:* je definováno jako podíl počtu osob, které byly během daného screeningového intervalu pozvány na screening a mezi osobami v celé cílové populaci.

*Míra účasti:* reflektuje účast pozvaných osob ke screeningu, tedy podíl ze všech pozvaných, kteří se zúčastnili.

*Pokrytí vyšetřením:* představuje další klíčový indikátor hodnotící úspěšnost náběru české populace. Jeho hodnocení je nezbytné i z důvodu zavedeného adresného zvaní, které je orientováno na neúčastníci se osoby. Představuje tedy podíl počtu všech vyšetřených osob prostřednictvím screeningové testu v rámci celé cílové populace.

Dalšími organizačními indikátory jsou časové intervaly, které sledují přiměřený časový interval mezi jednotlivými událostmi v průběhu testování, tedy např. interval mezi zhotovením testu a obdržením výsledku, mezi pozitivním výsledkem testu a navazujícím kolonoskopickým



vyšetřením, mezi pozitivním kolonoskopickým vyšetřením a následným zahájením terapeutickým opatřením nebo po sobě jdoucími screeningovými testy.

**(2) Indikátory výkonnosti provedených screeningových testů** vychází z klinických studií a jsou hodnoceny pro různé typy primárních screeningových testů – pro TOKS nebo pro endoskopické metody (kolonoskopie nebo flexibilní sigmoidoskopie), které jsou v některých státech EU rovněž využívány jako přímé screeningové metody. Tyto dvě screeningové metody se v řadě indikátorů shodují, takže budou v úvodu uvedeny společné indikátory:

*Podíl chybně provedených vyšetření:* podíl chybně/neadekvátně provedených testů mezi všemi osobami, kterým bylo provedeno dané screeningové vyšetření.

*Podíl úplných endoskopických vyšetření:* definuje podíl osob s adekvátní střevní přípravou a zejména kompletní vizualizací kolorekta (s ohledem na použitou endoskopickou metodu) mezi všemi vyšetřenými osobami s adekvátní přípravou střeva.

*Podíl pozitivních screeningového vyšetření:* ukazatel vyjadřuje podíl osob s pozitivním výsledkem testu mezi všemi adekvátně testovanými. Osoby s pozitivním výsledkem testu (zejména TOKS) jsou zpravidla zaslány na navazující (follow-up) kolonoskopii.

*Podíl osob s pozitivním výsledkem testu, které jsou pozvány na follow-up kolonoskopii:* reprezentuje osoby, které měly pozitivní TOKS a byly pozvány na následnou follow-up kolonoskopii (TOKS+ kolonoskopie) k definitivnímu určení diagnózy. U endoskopických vyšetření je tato míra mnohem nižší a v případě kolonoskopie minimální, a to z důvodu odstranění lézí již během screeningové kolonoskopie.

*Podíl osob pozvaných na follow-up kolonoskopii, kteří ji podstoupí:* plynule navazuje na výše zmíněný indikátor a odráží účast pozvaných osob na follow-up kolonoskopii. Na uvedený indikátor také navazují další dva ukazatele, které sledují *podíl úplných follow-up kolonoskopii* (podobná definice jak bylo zmíněno výše) a dále *míru endoskopických komplikací* vzhledem k počtu všech vyšetřených osob.

*Detekční míra:* představuje detekci abnormalit v tlustém střevě a konečniku prostřednictvím jakéhokoliv typu screeningového testu, tedy jinými slovy podíl osob, u kterých byla potvrzena určitá abnormalita (může představovat jakékoli léze, adenomy, pokročilé adenomy nebo zhoubné novotvary) mezi všemi adekvátně testovanými jedinci. V případě TOKS souvisí s tímto ukazatelem také hodnocení *pozitivní prediktivní hodnoty* testu, která má ve jmenovateli oproti detekční míře pouze osoby s pozitivním výsledkem testu, které podstoupily navazující kolonoskopii.

V tabulce 1 je popsána sada v současnosti doporučených časných ukazatelů výkonnosti pro monitoring screeningu kolorektálního karcinomu, která je aktuálně v ČR s ohledem na dostupnost datových zdrojů využívána. Vzhledem k výše zmíněné inovaci informační podpory screeningových programů (propojení jednotlivých datových zdrojů) lze očekávat hodnocení dalších indikátorů výkonnosti v českém prostředí.

Současně jsou v tabulce 2 prezentovány doporučené hodnoty vybraných indikátorů výkonnosti dle evropských doporučení pro zajištění odpovídajícího dopadu kolorektálního screeningu na populační zátěž tímto onemocněním.

**Tabulka 1.** Časné indikátory výkonnosti používané v Národním programu screeningu kolorektálního karcinomu v ČR

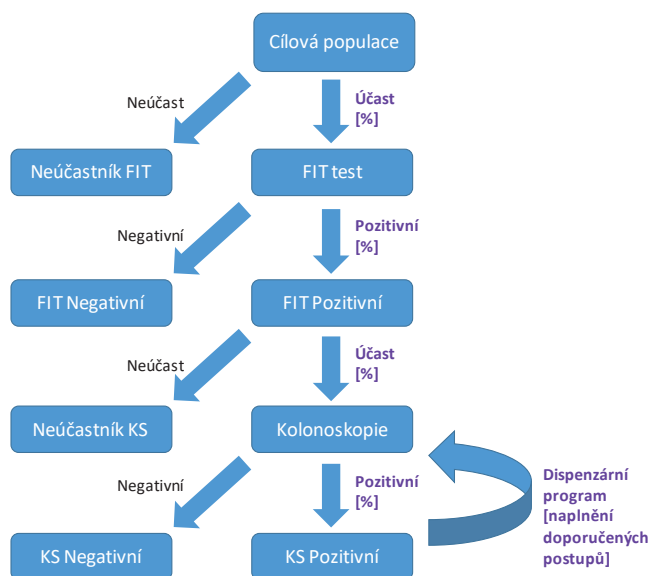
<p><b>1. Adresné zvaní a pokrytí cílové populace</b></p> <ul style="list-style-type: none"> <li>• Pokrytí pozváním (zvou se pouze neúčastníci se osoby od 50–70 let)</li> <li>• Míra účasti po pozvání</li> <li>• Pokrytí cílové populace screeningovým vyšetřením</li> </ul> <p><b>2. Výsledky screeningu prostřednictvím TOKS</b></p> <ul style="list-style-type: none"> <li>• Podíl pozitivních TOKS</li> <li>• Podíl úplných follow-up kolonoskopií</li> <li>• Pozitivní prediktivní hodnota TOKS pro adenomy, pokročilé adenomy, karcinomy</li> <li>• Míra endoskopických komplikací</li> </ul> <p><b>3. Výsledky screeningu prostřednictvím screeningové kolonoskopie</b></p> <ul style="list-style-type: none"> <li>• Podíl úplných kolonoskopií</li> <li>• Pozitivita kolonoskopie</li> <li>• Detekční míra pro adenomy, pokročilé adenomy, karcinomy (podíl pozitivních nálezů v celkovém počtu vyšetření)</li> <li>• Míra endoskopických komplikací</li> </ul> <p><b>4. Organizace screeningu</b></p> <ul style="list-style-type: none"> <li>• Časový interval pozitivním výsledkem TOKS a kolonoskopií</li> </ul>
--

**Tabulka 2.** Doporučené hodnoty vybraných indikátorů výkonnosti screeningu kolorektálního karcinomu dle evropských směrnic (Segnan et al., 2010)

INDIKÁTORY	Minimální hodnota	Preferovaná hodnota
Pokrytí pozváním	95 %	> 95 %
Míra účasti	45 %	> 65 %
Maximální doba mezi zhotovením testu a obdržením výsledků by měla být 15 dní	> 90 %	
Maximální doba mezi pozitivním výsledkem testu a navazujícím kolonoskopickým vyšetřením by měla být 31 dní	> 90 %	> 95 %
Maximální doba mezi pozitivní kolonoskopií nebo flexibilní sigmoidoskopií a zahájením léčebného opatření by měla být 31 dní	> 95 %	
Podíl chybně provedených TOKS	< 3 %	< 1 %
Podíl osob s pozitivním výsledkem testu, které jsou pozváni k navazující kolonoskopii	90 %	> 95 %
Podíl osob pozvaných k navazující kolonoskopii, které ji podstoupí	85 %	> 90 %
Podíl úplných kolonoskopií (screeningové a navazující kolonoskopie jsou hodnoceny zvlášť)	> 90 %	> 95 %

### 2.3. Vývojový diagram screeningové procesu kolorektálního karcinomu

Pro lepší porozumění onkologickým preventivním programům bývá často jejich screeningový proces popsán prostřednictvím vývojového diagramu (tzv. flow-chart). Tímto způsobem se sledují zejména podstatné procesní charakteristiky, respektive relevantní výše zmíněné indikátory výkonosti (např. míra účasti, pozitivita testu, návštěva navazujícího diagnostického vyšetření u pacientů s pozitivním výsledkem). Na obrázku 2 je znázorněn procesní diagram pro screeningový program kolorektálního karcinomu v ČR. V českých podmínkách je k dispozici dostatečná informační podpora pro sestavení uvedeného diagramu, který posléze často slouží ke kritické analýze současně nastaveného programu a identifikaci jeho konkrétních rezerv - například nedostatečné pokrytí a nepravidelnost návštěv, nestandardizovaný screeningový test, nedostatečná aktivita sítě primární péče atd. (Ngo et al., 2017). Současně mohou být také odhaleny možné rezervy v kontinuitě procesu a kapacitní nedostatky sítě kolonoskopických center. Ty mohou být způsobeny suboptimálním využíváním kapacity kolonoskopických center (neorganizovaná dispenzarizace pacientů s nálezem střevní neoplazie) a vedou k prodlužování čekacích časů na kolonoskopii, která pak může druhotně vést ke snížení účasti na kolonoskopii u FIT-pozitivních pacientů.



**Obrázek 2.** Procesní diagram screeningu kolorektálního karcinomu. Fialově jsou uvedeny příklady procesních indikátorů.

### 2.4. Hodnocení dopadů screeningových programů nádorových onemocnění

Jak již bylo v úvodu kapitoly zmíněno, hlavním cílem screeningových programů je snížení populační zátěže těmito onemocněními. U ZN tlustého střeva a konečníku lze zcela předejít vývoji onemocnění odstraněním prekancerózní kolorektální léze (zejména adenomový polyp) a současně jsme případně schopni zachytit onemocnění v časných stádiích. Kolorektální screening nevede tedy pouze ke snížení mortality onemocnění, ale také ke snížení celkové incidence. Zmíněné epidemiologické charakteristiky představují důležité veličiny pro vyhodnocování celkové efektivity screeningového programu kolorektálního karcinomu a je tedy podstatné jejich sledování.

Randomizované studie představují v této problematice zlatý standard a prostřednictvím precizního designu těchto klinických studií získáváme relevantní důkazy o účinnosti (efficacy) onkologických programů (Zvárová & Malý 2003). Příznivé výsledky z mnoha randomizovaných studií byly impulzem pro zavedení screeningových programů do klinické praxe řady zemí. Důkazy z klinických studií je však nezbytné doplnit i aktuálními studiiemi potvrzujícími praktickou účinnost (effectiveness) screeningu i v reálně implementovaných screeningových programech (Moss et al., 2012). Ověření jejich pozitivního vlivu v klinické praxi probíhá prostřednictvím tzv. observačních studií, jako jsou například studie případů a kontrol, incidenční studie nebo studie analyzující epidemiologické trendy (tyto studie v sobě zahrnují rozsáhlou škálu statistického hodnocení – zejména různé typy regresního modelování a také mikrosimulační modely).

### 3. Závěr

V zemích s velkou populační zátěží onkologickými onemocněními jsou screeningové programy důležitou složkou prevence. Česká republika dlouhodobě zaujímala přední příčky v incidenci a mortalitě těchto onemocnění při srovnání s ostatními evropskými zeměmi, a proto bylo o to více potřebné jejich zavedení do klinické praxe. V současné době jsou v České republice úspěšně implementovány plně organizované screeningové programy karcinomu prsu, tlustého střeva a konečníku a děložního hrdla. Od roku 2014 navíc došlo k zahájení adresného zvaní občanů ke screeningu a lze považovat tyto programy i za zcela populační.

Pro zajištění odpovídající výkonnosti a účinnosti programů je nezbytný jejich průběžný monitoring. Nepostradatelnou roli zde hraje komplexní informační podpora onkologických screeningových programů, která je v ČR na velmi dobré úrovni v souladu s evropskými doporučeními a jejím prostřednictvím lze hodnotit celý screeningový proces. Pro účely pravidelného monitoringu byla vyvinuta sada tzv. časných indikátorů výkonnosti. Tyto indikátory jsou schopny v raných fázích programu identifikovat jeho rezervy a včas stanovit opatření k jeho optimalizaci. Pouze díky takto nastavenému systému průběžného hodnocení výkonnosti a implementací nápravných opatření zajistíme kýžený dopad programů na populační zátěž těmito onemocněními. V rámci kolorektálního screeningu jsou v ČR tyto indikátory rutinně využívány k hodnocení celého programu a systém kontroly kvality jednotlivých center je postupně vyvíjen. Po zavedení kolorektálního screeningu sledujeme v posledních letech snižování populační zátěže kolorektálním karcinomem v ČR. I přes tento pozitivní dopad lze v organizaci programu identifikovat rezervy, jejichž řešení poskytuje příležitost k dalšímu zlepšování tohoto programu. V rámci pilotních projektů bude dopad těchto inovací zkoumat nově ustanovené Národní screeningové centrum v rámci Ústavu zdravotnických informací a statistiky.

## 4. Literatura

### 4.1. Knihy

- Abrahámová, J. (ed.) Možnosti včasného zachytu rakoviny prsu. Praha: Grada Publishing, 2003. 227 p.
- Arbyn, M., Anttila, A., Jordan, J., Ronco, G., Schenck, U., Segnan, N., et al. European guidelines for quality assurance in cervical cancer screening. Luxembourg: European Communities, 2008. 291 p.

- Karsa, L., Anttila, A., Ronco, G., Ponti, A., Malila, N., Arbyn, M., et al. Cancer Screening in the European Union: Report on the implementation of the Council Recommendation on cancer screening. Luxembourg: European Communities, 2008. 136 p.
- Perry, N., Broeders, M., de Wolf, C., Tornberg, S., Holland, R., von Karsa, L., et al. (ed.) European guidelines for quality assurance in breast cancer screening and diagnosis. Luxembourg: Office for Official Publications of the EC, 2006. 416 p.
- Segnan, N., Patnick, J., von Karsa, L. (ed.) European Guidelines for Quality Assurance in Colorectal Cancer Screening and Diagnosis. Luxembourg: Publications Office of the European Union, 2010. 386 p.
- Vainio, H., Bianchini, F. (ed.) Breast Cancer Screening. Lyon: IARC Press, 2002. 229 p.
- World Health Organisation. National cancer control programmes : policies and managerial guidelines, second edition. Geneva: World Health Organisation, 2002. 180 p.
- Zvárová, J., Malý, M. (ed.). Statistické metody v epidemiologii. Praha: Karolinum, 2003. 506 p.

## 4.2. Články

- Bretthauer, M., 2011. Colorectal cancer screening. *J. Intern. Med.* 270, 87–98.
- Dušek, L., Májek, O., Blaha, M., Daneš, J., Zavoral, M., Seifert, B., Dvořák, V., Skovajsová, M., Suchánek, Š., Jung, T., Brzková, M., Klika, P., Klimeš, D., Koptíková, J., Gregor, J., 2014. Koncepce populačního screeningu v České republice, metodika a první výsledky adresného zvaní občanů do preventivních onkologických programů. *Klin. Onkol.* 27, 2S59-2S68.
- Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int. J. Cancer* 136, E359-386.
- Frič DrSc, P., Ph. D, M.Z., Seifert, B., Pokorný, P., Suchánek, Š., 2007. Screening sporadického kolorektálního karcinomu v ČR. *Medicina Praxi* 4, 343–345.
- Gimeno Garcia, A.Z., 2012. Factors Influencing Colorectal Cancer Screening Participation. *Gastroenterol. Res. Pract.* 483417.
- Hardcastle, J.D., Chamberlain, J.O., Robinson, M.H., Moss, S.M., Amar, S.S., Balfour, T.W., James, P.D., Mangham, C.M., 1996. Randomised controlled trial of faecal-occult-blood screening for colorectal cancer. *The Lancet* 348, 1472–1477.
- Hewitson, P., Glasziou, P., Watson, E., Towler, B., Irwig, L., 2008. Cochrane Systematic Review of Colorectal Cancer Screening Using the Fecal Occult Blood Test (Hemoccult): An Update. *Am. J. Gastroenterol.* 103, 1541–1549.
- Kronborg, O., Fenger, C., Olsen, J., Jørgensen, O.D., Søndergaard, O., 1996. Randomised study of screening for colorectal cancer with faecal-occult-blood test. *The Lancet* 348, 1467–1471.
- Kronborg, O., Jørgensen, O.D., Fenger, C., Rasmussen, M., 2004. Randomized study of biennial screening with a faecal occult blood test: Results after nine screening rounds. *Scand. J. Gastroenterol.* 39, 846–851.
- Lansdorp-Vogelaar, I., Knudsen, A.B., Brenner, H., 2011. Cost-effectiveness of Colorectal Cancer Screening. *Epidemiol. Rev.* 33, 88–100.
- Mandel, J.S., Bond, J.H., Church, T.R., Snover, D.C., Bradley, G.M., Schuman, L.M., Ederer, F., 1993. Reducing Mortality from Colorectal Cancer by Screening for Fecal Occult Blood. *N. Engl. J. Med.* 328, 1365–1371.

- Moss, S.M., Nyström, L., Jonsson, H., Paci, E., Lynge, E., Njor, S., Broeders, M., Euroscreen Working Group, 2012. The impact of mammographic screening on breast cancer mortality in Europe: a review of trend studies. *J. Med. Screen.* 19 Suppl 1, 26–32.
- Ngo, O., Bučková, B., Suchánek, Š., Seifert, B., Zavoral, M., Dušek, L., Májek, O., 2017. Participation of the Czech population in the Colorectal Cancer Screening Programme – trends and current situation. *Gastroenterol. Hepatol.* 71, 377–383.
- Ransohoff, D.F., 2009. How Much Does Colonoscopy Reduce Colon Cancer Mortality? *Ann. Intern. Med.* 150, 50–52.
- Zavoral, M., Suchanek, S., Majek, O., Fric, P., Minarikova, P., Minarik, M., Seifert, B., Dusek, L., 2014. Colorectal cancer screening: 20 years of development and recent progress. *World J. Gastroenterol.* 20, 3825–3834.
- Zheng, S., Chen, K., Liu, X.Y., Ma, X.Y., Yu, H., Chen, K., Yao, K.Y., Zhou, L., Wang, L.B., Qiu, P.L., Deng, Y.C., Zhang, S.Z., 2003. Cluster randomization trial of sequence mass screening for colorectal cancer. *Dis. Colon Rectum* 46, 51–58.

# Interaktivní online reporting v praxi

Vojtěch Bulhart, Matěj Karolyi, Jakub Ščavnický, Martin Komenda

*Institut biostatistiky a analýz, Lékařská fakulta, Masarykova univerzita, Brno  
Ústav zdravotnických informací a statistiky ČR, Praha*

## Abstrakt

S rozvojem moderních technologií roste objem dat napříč Internetem. Vybrat relevantní a především garantovanou informaci není vždy snadné a často časově velmi náročné s ohledem na pestrost a počet dostupných online zdrojů. Vhodná forma prezentace dat pro vybrané cílové skupiny uživatelů, například v podobě dnes oblíbených vizualizací, může zásadním způsobem ovlivnit nejen výslednou přehlednost zveřejňovaných informací, ale také celkové pochopení souvislostí a tedy i správnou interpretaci dat. Z pohledu autora jakéhokoli webu je klíčové zamyšlení nad očekávanými výstupy, jejich formátem, rozsahem a v neposlední řadě také nad přidanou hodnotou pro uživatele. Tento příspěvek představuje obecný koncept datových vizualizací společně se zvoleným volně dostupným nástrojem pro rychlou a snadnou přípravu online tematicky ucelených reportů Google Data Studio.

## Klíčová slova

Vizualizace dat, interaktivní reporting, Google Data Studio

## 1. Úvod

S ohledem na dramatický nárůst objemu dostupných informací napříč Internetem ve všech oblastech lidského zájmu je otázka zpracování a následné vizualizace relevantních výsledků nad těmito daty velmi aktuální. Vhodná kombinace a případné propojení dostupných technologických řešení umožňující analytické zpracování a následnou prezentaci dat je zcela klíčové. V době, kdy se uživatelé bezproblémově pohybují napříč desítkami či stovkami online aplikací, je webová prezentace téměř nezbytností, a to jak v sektoru komerčním, tak ve státním či akademickém prostředí. Kompletní proces analýzy dat je netriviální a často je založen na jedné z ověřených metodiky pro získávání znalostí z různých zdrojů. Vždy se jedná o sekvenci kroků, které jsou logicky uspořádány a vzájemně provázány. Umožňují systematicky pokrýt všechny stěžejní fáze analýzy včetně často opomíjeného vyhodnocení. Příkladem může být například CRISP-DM, SEMMA nebo KDD přístupy (Azevedo and Santos, 2008). Pro analytika bez pokročilého inženýrského vzdělání v oblasti programování a vývoje je však interaktivní prezentace dosažených výstupů online mnohdy dost složitým problémem. Tento příspěvek si klade za cíl představit jeden z volně dostupných robustních webových nástrojů pro reporting dat, Google Data Studio, společně s obecným účelem a způsoby vizualizace dat. Na samém začátku každé analýzy bychom si měli určit, jaké výstupy vlastně očekáváme, v jakém formátu budou, pro jakou cílovou skupinu budou primárně určeny a v neposlední řadě také to, co mají uživatelé sdělit. Po osvojení si základního nastavení a ovládacích prvků aplikace je schopen běžný uživatel snadno a rychle vypublikovat online ucelený report, který bude přehledně prezentovat vybrané výstupy dané analýzy.

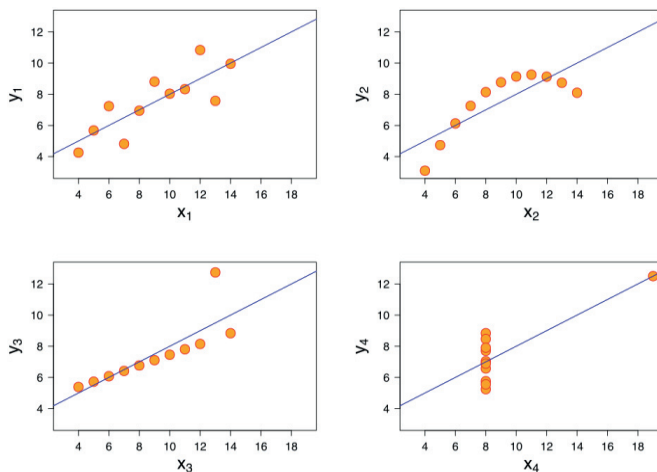
Vhodný způsob vizualizace odkrývá alternativní pohled na data. V mnoha případech jsou tyto vizualizace více informativní než tabulky nebo popisné statistiky. Příkladem může být tzv. *Amscombe's quartet*. Jedná se o čtyři různé dvourozměrné datové soubory, které jsou popsány

stejným lineárním modelem (až na rezidua), viz tabulka 1 (Tufte and Graves-Morris, 2014; Anscombe, 1973).

**Tabulka 1.** Amscombe's quartet

Data								
Pozorování	x1	y1	x2	y2	x3	y3	x4	y4
1,00	10,00	8,04	10,00	9,14	10,00	7,46	8,00	6,58
2,00	8,00	6,95	8,00	8,14	8,00	6,77	8,00	5,76
3,00	13,00	7,58	13,00	8,74	13,00	12,74	8,00	7,71
4,00	9,00	8,81	9,00	8,77	9,00	7,11	8,00	8,84
5,00	11,00	8,33	11,00	9,26	11,00	7,81	8,00	8,47
6,00	14,00	9,96	14,00	8,10	14,00	8,84	8,00	7,04
7,00	6,00	7,24	6,00	6,13	6,00	6,08	8,00	5,25
8,00	4,00	4,26	4,00	3,10	4,00	5,39	19,00	12,50
9,00	12,00	10,84	12,00	9,13	12,00	8,15	8,00	5,56
10,00	7,00	4,82	7,00	7,26	7,00	6,42	8,00	7,91
11,00	5,00	5,68	5,00	4,74	5,00	5,73	8,00	6,89
Statistiky								
Četnost	11,00	11,00	11,00	11,00	11,00	11,00	11,00	11,00
Průměr	9,00	7,50	9,00	7,50	9,00	7,50	9,00	7,50
Sm. odchylka	3,16	1,94	3,16	1,94	3,16	1,94	3,16	1,94
Korelace	0,82		0,82		0,82		0,82	
Model	Y = 3 + 0,5X		Y = 3 + 0,5X		Y = 3 + 0,5X		Y = 3 + 0,5X	

Dle statistik z tabulky 1 by se všechny datové sady mohly jevit stejné, po vizualizaci na obrázku 1 je možné v datech vidět značné rozdíly.



**Obrázek 1.** Bodové grafy dat z tabulky 1 spolu s regresní křivkou (Wikipedia, 2018).

## 2. Klíčové vlastnosti vizualizací

Vizualizace dat je důležitou součástí analýzy dat. Každý výstup výzkumu nebo pozorování musí být určitým způsobem prezentován cílové skupině. Datová vizualizace je vhodným a efektivním způsobem, jak efektivně předat informaci publiku v případě, že výstupem



pozorování je číselný soubor dat bez ohledu na zkoumanou doménu. Správná volba finální vizualizace je důležitá a netriviální činnost. Ale ještě více rozhodující je pochopit, že grafy a statistické výpočty jsou jen tak dobré jako data, na která jsou za nimi (Schneider et al., 2016). A kdykoli je to možné, informační grafika předkládaná čtenáři by měla:

1. Zobrazovat data
2. Přimět čtenáře, aby přemýšlel o podstatě datové sady spíše než o metodice, grafickém designu, technologiích grafické tvorby, nebo o něčem jiném
3. Nezkreslovat výsledky výzkumu (nevytvářet u čtenáře mylné domněnky)
4. Prezentovat mnoho čísel v malém prostoru (nebo pracovat dobře v proměnlivém webovém prostředí)
5. Učinit rozsáhlé datové soubory koherentní
6. Povzbudit oko ke srovnání různých dat (poukázat na zajímavé fakty a fenomény)
7. Odhalit údaje na několika úrovních detailů, od širokého přehledu k jemné struktuře
8. Sloužit rozumně jasnému účelu: popis, průzkum, tabulka, nebo dekorace
9. Být úzce spojena se statistickými a verbálními popisy souboru dat (Schneider et al., 2016)

## 2.1. Způsoby tvorby datových vizualizací

Zajímavou otázkou je výběr způsobu, jakým informace předána koncovému publiku. Existují mnoho způsobů, jak prezentovat informace, která je kódována kvantitativními nebo kvalitativními daty. Forma vizualizace a vybraný způsob vypravování má velký vliv na interpretaci dat i emocionální dopad na jedince. Je to také o úrovni znalostí publika o konkrétních tématech předložených zpráv. Užívání všechny tyto faktory nám pomáhají vyprávět příběh pomocí dat.

Vypravování pomocí dat ve webovém prostředí integruje vizualizace do příběhů stále častěji a dokonce umožňuje, aby vizualizace v určitých situacích nahradily písemnou formu předávání informace (Kosara and Mackinlay, 2013). Existuje více přístupů k sestavení vizualizace a každý z nich je vhodný v jiné situaci. Případně se dají kombinovat. Volba správného nosiče zpráv je jednou z největších povinností vypravěče. Hlavní dělení přístupů je následující<sup>1</sup>:

- **Autorem řízené vizualizace** – jedná se o vizualizace zcela vedené a usměrněné autorem. Z pravidla je sada statických obrazů předložena uživateli, který následně pasivně konzumuje obsah. Velkou výhodou tohoto přístupu je, že vizualizace může být velmi sofistikovaná a propracovaná do nejmenších detailů. Autor může sdělit dobře strukturovaný a komplexní příběh.
- **Uživatelé řízené vizualizace** – na opačné straně existuje přístup, u kterého je uživatel tvůrcem příběhu. Uživatelé zde mohou manipulovat s grafy a provádět vlastní průzkumné analýzy. Modifikovatelnost vizualizací může být obvykle přizpůsobena přidáním možnosti filtrovat podél více dimenzí. Různí uživatelé mohou odvodit různé závěry v závislosti na jejich procházení datové sady a jejich předchozí zkušenosti (s touto nebo jinou vizualizací).
- **Hybridní vizualizace** – poslední typ přebírá nejlepší součásti z předchozích dvou. Autor vybuduje část příběhu věnovaného tomuto problému, například poukáže na důležité okamžiky, odchylky nebo anomálie. A současně může uživatel podrobněji upravovat a filtrovat prezentované grafy a zobrazovat údaje do větší míry detailu. Typickými zástupci této skupiny jsou:
  - „Martini glass structure“ – začíná autorovým příběhem a následuje interaktivní částí (viz obrázek 2).

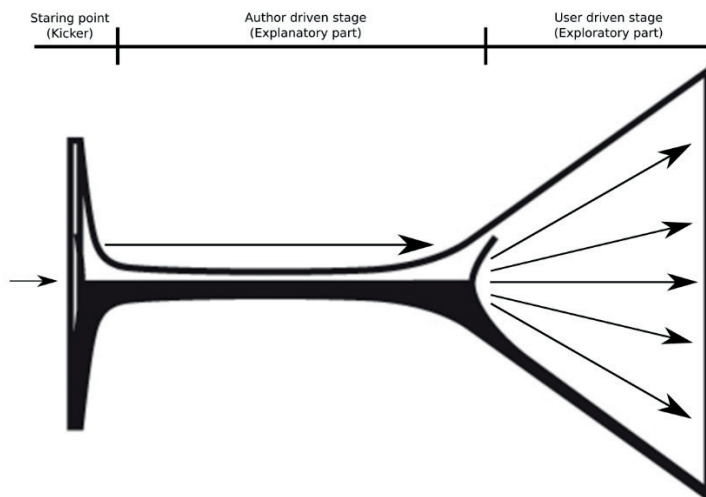
<sup>1</sup> <https://ieeexplore.ieee.org/abstract/document/5613452/>

- „Drill down story“ – nabízí přehled hlavních informací, ale dává uživateli svobodu výběru. Zvolí si tedy tu část příběhu, která ho zaujme.
- „Interactive slideshow“ – kopíruje strukturu tradiční prezentace, ale přidává do snímků interaktivní prvky.

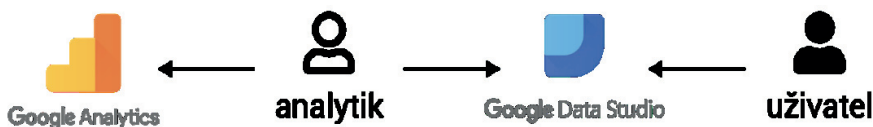
### 3. Technologie Google Data Studio

Google Data Studio ([datastudio.google.com](https://datastudio.google.com)) je online služba od společnosti Google k vytváření intuitivních a interaktivních přehledů dat. Designem se příliš neliší od svých sourozenců Google Apps se kterými je provázané, tudíž práce s tímto nástrojem je jednoduchá. Nejčastější použití může být vytvoření přehledu (dashboardu) analytikem a tu sdílet běžnému uživateli/manažerovi.

Mezi hlavní prvky Google Data Studia patří (i) propojení dat, (ii) vizualizace dat (tvorba dashboardu), (iii) spolupráce a sdílení<sup>2</sup>.



Obrázek 2. Martini glass structure (Karolyi, 2018)



Obrázek 3. Schéma možného využití Google Data Studio

<sup>2</sup> <https://support.google.com/datastudio/answer/7333350?hl=en>

### 3.1. Propojení dat

Primárním účelem služby Google Data Studio je zjednodušit prezentaci dat z ostatních služeb Googlu, proto při tvorbě nového přehledu (dashboard) je nabídka na propojení s ostatními službami bohatá. Propojit data můžeme ze služeb:

- Google Analytics
- Google Sheets
- Google Ads
- BigQuery
- YouTube Analytics

Další možností je využít propojení se službami třetích stran (Facebook Analytics, Amazon, Twitter, MailChimp, PayPal) nebo import vlastních souborů .csv (omezení do 100 MB)<sup>3</sup>. U importu vlastního .csv souboru je důležité věnovat pozornost kódování (UTF-8) a oddělovat hodnoty „čárkou“.

### 3.2. Tvorba přehledu

Přehled (dashboard) je hlavní prezentační část Google Data Studio. Po propojení s určitými daty můžeme začít vytvářet náš přehled dat. Největší výhodou je, že si lze celý přehled nastýlovat dle vlastního přání (např. využití prvků a barev jednotného vizuálního stylu). Další možností rychlé tvorby přehledu je využít již stávajících šablon vytvořených jinými uživateli<sup>4</sup>. Další šablony lze nalézt v rámci těchto zdrojů<sup>5, 6, 7</sup>. Základními prvky pro zobrazení dat jsou grafy, tabulky, rychlé přehledy a heatmapy. Ty se stávají interaktivní pomocí dodatečných filtrů, které lze uživatelsky měnit. Například filtr pro datum, konkrétní pohlaví, kraj apod. S každou změnou interaktivního filtru se všechny grafy, tabulky atd. ihned změní a zobrazí aktuální data s aplikovanými filtry. Dashboard lze kdykoliv upravovat, či přidat nová data. Vše je automaticky ihned ukládáno a aktualizováno.

### 3.3. Spolupráce a sdílení

Nedílnou součástí je i možnost sdílení a spolupráce na ostatních člen týmu. Tvorba prvků a veškeré změny při paralelní práci dvou a více uživatelů jsou ihned viditelné. Výsledný dashboard s daty lze komukoliv sdílet přes odkaz, aniž by měl založený účet u společnosti Google, popřípadě ho lze vložit do webové stránky. Uživatel, který má odkaz na daný přehled, nebo přístup na webovou stránku, kde je přehled vložený, může využít interaktivních filtrů, aby získal detailnější pohled na data dle jeho přání.

### 3.4. Nevýhody Google Data Studio

- Nemožnost exportovat report (např. do PDF).
- Nemožnost vytisknout report.
- Omezené možnosti nastavení grafů (ve srovnání s Excelem či nástroji jako je Tableau, MS PowerBI či KlipFolio)<sup>8</sup>.

---

<sup>3</sup> <https://support.google.com/datastudio/answer/7333350?hl=en>

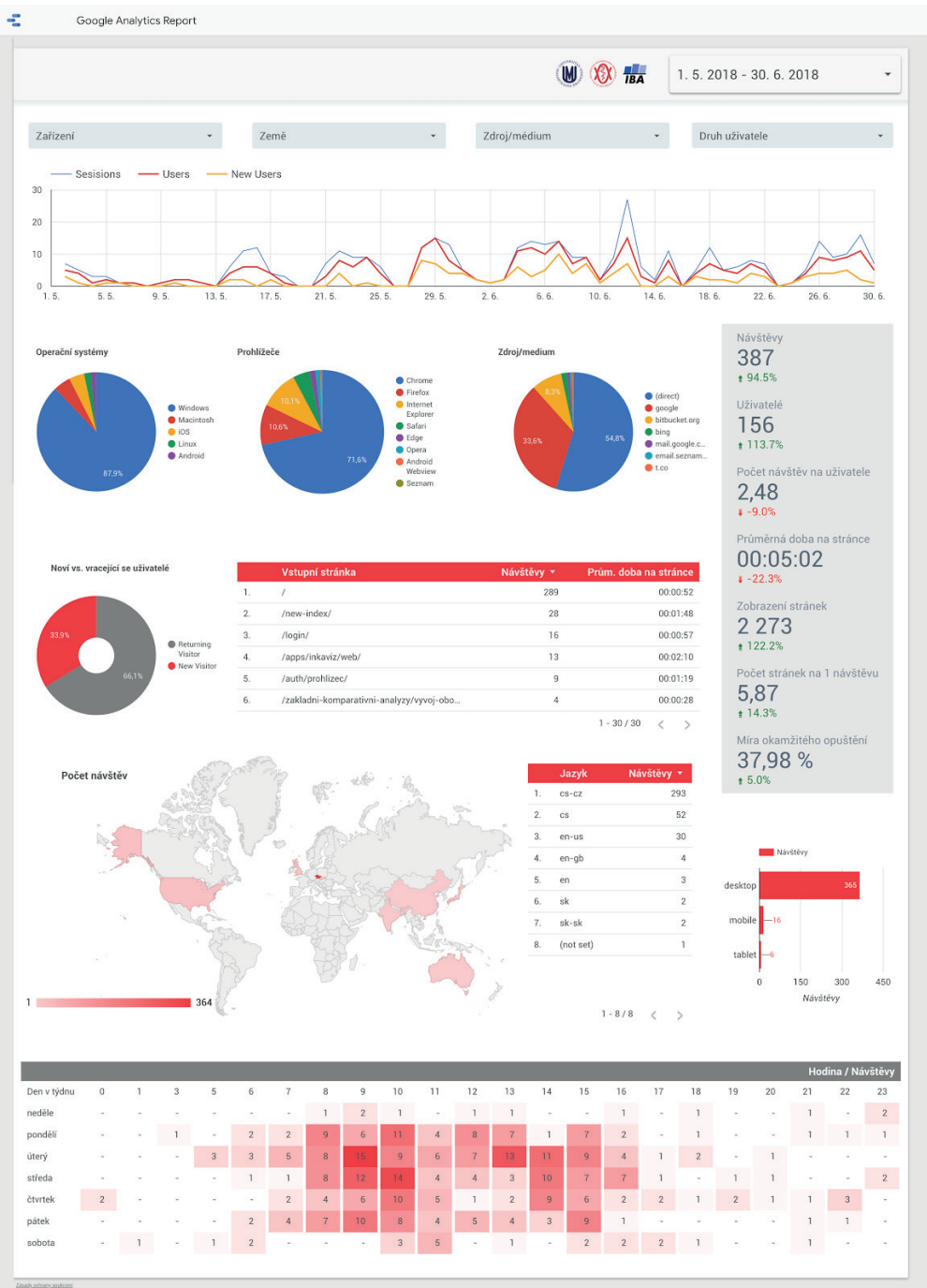
<sup>4</sup> <https://datastudiogallery.appspot.com/gallery>

<sup>5</sup> <https://www.lunametrics.com/blog/2016/08/25/free-data-studio-report/>

<sup>6</sup> <https://oneppcagency.co.uk/everything-else/data-studio-google-analytics-template/>

<sup>7</sup> <https://www.lunametrics.com/blog/2016/10/25/free-ecommerce-data-studio-template/>

<sup>8</sup> <https://miroslavpecka.cz/blog/google-data-studio-sablony/>



Obrázek 4. Výsledný přehled dat – Google Data Studio

## 4. Závěr

Příspěvek se zaměřuje na základní principy vizualizace dat společně s konceptem jejich systematického vytváření. Primární důraz musí být vždy kladen na porozumění datovému souboru a snaze najít co možná nejjednodušší cestu, jak data přehledně a srozumitelně prezentovat cílové skupině uživatelů. Z technologického pohledu je zde jako jeden z možných volně dostupných nástrojů představeno Google Data Studio, které svými širokými možnostmi poskytuje platformu pro online tvorbu a následné sdílení různých přehledů či reportů.

## 5. Literatura

- Anscombe, F. J. (1973). *The American Statistician* 27. *Graphs in Statistical Analysis*, (1), 17-21.  
doi:10.2307/2682899
- Azevedo, A. I. R. L., & Santos, M. F. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. *IADS-DM*.
- Karolyi, M. *Analýza chování uživatelů v systému pro správu výuky* [online]. Brno, 2018 [cit. 2018-08-14]. Dostupné z: <<https://is.muni.cz/th/m9sd8/>>.
- Kosara R.; Mackinlay J (2013). *Storytelling: The Next Step for Visualization*.
- Schneider, B., Ernst, C., & Wöpking, J. (Eds.). (2016). *Diagrammatik-Reader: Grundlegende Texte aus Theorie und Geschichte*. Walter de Gruyter GmbH & Co KG.
- Tufte, E., Graves-Morris, P. (2014). *The visual display of quantitative information.*; 1983.
- Wikipedia contributors. (2018, June 12). Anscombe's quartet. In *Wikipedia, The Free Encyclopedia*. Retrieved 03:56, August 21, 2018, from [https://en.wikipedia.org/wiki/Anscombe%27s\\_quartet#/media/File:Anscombe%27s\\_quartet\\_3.svg](https://en.wikipedia.org/wiki/Anscombe%27s_quartet#/media/File:Anscombe%27s_quartet_3.svg)

**Analýza klinických a epidemiologických dat: od přípravy dat až po vizualizaci výsledků**

11.–12. září 2018, Rakvice

Editoři: RNDr. Tomáš Pavlík, Ph.D., Mgr. Jakub Gregor, Ph.D.

Obálka: Mgr. BcA. Markéta Soukupová

Vydala Masarykova univerzita, Žerotínovo nám. 617/9, 601 77 Brno

První, elektronické vydání, 2018

ISBN 978-80-210-9031-6



MUNI  
MED

Institut  
biostatistiky  
a analýz

**muni**  
PRESS