

Komentář k výzkumné zprávě BISOP „Model šíření COVIDu-19 na základních školách“

Arnošt Komárek

Matematicko-fyzikální fakulta Univerzity Karlovy, Katedra pravděpodobnosti a matematické statistiky

Stanislav Katina

Přírodovědecká fakulta Masarykovy Univerzity, Ústavu matematiky a statistiky

Na webu [irozhlas.cz](https://www.irozhlas.cz) se 5. dubna objevil rozhovor s členem poradního orgánu MeSES Reném Levínským, ve kterém se s odkazem na výzkumnou zprávu BISOP (<https://www.bisop.eu/vyzkumna-zpravabisop-vytvoril-model-sireni-covidu-19-na-zakladnich-skolach/>), publikovanou 25. února, tvrdí, že „rotace ve školách sníží nákazu až o 90 procent“. Vzhledem k tomu, že zmíněná výzkumná zpráva BISOP byla dle všeho použita jako jeden z významných podkladů při zvažování o režimu otevírání škol, považujeme za nutné upozornit, že prakticky všechny závěry ohledně šíření nákazy virem SARS-CoV-2 ve školách učiněné v předmětné zprávě BISOP lze považovat za domněnky či hypotézy, které však **nelze podložit** žádnými daty, resp. jejich analýzou při respektování i těch nejzákladnějších principů statistické inference.

Zpráva BISOP se opírá o data získaná z dotazníkového šetření administrovaného na dvou vybraných školách (jedné vybrané základní a jedné vybrané střední škole). Cílem dotazníkového šetření bylo postihnout nejruznější typy kontaktů, ke kterým ve školském prostředí může docházet. Výsledkem šetření se stal model školních vazeb ve formě tzv. víceúrovňové sítě (*multilevel network*). Až potud se jedná o nepochybně zajímavou sociologickou studii, která zadavateli sdělí celou řadu poznatků o sociálních vazbách ve vybraných školách, případně ve školách s obdobnými charakteristikami.

Autoři studie však jdou dál. Jednotlivé vazby rozdělí do šesti kategorií dle „rizikivosti“ nákazy, které jsou nějakým způsobem kvantifikovány – z textu zprávy není úplně jasné jakým způsobem. Autoři zprávy hovoří o intenzitě vazeb, jež odpovídají délce a blízkosti kontaktů. Lze konstatovat, že o těchto aspektech mohou mít autoři studie jistou představu ze získaných dat. BISOP pokračuje: „*Průběh šíření epidemie ve škole je simulován prostřednictvím multi-agentního modelu na výše zmíněných grafech sociálních interakcí*“. Uvažovány jsou čtyři stupně epidemie (čtyři možné hodnoty denní incidence), kdy (patrně) uvažovaná denní incidence ovlivní „rizikovost“ toho kterého typu kontaktu (což je jistě pravda). Dále je uvažována řada opatření (všechno zavřeno, týdenní rotace atd.), kdy opět každé opatření skrze změnu intenzity ovlivňuje „rizikovost“ jednotlivých typů kontaktů (což je též jistě pravda). Simulace (Monte Carlo) studie se následně snaží ukazovat, jak se bude relativně lišit počet nakažených ve škole při jednotlivých opatřeních v porovnání se školou v obvyklém režimu (bez kroužků). Ze zprávy není úplně jasné, zda v rámci modelu mohou rizikovost ovlivnit i další faktory.

V diskuzi výzkumné zprávy BISOP sami autoři uvádějí: „*U všech výše uvedených výsledků je třeba mít na paměti několik jejich omezení. Prvním je to, že se v žádném smyslu nejedná o předpověď budoucího stavu škol či dokonce celého školství. Výsledky pouze ukazují, jaký je za daných podmínek a předpokladů pravděpodobný dopad jednotlivých opatření.*“ Spojení „**za daných podmínek a předpokladů**“ je přitom zcela klíčové. Při použití výsledků provedené studie k rozhodování by autoři měli též sdělit, jak moc reálné jsou právě ony dané podmínky a předpoklady, resp. zda je lze podložit nějakými reálnými daty.

Pro *realistické* modelování šíření epidemie je totiž zapotřebí též vyhodnotit rizikovost kontaktů vzhledem k rozdělení „nakažlivosti“ té které části populace. Zásadní roli zde nepochybně hrají věk a také stupeň imunizace příslušných částí populace. Například co se věku týká, současná odborná literatura, např. časopis *Science* (<https://science.sciencemag.org/content/371/6526/eabe2424>), se spíše přiklání k tezi, že děti do věku cca 10–12 let virus SARS-CoV-2 roznášejí v mnohem menší míře než starší děti a dospělí a většina menších dětí se spíše nakazí od dospělých než naopak. Otázkou v každém případě zůstává, jak tento aspekt kvantifikovat. BISOP k to-

muto nemá data žádná, v textu výzkumné zprávy se o tomto problému nepíše vůbec, domníváme se tedy, že byl ignorován a „rizikovitost“ jednotlivých typů kontaktů byla vyhodnocena čistě na základě intenzity kontaktů (ve spojení s uvažovanou denní incidencí nákazy). Alespoň částečnou odpověď na otázku, jaké je rozdělení a vývoj „nakažlivosti“ (incidence) v různých částech populace (včetně žáků) by přitom přinesla relativně jednoduchá longitudinální studie, na jejíž provedení bylo v uplynulých měsících času více než dost.

Taktéž nám není jasné, jakým způsobem byla vzata do úvahy míra imunizace studované populace, resp. jejich komponent (žáci, učitelé, uklízečky, školníci, družinářky, kuchařky apod.). Odhad stupně imunizace zaměstnanců škol na základě očkování jistě provést lze, o imunizaci žáků však nevíme nic. Ani BISOP, ani kdokoliv jiný v České republice neprovedl systematickou seroprevalenční studii.

Kromě výše zmíněných dvou aspektů jistě hrají roli i mnohé další faktory, jejichž vliv je tiše opomíjen. Kvantifikace „rizikovitosti“ jednotlivých vazeb je zcela zásadní pro validitu (a použitelnost) získaných výsledků a právě tento aspekt celého postupu je z velké části založen pouze na domněnkách autorů a nikoliv na existujících datech. I když se ve zprávě píše: „*Modelujeme tak nejen průměrný vliv jednotlivých opatření, ale i jejich rozptyl tak, abychom dobře popsali míru nejistoty modelu.*“, není pravdou, že by autoři svým postupem mohli jakkoliv popsat míru nejistoty svých výsledků. Toto mohou učinit pouze v rámci použitého scénáře (použité „nastavení“ modelu). BISOP nicméně **žádným způsobem nekvantifikuje nejistotu plynoucí z obecně neznámé rizikovitosti toho či onoho kontaktu**. S přihlédnutím k základním principům statistické inference tedy nelze v žádném případě doporučit, aby se jakékoliv závěry ze zprávy zobecňovaly a aby se jakékoliv rozhodnutí opíralo o výsledky ve zprávě uvedené.

Lze též podotknout, že je v podstatě jedno, jak moc použitelné jsou závěry předmětné studie BISOP. Opravdu nás zajímá šíření viru SARS-CoV-2 v rámci školy (a vesměs zanedbatelně rizikové populace)? **Není totiž důležitější vědět, jaký vliv má ten který režim otevřené, resp. zavřené školy na šíření tohoto viru směrem do rizikové populace, zejména prarodičů dětí?** Na tuto otázku studie BISOP neodpovídá vůbec a je otázkou, zda odpovědné orgány mají o tomto vůbec nějakou představu.

Musíme bohužel znovu zopakovat, že i nadále selháváme v procesu poznání založeném na datech. Ještě před tím, než dochází ke statistickému či matematickému modelování, je zapotřebí pro zadaný problém (*V jakém režimu mají být otevřeny školy?*) **položit správnou otázku** a následně **získat relevantní a správná data**. Pokud taková data nejsou k dispozici, je jakákoliv analýza pouze dalším vršením domněnek a hypotéz, které by měly zůstat uzavřeny v intencích čistě odborných debat a nikoliv být předkládány jako podklad k rozhodování o tak zásadní věci, jakým je zajištění prezenční školní docházky.

Arnošt Komárek je docentem statistiky na MFF UK, pravděpodobnost a statistiku tamtéž vystudoval (Mgr.) v roce 2000. Následně (2001) získal titul MSc. v oboru biostatistika na *Universiteit Hasselt* v Belgii a Ph.D. (2006) na *Leuven Biostatistics and Statistical Bioinformatics Centre Katholieke Universiteit Leuven* v Belgii. Autor je bývalým předsedou mezinárodní *Statistical Modelling Society* a editorem odborného časopisu *Statistical Modelling* (nadmediánový IF v sekci *Statistics and Probability*), publikoval 19 metodologických a 36 aplikačních odborných článků v časopisech s IF, jež byly více jak 1500-krát citovány a dosáhl h-indexu 20.

Stanislav Katina je docentem statistiky na PřF MU, Brno, zástupcem ředitele pro aplikovanou matematiku a statistiku, vedoucím týmu Modelování a statistika, *honorary research fellow* na *University of Glasgow*, předsedou komise *Statistika pro regulační záležitosti* v *Mezinárodní společnosti pro klinickou biostatistiku* a bývalým členem výkonného výboru téže společnosti. Vystudoval matematiku, biologii, pravděpodobnost a matematickou statistiku na PriF a FMFI UK v Bratislavě (Mgr.) v roce 1999, získal titul PaedDr. v oboru biologie vzdělávání (2000) na PriF UK, RNDr. (2001) a Ph.D. (2005) v oboru pravděpodobnost a matematická statistika na FMFI UK v Bratislavě, deset let působil na *University of Vienna* a *University of Glasgow* jako biostatistik. Publikoval 63 metodologických a aplikačních článků v časopisech s IF, jež byly více jak 850-krát citovány a dosáhl h-indexu 17.