Očkování, pozitivity, hospitalizace pro COVID-19, úmrtí, long covid a komorbidity u osob v ČR

Datová sada umožňuje hodnotit epidemiologické charakteristiky COVID-19 (incidence, prevalence, mortalita) v populaci České republiky od roku 2020. Uživateli je umožněno údaje stratifikovat dle demografických charakteristik pacientů (pohlaví, věk). Dále lze hodnotit počty pacientů s COVID- 19 léčených v akutní lůžkové péči, včetně detailů o průběhu infekce, symptomech, typech testů, očkování a hospitalizacích. Parametry datové sady zahrnují pořadí infekce, pohlaví, kategorii roku narození, týden detekce nákazy, týden výsledku testu, týden vyléčení, týden úmrtí, symptomy, typ testu, týdny očkování (první až sedmá dávka), druhy očkovacích látek, detaily hospitalizace (včetně JIP, kyslíku, HFNO, UPV/ECMO), a další relevantní údaje jako DCCI index a Long COVID. Datová sada obsahuje také informace o osobách, které do kontaktu s COVID-19 nepřišly. Tyto osoby pak nemají vyplněné datum pozitivity ani očkování proti COVID-19.

Téma	Časové pokrytí	Aktualizace	Související území
Zdraví	1. 9. 2020 - 31. 8. 2024	roční	Česká republika

Klíčová slova	Klasifikace dle EuroVoc	Kontakt	Organizace
ISIN, covid-19, pozitivita, očkování, hospitalizace, JIP, mutace, úmrtí, long COVID, DCCI	infekční nemoc, epidemie	ÚZIS ČR	Ústav zdravotnických informací a statistiky ČR

Distribuce CSV

Stáhnout výstup | Otevřít schéma

Formát: CSV
Velikost: 1351.68 MB
Velikost po rozbalení: 1351.68 MB
Počet záznamů: 12 580 942
Počet atributů: 53

Autoři: Šanca Ondřej, Jarkovský Jiří, Klimeš Daniel, Zelinková Hana, Klika Petr, Benešová Klára, Mužík Jan, Komenda Martin, Dušek Ladislav

Schvalování:

- porada vedení ÚZIS ČR dne 24. 9. 2024
- pověřenec pro ochranu osobních údajů dne 4. 9. 2024

Datum aktualizace dat: neuvedeno

Datum aktualizace dokumentace: 4. 9. 2024

Licence	Originální databáze	Zvláštní právo pořizovatele databáze	Osobní údaje
Autorské dílo licencované CC BY 4.0	Není autorskoprávně chráněnou databází	Není chráněna zvláštním právem	Neobsahuje osobní údaje

Citace

Šanca O., Jarkovský J., Klimeš D., Zelinková H., Klika P., Benešová K., Mužík J., Komenda M., Dušek L.. Očkování, pozitivity, hospitalizace pro COVID-19, úmrtí, long covid a komorbidity u osob v ČR. Národní zdravotnický informační portál [online]. Praha: Ministerstvo zdravotnictví ČR a Ústav zdravotnických informací a statistiky ČR, 2026 [cit. 2026-11-25]. Dostupné z: http://www.nzip.cz/data/2135-covid-19-prehled-populace. ISSN 2695-0340

Primární cíl

Datová sada umožňuje hodnotit epidemiologické charakteristiky COVID-19 (incidence, prevalence, mortalita) v populaci České republiky od roku 2020. Uživateli je umožněno údaje stratifikovat dle demografických charakteristik pacientů (pohlaví, věk). Dále lze hodnotit počty pacientů s COVID- 19 léčených v akutní lůžkové péči, včetně detailů o průběhu infekce, symptomech, typech testů, očkování a hospitalizacích. Parametry datové sady zahrnují pořadí infekce, pohlaví, kategorii roku narození, týden detekce nákazy, týden výsledku testu, týden vyléčení, týden úmrtí, symptomy, typ testu, týdny očkování (první až sedmá dávka), druhy očkovacích látek, detaily hospitalizace (včetně JIP, kyslíku, HFNO, UPV/ECMO), a další relevantní údaje jako DCCI index a Long COVID. Datová sada obsahuje také informace o osobách, které do kontaktu s COVID-19 nepřišly. Tyto osoby pak nemají vyplněné datum pozitivity ani očkování proti COVID-19.

Základní informace

V rámci zveřejňování dat z Národního registru hrazených zdravotních služeb (NRHZS) a Informačního systému infekčních nemocí vznikla datová sada přinášející informace průběhu epidemie COVID-19 v České republice ve formě anonymizovaných dat. Datová sada vzniká v souladu s národní metodikou tvorby otevřených dat a je dostupná v národním katalogu.

Jeden záznam (řádek) datové sady obsahuje informaci o jedné osobě a infekci s informacemi o její pozitivitě, hospitalizaci a očkování proti COVID-19. Pokud osoba neprodělala COVID-19 a ani nebyla očkována její řádek obsahuje pouze základní informace jako je pohlaví, kategorie let narození, případné úmrtí a DCCI.

Interpretace a informační hodnota

Datová sada je odvozena z dat Národního registru hrazených zdravotních služeb (NRHZS) (parametr DCCI a Long COVID), databáze zemřelých (Datum úmrtí) a Informačního systému infekčních nemocí. Jeden záznam (řádek) hodnotí jednu infekci pacienta. Pro zasazení záznamu do časového kontextu jsou uváděny roky a týdny jednotlivých sledovaných událostí, pokud nastaly. Osoba má v datové sadě tolik záznamů, kolikrát prodělala infekci COVID-19. Pakliže infekci neprodělala ani jednou, je uveden, a právě na jednom řádku.

Z důvodu ochrany osobních údajů fyzických osob je limitována úroveň detailu popisných charakteristik (sada neobsahuje geografickou stratifikaci a rok narození je agregován do rozsahu let).

V NZIS jsou evidované údaje určené explicitně pouze pro účely uvedené v § 70 odst. 1 zákona o zdravotních službách, zejména statistické účely. Nejedná se tedy o referenční údaje ve smyslu § 2 písm. b) zákona č. 111/2009 Sb., o základních registrech, ve znění pozdějších předpisů, které jsou dle § 4 odst. 4 téhož zákona považovány za správné, pokud není prokázán opak nebo pokud nevznikne oprávněná pochybnost o jejich správnosti. ÚZIS ČR neodpovídá za věrohodnost údajů evidovaných v NZIS, ani nemá k dispozici podklady na základě, kterých jsou tyto údaje do NZIS zapisovány a není možné z těchto dostupných údajů dovodit jejich správnost pro účely stanovené jinými právními předpisy.

Ústav zdravotnických informací a statistiky České republiky (ÚZIS) neposkytuje žádné záruky ohledně přesnosti, úplnosti nebo spolehlivosti dat. Jakékoli závěry, analýzy nebo interpretace vycházející z těchto dat jsou výhradní odpovědností uživatele. ÚZIS nenese odpovědnost za žádné škody nebo následky způsobené použitím těchto dat. V případě produkce chybných nebo zavádějících výsledků nemůže být ÚZIS použit jako odkaz či záruka správnosti těchto výsledků.

Příklady použití

Na základě této datové sady lze hodnotit epidemiologii COVID-19 v populaci ČR včetně časového vývoje. Kromě absolutních hodnot lze v kombinaci s dalšími datovými zdroji o populaci ČR počítat také standardizované hodnoty – např. přepočet na 100 000 obyvatel, a to souhrnně i s ohledem na pohlaví a věk pacienta.

Typické příklady využití dat:

- Incidenci lze odvodit z počtu záznamů (řádků), kde se hodnocený týden shoduje s týdnem detekce nákazy.
- Prevalenci lze odvodit na základě data pozitivity a vyléčení.
- Lze hodnotit počet osob hospitalizovaných pro COVID-19 v jednotlivých letech dle stavu jejich vakcinace a DCCI
- Příklad konkrétního výstupu:
- Uživatel hodnotí časový vývoj incidence COVID-19 v ČR (počet případů na 100 tisíc obyvatel v roce 2020 vs. v roce 2023), srovnává incidenci v různých věkových skupinách vzhledem k velikosti populace daného věku (tzv. věkově specifická prevalence).

Ochrana osobních údajů (GDPR)

Datová sada byla posouzena z hlediska souladu s následujícími právními předpisy v platném znění:

- Zákon č. 110/2019 Sb., o zpracování osobních údajů,
- Zákon č. 372/2011 Sb., o zdravotních službách a podmínkách jejich poskytování,
- Zákon č. 258/2000 Sb., o ochraně veřejného zdraví,
- 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,
- Zákon č. 106/1999 Sb., o svobodném přístupu k informacím,
- Směrnice (EU) 2019/1024 o otevřených datech a opakovaném použití informací veřejného sektoru (Směrnice o otevřených datech).

Pro odstranění rizika identifikace fyzických osob jsou data kompletně anonymizována. Časové údaje jsou prezentovány pouze po týdnech (rok), geografické údaje nejsou uváděny vůbec, rok narození pacienta je agregován na pětileté kategorie.

Zdroj dat

Zdrojem dat pro datovou sadu jsou následující součásti NZIS:

- databáze Národního registru hrazených zdravotních službách (LPZ), dle § 72 odst. 1 písm. g) zákona o zdravotních službách),
- databáze Informačního systému infekční nemoci (ISIN), dle § 72 odst. 1 písm. f) zákona o zdravotních službách,
- databáze Informačního systému List o prohlídce zemřelého (LPZ), dle § 72 odst. 1 písm. h) zákona o zdravotních službách.

Z technického hlediska vzniká tato datová sada kombinací z jednotlivých modulů ISIN (Hospitalizace, Vakcinace, Pozitivity), Informací o úmrtí z databáze LPZ a vybraných parametů (DCCI, Long COVID) z NRHZS.

Agregace a transformace dat

Detailní záznamy obsažené v datech ISIN byly agregovány a transformovány do podoby, která umožňuje hodnotit epidemii COVID-19 v České republice. Byly k nim následně připojeny data z NRHZS a databáze zemřelých.

Atributy datové sady

Název	Datový typ	Název sloupce	Popis	Možné hodnoty
ID	int	id	náhodný unikátní identifikátor řádku	přirozené číslo – pořadí infekce osoby
Infekce	int	Infekce	pořadí infekce pacienta	NULL – osoba neprodělala COVID, přirozené číslo – pořadí infekce osoby

Název	Datový typ	Název sloupce	Popis	Možné hodnoty
Pohlaví	text	pohlavi	pohlaví pacienta	1 – muž, 2 – žena, NULL – neznámé pohlaví
Rok narození	text	RokNarozeni	rok narození pacienta v 5letých kategoriích	Kombinace dvou let oddělené pomlčkou o Pomlčka – neznámý rok narození
Datum pozitivity	text	DatumPozitivity	rok a ISO týden pozitivity	od 2020-09, NULL – osoba nebyla pozitivní
Datum výsledku	text	DatumVysledku	rok a ISO týden výsledku testu	od 2020-09, NULL – osoba nebyla pozitivní
Vyléčen	text	Vylecen	rok a ISO týden vyléčení	od 2020-09, NULL – osoba nebyla pozitivní
Úmrtí	text	umrti	rok a ISO týden úmrtí	od 2020-09, NULL – osoba nezemřela
Symptom	bool	Symptom	informace o tom, zda byl pacient v době testu symptomatický	0 – ne, 1 – ano, NULL – Netestován/neurčeno
Typ testu	text	TypTestu	typ testu, který byl použit k identifikaci pozitivity	AG – Antigenní test, PCR – PCR test, NULL – Netestován/neurčeno
Datum první dávka	text	Datum_Prvni_davka	rok a ISO týden první dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum druhá dávka	text	Datum_Druha_davka	rok a ISO týden druhé dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum třetí dávka	text	Datum_Treti_davka	rok a ISO týden třetí dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum čtvrtá dávka	text	Datum_Ctvrta_davka	rok a ISO týden čtvrté dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum pátá dávka	text	Datum_Pata_davka	rok a ISO týden páté dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum šestá dávka	text	Datum_Sesta_davka	rok a ISO týden šesté dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Datum sedmá dávka	text	Datum_Sedma_davka	rok a ISO týden sedmé dávky očkování	od 2020-52, NULL – osoba nebyla očkována
Očkovací látka kód první dávka	text	OckovaciLatkaKod_Prvni_davka	kód druhu první dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Očkovací látka kód druhá dávka	text	OckovaciLatkaKod_Druha_davka	kód druhu druhé dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována

Název	Datový typ	Název sloupce	Popis	Možné hodnoty
Očkovací látka kód třetí dávka	text	OckovaciLatkaKod_Treti_davka	kód druhu třetí dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Očkovací látka kód čtvrtá dávka	text	OckovaciLatkaKod_Ctvrta_davka	kód druhu čtvrté dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Očkovací látka kód pátá dávka	text	OckovaciLatkaKod_Pata_davka	kód druhu páté dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Očkovací látka kód šestá dávka	text	OckovaciLatkaKod_Sesta_davka	kód druhu šesté dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Očkovací látka kód sedmá dávka	text	OckovaciLatkaKod_Sedma_davka	kód druhu sedmé dávky očkování	CO01–CO24 – viz číselník v kapitole 6 – Použité číselníky, NULL – osoba nebyla očkována
Primární příčina hospitalizace COVID	bool	PrimPricinaHospCOVID	pacient byl primárně hospitalizován z důvodu COVID	0 – ne, 1 – ano, NULL – nezadáno/neurčeno
Bin hospitalizace	bool	bin_Hospitalizace	Pacient byl hospitalizován s COVID	1 – ano, NULL – ne
Min Hospitalizace	text	min_Hospitalizace	rok a ISO týden začátku první hospitalizace při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní hospitalizace	int	dni_Hospitalizace	Počet dní hospitalizace. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	Od 2020-11, NULL – osoba nebyla hospitalizována
Max hospitalizace	text	max_Hospitalizace	rok a ISO týden konce poslední hospitalizace při dané pozitivitě	Od 2020-11 o NULL – osoba nebyla hospitalizována
Bin JIP	int	bin_JIP	Pacient byl hospitalizován na JIP s COVID	0 – pacient byl hospitalizován, nebyl hospitalizován na JIP lůžku, 1 – pacient byl hospitalizován, byl hospitalizován na JIP lůžku, NULL – pacient nebyl hospitalizován
Min JIP	text	min_JIP	rok a ISO týden začátku první hospitalizace na JIP při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní JIP	int	dni_JIP	Počet dní hospitalizace na JIP. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	libovolné přirozené číslo

Název	Datový typ	Název sloupce	Popis	Možné hodnoty
Max JIP	text	max_JIP	rok a ISO týden konce poslední hospitalizace na JIP při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Bin STAN	int	bin_STAN	pacient byl hospitalizován na standartním lůžku s COVID	0 – pacient byl hospitalizován, nebyl hospitalizován na standartním lůžku, 1 – pacient byl hospitalizován, byl hospitalizován na standartním lůžku, NULL – pacient nebyl hospitalizován
Min STAN	text	min_STAN	rok a ISO týden začátku první hospitalizace na standartním lůžku při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní STAN	int	dni_STAN	Počet dní hospitalizace na standartním lůžku. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	libovolné přirozené číslo
Max STAN	text	max_STAN	rok a ISO týden konce poslední hospitalizace na standartním lůžku při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Bin kyslík	int	bin_Kyslik	pacient byl hospitalizován s kyslíkem s COVID	0 – pacient byl hospitalizován, událost nenastala, 1 – pacient byl hospitalizován, událost nastala, NULL – pacient nebyl hospitalizován
Min Kyslík	text	min_Kyslik	rok a ISO týden začátku první hospitalizace s kyslíkem při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní kyslík	int	dni_Kyslik	Počet dní hospitalizace na JIP. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	libovolné přirozené číslo
Max kyslík	text	max_Kyslik	rok a ISO týden konce poslední hospitalizace s kyslíkem při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Bin HFNO	int	bin_HFNO	pacient byl hospitalizován s HFNO s COVID	0 – pacient byl hospitalizován, událost nenastala, 1 – pacient byl hospitalizován, událost nastala, NULL – pacient nebyl hospitalizován
Min HFNO	text	min_HFNO	rok a ISO týden začátku první hospitalizace s HFNO při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní HFNO	int	dni_HFNO	Počet dní hospitalizace na JIP. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	libovolné přirozené číslo
Max HFNO	text	max_HFNO	rok a ISO týden konce poslední hospitalizace s HFNO při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována

Název	Datový typ	Název sloupce	Popis	Možné hodnoty
Bin UPV ECMO	int	bin_UPV_ECMO	pacient byl hospitalizován s UPV/ECMO s COVID	0 – pacient byl hospitalizován, událost nenastala, 1 – pacient byl hospitalizován, událost nastala, NULL – pacient nebyl hospitalizován
Min UPV ECMO	text	min_UPV_ECMO	rok a ISO týden začátku první hospitalizace s UPV/ECMO při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Dní UPV ECMO	int	dni_UPV_ECMO	Počet dní hospitalizace na JIP. Hospitalizace mohla být přerušena a po tuto dobu nebyl počet dní hospitalizace počítán	libovolné přirozené číslo
Max UPV ECMO	text	max_UPV_ECMO	rok a ISO týden konce poslední hospitalizace s UPV/ECMO při dané pozitivitě	Od 2020-11, NULL – osoba nebyla hospitalizována
Mutace	text	Mutace	mutace byla určena na základě vyšetření diskriminačního PCR a zachycených lokusů	Viz číselník v kapitole 6 – Použité číselníky
Datum úmrtí LPZ	text	DatumUmrtiLPZ	rok a ISO týden úmrtí	Od 2020-10, NULL – osoba nebyla hospitalizována
Long COVID	text	Long_COVID	rok a ISO týden vykázání prvního Long COVID	Od 2020-48, NULL – osoba nebyla hospitalizována
DCCI	int	DCCI	DCCI = Deyova modifikace indexu komorbidit dle Charlsonové	0–4 – Počet bodů dle DCCI o 5 – 5 a více bodů dle DCCI 6 Použité číselníky