

Digital Epidemiology in the Global South: Harnessing AI and Infodemiology to Counteract Urban Adolescent Non Communicable Disease (NCD) Risks

By Dr.Krithiga Sekar, Dr. Nivetha S, Dr. Kumaresh Purushothaman, Madha Dental College and Hospital, India

Published by Global South Academia — October 2026

Abstract

While the Global South is experiencing rapid urbanization, non-communicable diseases (NCDs) are becoming the main health issue. At the same time, the availability of mobile connectivity is causing urban adolescents to be subtly marketed towards new nicotine products and carcinogenic substances. Since traditional surveillance methods, which depend on repeated cross-sectional surveys, are not appropriate for tracking these algorithmically driven behavioural changes, this article suggests that digital epidemiology and infodemiology should be introduced into the health systems of low- and middle-income countries. By using longitudinal search interest data together with machine learning (ML) classifiers such as Random Forest and supervised analysis, municipal authorities will be able to identify the hidden pathways of digital promotion, predict adolescent engagement, and determine disease risk patterns. We provide a practical plan for municipal health departments, and at the same time examine the methodological limitations of search data and the key ethical problems relating to data colonialism and algorithmic bias. Moving from a reactive approach based on clinical monitoring to one that makes use of AI for public health surveillance can help to protect adolescent populations and advance health equity in areas that are rapidly urbanizing.

Keywords: *Digital epidemiology, Infodemiology, Machine learning, Urban health, Adolescent, Non-communicable diseases*

The Changing Urban Health Landscape in the Global South

Urban centres in the Global South are experiencing extremely rapid urbanisation and digital change. In the past, public health systems in low- and middle-income countries (LMICs) had been set up with a focus on controlling infectious diseases and had been tailored to deal with sanitation and maternal health (Rossier et al., 2014). Today, these systems are facing a double challenge since non-communicable diseases (NCDs), such as cardiovascular diseases, metabolic syndromes and oral or head-and-neck cancers, have become the main features of the disease patterns (Galea & Vlahov, 2005; Mocumbi et al., 2019). The World Health Organization (2021) states that NCDs are responsible for more than 70% of all deaths before people reach old age, and nearly 85% of these deaths which are preventable take place in developing economies. The main causes are fundamental changes in the health of adolescents brought about by unregulated commercial environments (Kickbusch et al., 2016). The quick fall in the price of mobile data, together with the widespread availability of cheap smartphones, has meant that urban adolescents are now in constant contact with digital platforms. Companies that are promoting new nicotine products, electronic nicotine delivery systems (ENDS), novel smokeless tobacco and ultra-processed food items have redirected their advertising efforts from traditional broadcasting media to decentralised digital environments

(Soneji et al., 2018). Recent global surveys of young people show that a large number of urban adolescents in middle-income countries are regularly seeing promotional material about nicotine on social networks (Camenga et al., 2018; Vassey et al., 2025). These companies often avoid breaking national advertising bans by using partnerships with influencers, peer-to-peer distribution and algorithmic micro-targeting (Hoang, 2023; World Health Organization, 2021). Since these marketing activities take place silently on encrypted or temporary digital platforms, they lead to early experimentation, the start of use and the development of dependence among urban youth long before the health effects of this become evident clinically.

The Surveillance Blindspot in Conventional Public Health

Health systems at the municipal level in the Global South encounter a fundamental and operational problem in that they cannot keep up with algorithmic marketing. Traditional methods of health surveillance place great emphasis on periodic national family health surveys, paper questionnaires administered in schools and hospital discharge records. Although these tools offer rigorous and highly validated demographic baseline data, they are essentially retrospective and require a lot of resources, with several years typically passing between the first collection of the data, the statistical analysis and the following implementation of policy (Eysenbach, 2009; Ganasegeran et al., 2025). If a typical cross-sectional survey finds a statistically significant rise in the use of tobacco or nicotine among adolescents in a particular urban area, then the commercial factors responsible for this behaviour have already become firmly established in the community. Moreover, since conventional survey methods are unable to properly record the informal and decentralised digital channels that are influencing current adolescent peer norms, adolescents tend to underreport their exposure to online marketing or their use of new smokeless products when they are surveyed in formal school settings (Camenga et al., 2018). This delay together with the tendency to underreport results in a gap in surveillance which causes municipal health authorities to have to take reactive measures against quickly changing consumption patterns driven by algorithms.

Infodemiology and Machine Learning as Policy Tools

In order to overcome this important delay in information, urban health governance should include infodemiology, which is the study of monitoring web queries, metadata and online interactions in order to dynamically evaluate trends in population health. Real-time insights into public awareness, product uptake and local spikes in behaviour can be obtained by examining search volumes over time (Mavragani & Ochoa, 2019). This passive search analytics method is supplemented by supervised machine learning (ML) techniques, which add an important predictive ability to public health management. Algorithms of this type, for example the Random Forest classifier, are specially designed to analyse complex, non-linear interactions among a large number of demographic and socio-behavioural data items without suffering from the same degree of overfitting as simple regression models (Breiman, 2001; Shi et al., 2022). The ML models are able to accurately detect the exact risk factors that predict youth susceptibility to digital marketing by taking into account a variety of variables such as local search intensity, digital peer engagement measures, the demographics of school districts and patterns of online content consumption (Choi et al., 2021; Fu et al., 2022). Instead of having to wait many years for the effects of a disease to appear, local authorities can make use of this kind of computational modelling to predict consumption trends and take proactive steps to identify vulnerable urban sub-populations (Singh et al., 2023).

Methodological Limitations of Digital Surveillance

It is essential to recognise that infodemiological data is greatly influenced by the proprietary algorithms of search engines and social media platforms (Rovetta, 2024). In the case of low- and middle-income countries, services such as Google Trends suffer from considerable methodological limitations, particularly due to language diversity and the use of local dialects, which can cause the search volume data to be fragmented (Alibudbud, 2023; Nuti et al., 2014). Moreover, since internet access is incomplete in peri-urban or informal areas, infodemiological data tends to reflect mainly the behaviours of more affluent and better-connected groups, thereby risking the systematic underrepresentation of highly vulnerable populations (Nuti et al., 2014). Predictive machine learning models that are trained on this type of digital data are likely to pick up on these biases and could result in incorrect allocation of resources if they are not continually adjusted against actual clinical data (Shi et al., 2022).

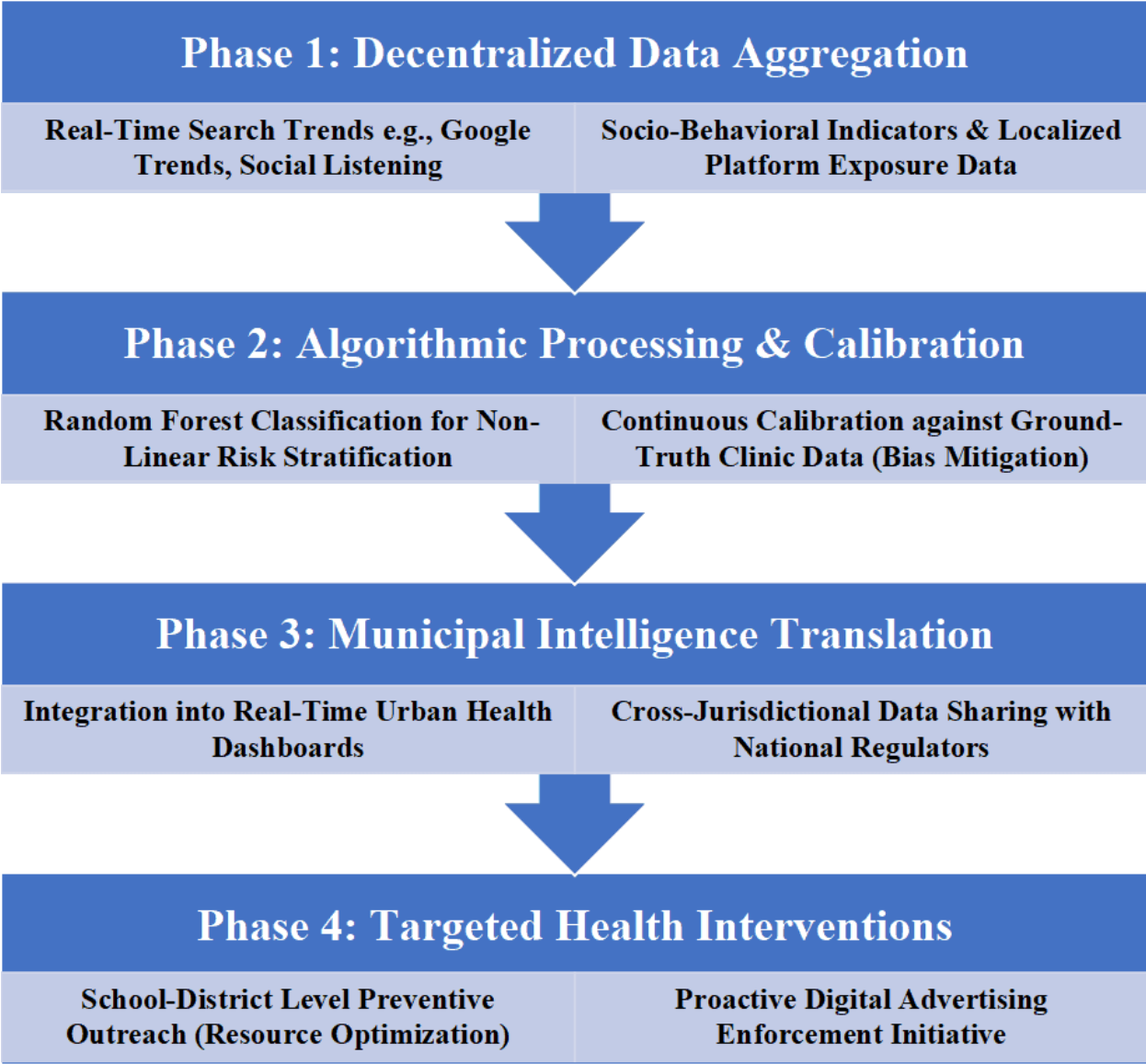
Ethical Imperatives and Governance Complexities

Introducing artificial intelligence into urban health systems presents municipal governments with great ethical and regulatory complexities which they have to deal with. In the first place, there are concerns about data privacy and whether young people give their consent when their online behaviour is automatically tracked. In order to maintain ethical standards, local government frameworks should use only aggregated, anonymised open-source metadata and must avoid collecting identifiable data (Ali et al., 2019). Second, when AI is deployed in the Global South it is necessary to prevent “data colonialism”, which is the situation where multinational technology companies take digital health data from populations in low- and middle-income countries without providing any local and equitable public health benefits (Flores et al., 2023). Third, there is a great risk of “function creep”, that is, the scenario in which surveillance systems which were set up for public health purposes are later put to use by state or local authorities for the punitive monitoring of marginalised young people (Rennie et al., 2023; Sergeant, 2025). In the end, local city governments have only limited authority over cross-border digital platforms that are hosted internationally. If they are to regulate algorithmic marketing, they will have to create new legal frameworks and enter into unprecedented data-sharing agreements with multinational technology companies, a procedure involving considerable geopolitical and commercial friction.

A Strategic Roadmap for Municipal Urban Governance

In order to apply digital epidemiology in urban areas with limited resources, we suggest a four-stage logic model (Figure 1) which links decentralised digital inputs to practical municipal governance.

Figure 1: Operational Logic Model for AI-Driven Municipal Health Surveillance



Executing this framework requires structured, interdisciplinary policy action:

1. Local public health departments should incorporate open-source infodemiology tracking alongside their conventional, clinic-based disease registries into the urban disease control rooms (Ganasegeran et al., 2025).
2. Municipal health directorates and academic public health institutions should work together to apply validated predictive algorithms to local data in order to guide preventive campaigns that are targeted at schools in areas identified as high risk.
3. Municipal authorities should work with the national regulatory agencies in order to keep an eye on and impose penalties on algorithmic marketing of new nicotine products, eliminating the jurisdictional loopholes which businesses have exploited.

The training in the field of public health should include in its primary prevention courses data science, machine learning literacy, and digital ethics, such as those found in health sciences curricula, for example in public health dentistry, community medicine and nursing, so as to prepare practitioners for the governance of public health through algorithms. The combination of fast urbanization and a deep level of digital access bring with it both

opportunities and risks for health governance in the Global South. If municipal authorities continue to depend only on outdated, backward-looking survey methods, then their preventive policies will always fall behind commercial digital marketing and changes in adolescent behaviour. However, by introducing infodemiological surveillance and artificial intelligence into urban health systems and by taking care to avoid ethical problems and algorithmic bias, local authorities can move from a reactive approach involving clinical monitoring to one that is proactive and based on real-time disease prevention, thus safeguarding adolescent populations and advancing health equity.

Declarations

Ethics Statement

This manuscript is a viewpoint/commentary and does not involve new human or animal subjects research. Therefore, institutional ethical approval was not required.

Funding

The authors declare that no financial support or funding was received for the research, authorship, or publication of this article.

Disclosure of Interest

The authors report there are no competing interests to declare.

Data Availability Statement

Data sharing is not applicable to this article as no new primary datasets were generated or analyzed during the preparation of this viewpoint.

Declaration of Generative AI and AI-Assisted Technologies

During the preparation of this work, the author(s) used Google Gemini solely to improve the language, clarity, and grammatical readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content and integrity of the published article. Generative AI was not used to create, alter, or manipulate data, results, or scientific interpretations.

References

1. Hyder AA, Pratt B, Ali J, Kass N, Sewankambo N. The ethics of health systems research in low- and middle-income countries: a call to action. *Glob Public Health*. 2014;9(9):1008-22. doi:10.1080/17441692.2014.931998.
2. Alibudbud R. Google Trends for health research: its advantages, application, methodological considerations, and limitations in psychiatric and mental health infodemiology. *Front Big Data*. 2023; 6:1132764. doi:10.3389/fdata.2023.1132764.
3. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/A:1010933404324.
4. McKelvey K, Popova L, Kim M, Chaffee BW, Vijayaraghavan M, Ling P, et al. Heated tobacco products likely appeal to adolescents and young adults. *Tob Control*. 2018;27(Suppl 1): s41-s47. doi:10.1136/tobaccocontrol-2018-054596.

5. Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health*. 2017;7(3):185-189. doi: 10.1016/j.jegh.2017.06.001.
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.
7. Carneiro HA, Mylonakis E. Google Trends: a web-based tool for real-time surveillance of disease outbreaks. *Clin Infect Dis*. 2009;49(10):1557-1564. doi:10.1086/630200.
8. Hoodbhoy Z, Jeelani SM, Aziz A, Habib MI, Iqbal B, Akmal W, et al. Machine learning for child and adolescent health: a systematic review. *Pediatrics*. 2021;147(1): e2020011833. doi:10.1542/peds.2020-011833.
9. Galea S, Vlahov D. Urban health: evidence, challenges, and directions. *Annu Rev Public Health*. 2005; 26:341-365. doi: 10.1146/annurev.publhealth.26.021304.144708.
10. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58. doi:10.7326/M18-1376.
11. Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating*. 2nd ed. Cham: Springer; 2019. doi:10.1007/978-3-030-16399-0.
12. Kickbusch I, Allen L, Franz C. The commercial determinants of health. *Lancet Glob Health*. 2016;4(12):e895-e896. doi:10.1016/S2214-109X(16)30217-0.
13. Mavragani A, Ochoa G. Google Trends in infodemiology and infoveillance: methodology framework. *JMIR Public Health Surveill*. 2019;5(2):e13439. doi:10.2196/13439.
14. Rothenberg WA, Bizzego A, Esposito G, et al. Predicting adolescent mental health outcomes across cultures: a machine learning approach. *J Youth Adolesc*. 2023; 52:1595-1619. doi: 10.1007/s10964-023-01767-w.link.springer
15. Nuti SV, Waymack JR, Ding Z, Cramer JA. The use of Google Trends in health care research: a systematic review. *PLoS One*. 2014;9(10): e109583. doi: 10.1371/journal.pone.0109583.
16. Soneji S, Yang J, Knutzen KE, Moran MB, Tan AS, Sargent J, et al. Online tobacco marketing and subsequent tobacco use. *Pediatrics*. 2018;141(2): e20172927. doi:10.1542/peds.2017-2927.
17. Donaldson SI, Dormanesh A, Perez C, Majmundar A, Allem JP. Association between exposure to tobacco content on social media and tobacco use: a systematic review and meta-analysis. *JAMA Pediatr*. 2022;176(9):878-885. doi:10.1001/jamapediatrics.2022.2223.
18. Collins L, Glasser AM, Abudayyeh H, Pearson JL, Niaura RS. E-cigarette marketing and communication: how e-cigarette companies market e-cigarettes and the public engages with e-cigarette-related information. *Nicotine Tob Res*. 2019;21(1):14-24. doi:10.1093/ntr/nty284.
19. World Health Organization. *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization; 2021. Available from: <https://www.who.int/publications/i/item/9789240029200>
20. Vogel EA, Barrington-Trimis JL, Vassey J, Soto D, Unger JB. Young Adults' Exposure to and Engagement With Tobacco-Related Social Media Content and Subsequent Tobacco Use. *Nicotine Tob Res*. 2024;26(Suppl 1): S3-S12. doi: 10.1093/ntr/ntad108.
21. Dunbar MS, Davis JP, Rodriguez A, Tucker JS, Seelam R, D'Amico EJ. Exposure and engagement with tobacco-related social media and associations with subsequent tobacco use among young adults: a longitudinal analysis. *Drug Alcohol Depend*. 2020; 213:108072. doi: 10.1016/j.drugalcdep.2020.108072.

22. World Health Organization. *WHO report on the global tobacco epidemic 2021: addressing new and emerging products*. Geneva: World Health Organization; 2021. Available from: <https://www.who.int/publications/i/item/9789240032095>