

# COVID 19: Datové sady pro prediktivní modelování

---

Metodický popis

Autoři:

Hlavní autor:

RNDr. Jiří Jarkovský, Ph.D., RNDr. Ondřej Májek, Ph.D.

Spoluautoři:

RNDr. Martin Komenda, Ph.D.,

Mgr. Matěj Karolyi, Bc. Petr. Panoška,

RNDr. Daniel Klimeš, Ph.D., RNDr. Jan Mužík, Ph.D.,

prof. RNDr. Ladislav Dušek, Ph.D.

Verze:

3.0

Datum:

20. 10. 2020

## Obsah

1	Základní popis.....	3
2	Technická charakteristika a dostupnost.....	3
3	Popis datových sad.....	3
3.1	Pozitivita, vyléčení a zemřelí na úrovni okresů .....	3
3.2	Časová efektivita testování .....	4
3.3	Opakované testování pozitivních pacientů .....	4
3.4	Hospitalizovaní pacienti .....	4
3.5	Hospitalizovaní pacienti (analýza) .....	5

## 1 Základní popis

---

Pro účely vývoje prediktivních modelů jsou publikovány sady ve formátu sdílených dat.

## 2 Technická charakteristika a dostupnost

---

- Aktualizace
  - 1 x týdně
- Časové období
  - Od začátku epidemie COVID-19 do dne o 1 týden předcházející datu exportu (účelem je získání kompletního a validovaného souboru; v datech po prvotním nahlášení positivity dochází v následujících dnech k dalšímu doplnění a zpřesnění některých údajů, např. o bydlišti pacienta)
- Zdroj informací
  - Zabezpečený server UZIS pro předávání dat [share.uzis.cz](https://share.uzis.cz)
- Formát datových sad
  - CSV, CSV schéma

## 3 Popis datových sad

---

### 3.1 Pozitivita, vyléčení a zemřelí na úrovni okresů

- Zdroj dat
  - ISIN, ztotožnění s ROB
- Popis
  - Datová sada poskytuje agregovaná data o počtech pozitivních, vyléčených a zemřelých pacientů v jednotlivých dnech stratifikovaná dle okresu pobytu, pohlaví a kategorií věku na základě hlášení krajských hygienických stanic, laboratoří a nemocnic.
- Obsah datové sady
  - Datum (positivity, vyléčení nebo úmrtí)
  - Okres ve formátu CZ-NUTS
  - Pohlaví
  - Věková kategorie 0-19, 20-64 a 65+ let
  - Počet pozitivních
  - Počet vyléčených
  - Počet zemřelých
  - Počet zemřelých za hospitalizace
  - Počet zemřelých se zaznamenaným komplikovaným stavem

### 3.2 Časová efektivita testování

- Zdroj dat
  - ISIN
- Popis
  - Datová sada na úrovni anonymizovaných záznamů pozitivních pacientů již zpracovaných KHS a obsahující časové údaje o odstupu mezi příznaky onemocnění, odběrem, výsledkem testů, hlášením KHS a izolací pozitivního pacienta na základě hlášení krajských hygienických stanic a laboratoří.
- Obsah datové sady
  - Datum prvního příznaku
  - Datum odběru
  - Datum výsledku prvního pozitivního PCR testu
  - Datum hlášení KHS
  - Datum izolace

### 3.3 Opakované testování pozitivních pacientů

- Zdroj dat
  - ISIN
- Popis
  - Datová sada poskytuje agregovaná data po jednotlivých dnech o počtech primárních pozitivních testů, všech testů u pozitivních pacientů a opakovaných testů u pozitivních pacientů na základě dat laboratoří.
- Obsah datové sady
  - Datum
  - Počet primárních pozitivních testů
  - Počet všech testů u pozitivních pacientů
  - Počet opakovaných testů u pozitivních pacientů

### 3.4 Hospitalizovaní pacienti

- Zdroj dat
  - ISIN, ztotožnění s ROB
- Popis
  - Datová sada poskytuje agregovaná data o počtech hospitalizovaných pacientů s agregací na okres pobytu, pohlaví, věkové kategorie, přítomnost komplikovaného stavu a výsledný stav pacienta dle dat nemocnic doplněných o údaje z KHS.
- Obsah datové sady
  - Okres ve formátu CZ-NUTS
  - Pohlaví
  - Věková kategorie 0-19, 20-64 a 65+ let
  - Komplikovaný stav za hospitalizace

- Výsledný stav k datu exportu propuštěn, dosud hospitalizován, úmrtí za hospitalizace
- Počet pacientů

### 3.5 Hospitalizovaní pacienti (analýza)

- Zdroj dat
  - ISIN
- Popis
  - Datová sada poskytuje řádková data hospitalizovaných pacientů popisující jejich onemocnění a průběh hospitalizace.
- Obsah datové sady
  - Věk (5 leté věkové kategorie)
  - Pohlaví
  - Kraj bydliště – kód
  - Kraj nemocnice při zahájení hospitalizace (bez překladů)
  - Datum prvního příznaku
  - Datum odběru
  - Datum positivity
  - Počáteční tíže stavu dle záznamu KHS
  - Datum zahájení hospitalizace
  - Datum posledního záznamu hospitalizace
  - Poslední tíže stavu dle hospitalizačních záznamů
  - Poslední záznam hospitalizace
    - Hospitalizace pokračuje (neuzavřený záznam)
    - COVID19negativní – sledování ukončeno (uzavřený záznam)
    - Propuštěn z hospitalizace (uzavřený záznam)
    - Úmrtí za hospitalizace (uzavřený záznam)
  - Nejtěžší stav v průběhu hospitalizace
  - Těžký stav ano/ne
  - Datum počátku těžkého stavu
  - Počet dní v těžkém stavu
  - Datum posledního záznamu těžkého stavu
  - JIP ano/ne
  - Datum zahájení pobytu na JIP
  - Počet dní na JIP
  - Datum posledního záznamu pobytu na JIP
  - Kyslík ano/ne
  - Datum zahájení podání kyslíku
  - Počet dní na kyslíku
  - Datum posledního záznamu podání kyslíku
  - UPV ano/ne
  - Datum zahájení UPV
  - Počet dní na UPV
  - Datum posledního záznamu UPV
  - ECMO ano/ne
  - Datum zahájení ECMO

- Počet dní na ECMO
- Datum posledního záznamu ECMO
- Úmrtí
- Datum úmrtí