

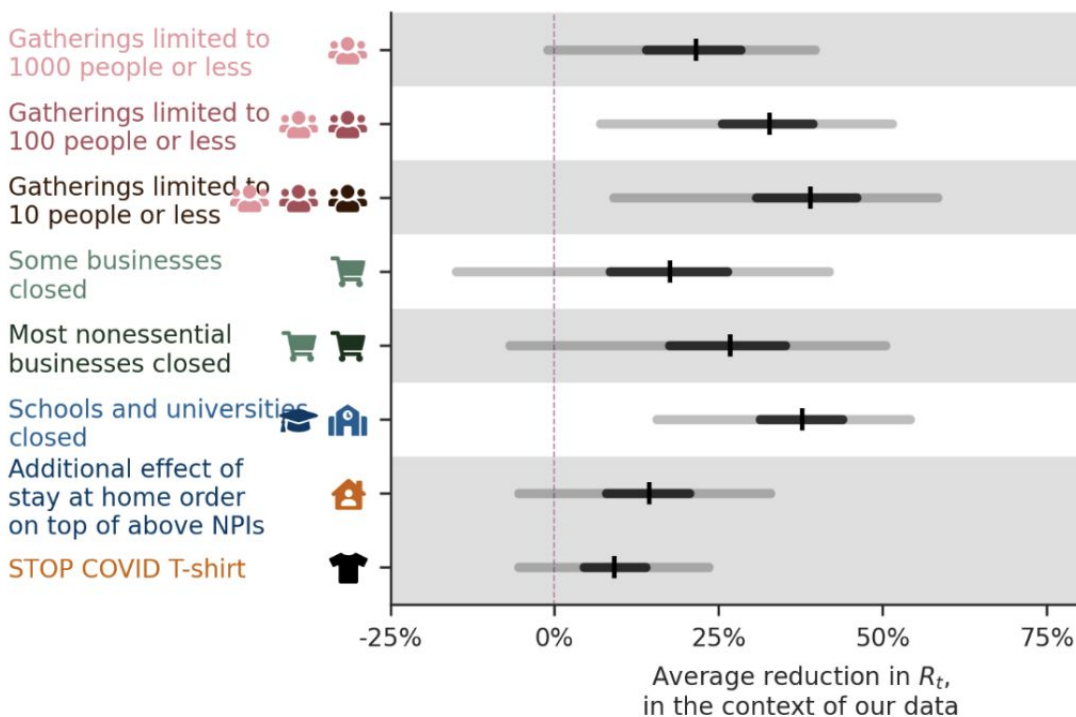
Renomovaný vědecký časopis Science nedávno otiskl [článek](#) prezentující naše empirické odhady velikosti dopadů protiepidemiologických opatření na rychlost šíření nákazy COVID-19. Naše odhady jsou založeny na bayesovském statistickém modelu a na datech z 41 zemí z první vlny. Tomáš Fürst, Jakub Dostál a Ondřej Vencálek nedávno na webu České statistické společnosti zveřejnili [krátký text](#) na pomezí odborné a politické kritiky. Běžný čtenář po jeho přečtení může podlehnout dojmu, že jejich kritika je namístě. V tomto krátkém komentáři jejich kritiku vyvracíme objasněním stochastického charakteru odhadů našeho modelu.

Podstatou jejich zpochybnění se zdá test našeho modelu, který kritici popisují takto:

Vymysleli jsme si vládní opatření, které nikdy neexistovalo. Naše fiktivní opatření vypadalo tak, že vláda k danému dni uzákonila pro všechny občany povinnost nosit tričko s nápisem „STOP COVID“. Z intervalu, po který byla každá země sledována, jsme náhodně vylosovali nějaké datum a k tomuto datu jsme předstírali, že bylo toto tričkové opatření zavedeno. Tedy přidali jsme k původním datům, která autoři použili, nový sloupec, který udává začátek zavedení

fiktivního tričkového opatření. Samozřejmě jsme nijak nezměnili počty případů ani počty úmrtí. Pro pořádek znovu opakujeme, že v žádné sledované zemi tohle opatření ve skutečnosti nikdy zavedeno nebylo, tedy nemohlo mít žádný efekt na průběh epidemie. Takto upravenou datovou sadu jsme prohnali modelem, který autoři laskavě zveřejnili. Ani jsme nesáhli na žádný z parametrů modelu.

Vyšlo jim toto



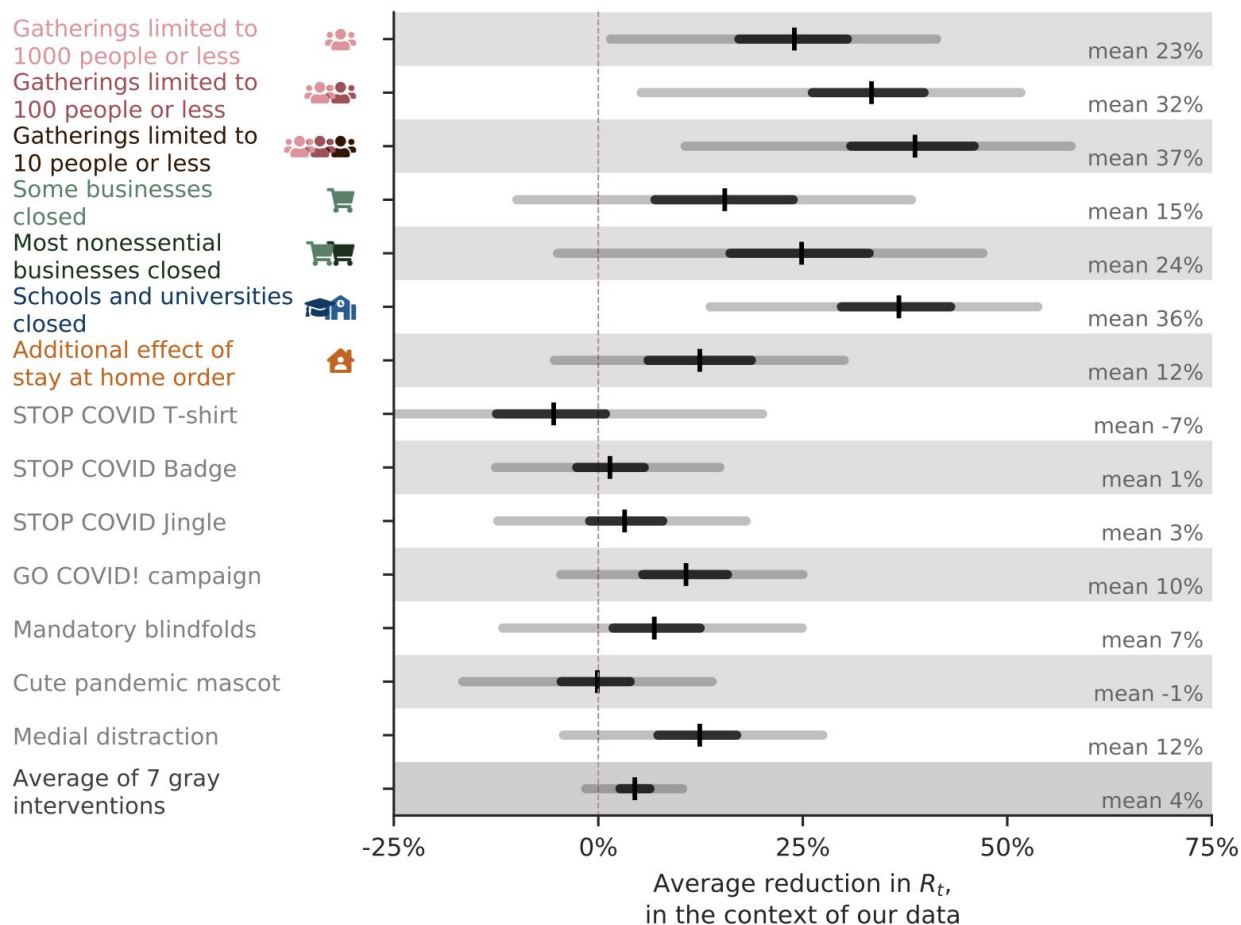
Výsledek autoři interpretují tak, že: *fiktivní nařízení nosit „STOP COVID“ trička mělo (dle modelu) poměrně výrazný pozitivní efekt na snížení reprodukčního čísla.* Tento podle nich zjevně špatný výsledek podle nich zpochybňuje všechny prezentované odhady efektů, a vede je k řadě silných závěrů. Například píší:

- *Na každou otázku, zda měl jakýkoliv zákaz efekt, totiž model odpoví „Jistě, pane ministře“!*
- *Tento model nám nijak nepomůže to zjistit a pochopit*
- *Takto pojatý model je dle našeho názoru zavádějící a nebezpečný.*
- *A výše citovaná studie nám rozhodně nepomůže, abychom [...] zjistili efekt opatření*

Výsledek jejich testu však není absurdním jak se může na první pohled jevit a určitě na jeho základě nelze dělat citované závěry.

Především je třeba správně číst výstupní graf: efekt intervencí nelze z dat odhadnout přesně, a výstupem modelu je vždy statistické rozdělení odhadu velikosti efektu. V grafické reprezentaci tmavý pás označuje oblast, kde se odhadovaný efekt nachází s 50% pravděpodobností. Světlejší pásy pak rozšiřují oblast tak, že se někde uvnitř ní skutečný efekt nachází s 95% pravděpodobností. Levý a pravý šedý pás tedy pokrývají každý necelou čtvrtinu pravděpodobnosti. Model tedy říká, že se zhruba čtvrtinovou pravděpodobností má falešná intervence dopad -5% až +4%, tedy skoro žádný. S poloviční pravděpodobností je odhadnutý efekt slabě pozitivní. Jediné, co model prakticky vylučuje je, že by “trička” měla silný pozitivní nebo negativní efekt.

Pokus jsme zkusili zopakovat s řadou sedmi zcela nových falešných intervencí vygenerovaných popsaným způsobem (nezávisle na datech od ČSS). Takhle to dopadlo:



Vidíme, že ze sedmi falešných intervencí někde vychází velmi slabě negativní efekt, někde efekt velmi blízky nule, a někde slabě pozitivní efekt. Efekt průměru všech falešných intervencí je jen +4%, tedy skoro žádný (např. ve srovnání s šumem měření).

U všech falešných intervencí je střední odhad efektu menší, než odhady u všech skutečných intervencí. Nejmenší rozdíl je u intervence "Dodatečný efekt zákazu vycházení v době platnosti ostatních intervencí". Že tento efekt "Zákazu vycházení" není jistý a nejspíš není velký, je jedním ze závěrů naší studie, a naopak často tento výsledek byl interpretován jako důvod k pochybnostem nad přínosem zákazů vycházení, pokud platí jiné silné intervence.

Důležité je i to, že relativní odhadovaný efekt skutečných intervencí se přidáváním falešných intervencí výrazně nemění, tedy že je model vůči tomuto robustní.

Tedy teze, že na jakoukoliv otázku model odpoví že intervence měla výrazný pozitivní efekt, nebo že nám o skutečném efektu skutečných intervencí neříká nic, je tak poměrně názorně vyvrácena. Náš model i výsledky byly před publikací samozřejmě podrobeny celé řadě dalších kontrol robustnosti modelu (dokonce v oboru nadstandardní) - více v našich článcích a níže.

Naivně by se mohlo zdát, že je chybou modelu, že u falešných intervencí neukazuje “přesně” nulový efekt s rozložením pravděpodobnosti těsně kolem 0. Je ale třeba uvážit, že model neví, že některé intervence jsou “pravé” a některé “falešné”, vidí pouze data o počtech případů a úmrtí. Správná interpretace je pak taková, že data nevyklučují slabě pozitivní nebo slabě negativní efekt. Náhodná “falešná” intervence se také může “trefit” tak, že se ve více zemích přibližně potkává s datem intervencí skutečných, a část efektu skutečných intervencí nebo dalších událostí pak model připíše intervenci falešné.

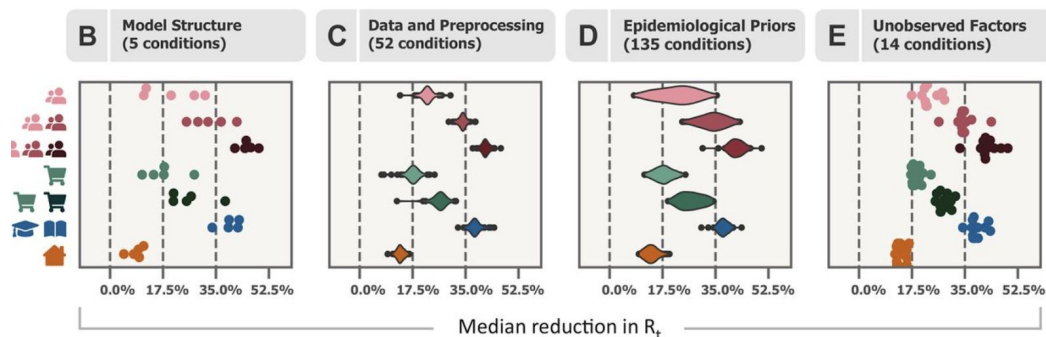
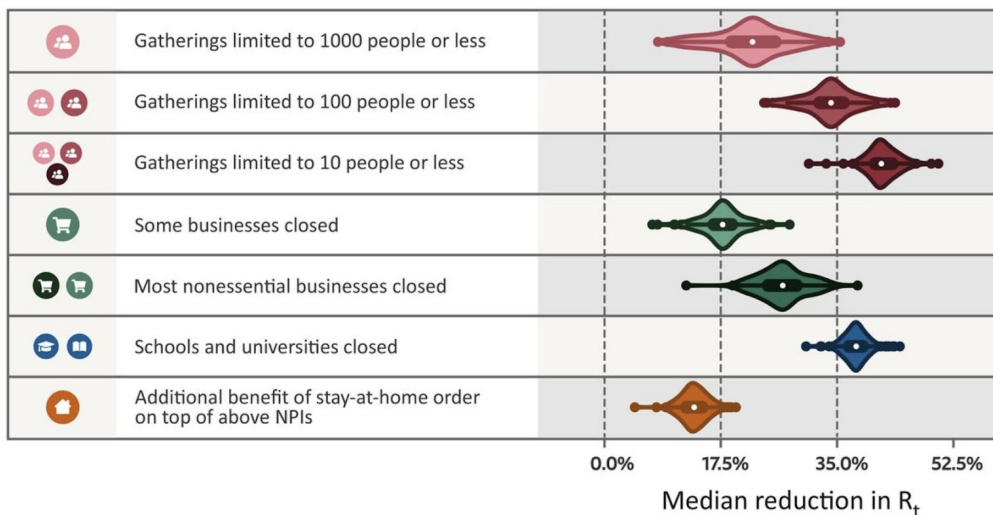
Obecně je třeba zdůraznit, že model pouze popisuje, jakou informaci obsahují data, a kvantifikuje neurčitost. V případě odhadu efektu jednotlivých protiepidemických intervencí je přitom tato neurčitost značná. (Jak již bylo řešeno, model například nevyklučuje, že v situaci, kdy jsou zavřené obchody, školy a zakázané shromažďování, nepřinesl zákaz vycházení nic navíc - alespoň v kontextu první vlny ve zkoumaných zemích).

Pokud bychom se ponořili do podrobností toho, jaký průměrný efekt model ukáže u neexistující intervence, zjistili bychom, že tento průměr není přesně nula, vysvětlení čehož ale přesahuje rámec tohoto textu. (Důvodem může být asymetrický prior, a směr nevysvětlené variance v datech. Zda použít prior symetrický či asymetrický bylo předmětem obsáhlé diskuse mezi autory článku v Science a recenzenty, kdy autoři se původně klonili spíše k prioru symetrickému. Toto téma je ovšem vhodné pro hlubší, skutečně odbornou debatu.)

Vážnějším zájemcům o pochopení, jak Bayesovské modely efektu intervencí fungují, a do jaké míry jsou odhadnuté parametry robustnostní vůči různým volbám modeláře velmi doporučujeme [náš příspěvek na konferenci NeurIPS](#), který na přesně na tuto otázku odpovídá.

Je zjevné, že různé volby parametrů modelu ovlivňují výsledky, proto v našem původním článku v Science provádíme 206 dalších analýz sensitivity modelu, a nakonec klasifikujeme intervence do kategorií podle toho, jak dopadly v tomto celém ensamble dvou stovek modelů. Analýzy citlivosti zahrnují i analýzy citlivost výsledků k faktorům nezahrnutým v modelování.

A All Sensitivity Analyses (206 conditions)



Experiment provedeny Tomášem Fürstem, Jakubem Dostálkem a Ondřejem Vencálkem a forma jeho následné prezentace bohužel do odborné diskuse nad bayesovským modelováním a interpretací jeho výsledků smysluplně nepřispívá.

Při použití modelů pro praktické politické rozhodování bychom důrazně varovali před zakládáním rozhodnutí na výsledku jediné studie. Rozhodnutí by vždy měla vycházet ze zohlednění celé relevantní literatury. (Která se v hrubých rysech v případě efektu intervencí shoduje.) O to důrazněji bychom varovali před rozhodováním na základě subjektivních dojmů na hranici politického komentáře a odborného textu, byť by šlo o text publikovaný na webu odborné společnosti.

J.K. s přispěním kolegů kteří by raději nebyli zataženi do zpolitizovaného sporu