

Odvození indexu rizika pro epidemii COVID-19 v České republice

Jan Kulveit * † Tomáš Gavenčiak *

Obsah

1	Úvod	2
2	Výpočet indexu rizika	2
2.1	Vstupní proměnné	3
2.2	Odvozené a další proměnné	3
2.2.1	Konstanty	3
2.3	Odhady počtu vážných případů	4
2.4	Index rizika	4
2.5	Navržené komponenty indexu rizika	4
2.6	Konkrétní bodové tabulky	5
3	Ascertainment rate	6

Tento dokument popisuje motivaci a teoretické odvození indexu rizika pro epidemii COVID-19 v České republice. Tento dokument má status technické zprávy skupiny EpidemicForecasting.org přibližující použité modely; nemá za cíl být kompletní vědeckou publikací a neprošel recenzním řízením. Autoritativní zdroj pro metodiku výpočtu indexu rizika (rizikového skóre) pro epidemii COVID-19 v ČR je metodicky popis vydaný ÚZIS ČR.

Na vývoji indexu se dále podíleli Jakub Šmíd, Ondřej Májek, Jan Mužík, Jiří Jarkovský, Ladislav Dušek, Luděk Berec, René Levínský, Daniel Prokop, Lenka Příbylová, Pavel Řehák, Jakub Steiner a Josef Šlerka.

*EpidemicForecasting.org project

†Future of Humanity Institute, University of Oxford

1 Úvod

Hlavním cílem indexu rizika je vystihnout jedním číslem nebezpečnost dané epidemické situace způsobem, který je srozumitelný pro širokou veřejnost. Index je proto založen na několika epidemiologicky relevantních a zároveň dobře pochopitelných kritériích.

Čtyři sledované hodnoty jsou počet pozitivních testů za posledních 14 dní (incidence), počet pozitivních testů za posledních 14 dní v populaci 65+, odhad růstového čísla R (z incidence posledních 12 dní „německou“ metodou) a pozitivita testů za poslední týden.

Index je možné spočítat i zpětně, je možné jej spočítat z veřejně dostupných dat a má prediktivní charakter – odhaduje budoucí vývoj místo hodnocení minulé či současné situace. Index je možné kromě ČR spočítat i pro další větší územní celky, jako jsou kraje a další země, a výsledné indexy rizika jsou srovnatelné. I když je možné spočítat index i pro menší celky než kraje, výsledky jsou obecně nestabilní a nemají velkou vypovídající hodnotu.

Použitá konstrukce indexu jakožto součtu čtyř faktorů má přitom hlubší teoretický význam – index se odvíjí primárně od odhadovaného počtu vážných případů za určité budoucí období délky D dní, pokud by nedošlo k žádné změně epidemiologických podmínek (například úpravou protiepidemických opatření). Index počítá s výhledem $D = 30$ (cca měsíc).

Index rizika RI je logaritmický ukazatel hodnoty Z – predikovaný počet vážných případů za nějaké budoucí období.

$$RI = C_0 + C_1 \log_2 Z,$$

kde C_0 a C_1 jsou libovolně zvolené konstanty škálující hodnotový rozsah indexu (viz níže).

Odhadované budoucí období D by nemělo být příliš dlouhé kvůli nepředvídatelnosti vývoje pandemie i vnějších okolností (sezónnost etc.) i kvůli aproximačnímu charakteru indexu, ale má být dost dlouhé, aby mimo jiné obsáhlo dobu potřebnou k realizaci a projevu intervencí i dostatečně zohlednilo růst či ústup epidemie. Vyšší hodnoty D dávají v indexu větší váhu rychlosti růstu epidemie (a naopak).

Index závisí na incidencích dvou subpopulací – celé populace a populace 65+. Toto činí index robustnější vůči různým scénářům a neznámým, především různému rozložení epidemie v populaci a různé efektivitě testování v subpopulacích.

2 Výpočet indexu rizika

Tato sekce stručně popisuje odvození a výpočet indexu od vstupních dat přes vnitřní proměnné indexu až k jeho konkrétní hodnotě a její aproximaci bodovými tabulkami.

$X[d]$ označuje hodnotu proměnné X v den d . Index pracuje pouze s denní granularitou dat.

2.1 Vstupní proměnné

$P[d]$	Incidence za posledních 14 dní (součet, na 100k populace)
$P_{65+}[d]$	Incidence 65+ let za posledních 14 dní (součet, na 100k populace 65+)
$T[d]$	Průměrná pozitivita testů za posledních 7 dní (škála 0—1)
$U[d]$	Podíl sedmidenních průměrů incidence posunutých o 5 dní. Odhad hodnoty R podle an der Heiden and Hamouda [2020].

$$U[d] = \text{avg}_{i=d-6..d}(P[i]) / \text{avg}_{i=d-11..d-5}(P[i])$$

2.2 Odvozené a další proměnné

Tyto hodnoty jsou později použity při výpočtu indexu rizika níže.

$g(U) = U^{1/5}$	Denní relativní přírůstek (násobek) nově nakažených
$G(D, U)$	Násobitel sumy nakažených v následujících D dnech jako násobek nakažených v počáteční den (pro $D < 0$ v D předchozích dnech)

$$G(D, U) = (1 - g(U[d])^D) / (1 - g(U[d])) \text{signum}(D)$$

$A(T)$	Ascertainment rate (podíl odhalených z nakažených) podle test positivity (viz níže)
$I[d]$	Očekávání nově nakažení v den d (na 100k populace)
$I^{(D)}[d]$	Očekávání nově nakažení za následujících D dní (na 100k populace)
$I_{65+}[d]$	Očekávání nově nakažení 65+ v den d (na 100k populace 65+)
$I_{65+}^{(D)}[d]$	Očekávání nově nakažení 65+ za následujících D dní (na 100k populace 65+)

$$I[d] = P[d] / A(PT[d]) / G(-14, U[d])$$

$$I^{(D)}[d] = I[d] G(D, U[d])$$

2.2.1 Konstanty

Na konkrétních hodnotách těchto konstant závisí pouze posun, resp. škálování výsledného indexu. U podílu mrtvých nad 65 let je podstatné primárně to, že ve většině scénářů jde o zásadní část očekávaných vážných případů, a tedy že je robustní počítat z nich odhad celkového počtu.

$S = \text{const.}$	Podíl vážných případů z nakažených.
$S_{65+} = \text{const.}$	Podíl vážných případů v subpopulaci 65+ let.
$Z_{65+}[\text{total}]/Z[\text{total}] > 0.9$	Odhadovaný celkový podíl vážných případů nad 65 let za celou pandemii (odhad ze smrtnosti věk. skupin).
$N = 10600000$	Populace.
$N_{65+} = 2000000$	Populace nad 65 let.

2.3 Odhady počtu vážných případů

Odhad vážných případů počítáme z geometrického průměru: A) odhadu provedeného pro celou populaci a B) odhadu provedeného pro populaci od 65 let (65+). V odhadu pro subpopulaci 65+ používáme pozitivitu testů i odhad rychlosti růstu z celé populace kvůli jednoduchosti indexu. Index by bylo možno rozšířit i o další populace (e.g. 75+), ale zvýšená komplexita nestojí za už jen mírné zvýšení robustnosti. Hodnoty Z_{\bullet} jsou počítané na 100 000 jedinců celé populace.

Odhad z nakažených celkem (Z_A).

$$Z_A[d] = S I^{(D)}[d]N/10^5$$

Odhad z nakažených 65+ let (Z_B).

$$Z_B[d] = S_{65+} I_{65+}^{(D)}[d] \frac{N_{65+}/10^5}{Z_{65+}[\text{total}]/Z[\text{total}]}$$

2.4 Index rizika

Hodnota Z pro index rizika je odvozena z geometrického průměru Z_A a Z_B .

$$Z = \sqrt{Z_A Z_B}$$

Samotný index rizika potom expandujeme do formy níže. Aditivní a multiplikativní konstanty jsou sdruženy do C_0 a C_1 , konstantu C_0 dále nevyčísľujeme. Index rozdělíme do 4 komponent $RI_{\bullet}[d]$, které jsou popsány níže.

$$\begin{aligned} RI[d] &= C_1(0.5 \log_2 P[d] + 0.5 \log_2 P_{65+}[d] + \\ &\quad + \log_2 (1/A(T[d])) + \log_2 (G(D, U[d])/G(-14, U[d]))) + C_0 = \quad (1) \\ &= RI_P[d] + RI_{P_{65+}}[d] + RI_A[d] + RI_U[d] \end{aligned}$$

2.5 Navržené komponenty indexu rizika

Níže navržené hodnoty počítají s $C_1 = 6$, $D = 30$. Konkrétní hodnotu C_0 nepočítáme. Pro přehlednost indexu je počet hodnotových kategorií (binů) omezen na 7–8, ale je smysluplné rozšiřovat jej na obě strany, zjemňovat etc.

Týdenní růstový trend. Pro některé ze složek indexu se započte bodový bonus, pokud sledovaná hodnota mezitýdenně rostla. Konkrétně pokud je hodnota před 7 dny menší, než současná hodnota. Růst $P[d]$ není bodován protože růst $P[d]$ je již přímo zahrnut v U . Růst U není bodován pro zvýšení stability (rychle se mění oproti ostatním ukazatelům a v modelových simulacích regulace epidemie přispívá k oscilacím). Vyšší stabilitu tohoto ukazatele zajišťuje to, že každý z parametrů již je spočten z týdenních či delších průměrů.

2.6 Konkrétní bodové tabulky

Skóre i hraniční hodnoty jsou vhodně konzistentně zaokrouhleny pro větší přehlednost a snadnost použití indexu, změna zaokrouhlováním je pod zamýšlenou přesností modelu. Skóre zhruba odpovídá RI_{\bullet} levého konce intervalu stupně.

V případě, že by se lišily, mají přednost hodnoty a metodika z publikace „Metodika pro výpočet rizikového skóre COVID-19“ od Kulveit et al. [2020]

$$RI_P[d] = 3 \log_2 P[d] - 10$$

P	< 10	$10 - 25$	$25 - 50$	$50 - 120$	$120 - 240$	$240 - 480$	$480 - 960$	> 960
RI_P	0	2	4	7	10	13	16	20

$$RI_{P_{65+}}[d] = 3 \log_2 P_{65+}[d] - 10$$

P_{65+}	< 10	$10 - 25$	$25 - 50$	$50 - 120$	$120 - 240$	$240 - 480$	$480 - 960$	> 960
$RI_{P_{65+}}$	0	2	4	7	10	13	16	20

Při mezitýdenním růstu P_{65+} (pokud $P_{65+}[d] > P_{65+}[d - 7]$): **+2 body**

$$RI_U[d] = 6 \log_2((1 - U[d]^{D/5})/(U[d]^{-14/5} - 1)) + 2$$

U	< 0.8	$0.8 - 1$	$1 - 1.2$	$1.2 - 1.4$	$1.4 - 1.6$	$1.6 - 1.9$	> 1.9
RI_U	0	5	10	15	20	25	30

$$RI_A[d] = -6 \log_2 A(T[d]) = 84 \log_2(T[d] + 1)$$

T	< 0.03	$0.03 - 0.07$	$0.07 - 0.11$	$0.11 - 0.15$	$0.15 - 0.19$	$0.19 - 0.23$	$0.23 - 0.26$	> 0.26
RI_A	0	3	7	11	15	20	25	30

Při mezitýdenním růstu T (pokud $T[d] > T[d - 7]$): **+2 body**

Součet maxim bez růstových bonusů je 100 bodů, avšak toto je arbitrárně zvolená hodnota i limit – index nemá význam procent ani jiný podobný. Jediná interpretace RI je, že 6-bodový nárůst je cca $2\times$ větší počet vážných případů za následujících $D = 30$ dní. Součet s růstovými bonusy je 104 bodů, ale maximální hodnota RI je vždy do 100 bodů, součty nad 100 se interpretují jako 100.

Poznámky pro možné úpravy indexu. Je možné posouvat každý ukazatel zvlášť přičtením konstanty k bodům a škálováním všech bodových ukazatelů najednou násobením konstantou. Není korektní škálovat ukazatele násobením konstantou jednotlivě – to odpovídá (nežádoucí) změně váhy ukazatele v indexu.

3 Ascertainment rate

Ascertainment rate (A , podíl odhalených ze všech nakažených) hrubě odhadujeme jako funkci pozitivitu testů T . Toto je sice jen hrubé přiblížení, ale umožňuje zohlednit dva faktory:

1. Vyšší T implikuje nižší očekávané A . Toto platí za předpokladů že a) osoby s vyšší pravděpodobností nákazy se obecně testují přednostně před těmi s nižší pravděpodobností a b) způsob testování se nebude měnit zásadně a skokově.
2. Vyšší T znamená i vyšší nejistotu situace a tedy i rozptyl A . Např. za předpokladu u předchozího bodu je minimální hodnota T skutečná prevalence a dosahuje se jen při $A = 1$ (testování celé populace). Nízké hodnoty A lze naopak dosáhnout více různými scénáři s velmi různými T .

Power-law model. Zvolený model pro A je power-law

$$A(T) = (1 + T)^{-C_A}$$

s parametrizací $C_A = 14$. Power-law model s exponentem $C_{A_0} = 8.5$ by odpovídal přibližnému fitu na český vývoj z podzimu 2020 (v rozsahu $T < 25\%$). Exponent $C_A = 14.0 \approx 8.5 \cdot 1.65$ zachycuje faktor nejistoty při vyšších hodnotách T (viz níže).

Pozn.: Krom teoretického modelu jsou ale zásadním kritériem volby $A(T)$ i expertní odhad přiměřenosti vlivu pozitivitu testů na vážnost situace a jednoduše komunikovatelná forma.

U vztahu A a T panuje poměrně velká nejistota co do modelu i jeho parametrizace. Nejsme si vědomi publikovaného modelu vztahu A a T , existující literatura ascertainment rate odhaduje pouze v konkrétních situacích. Naproti tomu, modelování zvýšené nejistoty ze zvýšeného T multiplikativním faktorem u Z (tedy aditivním faktorem u RI) odpovídá standardnímu modelu rozptylu náhodné proměnné a optimalizaci na interval nejistoty, zmíněný níže.

Model zvýšené nejistoty. Druhou roli RI_A – nejistotu v A – lze formálně modelovat tak, že v indexu rizika nesledujeme přímo očekávanou hodnotu $A = A(T)$, ale předpokládáme, že $A(T) \sim \mathbf{A}(T)$, kde $\mathbf{A}(T)$ je náhodná proměnná se střední hodnotou skutečné $A_0(T)$ a relativní směrodatnou odchylkou rostoucí s T . Poté v RI uvažujeme $A(T)$ jako sledující zvolený percentil q distribuce $\mathbf{A}(T)$ místo střední hodnoty jakožto horní mez konfidenčního intervalu RI.

Reference

- M. an der Heiden and O. Hamouda. Schätzung der aktuellen Entwicklung der SARS-CoV-2-Epidemie in Deutschland – Nowcasting. *Epidemiologisches Bulletin*, 2020(17): 10–15, 2020. doi: <http://dx.doi.org/10.25646/6692.4>.
- J. Kulveit, T. Gavenčiak, J. Šmíd, L. Berec, R. Levínský, D. Prokop, L. Příbylová, P. Řehák, J. Steiner, J. Šlerka, O. Májek, J. Mužík, and J. Jarkovský. Metodika pro výpočet rizikového skóre COVID-19. Technical report, Ústav zdravotnických informací a statistiky ČR, 2020.