From data to decisions: the vital role of digital epidemiology for strengthening public health

Syed Shah Areeb Hussain, R. Pushpa, Amit Sharma, C. P. Yadav

Submitted to: Journal of Medical Internet Research
on: February 20, 2025

Table of Contents

From data to decisions: the vital role of digital epidemiology for strengthening public health

Syed Shah Areeb Hussain[1, 2] BEng, MSc; R. Pushpa[1] BTECH, MSc; Amit Sharma[3] BSc, MSc, PhD; C. P. Yadav[1] BSc, MSc, PhD

[1] National Institute of Cancer Prevention and Research Noida IN
[2] Academy of Scientific and Innovative Research Ghaziabad IN
[3] International Centre for Genetic Engineering and Biotechnology New Delhi IN

Corresponding Author:
C. P. Yadav BSc, MSc, PhD

National Institute of Cancer Prevention and Research
ICMR-National Institute of Cancer Prevention and Research
Sector 39, Noida
Noida
IN

Abstract

The academic landscape has witnessed an overwhelming surge in published research, with over 64 million scholarly articles added since 1996 and a growth rate of over 22% in the last five years (2018-2022). This rapid expansion of data represents a vast repository of knowledge, yet it also poses significant challenges in synthesizing and extracting meaningful insights. Traditional methods like systematic reviews, while valuable, often capture only a fraction of the available information due to methodological constraints. In contrast, the rise of digital tools such as interactive dashboards offers a more dynamic and comprehensive approach to data integration. These tools enable the visualization, analysis, and dissemination of large datasets in an accessible format, expanding their reach and utility. Digital epidemiology, by leveraging such advancements, is poised to transform medical research, offering innovative solutions for data synthesis and contributing to more informed, evidence-based outcomes in healthcare.

(JMIR Preprints 20/02/2025:72882)
DOI: https://doi.org/10.2196/preprints.72882

Original Manuscript

From data to decisions: the vital role of digital epidemiology for strengthening public health

Sayed Shah Areeb Hussain[1, 3], R. Pushpa[2], Amit Sharma[4] & Chander Prakash Yadav[2, 3]

1-ICMR-National Institute of Malaria Research, New Delhi, India
2-ICMR-National Institute of Cancer Prevention and Research, Noida, India
3- Academy of Scientific and Innovative Research (AcSIR), India
4- Molecular Medicine Division, International Centre for Genetic Engineering and Biotechnology, New Delhi, India

*Corresponding author: Chander Prakash Yadav
Email ID: cpyadav123@gmail.com

Abstract

The academic landscape has witnessed an overwhelming surge in published research, with over 64 million scholarly articles added since 1996 and a growth rate of over 22% in the last five years (2018-2022). This rapid expansion of data represents a vast repository of knowledge, yet it also poses significant challenges in synthesizing and extracting meaningful insights. Traditional methods like systematic reviews, while valuable, often capture only a fraction of the available information due to methodological constraints. In contrast, the rise of digital tools such as interactive dashboards offers a more dynamic and comprehensive approach to data integration. These tools enable the visualization, analysis, and dissemination of large datasets in an accessible format, expanding their reach and utility. Digital epidemiology, by leveraging such advancements, is poised to transform medical research, offering innovative solutions for data synthesis and contributing to more informed, evidence-based outcomes in healthcare.

Background:

Digital epidemiology is set to revolutionize medical research by transforming data collection, presentation, storage, and analysis methods, leading to more meaningful outcomes. The foundations for digital epidemiology were laid down in the early 2000s with the rise in user-generated content on the internet as well as the rising trend of data digitalization. The advent of web 2.0 technologies led to a rapid increase in user interaction and content creation which prompted early applications of digital epidemiology such as Google Flu trends and social media monitoring models. With recent revolutions in the digital landscape, the scope of digital epidemiology has expanded from merely monitoring infectious diseases, to a broad range of health issues such as disease forecasting, early warning system, research project management, co-occurrence and environmental health.

What is Digital Epidemiology?

Broadly, digital epidemiology has been defined as epidemiology that employs digital methods for collection and analysis of data [1]. A more specific definition, however, identifies it as epidemiology based on data generated outside the traditional public health system [2]. The two definitions differ in their scope, nevertheless, they highlight two important defining features of digital epidemiology – the source of the data, and the methods employed in its collection, representation and analysis.

I. *Digital sources of data:* The strength of digital epidemiology lies in its ability to extract epidemiological information from non-conventional sources such as social media, search engines, mobile applications etc. User-generated content in social media websites can be analysed to track disease spread [3–6], health behaviour [7,8] as well as public sentiment [9–12], whereas health related searches on the internet may indicate emerging health issues or disease outbreaks [13–15]. However, it is important to note that data from these sources lack threat validation and cross-verification, often leading to population and information biases [16].
II. *Digital methods in epidemiology:* The rapid increase in digital data coincides with advancements in computing power, machine learning and artificial intelligence. These advancements have enabled rapid and efficient extraction, presentation and analysis of large amounts of data, thereby igniting the interest of researchers to glean what information may be available in the treasure trove of digital data. Hence, we have seen a significant upsurge in the use of digital technologies in health research in the past decade such as the use of machine learning for big data and language analysis [13,17,18], applying AI image recognition techniques

for vector/parasite identification [19,20], as well as the use of digital dashboards for dissemination of health information [21–23].

Transforming public health surveillance through digital epidemiology

Traditional public health surveillance relies heavily on data from health centres, paper-based surveys and disease registries which undergo rigorous validation, and provide fairly specific and reliable information for public health decision making. However, these approaches suffer from delays due to time taken for data collection, are costly and resource intensive, have limited scope and coverage, and only provide a snapshot of the health status at particular points of time. Collating epidemiological data from digital sources has the potential to transform these traditional epidemiological approaches, offering numerous benefits that can significantly improve public health surveillance:

I. **Real-Time & Cost-Effective data collection/analysis:** Online surveys, social media, online search trends, and health-related dashboards, can provide continuous streams of data that can be analysed to detect emerging health threats in real-time. For instance, Google Flu Trends[24] can track influenza activity weeks ahead of traditional surveillance systems. Digital disease dashboards were also used extensively during the Covid-19 pandemic, to track and visualize cases, deaths, recoveries, and other relevant data in real-time[25–30]. Bluetooth and GPS based contact tracing apps were used to provide alerts to individuals who have been in close proximity to a person diagnosed with COVID-19 in several countries [31–38]. These apps also assisted in determining whether medical care was necessary or if self-isolation based on symptoms and contact history was sufficient [39–43]. Therefore, digital epidemiology can reduce the need for extensive fieldwork and data collection efforts and minimize the resources required for manual data processing and interpretation.

II. **Enhanced surveillance and early warning:** Digital epidemiology enables the development of more sophisticated surveillance systems that can provide early warnings about potential disease outbreaks. By analysing data from multiple sources, a public health official can identify patterns and anomalies that may indicate the onset of an epidemic. For example, platforms like HealthMap[44] aggregate data from various online sources to provide real-time information on emerging health threats globally.

III. **Enhanced information dissemination:** Online databases and portals can provide access to vast amounts of health data, enabling researchers to conduct large-scale epidemiological studies and share findings quickly[23]. Digital epidemiology also empowers individuals by

providing them with the tools and information needed to monitor their own health and contribute to public health efforts.

Limitations of Internet Data used in Digital Epidemiology

Despite the strong potential of digital data derived from social media activity, internet searches etc. in transforming epidemiology, they suffer from some inherent flaws. Such datasets lack standardized control for validation and cross-verification which can often lead to lower accuracy and inconsistent conclusions. User-generated content can be fraught with internal biases of the users and may lead to misinformation, over-exaggeration or misrepresentation of information, thereby leading to false conclusions. At the same time, these datasets suffer from significant population bias and are not representative of the general population. For example, a vast majority of the users on social networking websites lie within a particular demographic based on their age and their socio-economic status. Factors such as the geographic location and internet access can also perpetuate inequalities in health outcomes in such datasets. Finally, while such data can be useful for interpreting the population concerns, attitudes and behaviour, their clinical relevance is often limited.

Health-Research Dashboards: The next step in Digital Epidemiology

In the realm of academic research, the influx of scholarly articles is staggering, with millions being added annually. An estimate suggests that over 64 million academic papers have been published since 1996, showcasing a cumulative growth of over 22% in the last five years alone (2018-2022), with a notable surge observed in 2021 (7.62%). China, the United States, India, the United Kingdom, and Germany collectively contributed to approximately 60% of the total published articles in 2022. The proliferation of academic journals has also witnessed a substantial rise. In 2010, there were approximately 24,552 journals worldwide, a number that has more than doubled to over 46,736 by 2020[45]. This exponential growth underscores the vast reservoir of information available, akin to an ever-expanding ocean. There is an urgent imperative to effectively harness this information for scientific analysis and clinical decision-making. As compared to the other non-conventional sources of data for digital epidemiology (social media, internet searches etc.), these provide a significantly more reliable and validated form of information.

The COVID-19 pandemic encouraged the use of digital dashboards for disseminating health information[25–27,30,46]. By integrating epidemiological & statistical methods into a digital interface, these dashboards can swiftly deliver tailored information to users, be it researchers or medical practitioners unfamiliar with complex methodologies. This approach enables data retrieval in various

formats, including time series graphs and charts, geospatial heat maps, real-time metrics as numerical indicators, and other interactive visuals [47]. However, the potential of a dashboard extends far beyond mere data retrieval. It facilitates proactive disease forecasting, enabling timely interventions, early outbreak detection, and continuous monitoring of population health trends. In recent years, the advent of digital epidemiology has marked a paradigm shift in public health surveillance and research methodologies, promising a future where data-driven insights drive actionable interventions for the benefit of society.

These dashboards can prove pivotal in enhancing our comprehension of disease dynamics as well as support their surveillance and control in a number of ways:

I. **Enhancing accessibility of Systematic Review data:** Systematic reviews are rigorous and complex methods for synthesizing or summarizing all research information on specific topics. Systematic reviews are considered the highest level of evidence and can be instrumental in policy-making or clinical decision-making. However, their reach to end users, such as program personnel and healthcare professionals, is often limited, likely due to complex methodologies, passive or non-interactive presentation of data, and fragmented information. A research dashboard can help overcome these barriers by presenting relevant information on prominent research questions within a disease domain from all systematic reviews available in an interactive and user-friendly manner. This can significantly enhance the accessibility and retrieval of data in systematic reviews according to the specific objectives and needs of the user.

II. **Integrating all disease associated information:** Integrating all related data about a particular disease is crucial for gaining a comprehensive understanding of its dynamics. For instance, malaria, a vector-borne disease, is influenced by various factors such as vector distribution (e.g., different species of Anopheles), ecological conditions (e.g., rainfall, humidity, temperature), malaria control activities (e.g., distribution of LLINs, frequency of IRS, mass surveys, treatment policies), drug resistance, and comorbidities (e.g., malnutrition). Such data can be extracted from research papers, reports, and public health surveys data and may be integrated in a health research dashboard to provide a comprehensive picture of malaria transmission. Such a holistic understanding of the disease and its related factors would be crucial for more effective decision-making and strategic planning for malaria control and prevention efforts[48].

III. **Collating data from public health surveys:** Many countries conduct periodic health surveys for various needs such as resource allocation, policy formation, identification of health trends, and planning. For example, in India, multiple surveys such as the National Family Health Survey (NFHS), District Level Household and Facility Survey (DLHS), Annual Health Survey (AHS), Global Adult Tobacco Survey (GATS), Clinical, Anthropometric, and Biochemical (CAB) Survey, and National Mental Health Survey (NMHS) are regularly conducted to meet diverse requirements. These surveys are rich in information and can be utilized in different contexts to enhance medical research. However, a significant barrier is the lack of awareness among researchers, especially medical and biomedical scientists, regarding the complex methodologies of these surveys. This includes understanding sampling weights, identifiers, merging different datasets, and identifying variables of interest among hundreds of variables. By utilizing digital dashboards, this vast amount of information can be summarized and presented to end users, enabling them to effectively use the data for better research outcomes[49–54].

IV. **Enhanced data collection and real-time monitoring:** Field based or patient data in medical research is often collected as paper-based Case Report Forms (CRF) which is then digitized manually. This process is highly time-consuming and inefficient, leading to significant delays in data presentation and analysis. Furthermore, in studies that require data collection over several months or years, with multiple follow-ups, conventional methods of data collection can become particularly cumbersome. With increased access to technology, it has become significantly easier to directly capture such data in digital forms using mobile hand-held devices. Such digital forms can be directly linked to the health research dashboard to facilitate real-time data entry and monitoring. Real-time data capture not only streamlines the data collection process but also enhances the ability to monitor study progress and make timely adjustments, ultimately leading to more accurate and actionable research outcomes.

Platforms for designing digital epidemiology dashboards

In the past, dashboards were developed using high-level programming languages such as HTML, CSS, JavaScript, and C. Today, however, there are many platforms available that streamline the dashboard development process, including Tableau, Power BI, Google Looker Studio (formerly Google Data Studio), R Shiny and Microsoft Excel[55–58]. Tableau, Power BI, and Google Looker

Studio are prominent business intelligence and data visualization tools that enable users to explore and analyze data in depth to draw meaningful conclusions.

These platforms support a wide range of data sources, including Excel, CSV, text files, Google Sheets, SQL Server, MySQL, PostgreSQL, Oracle, Google BigQuery, Snowflake, Amazon Redshift, and Microsoft Azure. Although primarily designed for business analytics, these tools can also be effectively used for health analytics, leveraging their robust data visualization and analysis features.

R Shiny is specialized frameworks used to develop customized web applications for study data. Recognized as leading tools in medical data analysis, R offer an extensive range of techniques and tools. These range from simple tabulations to highly advanced statistical and machine learning methods. Working with R Shiny can be challenging, but it provides unparalleled flexibility and options. Users can leverage numerous libraries to perform countless operations on their data, making these tools invaluable for sophisticated data analysis and visualization.

Conclusion

The volume of research material has increased exponentially in the last couple of days, however, its translation into decision-making and policy formation is still relatively limited, likely due to difficulty in accessing and interpreting this data. The present commentary explores how digital dashboards may be used to make information this information for accessible and comprehensible for policy makers as well as researchers. Research based health information dashboards can compile highly reliable and detailed health information for utilization by public health systems, that would otherwise require vast amounts of resources and efforts to collect. As we advance towards more intelligent and effective disease surveillance, monitoring and control, such systems can be instrumental in developing the knowledge base and analysis to facilitate better disease control and mitigation.

Declaration

Ethics Approval

Not applicable as it is review article

Consent for publication

All authors have given their consent for publication

Availability of data and material

All data are given in manuscript.

Competing interests
The authors declare that there is no competing interest.

Financial Support
This research received no specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author Contribution
CPY conceived, designed and reviewed the study. CPY, SSAH and RP analyzed the data and wrote the manuscript equally. AS critically appraised and edited the manuscript. All authors participated in finalizing the draft.

Conflict of Interest
There is no conflict of interest

Figure Permissions
No image has been taken from other sources; we create all images/illustrations for this study

Acknowledgment

We are also thankful to ICMR-NIMR and ICMR-NICPR for providing all the logistical support involved in the study.

References

1. Eckhoff, P. A. & Tatem, A. J. Digital methods in epidemiology can transform disease control. *Int Health* **7**, 77–78 (2015).

2. Salathé, M. Digital epidemiology: what is it, and where is it going? *Life Sci Soc Policy* **14**, 1 (2018).

3. Peng, Z., Wang, R., Liu, L. & Wu, H. Exploring Urban Spatial Features of COVID-19 Transmission in Wuhan Based on Social Media Data. *ISPRS International Journal of Geo-Information* **9**, 402 (2020).

4. van Heerden, A. & Young, S. Use of social media big data as a novel HIV surveillance tool in South Africa. *PLoS One* **15**, e0239304 (2020).

5. Ye, X., Li, S., Yang, X. & Qin, C. Use of Social Media for the Detection and Analysis of Infectious Diseases in China. *ISPRS International Journal of Geo-Information* **5**, 156 (2016).

6. Cowley, L. A. *et al.* Genomics, social media and mobile phone data enable mapping of SARS-CoV-2 lineages to inform health policy in Bangladesh. *Nat Microbiol* **6**, 1271–1278 (2021).

7. Huang, X. *et al.* Can online self-reports assist in real-time identification of influenza vaccination uptake? A cross-sectional study of influenza vaccine-related tweets in the USA, 2013–2017. *BMJ Open* **9**, e024018 (2019).

8. Legeby, A. *et al.* New urban habits in Stockholm following COVID-19. *Urban Stud* **60**, 1448–1464 (2023).

9. Han, X., Wang, J., Zhang, M. & Wang, X. Using Social Media to Mine and Analyze Public Opinion Related to COVID-19 in China. *Int J Environ Res Public Health* **17**, 2788 (2020).

10. Hung, M. *et al.* Social Network Analysis of COVID-19 Sentiments: Application of Artificial Intelligence. *J Med Internet Res* **22**, e22590 (2020).

11. Kim, I.-H., Feng, C.-C., Wang, Y.-C., Spitzberg, B. H. & Tsou, M.-H. Exploratory Spatiotemporal Analysis in Risk Communication during the MERS Outbreak in South Korea. *The Professional Geographer* **69**, 629–643 (2017).

12. Pruss, D. *et al.* Zika discourse in the Americas: A multilingual topic analysis of Twitter. *PLoS One* **14**, e0216922 (2019).

13. Milinovich, G. J. *et al.* Using internet search queries for infectious disease surveillance: screening diseases for suitability. *BMC Infect Dis* **14**, 690 (2014).

14. Aiken, E. L. *et al.* Real-time estimation of disease activity in emerging outbreaks using internet search information. *PLOS Computational Biology* **16**, e1008117 (2020).

15. Lopreite, M., Panzarasa, P., Puliga, M. & Riccaboni, M. Early warnings of COVID-19 outbreaks across Europe from social media. *Sci Rep* **11**, 2147 (2021).

16. Jing, F. *et al.* Using geospatial social media data for infectious disease studies: a systematic review. *Int J Digit Earth* **16**, 130–157 (2023).

17. Lu, F. S. *et al.* Accurate Influenza Monitoring and Forecasting Using Novel Internet Data Streams: A Case Study in the Boston Metropolis. *JMIR Public Health Surveill* **4**, e4 (2018).

18. Yuan, Q. *et al.* Monitoring Influenza Epidemics in China with Search Query from Baidu. *PLoS One* **8**, e64323 (2013).

19. Madhu, G., Mohamed, A. W., Kautish, S., Shah, M. A. & Ali, I. Intelligent diagnostic model for malaria parasite detection and classification using imperative inception-based capsule neural networks. *Sci Rep* **13**, 13377 (2023).

20. Kittichai, V. *et al.* Deep learning approaches for challenging species and gender identification of mosquito vectors. *Sci Rep* **11**, 4838 (2021).

21. Thorlund, K. *et al.* A real-time dashboard of clinical trials for COVID-19. *The Lancet Digital Health* **2**, e286–e287 (2020).

22. Cheng, C. K. *et al.* Digital Dashboard Design Using Multiple Data Streams for Disease Surveillance With Influenza Surveillance as an Example. *Journal of Medical Internet Research* **13**, e1658 (2011).

23. Yadav, C. P. & Sharma, A. National Institute of Malaria Research-Malaria Dashboard (NIMR-MDB): A digital platform for analysis and visualization of epidemiological data. *Lancet Reg Health Southeast Asia* **5**, 100030 (2022).

24. Google Flu Trends Estimates - Google Public Data Explorer. https://www.google.com/publicdata/explore?ds=z3bsqef7ki44ac_#!ctype=l&strail=false&bcs=d&nselm=h&met_y=flu_index&scale_y=lin&ind_y=false&rdim=region&idim=region:FR-A&ifdim=region&hl=en_US&dl=en_US&ind=false.

25. Published: Global COVID-19 Tracker. *KFF* https://www.kff.org/coronavirus-covid-19/issue-brief/global-covid-19-tracker/ (2024).

26. COVID - Coronavirus Statistics - Worldometer. https://www.worldometers.info/coronavirus/.

27. Covid19 Dashboard | Home. https://covid19dashboard.mohfw.gov.in/.

28. COVID-19 Map. *Johns Hopkins Coronavirus Resource Center* https://coronavirus.jhu.edu/map.html (2024).

29. COVID-19 advice and services. *nhs.uk* https://www.nhs.uk/covid-19-advice-and-services/ (2023).

30. COVID-19 cases | WHO COVID-19 dashboard. *datadot* https://data.who.int/dashboards/covid19/cases.

31. COVIDSafe. *Wikipedia* (2024).

32. Open-Source Project Corona-Warn-App. https://www.coronawarn.app/en/.

33. TraceTogether. *Wikipedia* (2024).

34. Aarogya Setu. https://aarogyasetu.gov.in/.

35. Apple releases new COVID-19 app and website based on CDC guidance. *Apple Newsroom* https://www.apple.com/newsroom/2020/03/apple-releases-new-covid-19-app-and-website-based-on-CDC-guidance/.

36. Correspondent, H. B. Coronavirus: S’pore Government to make its contact-tracing app freely available to developers worldwide. *The Straits Times* (2020).

37. Singapore wants everyone to download Covid-19 contact tracing apps. https://qz.com/1842200/singapore-wants-everyone-to-download-covid-19-contact-tracing-apps.

38. Covid-19: Germany launches smartwatch app to monitor coronavirus spread | The Star. https://www.thestar.com.my/tech/tech-news/2020/04/07/covid-19-germany-launches-smartwatch-app-to-monitor-coronavirus-spread.

39. Buoy Health: Check Symptoms & Find the Right Care. https://www.buoyhealth.com.

40. Health. Powered by Ada. *Ada* https://ada.com/.

41. Kent, C. Babylon Health launches Covid-19 symptom tracker. *Medical Device Network* https://www.medicaldevice-network.com/news/babylon-covid-19/ (2020).

42. Healthcare, I. Covid-19 symptom checker. https://info.isabelhealthcare.com/covid-19-symptom-checker.

43. HealthTap. HealthTap Announces Free COVID-19 Virtual Doctor Visits In Partnership With Doctors Across U.S. https://www.prnewswire.com/news-releases/healthtap-announces-free-covid-19-virtual-doctor-visits-in-partnership-with-doctors-across-us-301030189.html.

44. HealthMap | Flu Map | Contagious Disease Surveillance | Virus Awareness. http://healthmap.org.

45. Number of Academic Papers Published Per Year – WordsRated. https://wordsrated.com/number-of-academic-papers-published-per-year/ (2023).

46. Ravinder, R. *et al.* An adaptive, interacting, cluster-based model for predicting the transmission dynamics of COVID-19. *Heliyon* **6**, e05722 (2020).

47. Yadav, C. P. & Sharma, A. National Institute of Malaria Research-Malaria Dashboard (NIMR-MDB): A digital platform for analysis and visualization of epidemiological data. *Lancet Reg Health Southeast Asia* **5**, 100030 (2022).

48. Rahi, M. & Sharma, A. For malaria elimination India needs a platform for data integration. *BMJ Glob Health* **5**, e004198 (2020).

49. NFHS-5_Phase-II_0.pdf.

50. Annual Health Survey (AHS) Data Released! https://www.icpsr.umich.edu/web/DSDR/cms/3920.

51. District level household project | International Institute for Population Sciences (IIPS).

https://iipsindia.ac.in/content/district-level-household-project.

52. Global-Adult-Tobacco-Survey-Second-Round-India-2016-2017.pdf.
53. Open Government Data (OGD) Platform India. https://data.gov.in (2022).
54. National Mental Health Survey, 2015-16 - Mental Health Systems_0.pdf.
55. Power BI - Data Visualization | Microsoft Power Platform. https://www.microsoft.com/en-us/power-platform/products/power-bi.
56. Business Intelligence and Analytics Software | Tableau. https://www.tableau.com/.
57. Looker Studio. https://lookerstudio.google.com/navigation/reporting.
58. Shiny - RStudio. https://www.rstudio.com/products/shiny/.