

T

ТЕМА БРОЈА: ПАНДЕМИЈА И ПАНДЕМОНИЈУМ

Модели и подаци

—

Подаци су на разним серверима букнули преко ноћи, без јасно дефинисаних и документованих образложења о систематици њиховог прикупљања





ТЕКСТ:
Дарко Донеvски

ПАНДЕМИЈА SARS-COV-2, коронавируса који изазива болест Ковид-19, омогућава истраживање комплексних веза између отворености модерне академске науке, спремности на флексибилне реакције у реалном времену и начина на које јавност прихвата њене резултате и препоруке. Упркос знатно увећаном ангажовању научних и технолошких ресурса, актуелну ситуацију у вези са пандемијом многи сматрају и најзначајнијом глобалном кризом науке у последња два века. У наставку текста испитаћемо неколико важних узрочно-последичних веза ове тврдње. Важна напомена: научно мотивисане акције у јеку

ИЛУСТРАЦИЈЕ: Давид Билобрк

пандемије могу се грубо поделити у две категорије. Прву чине акције уже медицинске науке (нпр. епидемиологије) у примени решења која се тичу планирања акција превенције и третмана болести. Друга категорија обухвата анализу синергичности метода из других наука, и то како се отворена питања модерне науке у целини рефлектују на ефикасност решења, као и то где је простор за напредак. Овај текст бави се другом групом питања.

КРИЗА НАУКЕ КАО КРИЗА СУПЕРХЕРОЈА

У тренутку писања овог текста пандемија коронавируса обликује нову демографску слику планете. Док подаци који ту слику квантификују треба да буду стављени у шири контекст, један од најпосећенијих сајтова за статистику у реалном времену, *Worldmeter*, извештава да је укупан број оболелих од почетка пандемије 4,46 милиона, број умрлих 300.000, а број оних који су се опоравили 1,67 милиона. Откад је Светска здравствена организација (СЗО) алармирала међународну јавност о болести проузрокованој коронавирусом, притисак на истраживачке институције подигнут је на највиши могући ниво. То се манифестује како кроз очекивање моменталних решења проблема тако и кроз ефикасну комуникацију која би кризу искористила за промоцију научног дијалога. Данас, више од два месеца од проглашења пандемије, дајемо феноменолошки осврт на „пролазно време“ поменутих очекивања.

Пажљиво пратећи садржај и реакције посетилаца више од 50 водећих светских информативних сајтова и научних блогова, делује да је велики део људске популације збуњен урушавањем идеалистичне и нереалне представе о модерној науци као стално присутном суперхероју који се појављује „ниоткуда и ефикасно решава сваки проблем“. Због чега се поменути идеал створио, оставићемо по страни, јер захтева пажљиву и дугу дискусију. Овде, пре свега, анализирамо колико је модерна наука припремљена да, у контексту отворених података и метода, произведе моментални и видљиви пробој у решавању низа питања у вези са Ковидом-19 и његовим узроцима.

Да бисмо разумели комплексност проблема, његове градивне делове ставићемо у шири контекст. Последњих година интензиван је број радова који су протеклу декаду означили као „кризу репродукцибилности научних резултата“ (енг. *reproducibility crisis*, [1], [2], [3]). Ово је синоним за кризу методологија које се примењују у разним научним дисциплинама, а које заузврат не омогућавају да се објављени резултати лако провере и потврде. И док криза није драстично погодила истраживања у природним наукама

попут физике или хемије, она је у огромној мери присутна у психологији, економији и медицини. Исход бурних расправа на ову тему су оптужбе да је криза проузрокована са једне стране жељом да се државе са брзим научним растом (Индија, Кина) доминантније позиционирају на тржишту, а са друге стране, хиперпродукцијом радова који податке из истраживања не отварају за научну јавност [4]. У контексту актуелне пандемије важно је поменути да је чак 20 % медицинских студија, које су од 1990. до 2015. године имале преко 1000 научних цитата, оспорено као нерепродуктивно или статистички безначајно [5]. То је активирало неке перманентне проблеме. Један од њих је селекциони ефекат публикација научног рада (енг. *publication bias*). Он осликава како научни радови са снажним тврдњама и потенцијално високоимпактним резултатима добијају знатну предност у динамици објављивања у односу на резултате који неку почетну тезу оповргавају. Опасност овог ефекта је у свесном избегавању корекције резултата на комплетни сет података који је коришћен у открићу (тј. публикују се они резултати који „не ометају“ јак закључак). Један од најконтроверзнијих недавних примера је испитивање комбинације лекова хидроксицхлорохина и азитромицина у контексту лечења Ковида-19, где су снажне препоруке изведене на основу истраживања малог и изразито селективног круга пацијената [6]. Као последица притиска на тренутно проналажење адекватног лека за третман тешко оболелих, у комбинацији са ауторитетом институције и аутора студије, акциони тимови многих држава доведени су у опасност да без дубљих анализа одлучују о имплементацији предложеног третмана.

На истраживање француских лекара свет је реаговао једнако неусаглашено, те су многи државници позивали на хитну набавку „чудотворног лека“, а многи научници без експертизе у области на друштвеним мрежама масовно су делили линк ка раду уз коментаре „наука лако побеђује сваку кризу!“. Рад *Gautret et al.* објављен је по процесу хипербрзе ревизије у међународном часопису *International Journal of Antimicrobial Agents*. Међутим, само 20 дана након његовог објављивања, издавач Елсевијер објавио је саопштење у ком изражава сумњу у етичке и статистичке методе ове студије, тражећи независне потврде резултата [7].

Као реакција на кризу квалитета и репродуковања научних резултата, неки врхунски институти попут америчког Станфорда, предложили су радикалне промене у начину статистичке анализе клиничких истраживања [8]. Срећом, научне кризе неумитно са собом доносе и бројне позитивне аспекте, те је тако и ова позвала на ближу сарадњу хетерогених научних дисциплина. Пример једног конкретног предлога о

премошћавању постојећих недостатака дошао је из домена статистичког моделовања великих података у физици и математици. Наиме, за статистичко тестирање неке научне хипотезе, већина радова у друштвеним и медицинским истраживањима ослања се на тзв. p -вредност, те се често сматра да су резултати „статистички значајни“ уколико је тај број мањи од референтног (нпр. 0,05, што се често узима као канонска вредност). Међутим, многобројна истраживања у природним наукама показала су да закључивање на основу овако упрошћених критеријума може да буде погрешно или самообмањујуће, предложивши експресни прелазак на далеко прецизнију (али и компликованију) тзв. бајесијанску статистику [9]. С обзиром на недостатак простора да читаоцима детаљно опишемо математички апарат иза Бајесове статистике, остављамо линк где сами могу да провере на који начин се „лажни позитивни резултати“ могу квантификовати [10]. Оно што треба имати на уму пре него што наставите да читате текст јесте да пун консензус у примени бајесијанских статистичких метода није успостављен, нарочито у медицинским и друштвеним наукама.

ПРЕДНОСТИ И НЕДОСТАЦИ ОТВОРЕНИХ БАЗА ПОДАТАКА

Због свега набројаног, пандемија коронавируса интензивирала је захтеве за коришћењем „концепта науке отворених података“, тј. науке где је приступ истраживањима и њиховим резултатима отворен за све. Од почетка 2020. вртоглаво су порасли захтеви за претраживањем база података са обрађеним мерењима истраживачких студија у вези са Ковидом-19. Међутим, наука отворених података није исто што и наука са квалитетним подацима, и док модерни сервери немају више проблем да опслуже пентабајте мерних података, дотле је много упитније на који начин се њихов квалитет проверава. У преводу, остаје дилема колико ефикасно се отварање огромних података директно транслира у боље и прецизније доношење одлука. Даћемо малу аналогију из једне потпуно различите научне области. Када је европски програм за прикупљање података о милионима галаксија у свемиру почео са радом 2013. године, велики тим од 28 научника је наредних пет година био задужен за проверу квалитета података и развој симулација за њихову верификацију. Тек након дуготрајне и детаљне процедуре корисницима је омогућен приступ отвореној бази каталога и софтверима који се користе за испитивање. У случају пандемије коронавируса, подаци су на разним серверима букнули преко ноћи, без јасно дефинисаних и документованих детаљних образложења о систематички њиховог прикупљања, односу сигнал/шум,



корекцији селекционих ефеката и упутима шта урадити у случају недостајућих података. Под поменутим сајтовима овде подразумевамо и неке референтне тачке многих студија, попут сајта Универзитета „Донс Хопкинс“ или многих националних сајтова за Ковид-19, укључујући и онај у Републици Србији. Неподељено мишљење великог броја експерата из домена моделирања великих података је да су подаци на којима се заснивају одлуке некомплетни, са израженим структурним проблемима [11].

Срећом, ниједан проблем није нерешив. Корекција селекционих ефеката, као и моделирање (не)детекција, кључни су за моделовање еволуције како галаксија тако и ефеката пандемије коронавируса. Као што у свемиру не можемо увек да једним јединим телескопом откријемо где се сваки појединачни објекат налази (мале, једва видљиве звезде и галаксије су многобројније од оних великих), тако је и у случају коронавируса нејасно колики је број асимптоматских људи заражених вирусом. Ово подсећа на проблем игле скривене у пласту сена – односно ослањање на статистике које су засниване искључиво на бројању пацијената са јасно израженим симптомима исто је као бројање само оних звезда које ноћу видимо голим оком. У оба случаја, јасно је да пропуштамо да документујемо скривене објекте у крајњој статистици.

Питање је: како моделовати те недостајуће елементе комбинујући теорију и експерименте?

Најдиректнији начин да се приступи проблему је насумично скенирање (енг. *blind survey* = без претходне селекције) великог статистичког

узорка. Пример такве анализе одрађен је у малом италијанском месту по имену Во, близу Павове. У консултацијама са истраживачима из више области моделирања података (математике, астрофизике, биостатистике), тим медицинских стручњака је одлучио да на самом почетку ширења заразе у Италији тестира читав градић – Во. Резултати студије недавно су објављени, и примера ради, истраживање је открило да је чак код 40 % тестираних потврђено присуство вируса било асимптоматско [12]. Бенефити које овакви насумични тестови доносе су вишеструки: (1) могућност да се увиди значај скривене популације (асимптоматских носилаца заразе) која није детектована стандардним процедурама; (2) могућност да се потпуно исти, рандомизиран узорак тестира у различитим временским стадијумима еволуције вируса, (тј. пре и после епидемиолошких мера, рестрикција кретања, физичког дистанцирања итд.); (3) да се на основу модела заснованог на комплетном узорку изврши корекција селекционих ефеката, самим тим и прецизнија процена еволуције пандемије у гушће насељеним областима где су узорци сакупљани у изразито некомплетном маниру. То се односи на велике градове попут Рима, Њујорка, Париза. Иако једноставног концепта, пут до реализације читавог алгорита захтева висок ниво синхронизованих мултидисциплинарних акција подржаних како научним тако и политичким структурама. Нагласићемо две важности овог процеса:

1. Комплетно разумевање модела и резултата у временском домену немогуће је обавити изолованим истраживањима, било да су у питању епидемиолози или биостатистичари с једне стране, односно аналитичари података и статистички моделари, са друге стране. Проблем глобалних феномена/проблема у науци (попут овог са коронавирусом, или пак проблемом ефекта климатских промена) могуће је решавати *искључиво* ангажовањем различитих научних дисциплина и профила истраживача. Другим речима, све корекције модела који служе да се еволуира ефекат пандемије, не могу се ослањати на изоловани рад, макар он долазио из најпрестижнијих светских лабораторија. Они морају бити интердисциплинарног и међународног карактера, како би се динамика решавања проблема правилно балансирали. Учешћа кроз отворене дискусије о стратегијама су посебно значајне за слабије развијене земље које нису довољно компетитивне у тржишној научној трци.

2. Све наведено борбу са пандемијом коронавируса смешта у домен не само квалитетног научног рада, већ и *ефикасног научноколаборативног менаџмента*. С обзиром на то да је од почетка кризе прошло мање од шест месеци, још је незахвално доносити суд о томе да ли је интердисциплинарна повезаност довољно активирана

Време кризе уједно је и најпогоднији тренутак да се ојача поверење у науку

или не. Оно што је, међутим, јасно – државе су махом самостално доносиле одлуке о иницијалним акцијама, а у многим случајевима дијалог између ширих научних експертиза није инициран на унутардржавном нивоу. Као извесна реакција на ове проблеме почеле су да се јављају регионалне иницијативе које имају за циљ успостављање транспарентније дискусије о квалитету науке и научне комуникације у вези са пандемијом коронавируса. Једна од њих недавно је активирана и на територији Балкана, под називом Ковид-19 Балкан <https://covid19balkans.com/>. Иницијатива је окупила око шездесетак научника, лекара и инжењера са циљем да се понуди обједињена платформа за транспарентније виђење целокупног проблема.

ПРОБЛЕМ ОТВОРЕНОСТИ НАУЧНОГ МЕТОДА И КОМУНИКАЦИЈЕ

Паралелно са отварањем база података, многи издавачи попут Елсевијера привремено су скинули рестрикције са својих издања о Ковиду-19. Електронски „препринт“ сервери отвореног типа (енг. *Open Access*), који архивирају научне радове независно од њиховог статуса у процесу рецензије, забележили су енормни пораст примљених рукописа везаних како за клиничке анализе тако и комплексно моделовање ефеката пандемије. Примера ради, на најстаријој бази отвореног типа, *arXiv*, која је доминантно испуњена радовима из астрономије, космологије и физике, забележено је чак 800 чланака о статистичким ефектима пандемије коронавируса. Два најпопуларнија сервера за медицинска и биолошка истраживања, *medRxiv* и *bioRxiv*, забележили су преко 3000 примљених наслова о Ковиду-19 у последњих десетак недеља. Одређени филтери за прослеђивање радова ка глобалној мрежи су постојали, рецимо на *arXiv*-у се додатно проверавају наслови који стижу са мање респектабилних институција, али како одговорни за вођење ових платформи кажу, дешава се да чак и са високо ранжираних института стижу радови који су својим квалитетом испод стандарда. То је све резултирало знатно појачаном контролом од стране стручњака, нарочито оних радова који се тичу третмана у случају Ковида-19. Ипак, ствари нису

једноставне јер различите отворене платформе имају различите критеријуме за филтрирање. Као последица, довољно атрактивни наслови могу лако да задобију медијску пажњу и фокус јавности одведу на погрешну страну.

Додатни проблем науке у доба оваквих криза је брз процес ревизије научних радова. По статистици коју је недавно објавио сајт *Nature*, у домену медицинских наука просечан број дана од тренутка подношења до тренутка публикавања рада је пао са 150 на мање од 50. Јасно је да је тешко одржати у исто време контролу квалитета и брзину корекција, те су многи издавачи потегли за изнуђеним решењем: направили су додатну листу експерата који би волонтирали као рецензенти у процесу радова за брзу објаву. На основу изложеног, види се да је оптерећеност научника са врхунских научних институција у доба кризе коронавируса вишеструка.

Мотивисани хаотичним стањем на мултирелацији академска наука – државне институције – јавности, недавно је у часопису *Science* објављено отворено писмо са позивом на транспарентност Ковид-19 модела [13]. Шта то значи? Научни модели су кључни алати који се користе за предвиђање и који касније адаптирају владе земаља у предузимању одређених мера. Нажалост, много је чешћа имплементација нетранспарентних метода над колекцијом података, што се сматра ефектом „црне кутије“ (*black box modelling*). Не само резултати модела, већ и сами модели, морају да се јасно образложе и јавности, и да буду основа за отворени научни метод и сарадњу у другом кругу анализе глобалне кризе. То још није случај ни на глобалном, ни на регионалном нивоу, што је био и основни мотив захтева да се сви методи и програмски кодови обједињено доставе на релевантно место, платформу CoMSES (*Trusted Digital Repositories*).

Интензивирање је и стари проблем квалитетне комуникације научних решења са широм јавношћу. Канадска влада га је међу првима приметила и покушала да реши тзв. *crowdsourcing* методом, новим видом прикупљања података. Око 300.000 Канађана, у два различита временска интервала, радило је упитнике о свом виђењу и замеркама на предузете мере и научној комуникацији тих мера. У мање развијеним земљама евидентнија је неповерљивост ка научним резултатима и дијалогу на ту тему, што многи од научника дефинишу као „проблем затворене научне комуникације“. И док се активно истраживање спроводи на ову тему, за крај треба истаћи да добар део дисбаланса у релацији научни метод – јавност лежи управо у (не)спремности да се комплексност научног процеса и неодређености научних закључака прихвате у најширој јавности. Време кризе је уједно и најпогодније време да се на ефектан начин поверење у научни метод ојача. —E

Референце:

- [1] Schooler, J. (2014). *Nature*, 515, 7525
- [2] Open Science Collaboration (2015-08-28). "Estimating the reproducibility of psychological science". *Science*. 349(6251): aac4716.
- [3] Leichsenring F. et al. (2017). "Biases in research: risk factors for non-replicability in psychotherapy and pharmacotherapy research". *Psychological Medicine*. 47 (6), 1000–1011.
- [4] Daniele Fanelli, *PNAS* March 13, 2018 115 (11) 2628–2631;
- [5] Ioannidis JA, *JAMA*. 294 (2): 218–228.
- [6] Gautret et al. *Int J Antimicrob Agents*. 2020 Mar 20:105949. doi: 10.1016/j.ijantimicag.2020.105949
- [7] <https://www.isac.world/news-and-publications/isac-elsevier-statement>
- [8] Ioannidis, JA (2016). *PLOS Med*. 13 (6): e1002049.
- [9] Colquhoun, D (2015). "An investigation of the false discovery rate and the misinterpretation of *p*-values". *Royal Society Open Science*. 1 (3): 140216
- [10] <http://fpr-calc.ucl.ac.uk/>
- [11] <https://www.datanami.com/2020/04/21/coming-to-grips-with-covid-19s-data-quality-challenges/>
- [12] Lavezzo et al. <https://www.medrxiv.org/content/10.1101/2020.04.17.20053157v1>
- [13] Barton et al. *Science* 01 May 2020.; Vol. 368, Issue 6490, pp. 482–483

Дарко Донеvски је савремена италијанске владе у области космологије, у оквиру пројекта „Прашина у раном свемиру“. Доктрирао је 2018. на Универзитету Aix-Marseille, у Француској, са темом „Еволуција далеких галаксија“. Као савремена, борао је на истраживањима у Лајдену (Холандија) и Тулузу (Француска). Основне студије завршио је на Универзитету у Новом Саду. Савремена је сарадник часописа Елементи.