

СИСТЕМНИЙ ПІДХІД ДО АНАЛІЗУ ДАНИХ ЩОДО РОЗВИТКУ ТА ПРОГНОЗУВАННЯ ПАНДЕМІЇ

Анотація. Питання аналізу та прогнозування розвитку пандемії коронавірусної інфекції є актуальним не лише для України, але і для багатьох інших країн. Однією з ключових проблем є недостовірність наявних даних по основних показниках пандемії, зокрема, по кількості інфікованих, кількості летальних випадків, кількості важких випадків тощо. Це перешкоджає вчасному прийняттю обґрунтованих рішень щодо протидії поширенню інфекції. Тому актуальним є залучення непрямих даних для підвищення надійності результатів аналізу та прогнозів. У роботі запропоновано елементи системного підходу до аналізу наслідків та прогнозування розвитку пандемії на основі комплексного застосування офіційних даних і додаткових джерел інформації.

Ключові слова: пандемія, Covid-19, прогнозування, статистичні дані, захворюваність, смертність, IFR, математична модель, пошуковий запит.

Постановка проблеми. Пандемія коронавірусної інфекції стала одним з найбільших викликів людству за останні кілька десятиріч. Ефективне запобігання її розвитку є задачею прийняття складних за структурою рішень з багатьма критеріями, високою невизначеністю результатів того чи іншого вибору, недостатньо зрозумілими зворотними зв'язками. Протягом останнього року з'явилася велика кількість статистичної інформації про основні показники пандемії. Втім, на думку більшості дослідників, первинні дані щодо кількості інфікованих та летальних випадків є істотно заниженими, що робить недостовірними інші важливі показники, зокрема, ті, що характеризують характер динаміки, швидкість поширення інфекції та її летальність. Водночас протягом останнього року з'явилася велика кількість результатів досліджень, що дають підстави залучати додаткові до офіційних даних джерела інформації для підвищення якості аналізу і прогнозів розвитку пандемії. Були істотно розвинені математичні моделі поширення інфекцій, які дають змогу уточнювати окремі показники. Істотно покращилося розуміння будови і властивостей вірусу, природи коронавірусної інфекції, її впливу на організм людини та

відділених наслідків хвороби. Все це дає змогу поставити питання про необхідність створення більш комплексних підходів до аналізу і прогнозування розвитку пандемії.

Мета дослідження. Обґрунтування елементів системного підходу до аналізу і прогнозування розвитку пандемії, що враховує офіційні дані, результати математичного моделювання та непрямі дані, які можуть бути отримані з відкритих джерел.

Аналіз останніх досліджень і публікацій. Ще на початку пандемії в Україні автор зазначав актуальність створення і оприлюднення масивів надійних і достатньо повних даних не лише безпосередньо про поширення коронавірусної інфекції, але і про фактори, які можуть впливати на цей процес, що могло б надати змогу підключити до пошуку оптимальних варіантів вирішення проблем значну кількість фахівців з аналізу даних і за рахунок цього послабити наслідки пандемії [1]. На сьогодні у світі та у багатьох країнах такі масиви інформації вже є, і вони поступово наповнюються новими даними. В Україні ситуація з наявністю достовірних даних є значно гіршою. Зокрема, за оцінками Робочої групи НАН України, реальна кількість померлих внаслідок пандемії в Україні може бути втричі більшою за офіційні дані [2]. Ще більш істотною може бути недооцінка кількості інфікованих. Наслідком цього є істотний розкид оцінок показників летальності.

Одним з найважливіших показників інфекції є IFR (Infection fatality rate – відношення кількості померлих від інфекції до кількості інфікованих. На початку пандемії його оцінювали переважно за офіційною статистикою зареєстрованих випадків (CFR), які інколи перевищували 10 – 20%. Але пізніше, з'явилися дослідження, що враховують результати масових обстежень на наявність антитіл, уточнені показники смертності тощо [3; 4]. За різними даними найбільш імовірні середні значення перебувають у межах 0,2 – 1%. Але вони можуть істотно варіюватися залежно від вікової групи та супутніх захворювань. Після першої хвилі з'явилася можливість отримати нижні оцінки IFR на основі офіційних даних щодо кількості летальних випадків у розрахунку на 1 млн. мешканців [5]. Для різних країн з різною віковою структурою населення вони перебувають у межах від 0,05 до 0,2%. Водночас слід зазначити, що на такі оцінки у різний бік впливають обмежений час збереження антитіл в організмі людини, можливість повторного інфікування та неповнота даних про летальні випадки.

Для моделювання розвитку пандемії найбільш використовуваними є модифіковані SEIR моделі, з яких найбільш інформативними у контексті мети дослідження є моделі Imperial College London (ICL) (<https://mrc-ide.github.io/global-lmic-reports>) та The Institute for Health Metrics and Evaluation (IHME) (<https://covid19.healthdata.org/global>). Ці моделі надають ретроспективні оцінки загальної кількості інфікованих та прогнози за окремими показниками як для окремих країн, так і у глобальному вимірі. Також вони оцінюють вплив деяких запобіжних заходів на розвиток пандемії. Водночас, ці моделі базуються на офіційній статистиці смертності, що, як зазначено вище, не є достовірною, а також оцінках IFR, які мають істотний розкид. Крім того, вони не повною мірою враховують запізнення надходження інформації. Все це збільшує розкид модельних оцінок.

Розвиток Інтернет технологій і систем відкритих даних стимулював в останнє десятиріччя становлення нового наукового напрямку – інфоміології (інформаційної епідеміології), що передбачає застосування інформації і даних з Інтернет та інших ресурсів для прогнозування, прийняття рішень та інформування щодо спалахів епідемічних хвороб [6, 7]. Зокрема, це можуть бути аналіз запитів пошукових систем Інтернету, моніторинг оновлень статусів людей у Twitter та інших мікроблогах, виявлення та кількісна оцінка диспропорцій у доступності медичної інформації, виявлення та моніторинг публікацій, що стосуються охорони здоров'я, в Інтернет, автоматизовані інструменти для вимірювання поширення інформації та перекладу, відстеження ефективності маркетингових кампаній у галузі охорони здоров'я тощо. Останнім часом такі технології використовуються багатьма дослідниками для дослідження розвитку пандемії корона вірусної інфекції. Результати відповідних досліджень є важливими для підвищення ефективності і вчасності рішень, що приймаються урядами і міжнародними організаціями для протидії пандемії.

Згідно з [8] для США кількість пошукових запитів за ключовими словами «face masks», «Lysol», та «COVID stimulus check» має сильну кореляцію з щоденною кількістю нових випадків по всій території США (коефіцієнти кореляції Пірсона дорівнюють, відповідно, 0.86, 0.82 та 0.79). Кількість пошукових запитів за ще декількома ключовими словами має значну кореляцію для окремих штатів. Також було показано, що значна кореляція спостерігається за часового лагу до 16 діб, що дає змогу використовувати аналіз пошукових запитів для прогнозування нових спалахів.

Аналогічні результати отримано у Китаї для коефіцієнтів кореляції Спірмена між кількістю підозр та лабораторно підтверджених випадків з одного боку та кількістю запитів за ключовими словами «коронавірус» та «пневмонія» у декількох пошукових системах [9]. При цьому найбільші значення коефіцієнтів кореляції були в діапазоні 0,89 – 96 і спостерігалися для лагів 6 – 8 діб для кількості нових підозр та 8 – 12 діб для лабораторно підтверджених випадків.

В [10] встановлено наявність істотної кореляції між кількістю запитів, що стосуються запахів і початком спалахів пандемії в Великій Британії, Ірані, Іспанії, Італії, Нідерландах, Німеччині, США та Франції.

Аналогічні дані про наявність істотної кореляції між окремими видами пошукових запитів та показниками захворюваності та смертності отримано в [11] для низки європейських країн та регіонів, у [12] для Ірану та в низці інших публікацій. Водночас, у [13] зазначається, що з віддаленням від точки початку спалаху кореляція між кількістю пошукових запитів і показниками захворюваності зменшується і може ставати від'ємною через зменшення зацікавленості користувачів пошукових систем..

Викладення основного матеріалу дослідження. Відкриті масиви офіційних статистичних даних по багатьох країнах доступні на сайтах <https://www.worldometers.info/coronavirus>, <https://ourworldindata.org/coronavirus>, <https://github.com/owid/covid-19-data> та інших. Крім того, багато країн мають відкриті ресурси державних органів, університетів, наукових установ, громадських організацій з додатковою інформацією.

У таблиці 1 наведено середні модельні оцінки та офіційні показники щодо окремих спалахів захворюваності для деяких країн (вихідні дані взято з <https://ourworldindata.org/covid-models>). Як видно з таблиці, максимальні кількості щоденних випадків для модельних даних можуть в 50 разів перевищувати показники офіційних даних, а дати максимуму – відрізнятись більш, ніж на місяць. Якщо брати мінімальні та максимальні модельні оцінки, відмінності дат і висот максимумів можуть бути ще більшими. Такі дані є важливими для побудови середньо- і довгострокових прогнозів, де важливо знати загальну частку населення, яка перехворіла, а також для оцінювання летальності інфекції. На жаль, для першого показника на сьогодні по багатьох країнах і регіонах існує дуже великий розкид оцінок від кількох відсотків до

понад 50-70%. Це змушує розглядати при прогнозуванні і прийнятті рішень істотно різні сценарії подальшого розвитку пандемії.

Для уточнення даних про летальність можуть бути використані дані щодо загальної смертності з усіх причин. Їх можна отримати на сайті Державної служби статистики України (http://ukrstat.gov.ua/operativ/menu/menu_u/ds.htm) та з сайту «Економіка карантину» (<https://q.rating.zone>), який використовує офіційні дані Міністерства юстиції.

Таблиця 1

Показники спалахів захворюваності для деяких країн у 2020 р.

Країна	Максимальна щоденна кількість (zareestrovano)	Максимальна щоденна кількість (ICL)	Максимальна щоденна кількість (IHME)	Дата максимуму (zareestrovano)	Дата максимуму (ICL)	Дата максимуму (IHME)
Аргентина	14941	131134	71030	22.10	26.09	18.09
Австралія	383	2317	479	30.03	26.03	25.03
Бельгія	1453	66531	16276	15.04	31.03	24.03
Греція	2673	25049	5971	18.11	18.11	11.11
Індія	93199	427736	1,14·10 ⁶	16.09	20.08	07-08.08
Німеччина	5837	47110	20534	02.04	04.04	25.03
США	32284	515231	229647	10.04	07.04	27.03
Україна	14270	65129	37233	29.11	14.11	19.11
Чехія	12585	55458	19488	28.10	23.10	17.10

Інформація Державної служби статистики є помісячною і зазвичай з'являється із запізненням на два місяці. Інформація Мінюсту є щоденною і постійно поповнюється. Але дані останнього місяця містять примітку, що реєстр ще наповнюється. Насправді, навіть більш ранні дані постійно корегуються протягом щонайменше 2 місяців. Через це дані по загальній смертності непридатні для цілей оперативного аналізу і прогнозування. Але вони є важливими з погляду ретроспективного аналізу, а також оцінки загального стану і середньотривалого прогнозування.

При аналізі розвитку пандемії такі дані використовують, зокрема, для розрахунку так званої «надлишкової» смертності, яка є різницею між смертністю за певний місяць чи тиждень та середнім значенням аналогічного показни-

ка за декілька попередніх років. За даними Євростат (https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Excess_mortality_-_statistics#Excess_mortality_in_Europe_between_January_and_November_2020) за період з березня по листопад 2020 р. «надлишкова смертність» в країнах ЄС становила близько 450 тис. осіб. При цьому, розрахунок за даними (<https://www.worldometers.info/coronavirus>) свідчить, що загальна кількість зареєстрованих за цей період летальних випадків асоційованих з коронавірусною інфекцією становила у країнах ЄС близько 330 тис., або понад 73% від «надлишкової смертності».

За даними [14] за 23.02.2021 у розрахунку на 100 тис. мешканців найбільша «надлишкова смертність» від початку пандемії була зафіксована у Болгарії (272), Перу (261), Литві (258), Росії (252) та Сербії (233). Натомість для деяких країн з низьким рівнем захворюваності вона була навіть від'ємною, зокрема, для Нової Зеландії (-40), Монголії (-38), Тайваню (-21), Австралії (-20), Кіпру (-17). Виключенням тут є Кіпр, де кількість офіційно зареєстрованих випадків у розрахунку на 1 млн. мешканців майже вдвічі перевищує середньосвітовий показник.

Для України «надлишкову смертність» було розраховано за даними Державної служби статистики та сайту «Економіка карантину». За базовий рівень було взято щомісячні середні значення кількостей померлих у 2015 – 2019 р., скореговані на зміну загальної чисельності населення України за відповідний період. Найбільші значення «надлишкової смертності» (у відсотках до базового рівня) спостерігалися у вересні (15,8%), жовтні (20,3%), листопаді (33,7%) та грудні (29,2%). У січні 2021 р. за неповними даними сайту «Економіка карантину» на 27.02.2021 очікується істотне зменшення показника до значення менш, ніж 2%. У розрахунку на 100 тис. мешканців «надлишкова смертність» за 2020 р. становить близько 38,8 тис. (у перші 5 місяців спостерігалися від'ємні значення показника), а за період з червня 2020 р. по січень 2021 р. – близько 53,5 тис., або 14,1% від базового рівня за цей період. З огляду на типові значення інкубаційного періоду, часу від появи симптомів до тестування та часу від позитивного результату тестування до смерті, які сумарно становлять 2 – 4 тижні, наведені дані дають підстави припустити, що пік захворюваності реально мав місце наприкінці жовтня – початку листопада, тобто значно раніше за дату піку за офіційними даними та, імовірно, раніше, за дати модельних оцінок, наведені у табл. 1. Останнє може бути зумовлено тим, що для отримання модельних

оцінок можуть використовуватися оперативні щоденні дані МОЗ України щодо смертності, які не враховують наступне корегування за датами смерті.

Додаткову корисну інформацію про перебіг пандемії можуть надати пошукові запити за специфічними для неї темами та ключовими словами: «пневмонія», «нюх», «пульсоксиметр» тощо. Як свідчить такий аналіз, перші спалахи інтересу за окремими ключовими словами у багатьох країнах різних регіонів (Австралія, Іспанія, Китай, Малайзія, США, Японія та ін.) спостерігалися наприкінці грудня 2019 р. – січні 2020 р. В деяких випадках (країни Європи, США) вони збігалися за часом з щорічними піками зацікавленості. Але для багатьох інших країн сезонних піків немає, або вони припадають на інший час. З урахуванням інкубаційного періоду цей факт можна розглядати як свідчення того, що вже на початку січня 2020 р. пандемія поширилася на всі континенти (це може бути пов'язано з міграцією під час Різдвяних канікул), а запобіжні заходи вживалися з великим запізненням.

Початок спалаху інтересу до зазначених тем зазвичай відбувався істотно раніше, ніж починалося зростання кількості офіційно зареєстрованих випадків. А в деяких країнах спостерігалися спалахи інтересу до них задовго до початку фіксації випадків захворюваності. Приміром, для Сінгапуру та Японії перші спалахи захворюваності офіційно були зареєстровані наприкінці березня – у квітні. Але за даними пошукових запитів їм могли передувати більші за масштабом спалахи захворюваності у січні – лютому.

Для України зафіксовано три спалахи інтересу до теми «пневмонія». Найбільший за висотою припадав на тиждень 22-28.03.2020, другий – на 25-31.10.2020 (90% від першого) і третій – на 19-25.01.2020 (49%). Січневий пік міг відображати сезонний підйом зацікавленості, але за три попередні роки висоти сезонних піків ніколи не перевищували показника 35%. За темою «нюх» було лише 2 спалахи інтересу – 18-24.10.2020 та значно менший (14% від жовтневого) 22-28.03.2020. За пошуковим терміном «пульсоксиметр» теж було лише два спалахи інтересу – 15-21.11.2020 та менший за розміром (53%) 15-21.03.2020. За пошуковим терміном «цинк» основний пік був 08-14.11.2020, а менший за розміром (35%) перший пік – 15-21.03.2020.

Розрахований за даними <https://www.worldometers.info/coronavirus> щодо зареєстрованих випадків захворюваності і смертності показник CFR для більшості країн перебуває в межах 1 – 3%. Разом з тим, для деяких країн він протягом тривалого часу залишається меншим за 0,5%. Серед них Сінгапур (0,05%), Катар (0,16%), Мальдіви (0,31%), ОАЕ (0,31%), Таїланд (0,32%), Бахрейн (0,36%),

Малайзія (0,37%), Ісландія (0,48%). У більшості з цих країн кількість зареєстрованих випадків у розрахунку на 1 млн. мешканців є вищою за середньосвітовий показник (виключеннями є Сінгапур, Малайзія і Таїланд). Але вони вирізняються високими показниками тестування, зокрема, під час весняної хвилі пандемії. Це дає підстави припустити, що за відсутності перевантаження медичної системи показник IFR імовірно не перевищуватиме 0,5%.

З іншого боку, на сьогодні у низці країн, територій та штатів США кількість зареєстрованих летальних випадків перевищує 0,2% загальної чисельності населення. Зокрема, це: Гібралтар (0,27%), Нью-Джерсі (0,26%), Нью-Йорк та Род Айленд (по 0,24%), Массачусетс (0,23%), Сан-Марино (0,21%) та деякі інші. Слід зазначити, що у більшості штатів США з найвищою смертністю та у Сан-Марино на кінець лютого 2021 р. зберігається високий рівень щоденної кількості нових зареєстрованих випадків захворюваності. Це дає підстави припускати, що реальні середні значення IFR будуть у межах 0,3 – 0,5%.

Висновки. Отримані дані свідчать, що залучення додаткових джерел інформації дає змогу компенсувати недоліки офіційних оперативних даних і глибше зрозуміти закономірності виникнення і розвитку пандемії коронавірусної інфекції. Водночас, системний підхід до прийняття рішень з протидії розвитку пандемії, має комплексно враховувати як інформацію про статистику захворюваності і смертності, так і іншу інформацію, зокрема, щодо мутацій вірусу, випадків повторної захворюваності, вакцинації, тестування, соціально-економічних наслідків тощо. Це має бути предметом наступних досліджень.

ЛІТЕРАТУРА / ЛІТЕРАТУРА

1. Бахрушин В.Є. Несистемні думки системного аналітика стосовно засобів боротьби з пандемією коронавірусу. DOI: 10.13140/RG.2.2.22487.01444
2. Прогноз розвитку епідемії COVID-19 в Україні в період 11 – 25 січня 2021 р. «Прогноз РГ-35». URL: <http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7356>.
3. Streeck, H., Schulte, B., Kümmerer, B.M. et al. Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. Nat Commun 11, 5829 (2020). <https://doi.org/10.1038/s41467-020-19509-y>.
4. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. Eur J Epidemiol. 2020 Dec;35(12):1123-1138. doi: 10.1007/s10654-020-00698-1. Epub 2020 Dec 8. PMID: 33289900; PMCID: PMC7721859.

5. Bakhrushin V. Risks of Data Inconsistency in Information Systems Used for Predicting the Pandemics Development / Bakhrushin V., Bakurova A., Pasichnyk M, Tereschenko E. // 1st International Workshop on Computational & Information Technologies for Risk-Informed Systems (CITRisk 2020) co-located with XX International scientific and technical conference on Information Technologies in Education and Management (ITEM 2020). Kherson, Ukraine, October 15-16, 2020. - Kherson National Technical University: 2020. – Vol-2805. – P. 1-15.

URL: <http://ceur-ws.org/Vol-2805/invited1.pdf>.

6. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res 2009;11(1):e11. DOI: 10.2196/jmir.1157. PMID: 19329408. PMCID: PMC2762766.

7. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med. 2011 May;40(5 Suppl 2):S154-8. doi: 10.1016/j.amepre.2011.02.006. PMID: 21521589.

8. Shyam J. Kurian, Atiq ur Rehman Bhatti, Mohammed Ali Alvi, Henry H. Ting, Curtis Storlie, Patrick M. Wilson, Nilay D. Shah, Hongfang Liu, Mohamad Bydon. Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis, Mayo Clinic Proceedings, V. 95, Is. 11, 2020. P. 2370-2381, ISSN 0025-6196, <https://doi.org/10.1016/j.mayocp.2020.08.022>.

9. Li Cuilian, Chen Li Jia, Chen Xueyu, Zhang Mingzhi, Pang Chi Pui, Chen Haoyu. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill. 2020;25(10):pii=2000199. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000199>.

10. Walker, A, Hopkins, C, Surda, P. The use of Google Trends to investigate the loss of smell related searches during COVID-19 outbreak. Int Forum Allergy Rhinol. 2020; 10: 839– 847. URL:

<https://onlinelibrary.wiley.com/doi/pdf/10.1002/alr.22580>.

11. Mavragani A/ Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020;6(2):e18941. URL:

<https://publichealth.jmir.org/2020/2/e18941>. DOI: 10.2196/18941.

12. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020;6(2):e18828. URL: <https://publichealth.jmir.org/2020/2/e18828>.

DOI: 10.2196/18828.

13. Mavragani, A., Gkillas, K. COVID-19 predictability in the United States using Google Trends time series. *Sci Rep* 10, 20693 (2020). <https://doi.org/10.1038/s41598-020-77275-9>.

14. Tracking covid-19 excess deaths across countries. *The Economist* (27.02.2021).

URL: <https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker?fbclid=IwAR2WV90->

[WpDel4laHKzC3ELnVMoI1RAzIQlaWi9_9FQ_4XHZ6s6MGdgy0gM](https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker?fbclid=IwAR2WV90-WpDel4laHKzC3ELnVMoI1RAzIQlaWi9_9FQ_4XHZ6s6MGdgy0gM)

REFERENCES

1. Bakhrushyn, V. Unsystem thouts of system analyst on how to control the coronavirus pandemic. DOI: 10.13140/RG.2.2.22487.01444

2. Forecast of COVID-19 pandemic in Ukraine on January 11 – 25 2021 p. «Prognoz RG-35».

URL: <http://www.nas.gov.ua/UA/Messages/Pages/View.aspx?MessageID=7356>.

3. Streeck, H., Schulte, B., Kümmerer, B.M. et al. Infection fatality rate of SARS-CoV2 in a super-spreading event in Germany. *Nat Commun* 11, 5829 (2020). <https://doi.org/10.1038/s41467-020-19509-y>.

4. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G. Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications. *Eur J Epidemiol.* 2020 Dec;35(12):1123-1138. doi: 10.1007/s10654-020-00698-1. Epub 2020 Dec 8. PMID: 33289900; PMCID: PMC7721859.

5. Bakhrushin V. Risks of Data Inconsistency in Information Systems Used for Predicting the Pandemics Development / Bakhrushin V., Bakurova A., Pasichnyk M, Tereschenko E. // 1st International Workshop on Computational & Information Technologies for Risk-Informed Systems (CITRisk 2020) co-located with XX International scientific and technical conference on Information Technologies in Education and Management (ITEM 2020). Kherson, Ukraine, October 15-16, 2020. - Kherson National Technical University: 2020. – Vol-2805. – P. 1-15 .

URL: <http://ceur-ws.org/Vol-2805/invited1.pdf>.

6. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. *J Med Internet Res* 2009;11(1):e11. DOI: 10.2196/jmir.1157. PMID: 19329408. PMCID: PMC2762766.

7. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med.* 2011 May;40(5 Suppl 2):S154-8. doi: 10.1016/j.amepre.2011.02.006. PMID: 21521589.

8. Shyam J. Kurian, Atiq ur Rehman Bhatti, Mohammed Ali Alvi, Henry H. Ting, Curtis Storlie, Patrick M. Wilson, Nilay D. Shah, Hongfang Liu, Mohamad Bydon.

Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis, Mayo Clinic Proceedings, V. 95, Is. 11, 2020. P. 2370-2381, ISSN 0025-6196, <https://doi.org/10.1016/j.mayocp.2020.08.022>.

9. Li Cuilian, Chen Li Jia, Chen Xueyu, Zhang Mingzhi, Pang Chi Pui, Chen Haoyu. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill. 2020;25(10):pii=2000199. <https://doi.org/10.2807/1560-7917.ES.2020.25.10.2000199>.

10. Walker, A, Hopkins, C, Surda, P. The use of Google Trends to investigate the loss-of-smell-related searches during COVID-19 outbreak. Int Forum Allergy Rhinol. 2020; 10: 839– 847. URL:

<https://onlinelibrary.wiley.com/doi/pdf/10.1002/alr.22580>.

11. Mavragani A/ Tracking COVID-19 in Europe: Infodemiology Approach. JMIR Public Health Surveill 2020;6(2):e18941. URL:

<https://publichealth.jmir.org/2020/2/e18941>. DOI: 10.2196/18941.

12. Ayyoubzadeh S, Ayyoubzadeh S, Zahedi H, Ahmadi M, R Niakan Kalhori S. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. JMIR Public Health Surveill 2020;6(2):e18828. URL: <https://publichealth.jmir.org/2020/2/e18828>. DOI: 10.2196/18828.

13. Mavragani, A., Gkillas, K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 10, 20693 (2020). <https://doi.org/10.1038/s41598-020-77275-9>.

14. Tracking covid-19 excess deaths across countries. The Economist (27.02.2021).

URL: [https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker?fbclid=IwAR2WV9O-](https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker?fbclid=IwAR2WV9O-WpDel4laHKzC3ELnVMoI1RAzlQlaWi9_9FQ_4XHZ6s6MGdgy0gM)

[WpDel4laHKzC3ELnVMoI1RAzlQlaWi9_9FQ_4XHZ6s6MGdgy0gM](https://www.economist.com/graphic-detail/coronavirus-excess-deaths-tracker?fbclid=IwAR2WV9O-WpDel4laHKzC3ELnVMoI1RAzlQlaWi9_9FQ_4XHZ6s6MGdgy0gM).

Received 12.04.2021.

Accepted 15.04.2021.

Системный подход к анализу данных по развитию и прогнозированию пандемии

Проанализированы данные по развитию пандемии коронавирусной инфекции в Украине и мире. Предложен системный подход, предусматривающий привлечение к анализу динамики заболеваемости и смертности вместе с официальной статистикой Минздрава данных других государственных реестров, модельных оценок, результатов поисковых запросов и другой открытой информации.

System approach to data analysis of pandemic development and forecasting

The Covid-19 pandemic is one of the greatest challenges to humanity of last decades. Effective prevention of its development is the problem of decision-making with many criteria, high uncertainty of the results of a choice, insufficiently understood feedbacks. There has been collected statistics on the main indicators of the pandemic over the past year. However, according to most researchers, the initial data on the number of infected and fatal cases are significantly

underestimated. This makes inaccurate other important indicators, in particular, those that characterize the nature of the dynamics, the rate of infection and its mortality. At the same time, a large number of research results provide additional sources of information to improve the quality of pandemic analysis and forecasts. Mathematical models infections spreading have been significantly developed, which make it possible to refine individual indicators.

According to model estimates, the maximum number of daily cases can be 50 times higher than the official data, and the dates of maxima can be up to 4-5 weeks earlier. For Ukraine, the highest values of "excess mortality" (as a percentage of the average level in 2015 - 2016) were observed in September (15.8%), October (20.3%), November (33.7%) and December (29.2%). In January 2021, according to incomplete preliminary data it is expected the significant reducing of the indicator to the value less than 2%. Per 100 thousand inhabitants, the "excess mortality" in 2020 was about 38.8 thousands (in the first 5 months there were negative values of the indicator), and for the period from June 2020 to January 2021 - about 53.5 thousands, or 14.1% of the base level for this period. An analysis of Google's search queries gives grounds to assume that in early January 2020, the pandemic spread in many countries on different continents, and precautionary measures were taken too late.

The obtained data show that the involvement of additional sources of information makes it possible to compensate for the imperfections of official operational data and to understand better the patterns of occurrence and development of Covid-19 pandemic. At the same time, a system approach to decision-making on prevention the development of a pandemic should take into account the information on morbidity and mortality statistics as well as other information, in particular on virus mutations, re-infection, vaccination, testing, socio-economic consequences etc. This should be the subject of further research.

Бахрушин Володимир Євгенович – професор, доктор фізико-математичних наук, професор кафедри системного аналізу та обчислювальної математики Національного університету «Запорізька політехніка».

Бахрушин Владимир Евгеньевич - профессор, доктор физико-математических наук, профессор кафедры системного анализа и вычислительной математики Национального университета «Запорожская политехника».

Volodymyr Bakhrushyn – professor, D.Sc. (Phys. & Math.), professor of the System Analysis and Calculative Mathematics department of the Zaporizhzhia Polytecnic National University.