

Katedra pravděpodobnosti a matematické statistiky



**MATEMATICKO-FYZIKÁLNÍ
FAKULTA**
Univerzita Karlova



doc. RNDr. Arnošt Komárek, Ph.D.

o neÚnavném ZKreslování Informací a Statistik

Kostelec nad Černými Lesy, 7. listopad 2021

I.

Data a neúnavné Zkreslování Informací a Statistik

Zcela jsme selhali v procesu poznání založeném na datech.

- Nemáme data?
- ÚZIS, zdravotní pojišťovny, . . . jsou jich přeci plné.
- Kde je tedy problém?



Professor, Chair of Computational Statistics, School of Mathematics, University of Edinburgh, UK

Editor, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*

“Statistics is about the honest interpretation of data. It’s not always a popular subject: honest interpretation of data is difficult, and much less appealing than less honest interpretation.”

S.N. Wood (2021). [Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns?](#) *Biometrics* (early-view, published 30 March 2021), doi: 10.1111/biom.13462.

The number of new infections per day is a key quantity for effective epidemic management.

It can be estimated relatively directly by testing of random population samples.

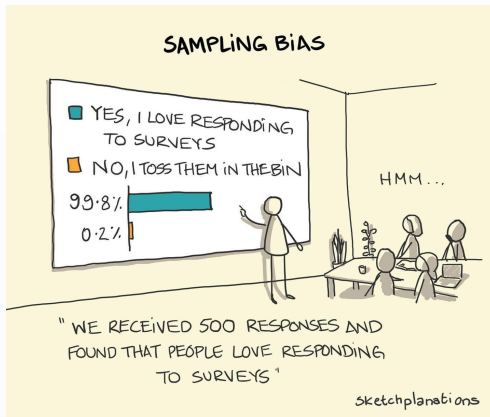
Without such direct epidemiological measurement, other approaches are required to infer whether the number of new cases is likely to be increasing or decreasing: for example, estimating the pathogen-effective reproduction number, R , using data gathered from the clinical response to the disease.

For coronavirus disease 2019 (COVID-19/SARS-CoV-2), such R estimation is heavily dependent on modelling assumptions,

because the available clinical case data are opportunistic observational data subject to severe temporal confounding...

- Prakticky vše **observační** data, která vznikají „sama od sebe“
- Např. **denní počty PCR pozitivních**
(\neq počet nových infekcí v populaci)
- Nejsou ani **reprezentativní** (pro populaci zájmu), ani **konzistentní** (význam se mění nejenom v čase)
- Nejsou bezcenná, ale...

- Extrémně náchylná k **dezinterpretacím** (až dezinformacím)...
- Vhodným výběrem, jaké charakteristiky dat **sdělit** a jaké **zamlčet**, lze dojít k **téměř libovolnému**, předem určenému a požadovanému **závěru**...



- Klíčový ukazatel pro rozhodování českých politiků v posledních týdnech/měsících: „**incidence**“ (počet případů na 100 tis. obyvatel)
- Počítaná z počtu **pozitivně otestovaných**
- Skupina **pozitivně otestovaných** = skupina **(nově) nakažených**?
- Nákaza virem SARS-CoV-2 je u každého (rychle) viditelná jako např. u viru MERS-CoV?
- Testujeme „rovnoměrně“ všechny segmenty populace?
 - **mladé, staré**
 - **očkované, neočkované**
 - **:**



A když si ještě trošinku **zakouzíme** s podmiňováním, **vykouzíme** ledasjakou iluzi. . .

Epidemie na populační úrovni roste, aktuálně se zátěž pohybuje na počtech nově diagnostikovaných případů $> 20 / 100\text{tis. obyvatel}$ za týden.



Průměrný počet nově prokázaných případů onemocnění se pohybuje kolem 270 denně. Relativní pozitivita indikovaných testů se stále drží bezpečných hodnot pod 4%, nicméně po 1.9. mírně roste.

V posledních týdnech se zvyšuje počet nově zachycených nákaz u mladých lidí, zejména ve věku 16 – 29 let.




Zátěž regionů ve vzájemném srovnání krajů k 8. 9.

Název kraje	7denní počet nových případů na 100 tis. obyv.	14denní počet nových případů na 100 tis. obyv.
Karlovarský kraj	38.5	55.2
Hlavní město Praha	35.9	63.7
Jihočeský kraj	23.2	39.6
Středočeský kraj	22.7	38.1
Jihomoravský kraj	21.1	31.5
Zlínský kraj	17.8	25.9
Ústecký kraj	17.7	27.1
Pardubický kraj	17.4	24.1
Liberecký kraj	17.0	28.5
Moravskoslezský kraj	16.7	25.7
Kraj Vysočina	14.9	25.9
Plzeňský kraj	12.9	22.5
Olomoucký kraj	10.5	18.7
Královéhradecký kraj	9.6	14.3
ČR	20.5	33.4

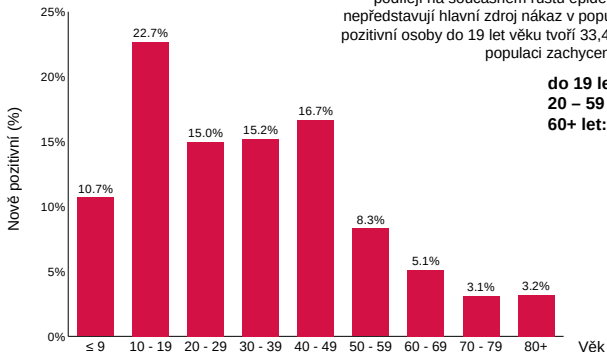
Virová zátěž dle věku:

- 6 – 11 let: 38 / 100 tis. za týden
- 12 – 15 let: 53 / 100 tis. za týden
- 16 – 19 let: 55 / 100 tis. za týden
- 20 – 29 let: 29 / 100 tis. za týden

 Hodnota za ČR je vyšší než 20 případů / 100 tis. obyv. za 7 dní

Nově pozitivní případy v celé populaci: 1.9.2021 – 6.9.2021

N = 1 271



Záchyty ve školách a u mladých lidí obecně se podílejí na současném růstu epidemie, avšak nepředstavují hlavní zdroj nálezů v populaci. Nově pozitivní osoby do 19 let věku tvoří 33,4% všech v populaci zachycených nálezů.

do 19 let : 33.4%
20 – 59 let: 55.1%
60+ let: 11.4%

„Statistiky nově nakažených naopak zcela zaplňují momentálně lidé do 30 až 40 let, přičemž díky plošným testům ve školách jich třetinu tvoří děti. Protože po čtvrtku se pokračovat celorepublikově v testování zatím nemá, jejich poměr patrně klesne. A s tím i celkový počet nově evidovaných jako nakažení.“

„Statistiky nově nakažených naopak zcela zaplňují momentálně lidé do 30 až 40 let, přičemž díky plošným testům ve školách jich třetinu tvoří děti. Protože po čtvrtku se pokračovat celorepublikově v testování zatím nemá, jejich poměr patrně klesne. A s tím i celkový počet nově evidovaných jako nakažení.“

Opravdu nás zajímá $\hat{P}(\text{věk } X \mid \text{pozitivně testovaný})$?

- Jaképak věkové rozdělení má **testovaná populace**?
- Zcela hypoteticky, jaká bude $\hat{P}(\text{věk } 0-40 \mid \text{pozitivně testovaný})$ pokud bychom testovali pouze lidi do 40 let?
- Jakápak je proporce osob do 40 let mezi testovanými?

„Statistiky nově nakažených naopak zcela zaplňují momentálně lidé do 30 až 40 let, přičemž díky plošným testům ve školách jich třetinu tvoří děti. Protože po čtvrtku se pokračovat celorepublikově v testování zatím nemá, jejich poměr patrně klesne. A s tím i celkový počet nově evidovaných jako nakažených.“

Opravdu nás zajímá $\hat{P}(\text{věk } X \mid \text{pozitivně testovaný})$?

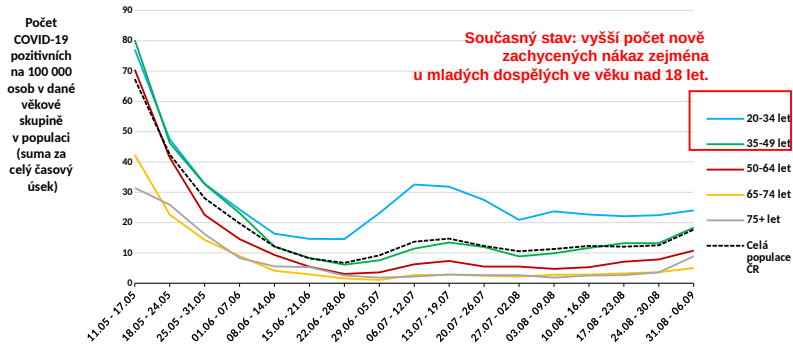
- Jaképak věkové rozdělení má **testovaná populace**?
- Zcela hypoteticky, jaká bude $\hat{P}(\text{věk } 0-40 \mid \text{pozitivně testovaný})$ pokud bychom testovali pouze lidi do 40 let?
- Jakápak je proporce osob do 40 let mezi testovanými?

Nebylo by účelnější vědět $\hat{P}(\text{pozitivně testovaný} \mid \text{věk } X)$?

Nebo ještě lépe $\hat{P}(\text{nakažený} \mid \text{věk } X)$?

Počty COVID-19 pozitivních v ČR na 100 000 v populaci

Populační zátěž v první polovině roku klesla ve všech věkových kategoriích dospělých, významně i v kategoriích seniorního věku – vliv očkování.



A co počty **nakažených** na 100 000 obyvatel?

- To je ale super, že v rizikových skupinách máme tak malou incidenci.
- To bude tím očkováním.
- Nebo spíš: Čím větší riziko těžkého průběhu, tím méně testování?

A to dává smysl! Ještě bychom se mohli (zavčasu) vyděsit a neudali bychom ty testy ve školách.

Bayesův vzorec a plošné testování

$$P(\text{VIR} | \text{Test}+) = \frac{\text{senzitivita} \cdot P(\text{VIR})}{\text{senzitivita} \cdot P(\text{VIR}) + (1 - \text{specifita}) \cdot P(\text{neVIR})}$$

Příklad

Test = Ag test made in China

VIR = PCR test najde RNA SARS-CoV-2

specificita $P(\text{Test} - \text{neVIR})$	senzitivita $P(\text{Test} + \text{VIR})$	prevalence $P(\text{VIR})$	PPV $P(\text{VIR} \text{Test}+)$
0,99	0,50	0,010	0,34
		0,005	0,20
0,999	0,50	0,005	0,72
		0,001	0,33

Zářijové školní testování

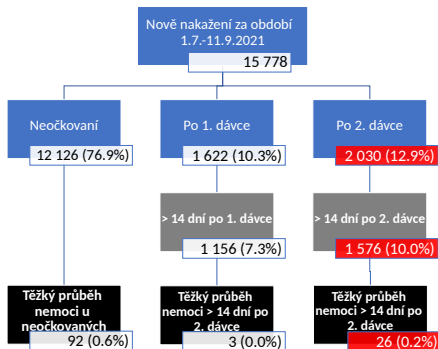
- 3 136 679 testů (za 63 MKč)
- 1 183 pozitivních ve škole
- 430 „potvrzených“ na PCR (na sliznici nalezena RNA viru SARS-CoV-2, nebožák ne nutně infekční)
- cena za nález jednoho PCR+ pozitivního (ne nutně infekčního!): 146 kKč

- **2/3 (753 dětí) pozitivních ve škole bylo falešně pozitivní** (vůči PCR testu, který sám nezaručuje infekčnost jedince)

Nově nakažení: 1.7.-11.9.

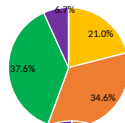


MINISTERSTVO ZDRAVOTNICTVÍ
ČESKÉ REPUBLIKY



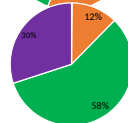
Věk všech nově nakažených

- Průměr: 31 let
- Medián: 28 let



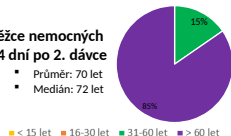
Věk nakažených > 14 dní po 2. dávce

- Průměr: 51 let
- Medián: 49 let



Věk těžce nemocných > 14 dní po 2. dávce

- Průměr: 70 let
- Medián: 72 let



Zajímá nás (povznesme se nad pofidérního jmenovatele)?

$$\hat{P}(\text{těžký průběh \& není očkovan}) = 92/15\,778 \doteq 0,006 (0,6 \%)$$

$$\hat{P}(\text{těžký průběh \& očkovan 2 dávky}) = 26/15\,778 \doteq 0,002 (0,2 \%)$$

Nebo spíš tohle (stále se povznášíme nad pofidérního jmenovatele a neznámou strukturu očkovaných, resp. neočkovaných vzhledem k dalším důležitým faktorům jako věk atd.)?

$$\hat{P}(\text{těžký průběh | není očkovan}) = 92/12\,126 \doteq 0,008 (0,8 \%)$$

$$\hat{P}(\text{těžký průběh | očkovan 2 dávky}) = 26/2\,030 \doteq 0,013 (1,3 \%)$$

**Kdepak byl šéf českých statistiků, když se probíraly dvou-
rozměrné kontingenční tabulky?**

- Potřebujeme znát „**incidenci**“ založenou z velké části na **testování nerizikové populace** (aby si politici zdůvodnili to, co potřebují zdůvodnit)?
- Nebo raději **incidenci** (zejména v rizikových skupinách), aby rozhodnutí o dalším postupu byla konečně účelná a racionální?

- Potřebujeme znát „**incidenci**“ založenou z velké části na **testování nerizikové populace** (aby si politici zdůvodnili to, co potřebují zdůvodnit)?
- Nebo raději **incidenci** (zejména v **rizikových skupinách**), aby rozhodnutí o dalším postupu byla konečně účelná a racionální?



Ústav zdravotnických informací
a statistiky?

Ústav **pro zkreslování** informací
a statistiky?

II.

Aktualita

Covid se stal epidemií neočkovaných. Vojtěch zmínil odstrašující Ostravsko

© 26. října 2021 7:03, aktualizováno 11:06

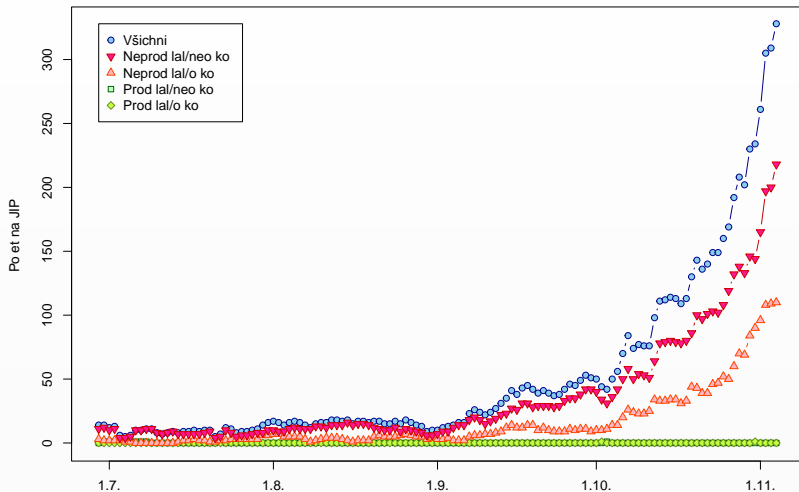


Epidemie se stala epidemií neočkovaných a vývoj sleduje rizikový scénář, řekl v úterý ministr zdravotnictví Adam Vojtěch. S ředitelem ÚZIS Ladislavem Duškem a hlavní hygieničkou Pavlou Svrčinovou dávají vinu za současný stav hlavně neočkovaným lidem. Těch je v celém Česku přes třicet procent. V nemocnicích také přibývá mladých, hospitalizovaní pacienti s deltou mají asi jen poloviční šanci na přežití.



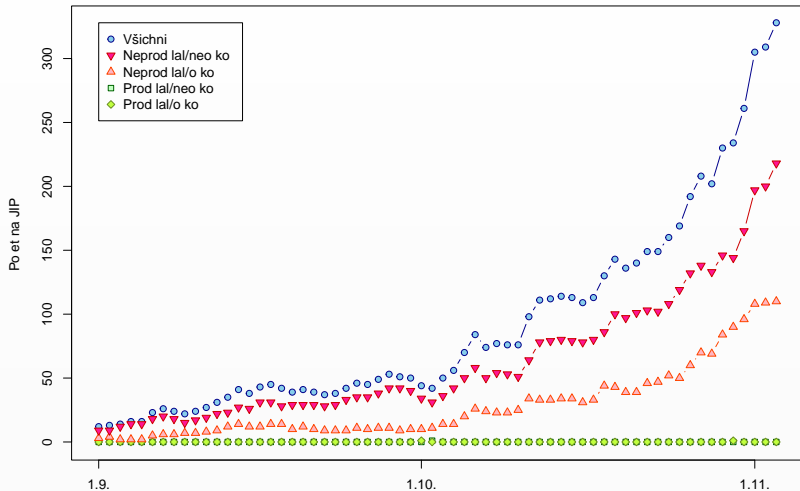
Kdopa nám s COVIDem plní JIPky?

COVID-19 pacienti na JIP od 28.6.2021



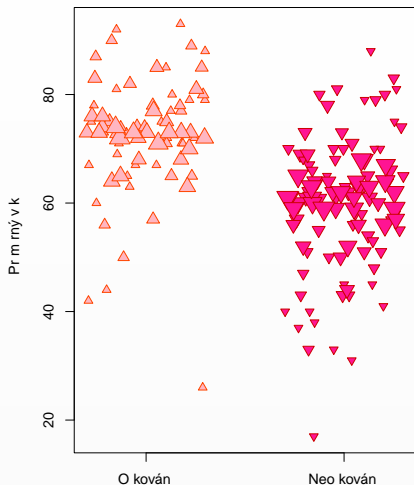
Kdopa nám s COVIDem plní JIPky?

COVID-19 pacienti na JIP od 1.9.2021



A japak jsou ti lidé staří?

Pacienti s COVID-19 na JIP od 28.6.2021

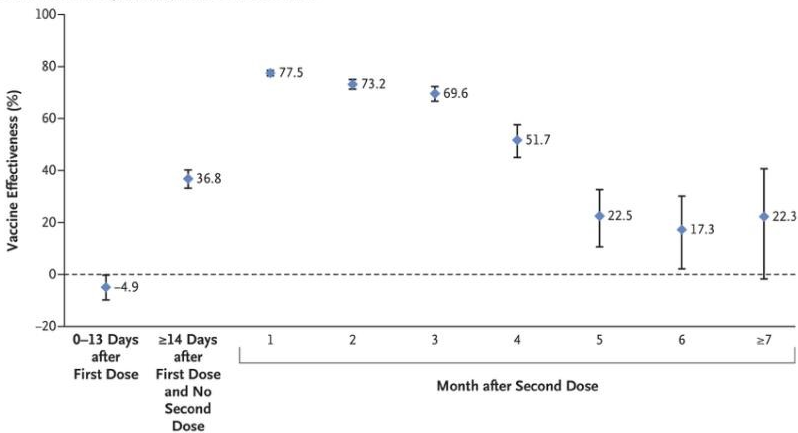


	Průměr	Směr. odch.	95% Interval spol.
Data od 28.6.2021			
Očkovaní	72,5	14,3	(70,9, 74,1)
Neočkovaní	61,0	17,0	(59,6, 62,3)
Data od 1.9.2021			
Očkovaní	72,6	14,7	(70,9, 74,4)
Neočkovaní	61,3	16,0	(59,9, 62,6)

A co ty vakcíny, po jejichž aplikaci se určitě nenakazíte, jak zjara tvrdil doc. MUDr. JB, Ph.D.?

Katarská studie: Chemaitelly a kol., *the New England Journal of Medicine*

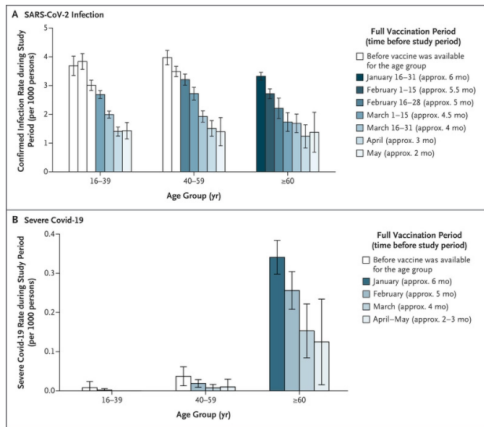
A Effectiveness against Any SARS-CoV-2 Infection



A co ty vakcíny, po jejichž aplikaci se určitě nenakazíte, jak zjara tvrdil doc. MUDr. JB, Ph.D.?

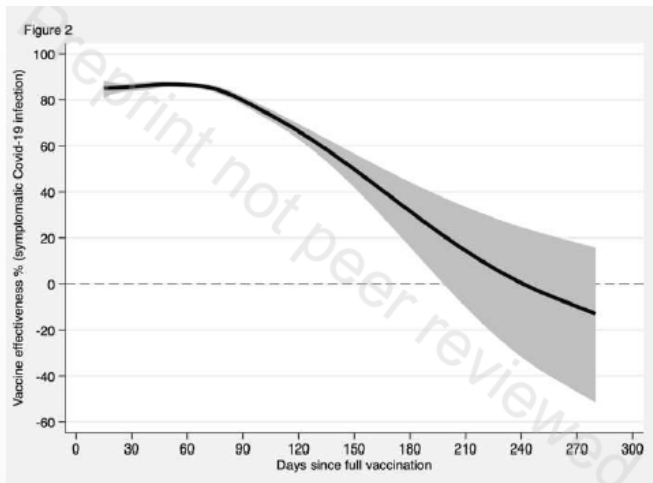
Izraelská studie: Goldberg a kol., *the New England Journal of Medicine*

Figure 3. Rates of Documented SARS-CoV-2 Infection and Severe Covid-19, July 11 to 31, 2021.



A co ty vakcíny, po jejichž aplikaci se určitě nenakazíte, jak zjara tvrdil doc. MUDr. JB, Ph.D.?

Švédská studie: Nordström a kol., posuzováno v *the Lancet*



- „Mladý“ (bez dnes již dobře popsanych komorbidit) se spíš utopí než skončí na JIP s COVIDem.
- Kdokoliv, kdo „prodělal“, bude spíš za bílého dne zastřelen před pražským orlojem než že skončí na JIP s COVIDem.
- I naočkovaný je na tom po cca půl roce vůči nakažení zhruba stejně jako neočkovaný.

Brutální fotky umírajících v covidářiích? Vláda spustí novou kampaň

5. listopadu 2021 16:42



Jedinou cestou z epidemie je podle vlády očkování. To však pokulhává především mezi mladými. Resort se proto rozhodl změnit kurz a stávající očkovací kampaň zaměřit na ně, i když právě oni nemocnice zrovna neplní. Tlak se nyní přesune na sociální sítě, včetně populárního TikToku. Tuto cestu poradila resortu společnost Semantic Visions. Premiér Andrej Babiš v pátek oznámil brutální kampaň, která je inspirována obrázky z cigaretových krabiček.



Kdyby se býval byl naočkovával.... I takto by mohla vypadat kampaň | foto: Úřední věstník Evropské unie

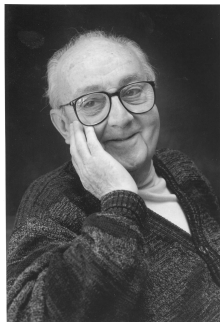
Reklama



III.

Všechny modely jsou špatné, ale některé...

Částečně vykradeno z prezentace RNDr. Tomáše Fürsta, Ph.D.,
v Karlíně vystudovaného **matematického modeláře**



(významný) britský statistik, 1919–2013

“Essentially, all models are **wrong, but** some are **useful**. The practical question is how wrong do they have to be to not be useful.”

„Jednodušší“ děje neživé přírody

- proudění za jasně daných podmínek, ...
 - „model“ = sada (nezpochybnitelných) fyzikálních zákonů
 - systému **dobře rozumíme**
 - téměř zanedbatelný vliv náhody
 - „data“ téměř nepotřebná
- **predikce** i „**kauzální**“ inference (kontrola procesu) celkem spolehlivé

Komplexní děje neživé přírody

- počasí, klima, . . .
 - „model“ = (větší) sada fyzikálních zákonů
 - systému rozumíme, ale již ne úplně do detailu
 - náhodné vlivy nelze úplně ignorovat, ale fyzikální zákony stále podstatným způsobem definují systém
 - „data“ potřebná, jsou k dispozici (reprezentativní i konzistentní)
- **predikce** poměrně spolehlivá krátkodobě
- „**kauzální**“ inference (kontrola procesu) téměř nemožná

„Jednodušší“ děje živé přírody

- pohyb kočky, . . .
 - „model“ = ???
 - systému **skoro vůbec nerozumíme**
 - náhodné vlivy obrovské
 - **reprezentativní a konzistentní** data nutná k samotnému pochopení systému
- **predikce**???
- „**kauzální**“ inference (kontrola procesu)???

matematické modelování → **statistické** modelování

Komplexní děje živé přírody

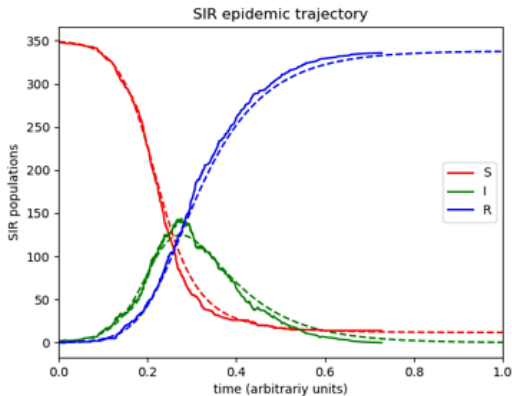
- interakce viru a lidské společnosti, ...
 - „model“ = ???
 - systému **skoro vůbec nerozumíme**
 - náhodné vlivy obrovské
 - **reprezentativní a konzistentní** data nutná k samotnému pochopení systému
- **predikce**???
- „**kauzální**“ inference (kontrola procesu)???

matematické modelování → **statistické** modelování

- Variace na kompartmentové (krabičkové) modely (SIR, SEIR, ...)
- Model chemické reakce: dokonale míchaný plyn, ignorují **prostor**.
- Produkují vždy **pouze jednu vlnu!**
- Nebezpečí **dezinterpretace**
 - model predikuje jednu vlnu
 - vidím dvě vlny, tři vlny, ...
 - musí to být důsledek **změny parametrů modelu** \equiv **vládních opatření**

Žádný (ani český ani světový) model nepředpověděl správně ani **bod zvratu probíhající vlny ani **nástup** další vlny.**

Modely epidemie

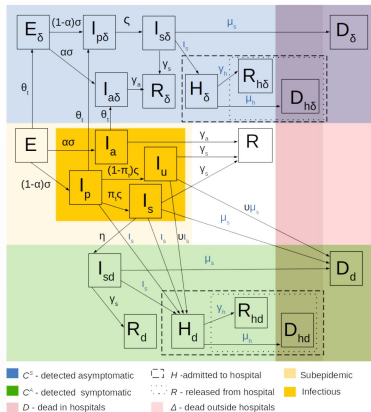


Susceptible

Infected

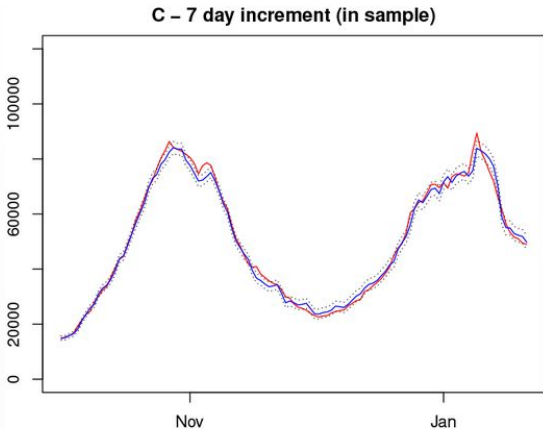
Recovered

Overfitting



Zdroj: Šmíd et al. SEIR Filter: A stochastic model of epidemics, model BISOP: 41 vstupních parametrů plus dalších 26 „faktorů“ Jeden z modelů J. Hamáčka.

- Víc parametrů = lepší model?
- Čím víc parametrů, tím lépe lze model „vyladit“ tak, aby (skoro) přesně kopíroval historická data.
- Každá šipka \equiv model \equiv „fyzikální zákon“, zde nijak ověřitelný předpoklad
- **Predikční schopnosti???**



Červeně: klouzavé 7denní součty počtu nově nakažených (*pozitivně testovaných*)

Modře: predikce na základě modelu

Vezmi 68 bodů a polynom 67. stupně a to by bylo, aby ses netrefil. . .

*Jeden **faktor** vládne všem, Jeden jim všem káže,
Jeden **viry odvede**, do temnoty sváže.*

*Jeden faktor vládne všem, Jeden jim všem káže,
Jeden viry odvede, do temnoty sváže.*



≡ intenzita kontaktů

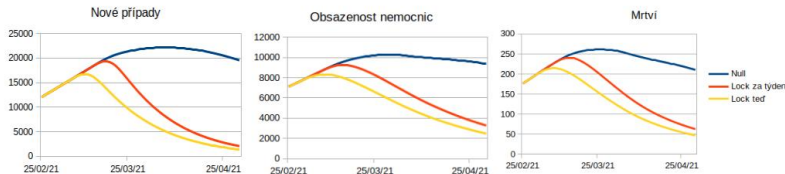
→ parametr modelu/faktor,
se kterým si lze hrát a koukat,
„co se stane, když“

„Scénáře“ prezentované (údajně) koncem února Janem Hamáčkem ve Sněmovně při hlasování o uvalení domácího vězení na všechny občany ČR. Údajně výstupy BISOP modelu.

Scénáře 24.2. - BISOP & PaQ

Null – snížení ze 13 na 12 kontaktů (např. homeoffice),

Lock – snížení na 6.5 kontaktu (zákaz návštěv domácností, zákaz pobytu na neesenc. pracovištích)



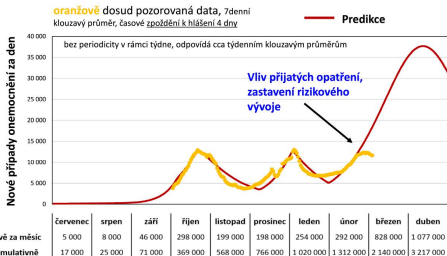
Poznámky

- scénář **lock teď** počítá se zavedením opatření neprodleně, scénář **za týden** se zavedením příští týden,
- velká míra nejistoty o nakažlivosti varianty B, zde spíše pesimistický odhad
- model počítá s očkováním a přirozenou imunizací (druhý efekt nejistý)
- použitý model <https://www.medrxiv.org/content/10.1101/2021.02.16.21251834v1> (preprint)
- předpovězené hodnoty jsou nejisté, zejména od měsíčního horizontu, grafy slouží hlavně pro srovnání
- v praxi bude nutno opatření upravovat dle aktuální situace

A když se to pak nestane. . .

Prezentace biologa Duška pro poslance PS PČR u příležitosti hlasování o (dalším) prodloužení nouzového stavu (okolo **20. března**)

SEIR model: scénář pro rozvolnění v březnu a eskalace šíření nákazy
Model pokračování výrazného nárůstu pozorovaného v únoru



Snímek prezentuje výsledky simulace prostřednictvím epidemiologického modelu SEIR, který zahrnuje vybrané předpoklady a slouží ke zkoumání dopadu změny různých parametrů epidemie. Vzhledem k neurčitostem ve struktuře modelu (například limitované znalosti o skutečné vnímavosti populace k viru a k jeho novým variantám) je nezbytné výsledky brát jako orientační, umožňující zejména celkové srovnávání jednotlivých scénářů, nikoli konkrétní předpovědi pro určité období.

Červeně: predikce učiněná ÚZISem **koncem února** před hlasováním o uvalení domácího vězení na všechny občany ČR
Oranžově: skutečnost

Uvolnění od 3.12.

- zvýšení pracovních kontaktů na 80 %
- zvýšení školních kontaktů na 50 %
- zvýšení jiných kontaktů na 100 %
- zvýšení základní reprodukce na původní úroveň

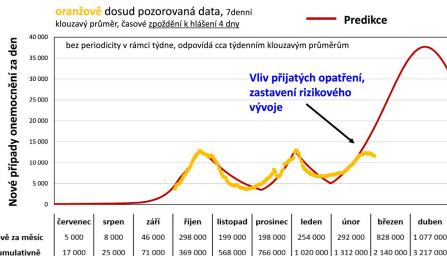
Efekt opatření od 28.12.

- snížení pracovních kontaktů na 40 %
- školní kontakty na 0 %
- snížení jiných kontaktů na 40 %
- snížení základní reprodukce o 10%
- školní kontakty na 30 % od 4.1.

Od 25.1.

- zvýšení základní reprodukce o 70 % (simulace šíření nakažlivějších forem viru, simulace nedodržování opatření nebo rozvolnění)
(uvedeny podíly normálních kontaktů)

SEIR model: scénář pro rozvolnění v březnu a eskalace šíření nákazy Model pokračování výrazného nárůstu pozorovaného v únoru



Snímek prezentuje výsledky simulace prostřednictvím epidemiologického modelu SEIR, který zahrnuje vybrané předpoklady a sází ke zkoumání dopadu změny různých parametrů epidemie. Vzhledem k neurčitostem ve struktuře modelu (například limitované znalosti o skutečné vnímavosti populace k viru a k jeho novým variantám) je nezbytné výsledky brát jako orientační, umožňující zejména celkové srovnávání jednotlivých scénářů, nikoli konkrétní předpověď pro určité období.

Uvolnění od 3.12.

- zvýšení pracovních kontaktů na 80 %
- zvýšení školních kontaktů na 50 %
- zvýšení jiných kontaktů na 100 %
- zvýšení základní reprodukce na původní úroveň

Efekt opatření od 28.12.

- snížení pracovních kontaktů na 40 %
- školní kontakty na 0 %
- snížení jiných kontaktů na 40 %
- snížení základní reprodukce o 10%
- školní kontakty na 30 % od 4.1.

Od 25.1.

- zvýšení základní reprodukce o 70 %**
(simulace šíření nakažlivějších forem viru, simulace nedodržování opatření nebo rozvolnění)
(uvedeny podíly normálních kontaktů)

Predikce nevyšla.
ANO, to bude **tými opatřeními!**

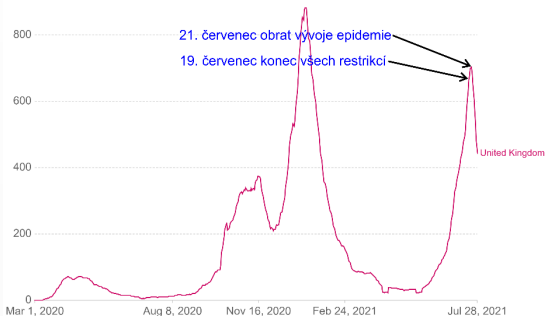
seznamzpravy.cz, 20.8.2021

I samotná londýnská vláda veřejně připustila, že po tzv. „dni svobody“ pravděpodobně dojde k výrazným nárůstům počtu nakažených covidem-19, které se mohou vyšplhat až na 100 000 denně, protože nemalá část populace ještě nebude očkována. Za „nebezpečné a neetické“ označilo v otevřeném [dopise](#) rozhodnutí 19. července zrušit téměř veškerá opatření víc jak 4000 vědců, lékařů a dalšího zdravotnického personálu. Britská vláda ale nakonec přesto rozvolnila a světe, div se, nárůst počtů nakažených se nekonal.

Daily new confirmed COVID-19 cases per million people

Shown is the rolling 7-day average. The number of confirmed cases is lower than the number of actual cases; the main reason for that is limited testing.

Our World
in Data



Source: Johns Hopkins University CSSE COVID-19 Data

CC BY

„Matematické“ modely epidemie (nejenom) v ČR

- Sada **mnoha** předpokladů, které nelze nijak ověřit
- Jsou krmena „daty“, které jenom vzdáleně odrážejí realitu (**reprezentativnost, konzistence**)
- Nejsou řádně **validovány**
- Autoři netuší, že **overfitting**, který činí jejich model tak atraktivním pro průměrného humanitně **nedovzdělaného** ministra, je ve skutečnosti **prů...**
- Žádný z modelů, na základě kterých čeští orgáni rozhodují, neprošel odbornou oponenturou **navzdory** faktu, že modelování dějů v živé přírodě při obrovské nejistotě a s apriori nulovou kvalitativní znalostí celého systému se velká část „vládních modelářů“ začala věnovat až na jaře 2020...
- Pokusy o použití **prediktivních** modelů ke **kauzální** inferenci jsou dálnicí do pekla



libanonsko-americký esejist, filozof, (finanční) matematik, naroz. 1960 (on September Eleven)

teorie „černé labutě“

“A wrong map does more harm than having no map.”

Za špatného počasí přistáváte ve **Vídni**, ale pilot nemá mapu. Chcete, aby přistával **bez mapy** nebo raději použil mapu (nedalekého) letiště v **Bratislavě**?

“Essentially, all models are **wrong, and**
some are **even dangerous. . .**”

“Essentially, all models are **wrong, and**
some are **even dangerous...**”

Ale když politici tak milují nebezpečí...

IV.

Efekty opatření?

S.N. Wood (2021). [Inferring UK COVID-19 fatal infection trajectories from daily mortality data: Were infections already in decline before the UK lockdowns?](#) *Biometrics* (early-view, published 30 March 2021), doi: 10.1111/biom.13462.

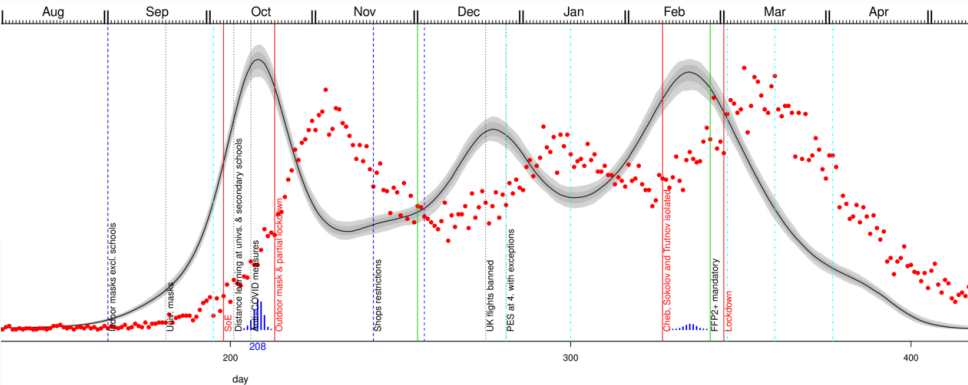
The number of new infections per day is a key quantity for effective epidemic management. It can be estimated relatively directly by testing of random population samples. Without such direct epidemiological measurement, other approaches are required to infer whether the number of new cases is likely to be increasing or decreasing: for example, estimating the pathogen-effective reproduction number, R , using data gathered from the clinical response to the disease. For coronavirus disease 2019 (COVID-19/SARS-CoV-2), such R estimation is heavily dependent on modelling assumptions, because the available clinical case data are opportunistic observational data subject to severe temporal confounding.

Given this difficulty, it is useful to retrospectively reconstruct the time course of infections from the least compromised available data, using minimal prior assumptions.

- S.N. Wood: průběh epidemie ve Velké Británii (+ Švédsku a Španělsku)
 - primární cíl článku: **metodologie**
- Robert Straka (*WIMiP AGH Krakow & FJFI ČVUT v Praze*) & AK:
 - použití Woodovy metodologie nad ÚZIS daty z České republiky

The least compromised available data

- denní počty úmrtí na/s COVID-19
 - v rámci jednoho státu poměrně konzistentní i reprezentativní data



- Úmrtí \equiv zpožděný následek nákazy
- **Kdy se zemřelý jedinec nakazil?**

- Úmrtí \equiv zpožděný následek nákazy
- **Kdy se zemřelý jedinec nakazil?**
 - obtížné zjistit u každého konkrétního
 - lze ale odhadnout **rozdělení** doby mezi nákazou a smrtí (pokud nastala) a „posunout“ v čase zemřelé do okamžiku, kdy se nakazili
 - zajímá nás, co se děje v **populaci**, tj. co možná nejlepší odhad **počtu nakažených** ten který den, nikoliv jestli se pan Novák, který zemřel 25.2. nakazil 24.1., 1.2. nebo 5.2.

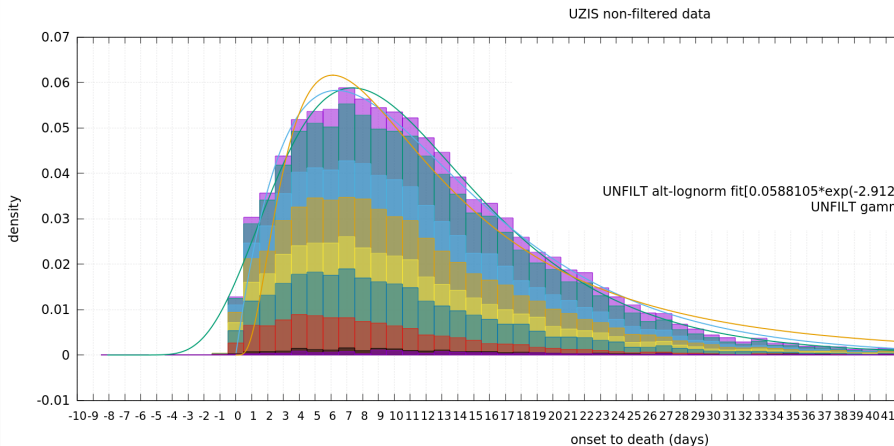
nákaza \rightarrow **příznaky** \rightarrow **smrt**

Čas nákaza → příznaky

- V ČR validní data prakticky neexistují, i jinde to je s daty horší, vyžadovalo by to **plánované prospektivní** sledování vhodně vybraných kohort populace. . .
- Odborná literatura (a observační data), např. McAloon et al. (2020). Incubation period of COVID-19: A rapid systematic review and meta-analysis of observational research. *BMJ Open*, **10**, e039652, doi: 10.1136/bmjopen-2020-039652.

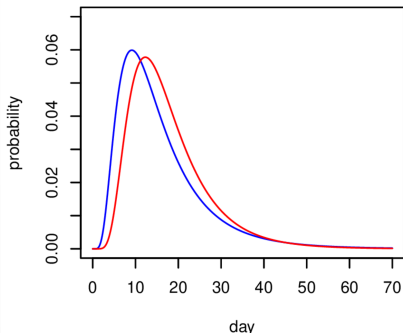
Čas příznaky → smrt

- Data v ČR, viz ÚZIS, „nemocniční“ databáze



- Trocha práce, **prakticky žádné** „modelovací“ předpoklady...

⇒ Odhad rozdělení doby/času **nákaza** → **smrt**



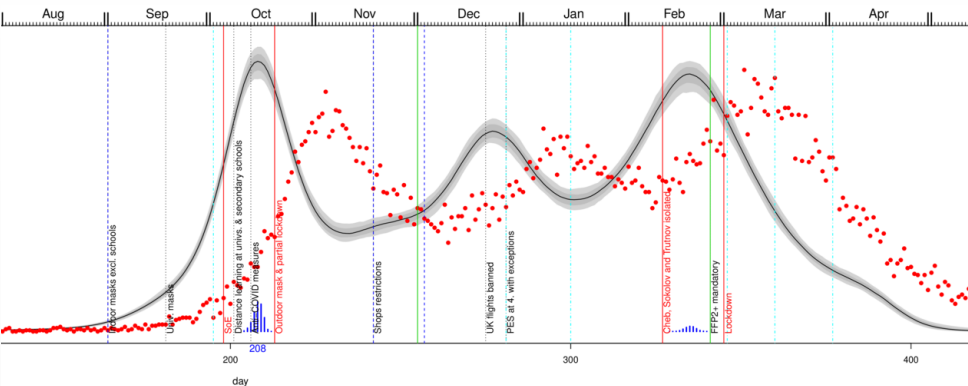
červeně: data 1.9.2020–1.5.2021

modře: data 1.3.2020–1.9.2020

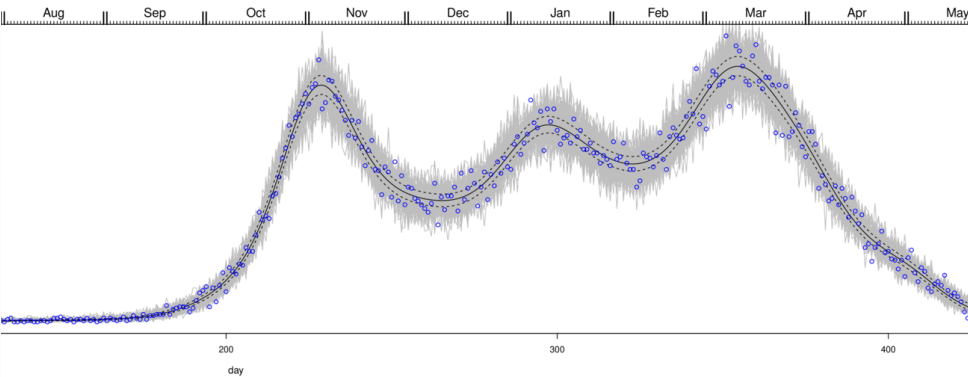
Odhadnutá střední doba mezi nákazou a smrtí

- ČR a data ÚZIS: 23,3 dne
- UK (analýza S.N. Wood): 26,7 dne

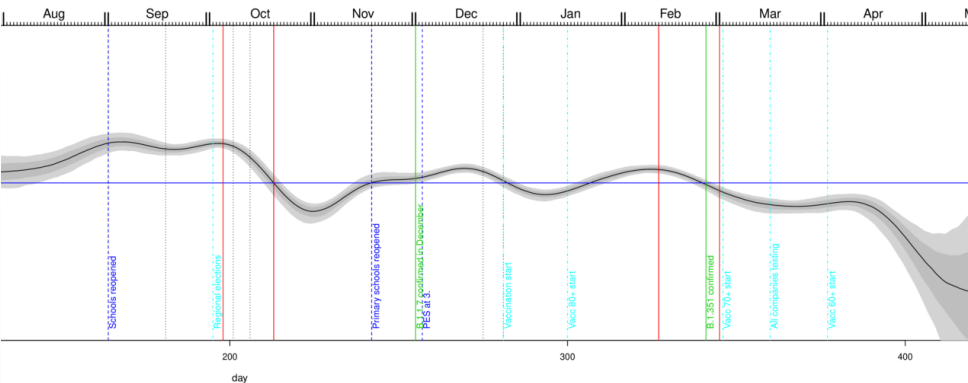
⇒ Odhad incidenční křivky fatálních infekcí



⇒ **Odhad lze zpětně validovat...**



⇒ Reprodukční číslo fatálních infekcí



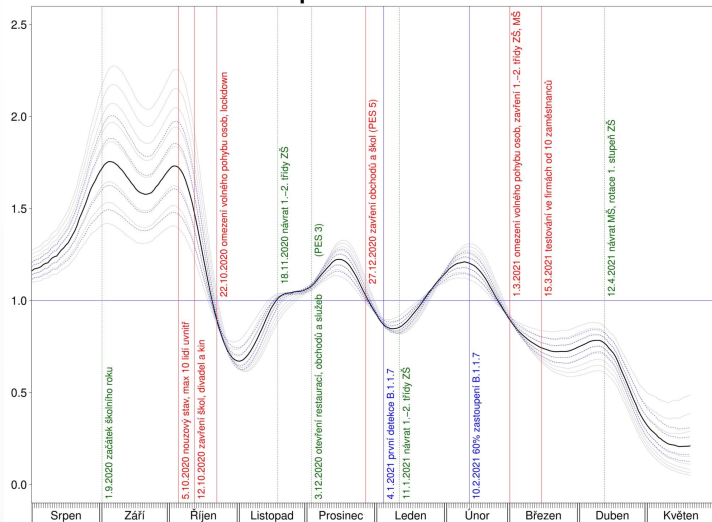
- Na obr. je $\log(R_t)$
- Horizontální modrá přímka $\equiv \log(R_t) = 0$, tj. $R_t = 1$

- Odhad incidenční křivky ani křivky R_t nijak nepoužil informaci o „opatřeních“
 - Ukazuje, jak to bylo. . .
 - **v ten který den**
- ⇒ skutečné počty **fatálně** nakažených
bez ohledu na aktuální testovací strategii atp.

- Odhad incidenční křivky ani křivky R_t nijak nepoužil informaci o „opatřeních“
 - Ukazuje, jak to bylo. . .
 - **v ten který den**
- ⇒ skutečné počty **fatálně** nakažených bez ohledu na aktuální testovací strategii atp.
- Zpětně můžeme přidat informaci o „opatřeních“ a podívat se, zda se při jejich zavedení (resp. krátce po jejich zavedení) „něco stalo“ s incidencí nebo reprodukčním číslem (fatálních infekcí).
 - Pokud má něco vliv, musí to být vidět **hned**, resp. s minimálním zpožděním.

„Matematické“ modely jako prostředek k rozhodování, „co udělat, aby“?

Odhad efektivního reprodukčního čísla fatálních infekcí



„Matematické“ modely jako prostředek k rozhodování, „co udělat, aby“?

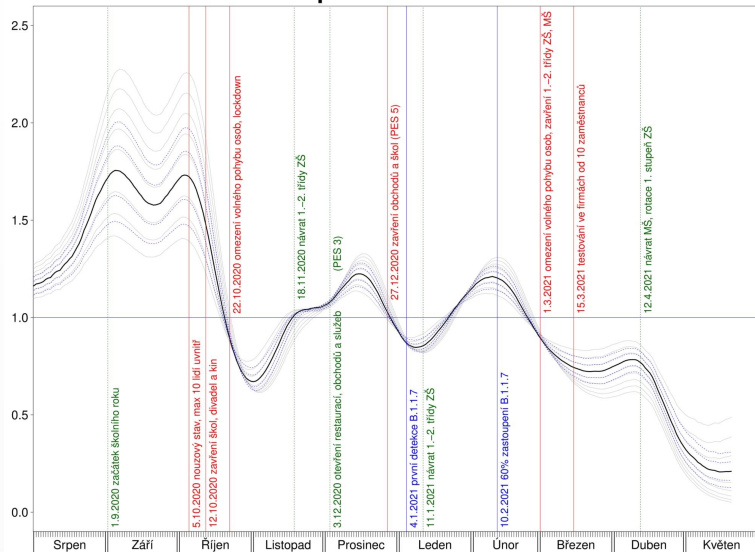
- PES = 3 velí rozvolňovat (začátek prosince 2020), když R_t roste...
- PES = 4, 5 velí zpřísnovat, když R_t doklesalo k hodnotě 1 a dále **nezměněným tempem** klesá...
- Modely ministra J. Hamáčka velí uzavřít obyvatele v katastrech obcí, když R_t již 14 dní klesá, doklesalo **pod** hodnotu 1 a dále klesá...

„Matematické“ modely jako prostředek k rozhodování, „co udělat, aby“?

- PES = 3 velí rozvolňovat (začátek prosince 2020), když R_t roste...
- PES = 4, 5 velí zpřísnovat, když R_t doklesalo k hodnotě 1 a dále **nezměněným tempem** klesá...
- Modely ministra J. Hamáčka velí uzavřít obyvatele v katastrech obcí, když R_t již 14 dní klesá, doklesalo **pod** hodnotu 1 a dále klesá...

**„Analýza“ dat jako alibi pro politická rozhodnutí nebo
prostředek k poznání reality?**

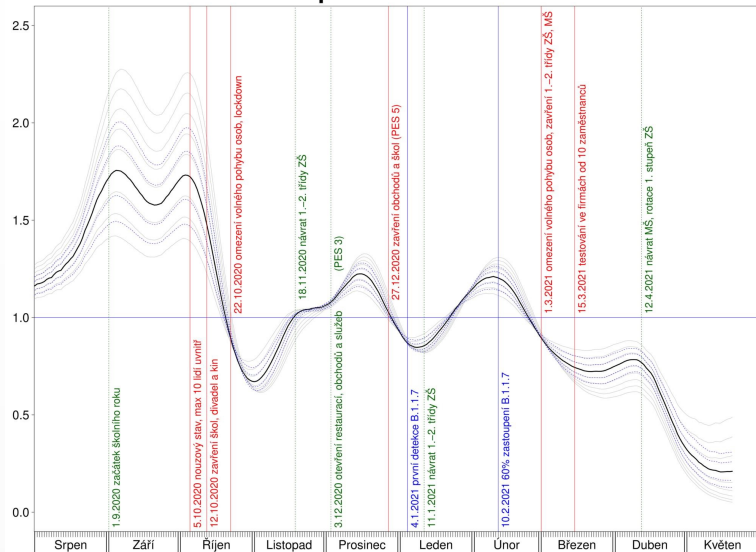
Odhad efektivního reprodukčního čísla fatálních infekcí



Viditelnější změny v tempu růstu/poklesu $R_t \equiv$ možná efektivní „opatření“

- Opatření přijatá po krajských volbách (říjen 2020)
 - změna pozvolného nárůstu R_t ve strmý pád
 - které/která opatření to mohla způsobit, když jich bylo v rychlém sledu přijato velké množství... ?
- Rozvolnění spojená s PES = 3 (prosinec 2020)
 - jisté zrychlení růstu R_t (které rostlo už před tím)
 - které/která rozvolnění to mohla způsobit... ?

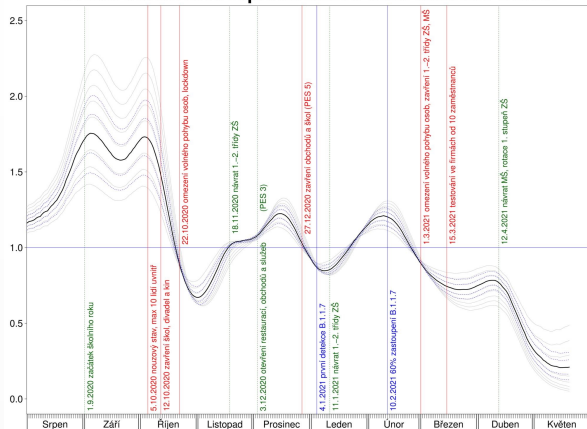
Odhad efektivního reprodukčního čísla fatálních infekcí



Zahájení/obnovení (alespoň částečné) školní docházky (září i listopad 2020): zpomalení růstu R_t

- Opravdu je tak důležité řešit přenos infekce **v rámci škol** (viz např. „výzkumná zpráva“ BISOP použitá MŠMT pro nastavení režimu provozu škol (rotace atp.) při jejich znovuotevření v květnu 2021)?
- Není podstatnější vědět, jaký vliv má ten který režim (ne)provozu škol **na pronikání nákazy směrem do rizikové populace?**
- Zavřená škola \implies více mrtvých (nikoliv dětí)...

Odhad efektivního reprodukčního čísla fatálních infekcí



Po většinu doby si epidemie žila vlastním životem...

- Dno či vrchol vlny či vlnky bez toho, že by se v okolí 14 i více dnů cokoliv měnilo na „opatření“...

Po většinu doby si epidemie žila vlastním životem...

- Woodovo srovnání vývoje epidemie ve Velké Británii a Švédsku, dvou zemích s diametrálně odlišným přístupem

Furthermore Sweden, which did not lockdown, saw declining infections only a day or two after the UK (Swedish GDP dropped about 3 % in 2020 against the UK's 10 % drop).

- Další (reference ve Woodově článku a jiné práce unikající „expertům“ Ministerstva zdravotnictví ČR)

V.

Pár slov na závěr

Umí nejenom

- sypat data do počítače
 - sypat data do „modelů“
 - vytvářet tabulky s čísílkami a barevné obrázky
- **popisná** statistika konkrétní sady čísílek svádějící k **dezinterpretacím a nepodloženým zobecněním** na populaci

Umí nejenom

- sypat data do počítače
 - sypat data do „modelů“
 - vytvářet tabulky s čísílkami a barevné obrázky
- **popisná** statistika konkrétní sady čísílek svádějící k **dezinterpretacím a nepodloženým zobecněním** na populaci

Ale hlavně má

- schopnost interpretovat výsledky s ohledem na **relevanci použitých metod** při vědomí faktu, jaká **data** byla použita a jak vznikla
- schopnost navrhnout **jaká data** a **jakým způsobem** je potřeba sbírat, abychom se dozvěděli, co se dozvědět chceme

JE TO MARNÝ...

JE TO MARNÝ...

JE TO MARNÝ...