

Incidence a prevalence zhoubných nádorů v regionech ČR

VERZE 2024-02.0

Metodický popis

Autorský kolektiv:	Jan Mužík ¹ , Denisa Krejčí ¹ , Michaela Cvanová ¹ , Petr Klika ¹ , Jiří Jarkovský ¹ , Klára Benešová ¹ , Daniel Klimeš ¹ , Martin Komenda ¹ , Ladislav Dušek ¹ ¹ Ústav zdravotnických informací a statistiky ČR
Recenzenti:	
Datum vytvoření dokumentu:	3. 8. 2023
Datum poslední aktualizace:	27. 11. 2024
Datum schválení pověřencem pro ochranu osobních údajů ÚZIS:	18. 3. 2024
Datum schválení poradou vedení ÚZIS:	30. 4. 2024
Typ výstupu:	Datová sada
ID výstupu:	NR-07-01
Klíčová slova:	NOR, Národní onkologický registr, novotvary, populace, incidence, prevalence
EuroVoc:	NT1 zdravotní statistika, NT1 rakovina
Citace:	Mužík J., Krejčí D., Cvanová M., Klika P., Jarkovský J., Benešová K., Klimeš D., Komenda M., Dušek L. Incidence a prevalence zhoubných nádorů v regionech ČR. <i>Národní zdravotnický informační portál</i> [online]. Praha: Ministerstvo zdravotnictví ČR a Ústav zdravotnických informací a statistiky ČR, 2023. Dostupné z: http://www.nzip.cz/data/1770-novotvary-incidence-prevalence-regiony-otevrena-data . ISSN 2695-0340
Datová sada dostupná na:	https://www.nzip.cz/data/1770-novotvary-incidence-prevalence-regiony-otevrena-data
Kontakt:	analyzy@uzis.cz

Popis datové sady

1.1 Základní informace

Datová sada „Incidence a prevalence zhoubných nádorů v regionech ČR“ umožňuje podrobné analýzy incidence a prevalence nádorů v ČR a krajích. Zdrojem dat jsou individuální anonymizované záznamy Národního onkologického registru (NOR) od roku 1977 do současnosti a obsahují základní sociodemografické údaje pacienta, údaje o diagnóze novotvaru a rozsahu onemocnění a údaj o případném úmrtí pacienta.

1.2 Primární cíl

Primárním cílem je poskytnout data pro komplexní hodnocení incidence a prevalence nádorů na úrovni celé ČR a krajů podle místa bydliště pacienta.

1.3 Interpretace a informační hodnota

Datová sada umožňuje hodnocení incidence a prevalence onkologických onemocnění na úrovni jednotlivých územních celků (celá ČR, kraje) včetně hodnocení epidemiologických trendů od roku 1977 do současnosti, to vše v absolutních počtech a případně stratifikovaně podle pohlaví, věkových skupin, klinických stadií.

Pro hodnocení prevalence (počty osob žijících s daným onemocněním nebo jeho historií k danému datu nebo v daném období) obsahuje datová sada údaje o pořadí novotvaru u osoby a údaje o případném úmrtí (úmrtí ano/ne, rok úmrtí a věk při úmrtí).

Ve spojení s populačními demografickými údaji (nejsou součástí této datové sady) umožňují tato data hrubé či věkově standardizované přepočty na populaci v daném čase na daném území.

1.4 Informační limity

Z důvodu ochrany osobních údajů fyzických osob je limitována úroveň detailu popisných charakteristik (geografická lokalizace bydliště do úrovně kraje, věk při diagnóze a věk při úmrtí na 5leté věkové kategorie, diagnózy na 3 znaky MKN10, rok stanovení diagnózy a rok úmrtí, nikoliv celé datumy), není uváděno anonymní ID osoby, je uvedeno pouze pořadí novotvaru u pacienta. Je uváděno pouze místo bydliště pacienta, nikoliv místo diagnostiky či léčby.

1.5 Příklad použití

Typickým příkladem využití dat je hodnocení epidemiologických charakteristik vybraných území včetně hodnocení trendů, např.:

- vývoj incidence a prevalence nádorových onemocnění ve vybraném regionu;
- srovnání incidence a prevalence skupin onemocnění v krajích ČR za zvolené období;
- věkově specifická incidence nádorových onemocnění a její změny v různých časových obdobích.

Příklad konkrétního výstupu:

Incidence zhoubných novotvarů kromě nemelanomových kožních (C00–C97 bez C44) je dlouhodobě vyšší u mužů než u žen. V České republice bylo ročně diagnostikováno průměrně 32 253 onemocnění u mužů (616,9 na 100 000 mužů) a 27 958 onemocnění u žen (518,8 na 100 000 žen).

1.6 Ochrana osobních údajů a identity právnických osob

Datová sada je připravována v souladu se zákonem 372/2011 Sb. (Zákon o zdravotních službách a podmínkách jejich poskytování) §73 odst. 8 a zákonem 110/2019 Sb. §16.

Pro odstranění rizika identifikace fyzických osob jsou data kompletně anonymizována a neobsahují žádné identifikátory umožňující propojení řádků mezi sebou, jednotlivé řádky souboru jsou tak zcela nezávislé.

Ze stejného důvodu jsou datumové údaje prezentovány pouze v ročním detailu a geografické údaje (bydliště pacienta) pouze do úrovně krajů.

1.7 Klíčová slova

NOR, Národní onkologický registr, novotvary, populace, incidence, prevalence

1.8 Klasifikace dle EuroVoc

NT1 zdravotní statistika, NT1 rakovina

2 Technická charakteristika

- Periodicita aktualizace
 - 1 x ročně za uzavřený kalendářní rok.
- Formát datové sady: csv
- Počet záznamů: 3 035 000
- Počet popisných atributů: 16
- Velikost: 193 MB
- Odborný garant: Jan Mužík

3 Zdroj dat

Zdrojem dat pro datovou sadu je databáze Národního onkologického registru (NOR) a data dalších databází Národního zdravotnického informačního systému (NZIS), které jsou ve správě ÚZIS ČR dle zákona 372/2011 Sb. (Zákon o zdravotních službách a podmínkách jejich poskytování). Výše uvedené zdroje jsou zpracovány do analytické databáze NOR, která je sestavena z vypořádaných dat primárních hlášení do NOR a doplněna z dat Národního registru hrazených zdravotních služeb (NRHZS), Národního registru hospitalizací (NRHOSP) a Listu o prohlídce zemřelého (LPZ). Zpravodajskými jednotkami primárních dat jsou jednotliví poskytovatelé zdravotních služeb (PZS), kteří diagnostikují anebo léčí pacienty s nově zjištěnými novotvary.

4 Agregace a transformace dat

Datová sada obsahuje řádková data nově zjištěných novotvarů. Použité transformace vybírají z datumu diagnózy a úmrtí pouze rok a kategorizují věk pacientů do 5letých věkových kategorií, je uváděno pořadí onemocnění u pacienta, nikoliv identifikátor pacienta; vše s cílem ochrany osobních údajů.

5 Atributy datové sady

1. ID řádku

- Název parametru: id
- Datový typ: numeric
- Popis parametru: anonymní pořadové id záznamu
- Možné hodnoty: 1 až n

2. Pohlaví

- Název parametru: pohlavi
- Datový typ: numeric
- Popis parametru: pohlaví pacienta
- Možné hodnoty:
 - 1 – muž
 - 2 – žena

3. Kraj bydliště

- Název parametru: kraj_kod
- Datový typ: string 5 znaků
- Popis parametru: kraj bydliště v době zjištění novotvaru
- Možné hodnoty:
 - CZ010 Hlavní město Praha
 - CZ020 Středočeský kraj
 - CZ031 Jihočeský kraj
 - CZ032 Plzeňský kraj
 - CZ041 Karlovarský kraj
 - CZ042 Ústecký kraj
 - CZ051 Liberecký kraj
 - CZ052 Královéhradecký kraj
 - CZ053 Pardubický kraj
 - CZ063 Kraj Vysočina
 - CZ064 Jihomoravský kraj
 - CZ071 Olomoucký kraj
 - CZ072 Zlínský kraj
 - CZ080 Moravskoslezský kraj

4. Diagnóza novotvaru

- Název parametru: diagnóza_kod
- Datový typ: string 3 znaky
- Popis parametru: diagnóza nádorového onemocnění – MKN kód na 3 místa
- Možné hodnoty: viz. číselník diagnóz MKN10 o délce 3 znaky v rozsahu C00–D48

5. Skupina diagnózy novotvaru

- Název parametru: diagnoza_skupina
- Datový typ: numeric
- Popis parametru: skupiny diagnóz novotvarů pro statistická hodnocení
- Možné hodnoty:
 - 1 – ZN hlavy a krku (C00–C14, C30–C31)
 - 2 – ZN jícnu (C15)
 - 3 – ZN žaludku (C16)
 - 4 – ZN tlustého střeva a konečníku (C18–C20)
 - 5 – ZN jater a intrahepatálních žlučových cest (C22)
 - 6 – ZN žlučníku a žlučových cest (C23, C24)
 - 7 – ZN slinivky břišní (C25)
 - 8 – ZN hrtanu (C32)
 - 9 – ZN průdušnice, průdušky a plic (C33, C34)
 - 10 – zhoubný melanom kůže (C43)
 - 11 – nemelanomový kožní ZN (C44)
 - 12 – ZN pojivových a měk. tkání a perif. nervů (C47, C49)
 - 13 – ZN prsu (C50)
 - 14 – ZN hrdla děložního (C53)
 - 15 – ZN dělohy (C54, C55)
 - 16 – ZN vaječníku (C56)
 - 17 – ZN prostaty (C61)
 - 18 – ZN varlete (C62)
 - 19 – ZN ledviny (C64)
 - 20 – ZN močového měchýře (C67)
 - 21 – ZN mozku, míchy a jiných částí CNS (C70–C72)
 - 22 – ZN štítné žlázy (C73)
 - 23 – Hodgkinův lymfom (C81)
 - 24 – non-Hodgkinův lymfom (C82–C86)
 - 25 – mnohočetný myelom (C90)
 - 26 – leukémie (C91–C95)
 - 27 – ostatní zhoubné novotvary
 - 28 – novotvary in situ (D00–D09)
 - 29 – novotvary nezhoubné a neznámého chování (D10–D36, D37–D48)

6. Pořadí novotvaru u osoby

- Název parametru: novotvar_poradi
- Datový typ: numeric
- Popis parametru: pořadí novotvaru u osoby
- Možné hodnoty:
 - 1 – první
 - 2 – další

7. Pořadí novotvaru v rámci diagnózy

- Název parametru: novotvar_poradi_dg
- Datový typ: numeric
- Popis parametru: pořadí novotvaru v rámci diagnózy na 3 místa u pacienta
- Možné hodnoty:

- 1 – první
- 2 – další

8. Pořadí novotvaru v rámci diagnostické skupiny

- Název parametru: novotvar_poradi_dg_skupina
- Datový typ: numeric
- Popis parametru: pořadí novotvaru v rámci diagnostické skupiny u pacienta
- Možné hodnoty:
 - 1 – první
 - 2 – další

9. Pořadí novotvaru v rámci zhoubných novotvarů

- Název parametru: novotvar_poradi_maligni
- Datový typ: numeric
- Popis parametru: pořadí novotvaru v rámci zhoubných novotvarů (C00–C97) u pacienta
- Možné hodnoty:
 - 1 – první
 - 2 – další

10. Pořadí novotvaru v rámci zhoubných novotvarů bez dg. C44

- Název parametru: novotvar_poradi_maligni_bez_C44
- Datový typ: numeric
- Popis parametru: pořadí novotvaru v rámci zhoubných novotvarů (C00–C97) kromě nemelanomových kožních nádorů (C44) u pacienta
- Možné hodnoty:
 - 1 – první
 - 2 – další

11. Stadium onemocnění

- Název parametru: stadium
- Datový typ: string 1 znak
- Popis parametru: stadium onemocnění podle TNM klasifikace včetně neznámých stadií rozdělených podle úplnosti záznamu novotvaru
- Možné hodnoty:
 - 1 – stadium I
 - 2 – stadium II
 - 3 – stadium III
 - 4 – stadium IV
 - X – stadium neuvedeno z objektivních důvodů (časné úmrtí / bez protinádorové léčby / není zavedena TNM klasifikace)
 - Y – chybějící údaj o stadiu – neúplný záznam

12. Rok diagnózy

- Název parametru: rok_dg
- Datový typ: numeric
- Popis parametru: rok stanovení diagnózy novotvaru
- Možné hodnoty: od 1977 dále

13. Kód věkové kategorie při diagnóze

- Název parametru: vek_kategorie_kod_dg

- Datový typ: numeric
- Popis parametru: kód věkové skupiny pacienta při diagnóze novotvaru vycházející z číselníku věkových skupin
- Možné hodnoty: kódy věkových skupin vycházející z [číselníku věkových skupin](#); kódy odpovídají hodnotám věku 0–84, 85+

14. Rok úmrtí

- Název parametru: umrti_rok
- Datový typ: numeric
- Popis parametru: rok úmrtí pacienta
- Možné hodnoty: od 1977 dále

15. Kód věkové kategorie při úmrtí

- Název parametru: vek_kategorie_kod_umrti
- Datový typ: numeric
- Popis parametru: kód věkové skupiny pacienta při úmrtí vycházející z číselníku věkových skupin
- Možné hodnoty: kódy věkových skupin vycházející z [číselníku věkových skupin](#); kódy odpovídají hodnotám věku 0–84, 85+

16. Úmrtí pacienta

- Název parametru: umrti
- Datový typ: numeric
- Popis parametru: úmrtí pacienta
- Možné hodnoty:
 - 0 – žije
 - 1 – zemřel