

# Porovnání algoritmu pro definici významného vzorku s reálnou situací

Pilotní projekt analyzující 3 vybrané měsíce s daty  
získanými z databáze GISAID

David Vostřák, Alena Fialová, Martina Leppová, Monika Liptáková,  
Jan Moskalyk, Jan Kynčl, Helena Jiřincová

**Ref. Grant/2021/PHF/23776**



V rámci grantu HERA, v souladu s doporučeními WHO, EU a dalších institucí se Česká republika i Státní zdravotní ústav zavázali posílit sekvenační kapacitu ČR, která vychází z principů surveillance.

Cílem algoritmu je nastavení surveillance. To znamená nastavit výběr vzorků pozitivních na SARS-CoV-2 k celogenomové sekvenaci tak, aby byl výběr reprezentativní a zahrnoval všechny epidemiologicky významné kategorie. Na základě celorepublikových dat poskytnutých z ÚZIS budou vybírány tzv. významné vzorky které splňují požadavky na:

## 1. Cílenou surveillanci:

- selhání vakcinace
- reinfekce
- hospitalizace
- importy
- ohniska
- neobvyklé klinické příznaky

## 2. Plošnou surveillanci, v jejímž rámci vybíráme reprezentativní soubor napříč populací s rovnoměrnou distribucí pohlaví, věku a geografické lokality.

Připravovaný algoritmus lze v principu použít pro jakoukoliv surveillance infekčního onemocnění nejen pro onemocnění covid-19.

Cílem pilotní analýzy dat bylo zjistit zda:

- Pro reprezentativní výběr sekvenovaných vzorků je vhodné implementovat algoritmus zajišťující tento výběr.
- Míru s jakou sekvenovaná data rovnoměrně pokrývají požadavky na věkovou distribuci, geografickou distribuci a pohlaví.
- Datová sada z GISAID byla použita pouze pro ověření reprezentativnosti výběru, bez ohledu na další parametry. Pro analýzu cílené surveillance nejsou v GISAID databázi potřebné údaje. Tato analýza bude provedena s datovou sadou získanou z ÚZIS v dubnu 2022.
- Pro pilotní ověření byly vybrány měsíce:
  - Září 2021 – dominance delta varianty, testování škol, nárůst incidence
  - Listopad 2021 – introdukce varianty omikron, povinné testování zaměstnanců, vysoká incidence
  - Únor 2022 – dominance varianty omikron, sestupný trend incidence.

**Ve sledovaných měsících se na sekvenaci SARS-CoV-2 v ČR podílely laboratoře těchto institucí:**

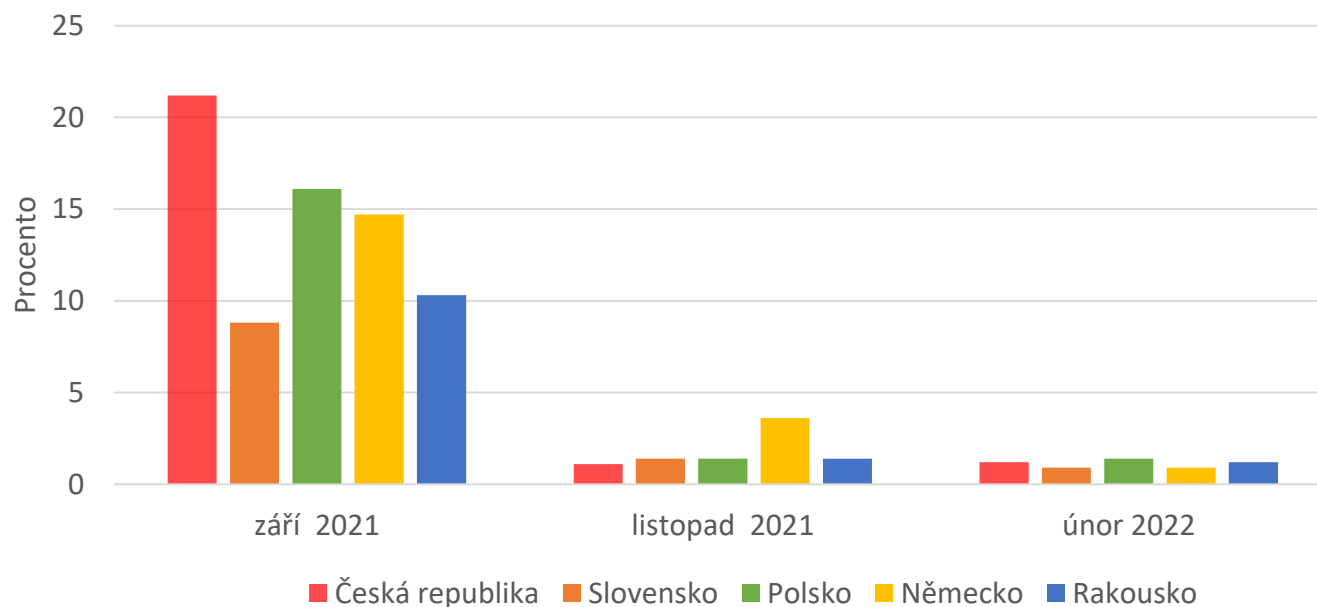
- CMBG FN Brno
- FN Ostrava
- FN Plzeň
- FN Sv. Anny Brno
- IMTM Olomouc
- KN Hradec Králové
- KN Liberec
- NRL SZÚ
- PŘF JČU
- PŘF UK / Biocev
- TU Dresden Dresden (Cross border Saxony cooperation, ukončena konec března 2022).

Zdroj dat: GISAID (<https://www.gisaid.org/>)

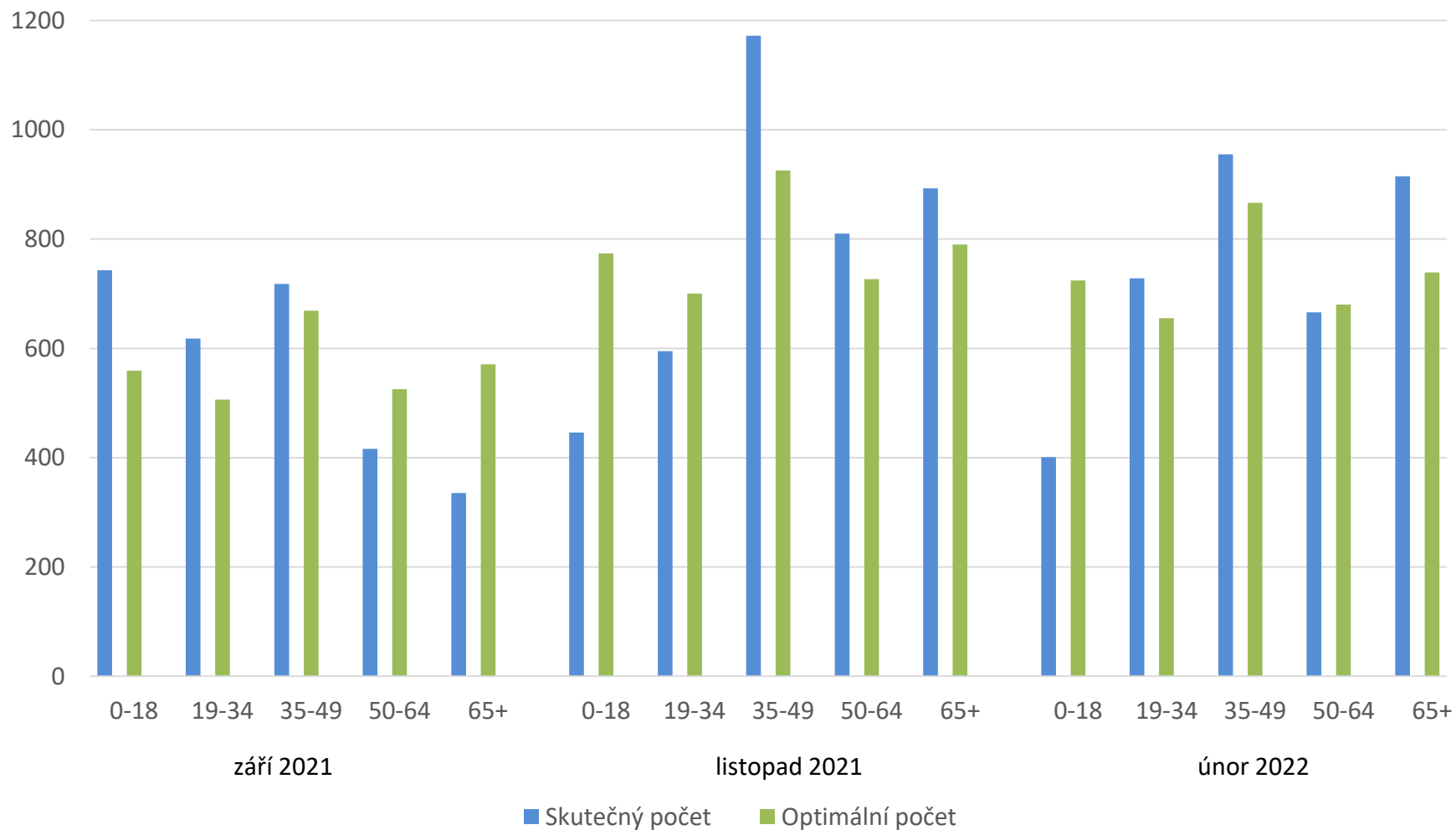
**Počet sekvenací během vybraných měsíců v ČR:**

měsíc/rok	počet sekvenací
září 2021	2845
listopad 2021	4025
únor 2022	3672

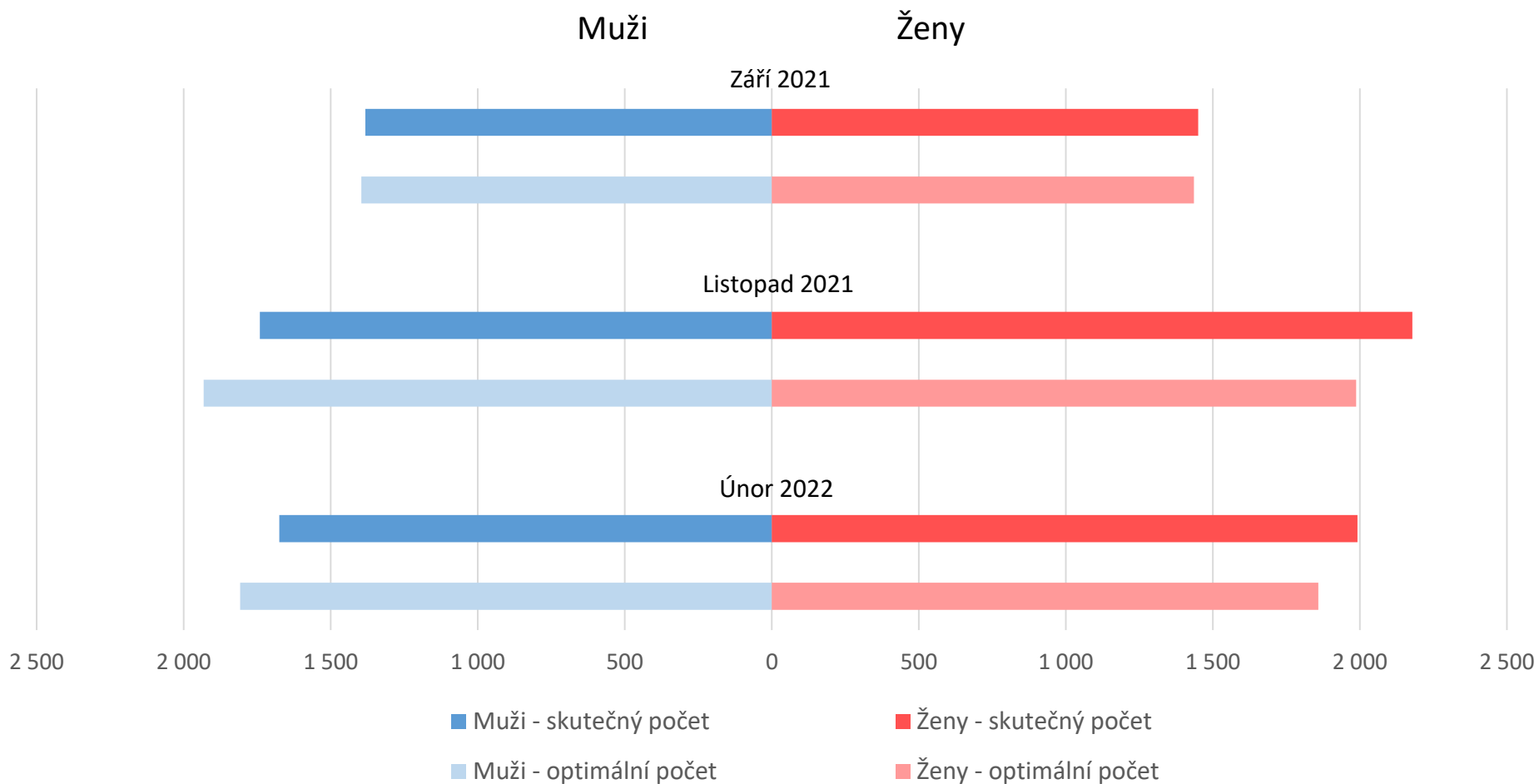
**Podíl (v %) sekvenovaných vzorků ze všech SARS-CoV-2 pozitivních vzorků během vybraných měsíců v ČR a sousedních zemích:**



**Graf 1: Porovnání skutečného a optimálního počtu sekvencí dle věkových skupin za 3 vybrané měsíce. Optimální počet je vztažen ke skutečnému podílu věkové skupiny v populaci.**

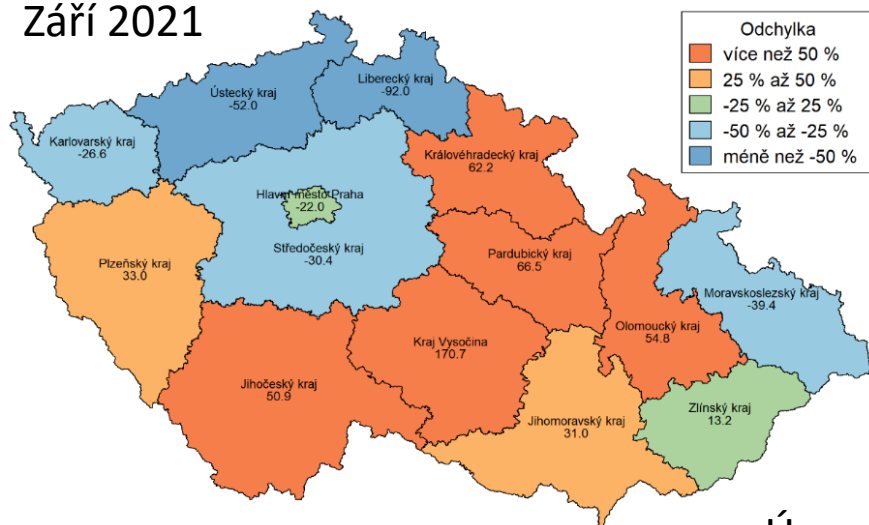


**Graf 2: Porovnání skutečného a optimálního počtu sekvenací dle pohlaví za 3 vybrané měsíce**  
**Optimální počet je vztažen ke skutečnému podílu mužů a žen v populaci.**

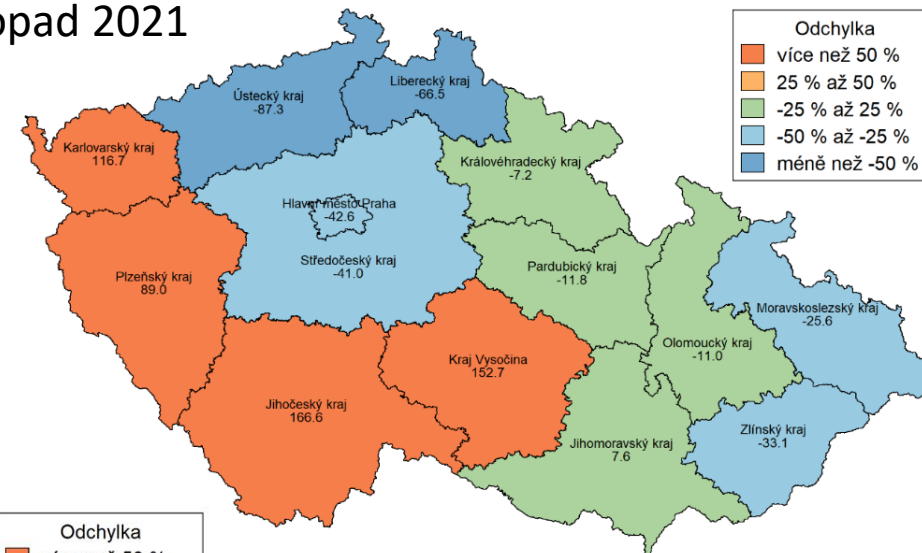


**Obrázek 3: Porovnání skutečného a optimálního počtu sekvencí dle kraje za vybrané měsíce.  
Optimální počet je vztážen ke skutečnému podílu všech pozitivních vzorků v kraji za vybraný měsíc.**

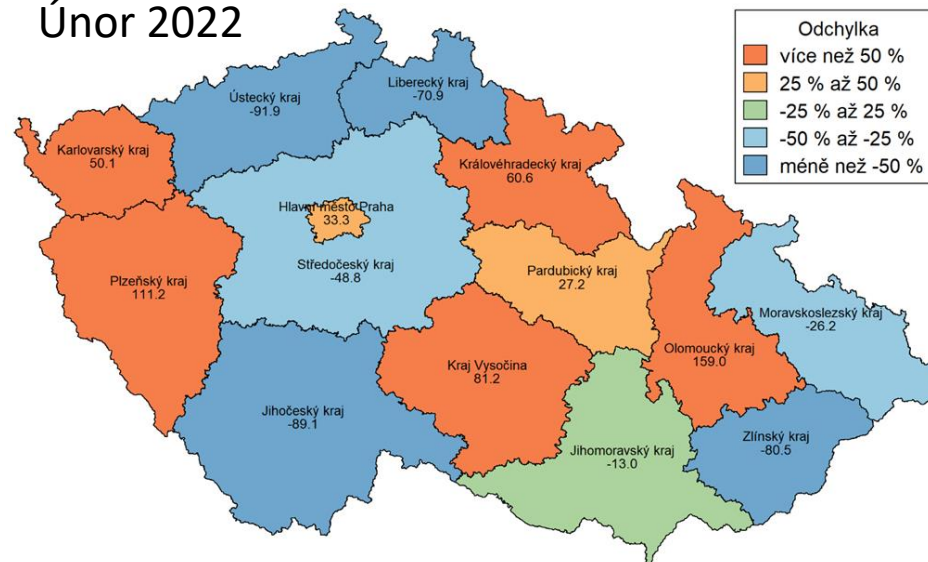
Září 2021



Listopad 2021



Únor 2022



Komentář ke grafům 1, 2 a obrázku 3:

- Odchytky od optimálního stavu z hlediska věkových skupin a pohlaví nejsou příliš velké s výjimkou věkové skupiny 0 – 18 let.
- V případě věkových skupin se do sekvenovaných vzorků přelévá počet vyšetřených vzorků v dané věkové skupině, který je spíše závislý na testovací strategii než na incidenci (testování ve školách – září 2021, testování v zaměstnání – listopad 2021).
- Poměrně významná neprosekvenovatelnost v kategorii 0 – 18 let byla v listopadu 2021 a v únoru 2022.
- Velká disproporce je v distribuci mezi kraji.
- Zavedení národního surveillančního algoritmu by mělo vyřešit vyvážení dle **všech** parametrů.



Při analýze dat z GISAID, kde nelze oddělit typ výběru vzorků (reprezentativní/cílený výběr), bylo zjištěno:

- V rámci věkové distribuce je trvale podhodnocena kategorie 0 – 18 let, vyjma období cíleného na testování ve školách.
- Rozdělení vzorků dle pohlaví osob je zastoupeno rovnoměrně.
- Regionální distribuce neodpovídá požadavkům na rovnoměrné zastoupení a nezohledňuje relativní incidenci.
- Trvale málo oskenované kraje jsou Ústecký a Liberecký kraj s více než 50 % zápornou odchylkou od požadovaného počtu vyšetření, mezi tyto kraje může spadat i Zlínský kraj.
- Trvalé nedostatečně oskenované kraje s 25 – 50 % zápornou odchylkou jsou kraje Středočeský a Moravskoslezský. Zařazení PŘF UK v rámci grantu HERA mezi sekvenující laboratoře zlepšilo prosekvenovanost v Praze v únoru 2022.
- Kraj Vysočina je naopak trvale sekvenován s více než 50 % pozitivní odchylkou. Plzeňský kraj vykazuje až + 80 % odchylku.
- Pro zajištění rovnoměrné prosekvenovanosti je zařazení algoritmu pro definici významného vzorku užitečné.
- Porovnání s požadavky na cílenou surveillance bude součástí podrobné zprávy dodané do konce května 2022 po zpracování dat z ÚZIS.

- Algoritmus je model, realizovaný v programovacím jazyce Python, který umožní zefektivnit výběr vzorků, usnadnit logistiku a realizovat produkci podrobných sekvenačních dat.
- Implementace algoritmu slibuje velký posun na poli sekvenace, kdy jsou vybírány vhodné vzorky a pouze v počtu, která sekvenační centra zvládnou zpracovat ve standardním čase.
- V souhrnu lze říci, že implementace algoritmu zajistí lepší dohled nad epidemiologickou situací v České republice a poskytne podklady pro efektivnější rozhodování vládě na základě dat a evidence/důkazů.
- Zároveň poskytne podklady pro další vědeckou práci a umožní tvorbu vyvážených reportů pro veřejnost.
- Cílem je také produkovat otevřenou datovou sadu na základě výstupů z algoritmu, a to tak, aby byl zajištěn rovný přístup k anonymizovaným informacím.

Project „Enhancing Whole Genome Sequencing (WGS) and/or Reverse Transcription Polymerase Chain Reaction (RT-PCR) national infrastructures and capacities to respond to the Covid-19 pandemic in the European Union and European Economic Area“ had received funding from the European Centre for Disease Prevention and Control under the Grant Agreement number ECDC/HERA/2021/004 ECD.12218.



More information about the project: <http://www.szu.cz/ecdc-1>

NOTE:

Information included reflects only the authors' view and the Centre (ECDC) is not responsible for any use that may be made of the information it contains.