

# Metodický popis užívaných stochastických modelů pro predikce pravděpodobného vývoje epidemie COVID-19 v ČR

---

Autoři: Ondřej Májek, Ondřej Ngo, Jiří Jarkovský, Monika Ambrožová,  
Barbora Budíková, Ladislav Dušek, Tomáš Pavlík

Verze: 1.0

Datum: 29. 4. 2020

## 1 Obsah

---

1	Obsah.....	2
2	Modely S(E)IR: využití pro krátkodobé predikce .....	3
2.1	Východiska – k čemu jsou modely využívány? .....	3
2.2	Podstata a vědecká relevance tohoto typu modelů .....	3
2.3	Metodické citace .....	3
2.4	Vstupní data .....	4
2.5	Výstupy modelu.....	6
2.6	Postup výpočtu.....	7
2.7	Proč jsou modely rekalibrovány cca 1x týdně .....	8
2.8	Jak jsou výstupy modelu verifikovány .....	9
3	Dlouhodobý S(E)IR model využívaný ÚZIS ČR.....	10
3.1	Východiska – k čemu je model využíván .....	10
3.2	Příklady metodických možností komplexních modelů, aplikace v jiných státech.....	10
3.3	Podstata a vědecká relevance zvoleného typu modelů.....	10
3.4	Metodické citace .....	10
3.5	Vstupní data .....	11
3.6	Výstupy modelu.....	11
3.7	Postup výpočtu.....	12
3.8	Jak jsou výstupy kalibrovány a verifikovány.....	13
4	Publikace datových sad umožňujících práci s modely.....	14
4.1	Jaká data se do modelů používají.....	14
4.2	Jaké datové sady byly navrženy pro daný typ modelování .....	14
4.3	Kde jsou tyto datové sady zveřejněny.....	14

## 2 Modely S(E)IR: využití pro krátkodobé predikce

---

### 2.1 Východiska – k čemu jsou modely využívány?

- Krátkodobá předpověď počtu nově potvrzených případů onemocnění
- Podklad pro předpovědi dalších klíčových charakteristik: počet hospitalizovaných, počet osob v těžkém stavu nebo s vysoce intenzivní péčí, počet úmrtí
- Vyhodnocení dynamiky epidemie, stanovení reprodukčního čísla, hodnocení souvislostí se zavedenými opatřeními

### 2.2 Podstata a vědecká relevance tohoto typu modelů

Jedná se o rozšíření epidemiologických SIR modelů, které jsou standardem v modelování vývoje akutních infekčních onemocnění<sup>1</sup>.

Model reprezentuje dynamiku infekčního onemocnění v populaci. Jedinci v populaci jsou v každém okamžiku uvažováni v jednom ze stavů (kompartimentů), jedná se tedy o tzv. stavový (kompartimentový) model S(E)IR:

- S            náchylní jedinci
- I            infikovaní jedinci
  - Prakticky se tento stav dělí v modelech COVID na
    - E            latentní onemocnění, před infekčností
    - Vlastní I    infekční jedinci
- R            jedinci po konci infekčnosti

S modely je spjata parametrická a strukturní nejistota a nejistota ohledně budoucích opatření a jejich dopadu na dynamiku epidemie.

### 2.3 Metodické citace

Příklad recentních aplikací S(E)IR modelu v odborné literatuře

- WU, Joseph T.; LEUNG, Kathy; LEUNG, Gabriel M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *The Lancet*, 2020, 395.10225: 689-697.
- KUCHARSKI, Adam J., et al. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*, 2020.
- JIA, Wangping, et al. Extended SIR prediction of the epidemics trend of COVID-19 in Italy and compared with Hunan, China. *medRxiv*, 2020. (článek zatím neprošel recenzním řízením)

---

<sup>1</sup> LIU, Xinzhi; STECHLINSKI, Peter. *Infectious Disease Modeling*. Springer, 2017.

## 2.4 Vstupní data

### Inkubační doba

- představuje časový interval mezi infekcí a prvními příznaky onemocnění
- v modelovém hodnocení umožňuje modelovat dosud nepotvrzené případy nákazy a šíření nákazy v infekční fázi inkubační doby pacienta

### Infekční doba

- představuje časový interval nakažlivosti infikovaného pacienta
- v modelovém hodnocení představuje nepostradatelnou informaci o časovém období možného šíření nákazy

### Sériový interval

- představuje interval mezi nástupem příznaků nakažlivého jedince a prvním nástupem příznaků nakažené osoby
- v modelovém hodnocení představuje rychlost šíření nákazy mezi jedinci
- lze odvodit z inkubační a infekční doby

### Podíl subklinických případů

- představuje infekci u pacienta, který nemá žádné či minimální příznaky onemocnění, ale může nakazit další osoby

### Efektivita testování

- představuje dobu od nástupu symptomů po potvrzení výsledku a zaznamenání do databáze ISIN
- v modelovém hodnocení ovlivňuje souvislost mezi průběhem onemocnění a reportováním do databáze ISIN

### Importované případy

- představují potvrzené nákazy onemocnění u pacientů v ČR, kteří byli infikováni v zahraničí
- v modelovém hodnocení představují importované případy iniciální ohniska pro modelování šíření nákazy v ČR

### Vstupy do použitého modelu

Parametr	Hodnota	Zdroj	Užití
Inkubační doba	6 dní	Literární přehled	zohlednění ve struktuře modelu
Infekční doba	3 dny	Předpoklad v předchozích publikacích, např. Kucharski et al., 2020	zohlednění ve struktuře modelu
Sériový interval	5 dní	Literární přehled	soulad s údaji o inkubační a infekční době
Podíl subklinických případů	10 % (souběžně využita i alternativa 30 %)	Předpoklad v předchozích publikacích, např. Hellewell et al., 2020 <sup>2</sup>	Parametr modelu
Efektivita testování	Datová sada	Kalibrace modelu, validace dle ISIN	Parametr modelu
Počty importovaných případů	Datová sada	ISIN	Pomoc při kalibraci počátečních parametrů modelu

<sup>2</sup> HELLEWELL, Joel, et al. Feasibility of controlling COVID-19 outbreaks by isolation of cases and contacts. *The Lancet Global Health*, 2020.

## 2.5 Výstupy modelu

Z modelu lze získat markovskou stopu, časově podmíněný záznam všech uvažovaných charakteristik. Jako klíčové výstupy modelu jsou uvažovány:

- Kumulativní počet potvrzených případů onemocnění COVID-19
- Počet nově potvrzených případů
- Počet dosud nepotvrzených infikovaných jedinců

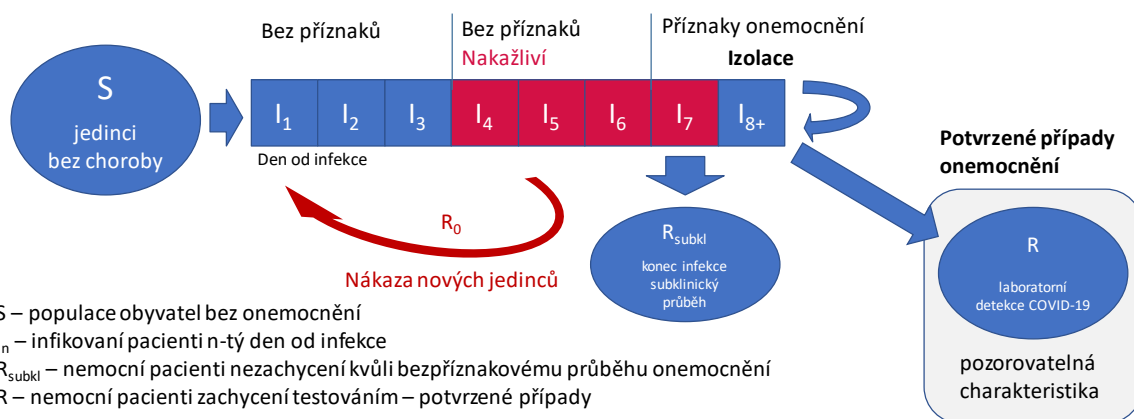
Jako výstup modelu zároveň slouží kalibrované hodnoty parametrů modelu, zejména základní/efektivní reprodukční číslo, standardně označované jako  $R$ :

- představuje průměrný počet dalších osob, které přímo nakazí jeden infikovaný jedinec
- základní reprodukční číslo udává počáteční hodnotu v dané populaci před přijetím ochranných opatření, postupně by se mělo snižovat na tzv. efektivní reprodukční číslo, které odpovídá pomalejšímu šíření epidemie díky přijatým opatřením
- v modelovém hodnocení představuje klíčový parametr infekčnosti onemocnění a zároveň zohledňuje přijatá opatření v čase

Intervalový odhad efektivního reprodukčního čísla lze získat jako součást kalibračního procesu. Aplikací tohoto intervalového odhadu v pravděpodobnostní analýze senzitivity lze získat „interval senzitivity“ pro kumulativní počet případů k danému referenčnímu datu.

## 2.6 Postup výpočtu

Struktura konkrétního modelu je vysvětlena na následujícím schématu



Model predikuje průchod pacientů průběhem onemocnění, s definovanou délkou inkubační doby. Noví pacienti přicházejí do modelu importem nebo nákazou, končí se subklinickým průběhem nebo jako potvrzený případ (**jediná přímo sledovaná charakteristika**). Klíčovým parametrem modelu je **tzv. reprodukční číslo**: průměrný počet osob, které nakazí 1 nakažená osoba.

- do modelu vstupují importované případy (nakažení je datováno 6 dní před datem počátku symptomů dle ISIN, kalibrace provedena na konci března)
- setrvání ve stavech  $I_1$ - $I_7$  trvá vždy právě 1 den, setrvání ve stavu  $I_{8+}$  je podmíněno efektivitou testování (pravděpodobnost přechodu, resp. průměrná doba setrvání, byla orientačně kalibrována na reálná data ISIN)
- počet nově nakažených (vstup do  $I_1$ ) odpovídá reprodukčnímu číslu
- šestidenní inkubační doba (Hellewell et al., 2020), druhá polovina intervalu představuje infekční období (alternativní předpoklad z Kucharski et al., 2020, odpovídá sériovému intervalu 5 dní)
- druhý den po projevu příznaků izolace, předpoklad konce nakažlivosti pro populaci
- efektivita testování kalibrována na pozorovaná data, předpoklad postupného zlepšování efektivit (zkracování doby do diagnózy)
- model tedy předpokládá testování pouze osob s příznaky
- subklinický průběh: uvažováno 10 % (Hellewell et al., 2020), variantně i 30 %

Výpočet je implementován jako markovský model v programu MS Excel. Modul pro kalibraci modelu je naprogramován v jazyce Microsoft Visual Basic for Applications.

Kalibrace dne 22. 4. 2020 zahrnovala následující parametry:

- reprodukční číslo identifikované v období 16. 3. – 31. 3.
- reprodukční číslo od 1. 4.
- efektivita testování od 1. 4.

(ostatní hodnoty byly ponechány dle předchozích edicí)

Kalibračním cílem je kumulativní počet potvrzených případů COVID-19. Metrikou shody modelu s pozorováním je metoda nejmenších čtverců. Parametr byl hledán metodou náhodného prohledávání prostoru daného přípustnými hodnotami kalibrovaných parametrů. Akceptováno bylo 5 % iterací s nejlepší shodou. Z tohoto souboru byly stanoveny bodové odhady parametrů, včetně intervalového odhadu reprodukčního čísla od 1. 4. k použití do analýzy senzitivity.

Reprodukční číslo je odhadováno samostatně v segmentech se zlomovými body danými zaváděním vládních opatření: 7. 3., 12. 3., 16. 3. a 1. 4. Efektivita testování byla kalibrována ve třech segmentech se zlomovými body 23. 3. a 1. 4. V prvním segmentu dosahuje pravděpodobnost přechodu  $I_{8+} \rightarrow R$  14 %, ve třetím 49 %, s lineárním růstem v prostředním segmentu (kalibrace ze dne 22. 4.).

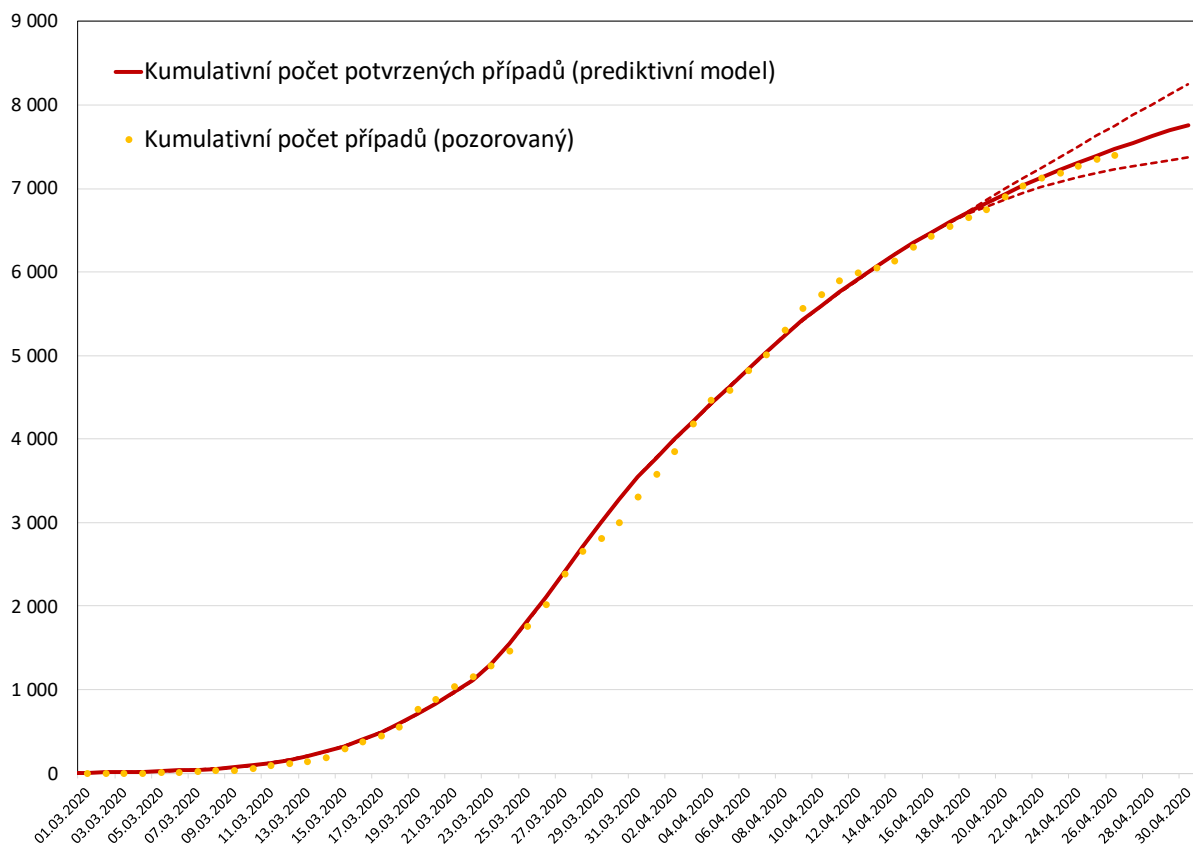
## 2.7 Proč jsou modely rekalibrovány cca 1x týdně

Modely jsou kalibrovány v cca týdenním intervalu, který lze považovat za minimální interval, v němž je možné věrohodně identifikovat změnu trendu.



## 2.8 Jak jsou výstupy modelu verifikovány

Modelové predikce jsou vždy zpětně srovnávány se skutečnými počty potvrzených případů dle referenčních dat ISIN.



Obr. 1: Kumulativní počet potvrzených případů (predikce na datech do 21. 4., pozorovaná data do 26. 4. 2020)

## 3 Dlouhodobý S(E)IR model využívaný ÚZIS ČR

---

### 3.1 Východiska – k čemu je model využíván

- Odhad dynamiky epidemie (čas a velikost vrcholu epidemie) v delším časovém horizontu
- Odhad dopadu intervencí pro snižování kontaktů populace, odhad dopadu rozvolnění těchto intervencí
- Dlouhodobé plánování kapacit zdravotnického systému

### 3.2 Příklady metodických možností komplexních modelů, aplikace v jiných státech

#### Německo

- Model ve spolupráci Institutu Roberta Kocha (vládní zdravotní ústav) a Humboldtovy univerzity
- Modifikovaný SIR model (SIR-X, kompartment pro izolované jedince)
- <http://rocs.hu-berlin.de/corona/docs/forecast/model/>

#### Slovensko

- Model zveřejněný Institutem zdravotní politiky (útvár MZ SR)
- Modifikovaný SEIR model (se zohledněním geografie a mobility)
- <https://izp.sk/covid-19/>

### 3.3 Podstata a vědecká relevance zvoleného typu modelů

Pro implementaci byl zvolen kompartmentový věkově strukturovaný SEIR model. Byl adaptován model London School of Hygiene & Tropical Medicine, publikován 25. 3. 2020 v Lancet Public Health. Model pracuje s místně-specifickými kontaktními vzorci (domácnost, zaměstnání, škola, jiné). **Model umožňuje pracovat s četností kontaktů (a jejich omezení) ve specifických prostředích, a tak umožňuje odhadovat dopad opatření k zamezení kontaktů v různých prostředích (škola, práce, veřejné prostory).**

### 3.4 Metodické citace

- PREM, Kiesha, et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *The Lancet Public Health*, 2020.
- Zdrojové kódy dostupné:  
<https://github.com/kieshaprem/covid19-agestructureSEIR-wuhan-social-distancing>

### 3.5 Vstupní data

- předpoklady převzaty z původní studie (zatím je totiž nelze validovat na datech na národní úrovni)
  - pravděpodobnost klinické manifestace pro mladší 20 let: 40 % (tj. 60 % subklinických)
  - pravděpodobnost klinické manifestace pro starší 20 let: 80 % (tj. 20 % subklinických)
  - snížená infekčnost subklinického případu: 25 %
- předpoklady adaptovaného modelu (vychází se z charakteristik modelu ÚZIS pro krátkodobé predikce, viz kapitola 2.5)
  - základní reprodukční číslo: 2,8 (předpoklad, kalibrace na nový model)
  - průměrná délka latentního období (první půlka inkubační doby): 3 dny
  - průměrná délka infekčního období: 4 dny
  - výchozí počet infekčních orientačně kalibrován na výchozí dynamiku onemocnění v ČR
- výchozí vzorce kontaktů (matice kontaktů) převzaty z mezinárodní studie, které provedla extrapolaci na ČR:
  - PREM, Kiesha; COOK, Alex R.; JIT, Mark. Projecting social contact matrices in 152 countries using contact surveys and demographic data. PLoS computational biology, 2017, 13.9: e1005697.

### 3.6 Výstupy modelu

Z modelu lze získat časově podmíněný záznam všech uvažovaných charakteristik. Jako klíčový výstup modelu jsou uvažovány:

- Kumulativní počet případů COVID-19
- Počet nově potvrzených případů

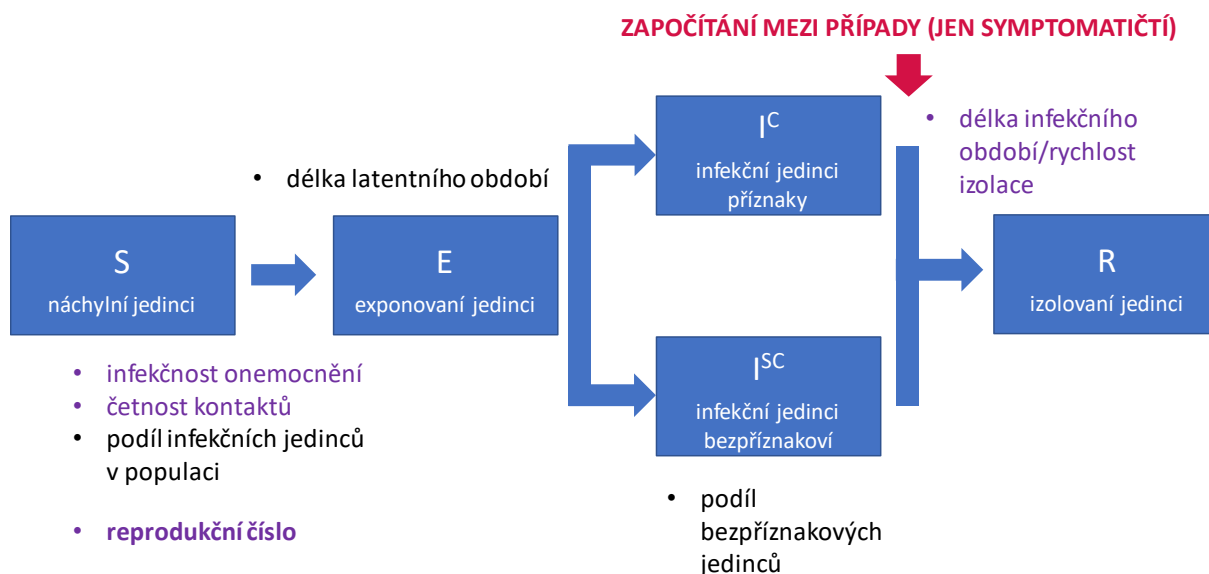
Model uvažuje započítání případu k datu skončení infekčního období (izolace). K započítání případu do pozorovaných statistik dochází s několikadenním odstupem s ohledem na efektivitu testování. V recentních statistikách je již patrný krátký (cca 1-2 denní) odstup mezi izolací a konfirmací.

Jako výstup modelu zároveň mohou sloužit kalibrované hodnoty parametrů modelu, zejména efektivní reprodukční číslo.

Model byl prozatím prezentován jen s uvažováním scénářové neurčitosti, tedy s variováním změn intenzity kontaktů v různých časových obdobích. V případě potřeby je možné provést pravděpodobnostní analýzu senzitivity se simulací výsledku dle uvažovaného rozdělení použitých parametrů.

### 3.7 Postup výpočtu

Struktura konkrétního modelu je vysvětlena na následujícím schématu:



Model je detailně popsán v citovaném odborném článku.

Nastavení scénářů je vždy součástí publikovaných výsledků modelu.

Oproti původní verzi modelu bylo upraveno započítávání jedinců do incidence onemocnění.

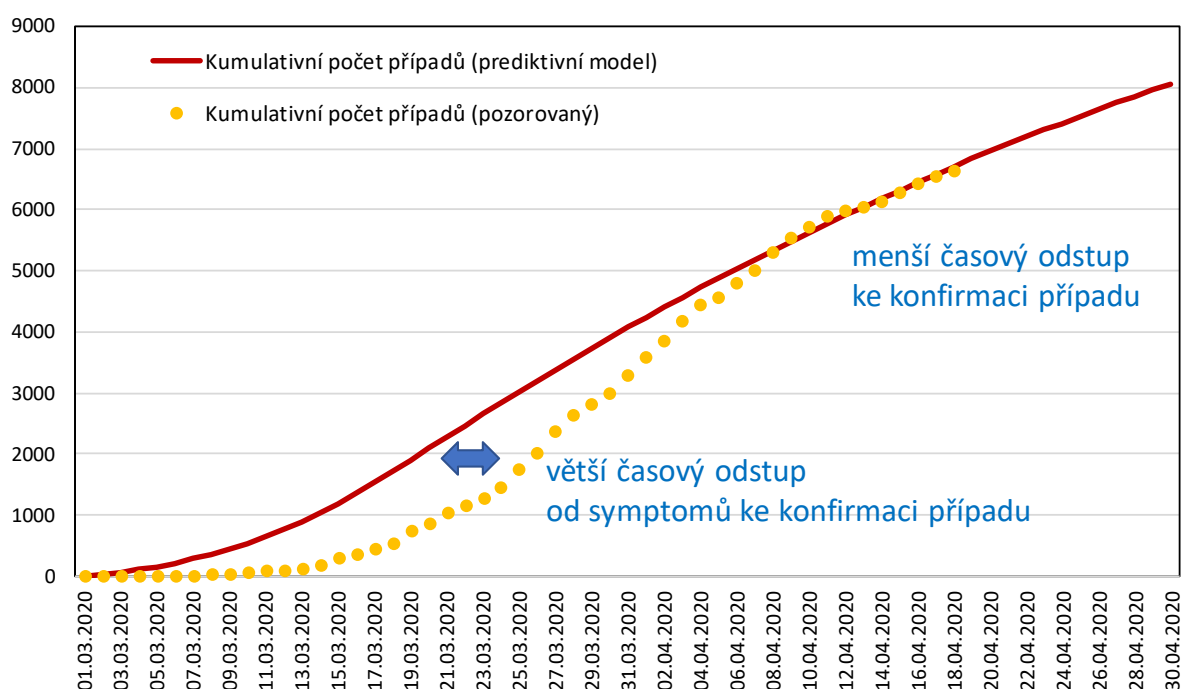
- Jedinci, u kterých onemocnění proběhne symptomaticky
  - pacienti jsou na konci infekčního období izolováni
  - přesun do kompartmentu R
  - **započítání mezi jedince s potvrzeným onemocněním**
- Jedinci, u kterých onemocnění proběhne s mírnými nebo žádnými symptomy
  - pacienti jsou na konci infekčního období přesunuti rovněž do kompartmentu R
  - takoví pacienti nejsou započítáni mezi jedince s potvrzeným onemocněním
  - incidence/prevalence těchto osob není přímo pozorovatelná

Výpočty jsou prováděny v programu R version 3.6.3 (2020-02-29).

### 3.8 Jak jsou výstupy kalibrovány a verifikovány

Model byl v iniciální fázi kalibrován na data ISIN. Pro tento účel bylo uvažováno:

- výchozí reprodukční číslo R ve výši 2,8
- uzavření škol od 11. 3. 2020 (nulové školní kontakty, omezení pracovních kontaktů na 60 %, jiných kontaktů také na 60 %)
- intenzivní opatření od 16. 3. 2020 (nulové školní kontakty, omezení pracovních kontaktů na 30 %, omezení jiných kontaktů také na 30 %)



Obr. 2: Znárodnění orientační kalibrace modelu na pozorovaná data

Pozn. Model uvažuje započítání případu k datu skončení infekčního období (izolace). K započítání případu do pozorovaných statistik dochází s několikadenním odstupem s ohledem na efektivitu testování. **V recentních statistikách je již patrný krátký (cca 1-2denní) odstup mezi izolací a confirmací.**

## 4 Publikace datových sad umožňujících práci s modely

---

### 4.1 Jaká data se do modelů používají

- Rešerše odborné literatury, výzkumných zpráv
- Počet osob ve věkových skupinách v ČR (věkově strukturovaný model)
- Počet osob s prokázanou nákazou (data ČR, registr ISIN)

Pro nastavení iniciačních parametrů krátkodobého modelu byly dále využity

- Analýza časového odstupu data výsledku testu od data prvního příznaku
- Počet případů importovaných ze zahraničí (kalibrace počáteční dynamiky byla provedena na konci března)

### 4.2 Jaké datové sady byly navrženy pro daný typ modelování

- COVID-19: Celkový (kumulativní) počet osob s prokázanou nákazou dle krajských hygienických stanic včetně laboratoří
- Počet osob s prokázanou nákazou dle věku
- Údaje ISIN-KHS pro odvození počtu případů importovaných ze zahraničí a efektivity testování

### 4.3 Kde jsou tyto datové sady zveřejněny

COVID-19 v ČR: Otevřené datové sady

- <https://onemocneni-aktualne.mzcr.cz/api/v1/covid-19>