

Křišťálová koule ředitele Duška

Odborné stanovisko k predikcím vývoje počtu nakažených, kterou prezentoval 11. prosince 2020 ředitel Ústavu zdravotnických informací a statistiky Ladislav Dušek.

Ondřej Vencálek: Přírodovědecká fakulta, Univerzita Palackého, Olomouc

Tomáš Fürst: Přírodovědecká fakulta, Univerzita Palackého, Olomouc

Arnošt Komárek: Matematicko-fyzikální fakulta, Univerzita Karlova, Praha

Michal Kulich: Matematicko-fyzikální fakulta, Univerzita Karlova, Praha

Stanislav Katina: Přírodovědecká fakulta, Masarykova Univerzita, Brno

Gejza Dohnal: Fakulta strojní, České vysoké učení technické, Praha

Média nedávno věnovala velkou pozornost predikci vývoje počtu nakažených, kterou prezentoval ředitel Ústavu zdravotnických informací a statistiky ČR (ÚZIS) Ladislav Dušek. Dozvěděli jsme se, že „v lednu hrozí exploze COVID“ (1).

Postupně ukážeme, že

- 1) ÚZIS opakuje chybu Pavla Řeháka z března (2) a Jaroslava Flégra z října tohoto roku (3), když *predikuje pomocí exponenciálního růstu*. Připomeneme, že obě tyto predikce naprosto selhaly a řekneme také, že důvodem tohoto selhání nemusely být restriktivní kroky vlády.
- 2) Horizont predikce zveřejněné ÚZIS je volen tak, aby hodnoty predikce vzbudily pozornost, ale přitom nebyly zjevně nerealistické.
- 3) ÚZIS sice deklaruje, že k tzv. krátkodobým predikcím využívá SIR model, ve skutečnosti však jde o model exponenciálního růstu.
- 4) Predikce vývoje epidemie sestavené ÚZIS jsou bezcenné, neboť zahrnují všechny možnosti budoucího vývoje.
- 5) Model používaný ÚZIS vytváří nebezpečnou iluzi, že pro změnu dynamiky šíření epidemie je nutný zásah vlády.

Predikce pomocí exponenciálního růstu

K predikci, kterou představil L. Dušek (4), netřeba žádných dat s výjimkou aktuálního denního počtu nových případů. Pro 9. prosinec (poslední z čísel v grafech prezentovaných L. Duškem) šlo o číslo 4465. Predikce vznikne tak, že počítáme s exponenciálním růstem s koeficientem $R=1,2$, který je však „rozložen“ do 4 dnů. To odpovídá dennímu nárůstu o necelých 4,7 % (technicky vzato spočteme čtvrtou odmocninu z čísla 1,2, která se rovná přibližně 1,047).

Denní počty nových případů tedy „predikujeme“ tak, že postupně násobíme $4465 \cdot 1,047$, pak toto číslo opět vynásobíme 1,047 atd. Veřejnost je už poučena, že tento typ růstu se nazývá exponenciální. Stačí jednoduchá tabulka třeba v Excelu a zjistíme, že na konci roku bude počet nových případů za den kolem 12 000 a v posledním týdnu ledna se dostaneme přes 40 000 nových případů za den. Celkový počet nových případů za leden pak vychází přes 800 000, tedy téměř přesně stejně jako predikce ÚZIS. Výpočet přiřkládáme v tabulce Excelu.

Volba horizontu predikce

Predikce učiněné v březnu P. Řehákem, v říjnu Jaroslavem Flégrem a nyní pracovníky ÚZIS mají jedno společné – *časový horizont predikce je volen tak, aby hodnoty predikce vzbudily pozornost, ale přitom nebyly zjevně nerealistické*. Nakolik jde o záměr, lze pouze spekulovat.

Řehákova predikční tabulka počítala s koeficientem růstu $R=1,45$ (odhadnutým na datech od 1. do 8. března) a predikovala počty nových denních případů na období 9. března – 1. dubna. Pro 1. dubna bylo přitom predikováno přes 230 000 nových případů za den. Kdyby byla prezentována predikce v delším časovém horizontu, viděli bychom, že již kolem 11. dubna 2020 mělo dle tohoto modelu být *denně* nakaženo přes 10 miliónů Čechů. Taková predikce by však byla očividně nesmyslná. Naše babičky říkávaly, že žádné stromy nerostou do nebe. A skutečně, žádný exponenciální růst v přírodě netrvá dlouho. Každý model založený na exponenciále tedy dříve či později ztratí platnost. Jediná otázka je, kdy.

Predikce ÚZIS počítá s podstatně pomalejším růstem než Řehákův model (koeficient růstu $R=1,047$) a proto si „může dovolit“ delší časový horizont, třeba až 2 měsíce, než začnou být predikce nesmyslné. 40 000 nových případů denně na konci ledna je číslo děsivé, ale pro naši mysl uvyklou číslům kolem 10 000 denně z října tohoto roku jde o predikci představitelnou. Pokud by ovšem predikce pokračovala ($R=1,2$ by zůstalo), byli bychom svědky stotisícových denních přírůstků již v únoru.

Špatný model (nebo jen špatná dokumentace?)

Dokumentace k epidemiologickému modelu, který ÚZIS používá pro krátkodobé predikce, ale také pro odhad tzv. reprodukčního čísla (5), budí řadu pochybností. Model je popisován jako SIR model. Skutečnost je ovšem jiná. SIR modely (ač velmi zjednodušující) jsou alespoň natolik realistické, že respektují zachování celkového počtu osob v modelu. Každý člověk je v dané chvíli buďto náchylný (susceptible) nebo nakažený (infectious) nebo imunní/mrtvý (removed) a model se snaží zachytit rychlost přechodu mezi těmito třemi „kompartmenty“. Konstantní součet velikostí těchto tří kompartmentů zaručuje, že šíření nemoci má exponenciální charakter pouze po velmi krátkou dobu, poté růst zpomalí a následuje pokles. Model používaný ÚZIS však nesplňuje ani tento triviální požadavek konstantního součtu, a proto *může v desetimilionové zemi klidně predikovat i dvacet milionů nakažených*.

Modelu lze vytknout i mnoho dalších vad. Například přepočítávání jeho parametrů v blíže nezjistitelných časových bodech. V původní dokumentaci (5) jsou časy přepočtů uvedeny, z aktualizované verze (6) již časy přepočtů nezjistíme. Občasné přepočtení tzv. reprodukčního čísla (které se mimochodem v modelu skokově mění) zajistí, že model velmi dobře popisuje historická data, ovšem k predikci se historická data nijak nevyužívají, neboť o budoucí hodnotě klíčového parametru (reprodukčním číslem) model neříká vůbec nic, a proto nezbyvá než vytvářet tzv. scénáře. Opět se tak potvrzuje slavné rčení Yogiho Berra, že predikovat je těžké, zejména pokud jde o budoucnost.

Bezenné predikce

Skoro jako kdyby si nesmyslnosti exponenciálního růstu byl L. Dušek vědom, prezentoval i další scénáře, které odpovídají hodnotě reprodukčního čísla $R=1$ (počty by se ustáleně pohybovaly kolem současných hodnot) a hodnotě $R<1$ (v tom případě by počty denních případů klesaly). Kromě vágní informace, že scénáře jsou různě pravděpodobné, máme tedy tři predikce, které říkají „počty mohou dramaticky růst, ale taky mohou zůstat stejné, či klesat“. *Taková predikce ve stylu „dva národy budou válčit, ale jen jeden zvítězí“ nemůže nevyjít*. Je ovšem stejně (bez)cenná jako věštby delfské Pýthie.

Co říká model používaný ÚZIS o změně dynamiky šíření epidemie

Je důležité si uvědomit, že zastavení růstu počtu nově infikovaných – jak v SIR modelu, tak typicky i ve skutečnosti – je důsledkem konečnosti populace, a nikoliv důsledkem nějakých „vládních opatření“. Model používaný ÚZIS – protože konečnost populace nerespektuje – nemůže nikdy „spontánně“ přejít z fáze exponenciálního růstu do fáze poklesu. Takový přechod v tomto modelu vyžaduje změnu parametrů zvnějšku. Vytváří tím velmi *nebezpečnou iluzi, že pro změnu dynamiky šíření infekce je nutný zásah vlády*. To rozhodně není pravda, jakkoliv z toho samozřejmě neplyne, že by zásahy vlády dynamiku nemoci neovlivňovaly.

Závěr

Slavný citát George Boxe říká, že všechny modely jsou chybné, některé jsou však užitečné. Jakožto lidé, kteří se práci s daty a matematickým modelováním reálných procesů mnoho let profesionálně zabývají, konstatujeme, že „model“, který ÚZIS používá, je nejen chybný, ale i matoucí a potenciálně nebezpečný. Model je:

1. Zcela nerealistický, neboť předpokládá, že populace naší vlasti je nekonečná.
2. Bezcestný, neboť je používán k tautologickým „predikcím“, že počet nově nakažených bude buďto klesat, nebo stagnovat nebo růst.
3. A navíc nebezpečný, neboť vytváří falešný dojem, že ke změně dynamiky vývoje nemoci je třeba vládních zásahů.

Je načase, aby odborná veřejnost tuto nesmyslnou a nebezpečnou hru s čísly důrazně odmítla. Od ÚZIS očekáváme především odpovědný sběr *relevantních* dat a jejich včasné zpřístupnění, samozřejmě při zachování ochrany osobních údajů.

Reference

1. Pergler, Tomáš. V lednu hrozí exploze covidu-19, varují propočty statistiků. [Online] <https://www.seznamzpravy.cz/clanek/v-lednu-hrozi-exploze-covidu-varuji-propocety-statistiku-133376>.
2. Tajemný muž, který na jaře „zachránil Česko“? Exředitel České pojišťovny. [Online] <https://www.seznamzpravy.cz/clanek/tajemny-muz-ktery-na-jare-zachranil-cesko-exreditel-ceske-pojistovny-124767>.
3. [Online] <https://video.aktualne.cz/dvtv/vse-zametal-pod-koberec-zivot-se-zastavi-vubec-se-to-nemuse/r~a857cc2e13d711eba7deac1f6b220ee8/>.
4. Dušek, Ladislav. Vývoj epidemie COVID-19 v ČR Souhrn hlavních statistických ukazatelů Aktuální epidemiologická situace a pravděpodobný vývoj. [Online] <https://www.uzis.cz/res/file/covid/20201211-dusek.pdf>.
5. Ondřej Májek, Ondřej Ngo, Jiří Jarkovský, Monika Ambrožová, Barbora Budíková, Jan Kouřil, Ladislav Dušek, Tomáš Pavlík. Dokumentace k epidemiologickému modelu ÚZIS ČR pro krátkodobé predikce. [Online] 8. 6 2020. <https://onemocneni-aktualne.mzcr.cz/doc/dokumentace-modely.pdf>.
6. Ondřej Májek, Ondřej Ngo, Renata Chloupková, Jiří Jarkovský, Monika Ambrožová, Barbora Budíková, Jan Kouřil, Tomáš Pavlík, Ladislav Dušek. Dokumentace k epidemiologickému modelu ÚZIS ČR pro krátkodobé predikce. *Metodická dokumentace prediktivních modelů*. [Online] 18. 10 2020. <https://onemocneni-aktualne.mzcr.cz/covid-19/zpravy-a-metodiky>.