**Saudi Journal of Medicine and Public Health**

https://saudijmph.com/index.php/pub

https://doi.org/10.64483/202522539

# Algorithmic Fairness in Clinical Predictive Models: A Review of Bias Audits and Mitigation Strategies in Epidemiological Research

**Redha Jaffar Albaqshi [(1)] , Hassan Ali Ahmad Najaei [(2)] , Hani Ghazi Abdulmalik [(3)] , Salman Saeed Saad Alzamaa [(4)] , Saeed Oudah Hammad Alshahrani [(5)] , Mohammad Mosad M Alshsarari [(6)] , Abdullah Saleh Essa Aldurayhim [(7)] , Ahmad Mohammad AlGhazal [(8)] , Khaled Ibrahim Sohail [(9)] , Ibtihal Abdullah Alotibi [(10)] , Ali Essa Dallak [(11)] , Ahmad Ali Ahmad Sahli [(12)] , Mohammed Saleh Najdi [(13)]**

*(1) King Fahad Hospital – Al-Hofuf, Al-Ahsa, Ministry of Health, Saudi Arabia,*
*(2) Damad General Hospital – Jazan Region, Ministry of Health, Saudi Arabia,*
*(3) Maternity and Children Specialized Hospital – Jeddah, Ministry of Health, Saudi Arabia,*
*(4) Al-Harjah General Hospital, Aseer, KSA, Ministry of Health, Saudi Arabia,*
*(5) Al-Maddah General Hospital, Aseer, KSA, Ministry of Health, Saudi Arabia,*
*(6) Security Forces Hospital, Al-Qurayyat, Medical Services, Ministry of Interior, Saudi Arabia,*
*(7) Public Health Department, Riyadh First Health Cluster, Ministry of Health, Saudi Arabia,*
*(8) Salwa General Hospital, Ministry of Health, Saudi Arabia,*
*(9) Salwa Hospital, Ministry of Health, Saudi Arabia,*
*(10) Imam Abdulrahman Al-Faisal Hospital, Riyadh, Ministry of Health, Saudi Arabia,*
*(11) Sabya General Hospital, Sabya, Ministry of Health, Saudi Arabia,*
*(12) Prince Mohammed bin Nasser Hospital, Jazan, Ministry of Health, Saudi Arabia,*
*(13) Al Mahd General Hospital, Ministry of Health, Saudi Arabia*

## Abstract

**Background:** Digital Epidemiology has emerged as a transformative approach to infectious disease surveillance, leveraging digital data streams such as social media, search queries, and mobility patterns. While these methods offer speed and scale, they introduce significant statistical and ethical challenges, particularly bias and fairness concerns in predictive modeling.

**Aim:** This review aims to examine algorithmic fairness in clinical predictive models within epidemiological research, focusing on bias audits and mitigation strategies in the context of Digital Epidemiology.

**Methods:** A comprehensive literature review was conducted, analyzing methodological differences between classical and digital approaches, sources of bias, and corrective strategies. Key themes include representativeness, measurement error, and algorithmic bias in machine learning models trained on digital data.

**Results:** Findings reveal that Digital Epidemiology offers real-time, large-scale data collection but suffers from structural biases due to self-selection, platform design, and digital divides. Bias mitigation is often retrospective, relying on weighting, normalization, and cross-validation. Ethical concerns such as privacy and informed consent intersect with fairness, as predictive models risk amplifying inequities. Integration of classical rigor with digital flexibility and continuous bias audits is essential for equitable outcomes.

**Conclusion:** Digital Epidemiology complements classical methods but requires robust frameworks for bias detection, ethical governance, and algorithmic transparency. Sustained collaboration, standardization, and inclusive data practices are critical to ensure predictive models support fair and actionable public health decisions.

**Keywords:** Digital Epidemiology, Algorithmic Fairness, Bias Mitigation, Infectious Disease Surveillance, Predictive Modeling, Public Health Ethics.

## Introduction

Epidemiology examines the distribution of health events and the factors that shape them within defined populations [1]. Its analytical foundation depends on the systematic collection and interpretation of heterogeneous data sources that include questionnaires, clinical examinations, laboratory findings, and sociodemographic indicators. These data allow epidemiologists to identify risk factors, estimate disease burden, and inform prevention strategies. Over time, the field has expanded in scope and methodology as new forms of data have become available. This expansion has accelerated with advances in computing capacity and the widespread integration of digital technologies into everyday life, giving rise to what is commonly termed Digital Epidemiology [2]. This development has reshaped how population health is studied and has introduced new analytical possibilities alongside substantial methodological and ethical concerns.


*Corresponding author e-mail: redha360@hotmail.com (Redha Jaffar Albaqshi).
Receive Date: 06 December 2025, Revise Date: 28 December 2025, Accept Date: 31 December 2025

Early conceptualizations of Digital Epidemiology emphasized the exploitation of novel digital traces such as mobile phone metadata, online search queries, social media interactions, and other platform-generated information to observe and model health-related phenomena at scale [3]. These data streams offered unprecedented temporal resolution and population coverage, often at lower cost than traditional surveillance systems. Subsequent refinements of the concept distinguished between broad and narrow interpretations. In the broad sense, Digital Epidemiology referred to any epidemiological inquiry that relies on digital data. In the narrow sense, it focused specifically on digital data generated outside formal public health infrastructures and not originally designed for epidemiological analysis [4]. These definitional efforts placed primary attention on where the data originate and whether they exist in digital form, framing the field largely in terms of technological novelty.

Such distinctions have become less analytically useful as digitalization has permeated nearly all domains of data production. Clinical records, population surveys, administrative registries, and even classical field studies are now routinely digitized. The digital format alone no longer differentiates data collected within public health systems from those generated elsewhere. Moreover, epidemiology has historically relied on secondary data sources that were not created exclusively for health research. Long before the digital era, investigators incorporated information on housing conditions, population density, climate patterns, transportation networks, and geographic identifiers to explore associations between environment and disease. The reuse of existing data has therefore always been integral to epidemiological practice, even when such data were collected for administrative or logistical purposes rather than scientific inquiry [4]. In addition, contemporary epidemiology increasingly depends on large-scale clinical and quasi-clinical databases that blur the boundary between traditional and digital approaches. Electronic medical records, prescription databases, insurance claims, and call-center triage logs now support disease surveillance, outcome prediction, and health services research. These resources exemplify the long-standing practice of leveraging routine data for epidemiological aims. The Oxford Record Linkage Study provides a historical illustration, demonstrating how systematically collected hospital records could be repurposed for population-based research with careful methodological oversight [5]. This example underscores that the reuse of data, when guided by rigorous planning and analytical discipline, has been compatible with core epidemiological principles for decades.

The critical distinction between classical and digital approaches therefore lies less in the digital nature of the data and more in how issues of bias are anticipated, identified, and addressed. Bias denotes a systematic deviation from the true population parameters that arises through flaws in study design, data collection, or analysis [6]. Classical Epidemiology has traditionally prioritized bias control through prospective planning. Studies are designed with explicit public health objectives, predefined sampling frames, and standardized measurement protocols. Randomization, stratification, and careful eligibility criteria aim to minimize selection bias and confounding before data are collected. Even when secondary data are used, they often originate from systems developed with statistical representativeness or standardized reporting in mind, such as censuses or meteorological services. By contrast, Digital Epidemiology frequently relies on data that emerge as by-products of digital systems optimized for purposes unrelated to health research. Social media platforms, smartphone applications, wearable devices, and online services generate vast quantities of information, but their user bases are shaped by access to technology, cultural practices, economic status, and platform design. These factors influence who contributes data, how frequently, and in what form. As a result, representativeness cannot be assumed, and systematic biases often become apparent only after analysis has begun. Bias identification and correction in this context are typically retrospective, relying on statistical adjustments, sensitivity analyses, or model-based corrections applied a posteriori [6].

This methodological asymmetry has significant implications for the development of clinical predictive models within epidemiological research. Predictive algorithms trained on digitally derived data may encode and amplify existing social and structural inequalities. If certain groups are underrepresented or misrepresented in the data, model outputs may systematically disadvantage them. Algorithmic bias thus becomes an extension of data bias, translating unequal data generation processes into unequal predictive performance. In classical settings, bias mitigation is embedded in the study design, whereas in digital contexts it often becomes a downstream corrective exercise. This shift challenges established norms of epidemiological validity and raises concerns about fairness, especially when predictive models inform clinical or public health decisions. Opportunities for a priori bias control in Digital Epidemiology do exist but remain limited. In some cases, researchers can influence recruitment strategies, platform design, or user engagement to reduce dropout and improve coverage. However, such control is feasible only when investigators collaborate closely with data-generating systems. More often, researchers inherit data whose structure, quality, and population coverage they cannot modify. Furthermore, digital data frequently

capture distinct subpopulations defined by age, socioeconomic status, geographic location, or health-seeking behavior. These subpopulations may not align with the target populations of epidemiological inference, complicating generalizability and fairness assessments [1][2][3].

Ethical considerations further differentiate digital from classical approaches. Digital data are sometimes collected without explicit informed consent for research use, relying instead on broad user agreements or passive data capture. This practice raises concerns about autonomy, privacy, and trust. When such data feed into predictive models, ethical issues intersect with algorithmic fairness. Models may produce accurate predictions for the majority while systematically underperforming for marginalized groups, reinforcing disparities in diagnosis, treatment, or resource allocation. Addressing these risks requires methodological frameworks that integrate bias audits, transparency, and mitigation strategies throughout the model lifecycle. These considerations motivate a redefinition of Digital Epidemiology that shifts attention away from the digital format or the institutional origin of data and toward their statistical properties. Defining Digital Epidemiology as the use of data not originally collected with epidemiological statistical rigor foregrounds the core methodological challenge. Such a definition recognizes that the principal difficulty lies in adapting repurposed data to answer population health questions without reproducing or exacerbating bias. It also aligns Digital Epidemiology with broader debates on algorithmic fairness, emphasizing the need for systematic evaluation of how data generation processes influence model behavior.

Under this perspective, two interrelated challenges become central. The first concerns the effective analytical use of secondary digital data while explicitly accounting for the biases embedded in their collection and processing. This task requires robust bias audits, including assessments of representativeness, measurement error, and differential model performance across subgroups. It also demands mitigation strategies that go beyond technical fixes to consider structural determinants of data inequality. The second challenge involves the development of ethical and privacy-preserving methodologies that reconcile the strengths of classical epidemiological design with the scale and granularity of digital data. Achieving this integration is essential for ensuring that predictive models support equitable public health outcomes rather than undermining them [4][5]. In this context, algorithmic fairness is not an ancillary concern but a defining criterion of methodological quality in modern epidemiological research. Predictive models increasingly influence clinical decision-making, risk stratification, and policy planning. If these models rest on biased data and unexamined assumptions, they risk institutionalizing inequity under the guise of objectivity. A refined understanding of Digital Epidemiology, grounded in statistical rigor and ethical accountability, provides a framework for addressing these risks. It emphasizes that fairness must be assessed and enforced at every stage, from data sourcing to model deployment.

Reframing Digital Epidemiology in this way does not reject technological innovation. Instead, it situates innovation within a tradition of critical methodological scrutiny that has long characterized Epidemiology. By acknowledging that many digital data sources lack the safeguards of classical study design, researchers can more transparently evaluate limitations and implement corrective strategies. This approach supports the development of clinical predictive models that are both analytically robust and socially responsible. Ultimately, aligning digital methods with principles of fairness and bias control strengthens the contribution of epidemiological research to public health, ensuring that advances in prediction translate into benefits that are shared across populations rather than unevenly distributed [5][6]. In this paper, infectious diseases are selected as the primary focus because they constitute more than half of published work within Digital Epidemiology [7]. Historically, the surveillance of infectious diseases has depended on resource-intensive and time-consuming infrastructures, including sentinel physician networks, population survey teams, and centralized diagnostic laboratories [8]. These conventional systems, while methodologically rigorous, often suffer from reporting delays, limited geographic coverage, and high operational costs. Such constraints can hinder timely outbreak detection and delay public health responses, particularly in rapidly evolving epidemic contexts. The emergence of digital data streams therefore appeared especially promising for infectious disease epidemiology, as they offered the potential for faster, broader, and more flexible monitoring of disease dynamics.

The early phase of 2020 marked a pivotal moment for Digital Epidemiology in the context of infectious diseases. Initial applications relied on online search queries, social media activity, mobility data, and participatory surveillance platforms to track emerging signals related to COVID-19 [9]. These approaches enabled near real-time observation of public concern, symptom reporting, and behavioral changes at population scale. The COVID-19 pandemic subsequently accelerated the integration of digital data into mainstream epidemiological practice. During this period, digital methods were not merely experimental complements but became essential tools for situational awareness, modeling transmission dynamics, and informing public health interventions. The pandemic demonstrated how traditional surveillance systems could be augmented through the integration of heterogeneous data streams, allowing

faster identification of outbreaks, earlier detection of shifts in transmission, and more adaptive response strategies [10]. At the same time, the pandemic exposed fundamental limitations of both classical and digital systems, particularly with respect to statistical bias. While digital data offered speed and scale, they also revealed deep structural inequalities in data generation and access. Patterns observed in digital traces often reflected differential access to technology, variation in health-seeking behavior, and context-specific platform use rather than true disease incidence. COVID-19 thus served as a natural experiment that highlighted the dual nature of Digital Epidemiology. It demonstrated its capacity to complement established surveillance infrastructures while simultaneously underscoring the need for rigorous bias assessment and mitigation. The crisis reinforced the argument that methodological scrutiny must evolve alongside technological innovation if digital tools are to contribute reliably and fairly to epidemiological inference.

Differences between Classical and Digital Epidemiology become especially pronounced when examined through the lens of bias. In classical frameworks, sampling and representation are addressed through structured study designs. Selection bias and coverage bias arise when participation is non-random or when the sampling frame fails to encompass the target population [74–77]. Clinic-based studies, for example, tend to overrepresent individuals who seek care while excluding healthier populations or those without access to healthcare services. These biases are typically anticipated and mitigated through a priori strategies such as random or stratified sampling and expansion of sampling frames. When biases persist, a posteriori techniques including statistical adjustment, dataset linkage, and reliance on validated self-reported outcomes are applied. In Digital Epidemiology, sampling and representation challenges are often more severe. Participation in online surveys, mobile applications, or social media platforms is driven by self-selection, which disproportionately favors younger, more technologically literate individuals while marginalizing older adults and those with limited internet access. These coverage gaps reflect broader digital divides and complicate generalizability. Bias mitigation in this context requires a combination of proactive and retrospective strategies. Researchers may attempt to analyze random samples within digital platforms or recruit structured cohorts, but more often they rely on weighting schemes, data integration across sources, and continuous audits to assess representativeness. Ethical oversight, transparent data practices, and stakeholder engagement become essential components of bias mitigation, particularly when data are collected through opt-in mechanisms.

Detection and surveillance bias further differentiate classical and digital approaches. In traditional epidemiology, variations in diagnostic practices or monitoring intensity across populations can lead to overestimation of associations between exposures and outcomes [6,77–80]. Such biases are addressed through standardized diagnostic criteria, protocol harmonization, and statistical controls for visit frequency or disease severity. Digital systems face analogous but amplified challenges. Data intensity varies widely across platforms, with heavy technology users generating more frequent signals than less engaged individuals. Wearable devices, symptom-tracking apps, and social media activity can therefore exaggerate disease detection in specific subgroups. Mitigation strategies often rely on normalization techniques, inverse probability weighting, multiple imputation, and cross-validation with independent datasets. Integrating digital signals with traditional surveillance data remains a key strategy for reducing detection asymmetries. Measurement bias represents another critical area of divergence. In classical studies, systematic measurement error can arise from uncalibrated instruments, inconsistent protocols, or observer variability [80]. Although these issues persist, they are typically addressed through standardized tools, personnel training, and calibration procedures. Digital Epidemiology introduces new measurement challenges linked to device heterogeneity and user-generated data. Wearables may differ in accuracy, and self-reported information collected through apps or online platforms can vary widely in reliability. While calibration and labeling standards can be established a priori, much of the corrective work occurs after data collection through cleaning, cross-validation, and statistical correction techniques. Advanced machine learning methods may assist in identifying patterns of measurement error, though their use must be carefully validated to avoid introducing additional bias.

Information and recall bias further illustrate contrasting strengths and weaknesses. Classical epidemiology often relies on retrospective self-reporting, which is vulnerable to inaccurate recall and misclassification [6,77–79]. Digital methods, by capturing data in real time, can reduce recall bias by recording behaviors and symptoms as they occur. However, digital data are frequently unstructured and influenced by social desirability, platform norms, or misinformation. Natural language processing and cross-referencing with passive data sources such as location or mobility records can partially address these issues, but variability in information quality remains a defining challenge. Technological and platform-related biases also play a prominent role. Availability bias occurs when researchers select data sources based on convenience rather than scientific relevance, while platform bias reflects differences in

data generation across systems [3,81,82]. In classical settings, mixed data collection methods can be deliberately designed to balance response rates and data quality. Digital Epidemiology often operates within proprietary and rapidly changing platforms, where data access and structure are shaped by commercial priorities. Mitigating platform bias therefore requires multidisciplinary collaboration, integration of multiple platforms, and post hoc weighting or correction strategies informed by external benchmarks. Attrition and behavioral biases further complicate longitudinal analysis. In traditional studies, participant dropout and socially desirable reporting can distort findings [2,83,84]. Retention strategies and statistical adjustments are commonly used to address these issues. Digital environments amplify attrition risks because users can disengage at any time, often in ways correlated with health status, motivation, or digital literacy. Online behavior is also highly sensitive to external events, such as pandemics, which can trigger transient spikes in searches or posts driven by anxiety rather than actual disease trends. Addressing these dynamics requires flexible analytical approaches, including time-series modeling, data imputation, and integration of complementary data streams.

Causal inference presents another area of contrast. Confounding and temporal biases undermine the ability to establish causal relationships in both classical and digital studies [6,85,86]. Classical epidemiology relies on randomization, matching, and hypothesis-driven designs to limit these biases. Digital data, while often lacking controlled design, offer dense temporal information that can support retrospective and longitudinal analyses at scale. Nevertheless, socioeconomic and contextual factors frequently confound digital signals, and imprecise timing of exposures and outcomes complicates interpretation. Combining time-stamped digital data with structured longitudinal datasets can partially address these challenges. Cognitive biases, including confirmation and anchoring, affect researchers across methodological traditions [87]. Classical studies mitigate these risks through preregistration, blinding, and standardized protocols. Digital Epidemiology faces heightened vulnerability due to rapid data influx and the simultaneous exploration of multiple hypotheses. Early digital signals, often noisy or incomplete, can disproportionately shape subsequent analyses. Preregistration of digital workflows, cross-platform peer review, and continuous validation against independent datasets are essential to counteract these tendencies.

Finally, algorithmic bias has emerged as a defining concern in Digital Epidemiology [80,88]. Classical epidemiology typically employs transparent statistical models whose assumptions can be scrutinized directly. In contrast, machine learning models trained on biased or unrepresentative digital data may systematically disadvantage smaller or marginalized groups. Black box algorithms further obscure bias detection and correction. Mitigation requires diverse and representative training data, external validation, transparency, and continuous model updating. Integrating traditional epidemiological data and leveraging emerging tools, including large language models, may support the identification of anomalies and unexpected patterns, but such tools must be applied within robust ethical and methodological frameworks. Together, these contrasts underscore that Digital Epidemiology offers transformative potential for infectious disease research while posing significant statistical and ethical challenges. Addressing bias is not a secondary technical task but a central requirement for ensuring that digital methods contribute to accurate, fair, and actionable public health knowledge.

**Digital Epidemiology:**

Digital Epidemiology evolved well before the COVID-19 pandemic, driven by the aim to improve infectious disease surveillance by increasing speed, reducing cost, and expanding population coverage. Early initiatives demonstrated that digital tools could complement traditional surveillance when designed with epidemiological rigor. A prominent example is the French sentinel doctor network established in 1984 for influenza monitoring, which relied on representative sampling and bias correction strategies applied both before and after data collection [11]. This contrasted sharply with later approaches such as Google Flu Trends, which depended on online search behavior rather than structured sampling. Although initially promising, Google Flu Trends failed to detect key epidemic dynamics and substantially overestimated influenza cases due to overfitting, spurious correlations, and sensitivity to media coverage [12–14]. These limitations highlighted the risks of relying on opportunistic digital data without sufficient validation. At the same time, experiences such as early media detection of the 2009 influenza outbreak in Mexico demonstrated the potential value of nontraditional data sources, particularly when they capture signals outside formal health systems [16]. Subsequent initiatives, including the CDC FluSight challenge, showed that combining traditional surveillance data with digital and environmental sources could improve predictive performance [17,18]. Participatory surveillance systems such as InfluenzaNet further expanded the field by enabling voluntary self-reporting of symptoms across multiple countries, often detecting outbreaks earlier than official systems [19]. While these approaches offered cost-effective and ethical means of data collection, they remained vulnerable to sampling, attrition, and self-selection biases that required retrospective correction.

Before 2020, Digital Epidemiology research steadily increased, focusing largely on infectious

diseases such as influenza. Studies leveraged data from social media, search engines, news media, and online knowledge platforms to forecast disease trends, analyze seasonality, and detect outbreaks [35]. Despite this growth, large-scale adoption by public health authorities remained limited. Barriers included insufficient validation, lack of funding, fragmented digital infrastructure, and poor interoperability between systems [36]. Although platforms such as ProMED, HealthMap, and GPHIN routinely identified thousands of potential outbreak signals worldwide [39–41], their influence on decision-making remained constrained. The discontinuation of key systems like GPHIN in 2019 exemplified the fragile institutional support for Digital Epidemiology prior to COVID-19 [42]. The COVID-19 pandemic marked a turning point. The global health emergency triggered unprecedented data generation, sharing, and analysis efforts. Governments, academic institutions, and technology companies collaborated to produce detailed case counts, mortality statistics, mobility indicators, and contact tracing tools. Lockdowns shifted much of daily life online, creating vast new sources of health-relevant digital data. As a result, Digital Epidemiology research activity surged during the pandemic [9]. This expansion was fueled by increased data availability, a new culture of openness, and strong political and societal pressure to deploy technological solutions. However, the pandemic also exposed deep methodological weaknesses. Widely used datasets, such as those from the Johns Hopkins University Center for Systems Science and Engineering, enabled rapid modeling but suffered from inconsistencies in testing practices, reporting standards, and death attribution across countries [44,45]. These disparities distorted comparisons and undermined model reliability. Many predictive models developed during the crisis showed high risk of bias, overfitting, and limited clinical applicability [47]. Sampling biases also became evident in mobility data, wearable device usage, and contact tracing apps, which disproportionately excluded older, poorer, and less digitally connected populations [60,61].

Despite significant innovation, Digital Epidemiology tools proved insufficiently mature to fully support large-scale public health action. Although the pandemic accelerated adoption of telemedicine, digital triage, and health information platforms [62–66], sustaining these advances requires long-term investment and structural reform. Following the pandemic peak, research output declined as data access diminished and attention shifted elsewhere. Overall, COVID-19 confirmed both the transformative potential of Digital Epidemiology and the central importance of addressing bias, representativeness, and implementation challenges to ensure durable public health impact.

**Strengths and challenges in Digital Epidemiology:**

Digital Epidemiology is often viewed as inferior to Classical Epidemiology because of concerns related to bias, privacy, ethics, and data quality [67,68]. This perception overlooks the fact that both approaches face structural limitations and offer distinct advantages. Classical Epidemiology benefits from carefully designed studies where bias prevention is embedded at the planning stage, whereas Digital Epidemiology relies heavily on secondary and unstructured data, making bias identification and correction largely retrospective. Despite these challenges, Digital Epidemiology provides access to large-scale, real-time data and captures health-related behaviors and signals that traditional methods cannot easily observe. A central strength of Digital Epidemiology lies in its scope and timeliness. Digital data can cover wide geographic areas, capture rapid behavioral changes, and support real-time surveillance. In contrast, classical studies often suffer from limited sample sizes, restricted spatial coverage, and delayed reporting. However, these advantages come at the cost of weaker validation. Digital models may suffer from overfitting, lack of external validation, and sensitivity to noise. While no epidemiological dataset is free from bias [69], correction strategies are better established and more controllable in classical research than in digital contexts. Post-collection bias correction in Digital Epidemiology is particularly challenging for several reasons. First, biases embedded in digital platforms are difficult to anticipate. User demographics, engagement patterns, and cultural norms vary across platforms and change rapidly. Second, the scale of digital data and the sensitivity of modern computational models can amplify existing biases, making them appear as meaningful signals. Third, many digital studies rely on proxy variables rather than direct measures of health outcomes, and the choice of proxies introduces additional bias [70,71]. These issues are compounded by researchers' limited control over data access, which often depends on private companies or government agencies. Some landmark studies were only possible through direct collaboration with platforms [72], highlighting structural dependencies that shape what research can be conducted.

Despite these limitations, Digital and Classical Epidemiology are not competing paradigms but complementary approaches. Combining the methodological rigor of classical studies with the flexibility and scale of digital data offers a path forward [73]. Focusing on statistical bias provides a practical framework for identifying gaps, prioritizing resources, and guiding methodological development, even though bias categories overlap and are not exhaustive. Future progress depends on deliberate integration strategies. Validation remains essential. Digital data should inform hypothesis generation and

rapid situational awareness, while classical methods should be used to validate findings, especially for populations with limited digital access. Standardization of data collection, metadata, and sharing practices is critical. International initiatives such as the European Health Data Space and updated ECDC directives emphasize interoperability, transparency, and privacy while requiring clearer documentation of potential biases [89,90]. Long-term collaboration with the private sector is necessary to ensure sustainable data access beyond emergencies, while actively managing conflicts of interest [91]. Advances in artificial intelligence offer opportunities to improve disease identification through multimodal data integration, but predictive accuracy must not replace causal reasoning in decision-making [92,93]. Addressing digital exclusion is equally important. Communities that could benefit most from digital tools often face the greatest barriers, as demonstrated during COVID-19 [94]. Public participation can improve data literacy, empower communities, and reduce global inequities in research [69,95,96]. Stronger collaboration across disciplines and institutions is needed to translate research into practice, with particular attention to algorithmic fairness, transparency, and explainability. Clear communication strategies, safeguards against misinformation [97], and transparent model governance are essential to build trust. Continuous performance assessment and real-world testing outside crisis periods ensure that digital tools are reliable, equitable, and ready for deployment when needed.

**Discussion of Digital Epidemiology:**

This discussion emphasizes that effective integration of Classical and Digital Epidemiology requires combining data collected with a priori statistical rigor with heterogeneous data sources created for non-epidemiological purposes, which demand extensive a posteriori debiasing. Surveys, censuses, and environmental data must therefore be analytically aligned with insurance records, helplines, and digital traces such as social media posts. This integration highlights the complementary nature of the two approaches rather than a hierarchy between them. Three conditions are identified as essential for advancing this transition. First, Classical and Digital Epidemiology should be treated as mutually validating systems. Each can compensate for the limitations of the other through cross-checking and triangulation. Second, bias mitigation cannot rely on uniform solutions. Different data sources generate distinct biases that require tailored statistical methods alongside social, ethical, and community engagement strategies. Third, progress depends on sustained multidisciplinary collaboration and appropriate infrastructures that support data quality, validation, and responsible implementation. A key concern involves bias amplification through online data and machine learning. Underrepresented groups in datasets often remain underrepresented in analyses, which can both worsen inequities and make them more visible. While increased visibility may support corrective action, current debiasing techniques can only address known biases. Unknown biases remain undetected, underscoring the need for broader demographic inclusion, cross-validation, and methodological innovation. Similar challenges arise in separating meaningful epidemiological signals from noise in digital data, where online behavior may reflect fear or media influence rather than actual disease incidence. This requires behavioral and multimodal models capable of contextual interpretation [98-101].

The discussion also highlights limits of predictive modeling. While advanced AI can improve case definitions and surveillance, correlation-based predictions often lack the causal depth needed for targeted interventions and may ignore ethical implications. Addressing infectious diseases as complex systems requires integrated strategies spanning data, models, governance, and response mechanisms. Finally, the paper stresses the importance of durable institutional infrastructures. Emerging pandemic intelligence hubs demonstrate promise, but many regions remain underserved, and global momentum may be waning. Without sustained investment, standards, and data sharing, future preparedness will remain inadequate [102][103].

**Conclusion:**

Digital Epidemiology represents a paradigm shift in public health research, offering unprecedented opportunities for rapid disease surveillance and predictive modeling. However, its reliance on secondary, unstructured digital data introduces profound methodological and ethical challenges. Bias—whether in sampling, measurement, or algorithmic processes—remains the most critical threat to validity and fairness. Unlike classical epidemiology, where bias control is embedded in study design, digital approaches often depend on retrospective correction, which is inherently limited and vulnerable to unknown biases. To realize the full potential of Digital Epidemiology, integration with classical methods is imperative. This requires harmonizing structured epidemiological data with heterogeneous digital sources through rigorous validation and tailored debiasing strategies. Ethical considerations, including privacy, informed consent, and transparency, must be prioritized to maintain public trust. Furthermore, algorithmic fairness should be treated as a core quality criterion, ensuring predictive models do not perpetuate health inequities. Future progress depends on sustained investment in interoperable infrastructures, multidisciplinary collaboration, and inclusive data practices that address digital divides. By embedding fairness and methodological rigor into every stage—from data sourcing to model deployment—Digital

Epidemiology can evolve into a reliable and equitable tool for global health preparedness and response.

**References:**


1. Rothman KJ, Huybrechts KF, Murray EJ.(2024). *Epidemiology*: *An introduction*. Oxford University Press.
2. Aiello AE, Renson A, Zivich P. Social media- and internet-based disease surveillance for public health. Annu Rev Public Health. 2020;41:101. doi: 10.1146/annurev-publhealth-040119-094402
3. Salathe M, Bengtsson L, Bodnar TJ, Brewer DD, Brownstein JS, Buckee C, et al. Digital epidemiology. PLoS Comput Biol. 2012;8(7):e1002616. doi: 10.1371/journal.pcbi.1002616
4. Salathé M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy. 2018;14(1):1–5. doi: 10.1186/s40504-017-0065-7 [
5. Acheson ED. "Oxford record linkage study: a central file of morbidity and mortality records for a pilot population,". Br J Prev Soc Med. 1964;18(1):8. BMJ Publishing Group.
6. Szklo M, Nieto FJ. *Epidemiology*: *Beyond the Basics*. Jones & Bartlett Publishers; 2014.
7. Park H-A, Jung H, On J, Park SK, Kang H. Digital epidemiology: use of digital data collected for non-epidemiological purposes in epidemiological studies. Healthc Inform Res. 2018;24(4):253–262. doi: 10.4258/hir.2018.24.4.253
8. Velasco E. Disease detection, epidemiology and outbreak response: the digital future of public health practice. Life Sci Soc Policy. 2018;14(1):1–6.
9. Milne R, Costa A. Disruption and dislocation in post-COVID futures for digital health. Big Data Soc. 2020;7(2):2053951720949567. doi: 10.1177/2053951720949567
10. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med. 2020;26(8):1183–1192. doi: 10.1038/s41591-020-1011-4
11. Valleron A-J, Bouvet E, Garnerin P, Menares J, Heard I, Letrait S, et al. A computer network for the surveillance of communicable diseases: the French experiment. Am J Public Health. 1986;76(11):1289–92. doi: 10.2105/ajph.76.11.1289
12. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. 2009;457(7232):1012–1014. doi: 10.1038/nature07634
13. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google Flu: traps in big data analysis. Science. 2014;343(6176):1203–1205.
14. Olson DR, Konty KJ, Paladini M, Viboud C, Simonsen L. Reassessing Google Flu Trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. PLoS Comput Biol. 2013;9(10):e1003256. doi: 10.1371/journal.pcbi.1003256
15. Tizzoni M, Panisson A, Paolotti D, Cattuto C. The impact of news exposure on collective attention in the United States during the 2016 Zika epidemic. PLoS Comput Biol. 2020. Mar;16(3):e1007633. doi: 10.1371/journal.pcbi.1007633
16. Freifeld CC, Mandl KD, Reis BY, Brownstein JS. HealthMap: global infectious disease monitoring through automated classification and visualization of Internet media reports. J Am Med Inform Assoc. 2008;15(2):150–7. doi: 10.1197/jamia.M2544
17. Santillana M, Nguyen AT, Dredze M, Paul MJ, Nsoesie EO, Brownstein JS. Combining search, social media, and traditional data sources to improve influenza surveillance. PLoS Comput Biol. 2015;11(10):e1004513. doi: 10.1371/journal.pcbi.1004513
18. Federal Register. Announcement of Requirements and Registration for the Predict the Influenza Season Challenge [Internet]. 2013 [cited 2023 Aug 6].
19. Koppeschaar CE, Colizza V, Guerrisi C, Turbelin C, Duggan J, Edmunds WJ, et al. Influenzanet: citizens among 10 countries collaborating to monitor influenza in Europe. JMIR Public Health Surveill. 2017;3(3):e7429. doi: 10.2196/publichealth.7429
20. Lwin MO, Yung CF, Yap P, Jayasundar K, Sheldenkar A, Subasinghe K, et al. FluMob: enabling surveillance of acute respiratory infections in health-care workers via mobile phones. Front Public Health. 2017;5:49. doi: 10.3389/fpubh.2017.00049
21. 21.Smolinski MS, Crawley AW, Baltrusaitis K, Chunara R, Olsen JM, Wójcik O, et al. Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. Am J Public Health. 2015;105(10):2124–2130. doi: 10.2105/AJPH.2015.302696
22. Moberley S, Carlson S, Durrheim D, Dalton C, et al. Flutracking: Weekly online community-based surveillance of influenza-like illness in Australia, 2017 Annual Report. Commun Dis Intell. 2019;43. doi: 10.33321/cdi.2019.43.31
23. Won M, Marques-Pita M, Louro C, Gonçalves-Sá J. Early and real-time detection of seasonal influenza onset. PLoS Comput Biol. 2017;13(2):e1005330. doi: 10.1371/journal.pcbi.1005330
24. .Danquah LO, Hasham N, MacFarlane M, Conteh FE, Momoh F, Tedesco AA, et al. Use of

a mobile application for Ebola contact tracing and monitoring in northern Sierra Leone: a proof-of-concept study. BMC Infect Dis. 2019;19(1):1–2.
25. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. PLoS ONE. 2014;9(5):e95133. doi: 10.1371/journal.pone.0095133
26. Yoneki E. Fluphone study: Virtual disease spread using haggle. In: Proceedings of the 6th ACM Workshop on Challenged Networks. 2011. p. 65–66.
27. Vorovchenko T, Ariana P, Loggerenberg FV, Amirian P. #Ebola and Twitter. What insights can global health draw from social media? In: Big Data in Healthcare. Springer; 2017. p. 85–98.
28. Albinati J, Meira Jr W, Pappa GL, Teixeira M, Marques-Toledo C. Enhancement of epidemiological models for Dengue fever based on Twitter data. In: Proceedings of the 2017 International Conference on Digital Health; 2017. p. 109–118.
29. McGough SF, Brownstein JS, Hawkins JB, Santillana M. Forecasting Zika incidence in the 2016 Latin America outbreak combining traditional disease surveillance with search, social media, and news report data. PLoS Negl Trop Dis. 2017;11(1):e0005295. doi: 10.1371/journal.pntd.0005295
30. de Lima CL, da Silva ACG, da Silva CC, Moreno GMM, da Silva Filho AG, Musah A, et al. Intelligent Systems for Dengue, Chikungunya, and Zika Temporal and Spatio-Temporal Forecasting: A Contribution and a Brief Review. In: Assessing COVID-19 and Other Pandemics and Epidemics using Computational Modelling and Data Analysis. Springer. 2022. p. 299–331.
31. Hargittai E. Potential biases in big data: Omitted voices on social media. Soc Sci Comput Rev. 2020;38(1):10–24.
32. Charaudeau S, Pakdaman K, Boëlle PY. Commuter mobility and the spread of infectious diseases: application to influenza in France. PLoS ONE. 2014;9(1):e83002. doi: 10.1371/journal.pone.0083002
33. Tizzoni M, Bajardi P, Decuyper A, King GKK, Schneider CM, Blondel V, et al. On the use of human mobility proxies for modeling epidemics. PLoS Comput Biol. 2014;10(7):e1003716. doi: 10.1371/journal.pcbi.1003716
34. Bharti N, Tatem AJ, Ferrari MJ, Grais RF, Djibo A, Grenfell BT. Explaining seasonal fluctuations of measles in Niger using nighttime lights imagery. Science. 2011;334(6061):1424–1427. doi: 10.1126/science.1210554
35. Shakeri Hossein Abad Z, Kline A, Sultana M, Noaeen M, Nurmambetova E, Lucini F, et al. Digital public health surveillance: a systematic scoping review. NPJ Digit Med. 2021;4(1):1–13.
36. Yavuz M, Savaskan N. A European roadmap to a digital epidemiology in public health system. Front Digit Health. 2024;6:1284426. Frontiers Media SA. doi: 10.3389/fdgth.2024.1284426
37. Paolotti D, Carnahan A, Colizza V, Eames K, Edmunds J, Gomes G, et al. Web-based participatory surveillance of infectious diseases: the Influenzanet participatory surveillance experience. Clin Microbiol Infect. 2014;20(1):17–21. doi: 10.1111/1469-0691.12477
38. Neto OL, Cruz O, Albuquerque J, de Sousa MN, Smolinski M, Cesse ÉAP, et al. Participatory surveillance based on crowdsourcing during the Rio 2016 Olympic Games using the guardians of health platform: descriptive study. JMIR Public Health Surveill. 2020;6(2):e16119. doi: 10.2196/16119
39. Blench M. Global public health intelligence network (GPHIN). In: Proceedings of Machine Translation Summit XI: Papers, 2007.
40. Tarkoma S, Alghnam S, Howell MD. Fighting pandemics with digital epidemiology. EClinicalMedicine. 2020. Aug;26:100497. doi: 10.1016/j.eclinm.2020.100512
41. Sridhar D. COVID-19: what health experts could and could not predict. Nat Med. 2020;26(12):1812. doi: 10.1038/s41591-020-01170-z
42. The Globe and Mail. Federal documents show sharp decline of Canada's pandemic warning. The Globe and Mail; 2023.
43. Wagner Peter. The lasting significance of viruses: COVID-19, historical moments and social transformations. Thesis Eleven. 177(1):122–132, 2023. SAGE Publications Sage UK: London, England.
44. Dong E, Ratcliff J, Goyea TD, Katz A, Lau R, Ng TK, et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. Lancet Infect Dis. 2022.
45. Singh B. International comparisons of COVID-19 deaths in the presence of comorbidities require uniform mortality coding guidelines. Int J Epidemiol 2021;50(2):373–377. doi: 10.1093/ije/dyaa276
46. Van Haute M, Agagon A, Gumapac FF, Anticuando MA, Coronel DN, David MC, et al. "Determinants of differences in RT-PCR testing rates among Southeast Asian countries during the first six months of the COVID-19 pandemic. PLOS Global Public Health. 2023;3(11):e0002593. Public Library of Science, San Francisco, CA, USA. doi: 10.1371/journal.pgph.0002593
47. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19:

systematic review and critical appraisal. *BMJ*. 2020;369. British Medical Journal Publishing Group. doi: 10.1136/bmj.m1328
48. Vaidheeswaran S, Karmugilan K. Consumer buying behaviour on healthcare products and medical devices during COVID-19 pandemic period-a new spotlight. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal 2021:9861–72.
49. Pandit JA, Radin JM, Quer G, Topol EJ. Smartphone apps in the COVID-19 pandemic. Nat Biotechnol. 2022;40(7):1013–22. doi: 10.1038/s41587-022-01350-x
50. Ojokoh BA, Aribisala B, Sarumi OA, Gabriel AJ, Omisore O, Taiwo AE, et al. Contact Tracing Strategies for COVID-19 Prevention and Containment: A Scoping Review. Big Data Cogn Comput. 2022;6(4):111.
51. Wymant C, Ferretti L, Tsallis D, Charalambides M, Abeler-Dörner L, Bonsall D, et al. The epidemiological impact of the NHS COVID-19 app. Nature. 2021;594(7863):408–412. Nature Publishing Group UK London. doi: 10.1038/s41586-021-03606-z
52. Sharma T, Bashir M. Use of apps in the COVID-19 response and the loss of privacy protection. Nat Med. 2020;26(8):1165–1167. doi: 10.1038/s41591-020-0928-y
53. Seto E, Challa P, Ware P, et al. Adoption of COVID-19 contact tracing apps: A balance between privacy and effectiveness. J Med Internet Res. 2021;23(3):e25726. doi: 10.2196/25726
54. Ng A. Google promised its contact tracing app was completely private—But it wasn't. 2021.
55. Bedson J, Skrip LA, Pedi D, Abramowitz S, Carter S, Jalloh MF, et al. A review and agenda for integrated disease models including social and behavioural factors. Nat Hum Behav. 2021;5(7):834–846. doi: 10.1038/s41562-021-01136-2
56. Salathé M. Privacy-preserving contact tracing curbed COVID. Nature. 619(7968):31–33, 2023. Nature Portfolio.
57. Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the effect of demographic factors, socioeconomic factors, and risk aversion on mobility during the COVID-19 epidemic in France under lockdown: a population-based study. Lancet Digit Health. 2020;2(12):e638–e649. doi: 10.1016/S2589-7500(20)30243-0
58. Pepe E, Bajardi P, Gauvin L, Privitera F, Lake B, Cattuto C, et al. COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Sci Data. 2020;7(1):1–7.
59. Lemey P, Ruktanonchai N, Hong SL, Colizza V, Poletto C, Van den Broeck F, et al. Untangling introductions and persistence in COVID-19 resurgence in Europe. Nature. 2021;595(7869):713–717. doi: 10.1038/s41586-021-03754-2
60. Levy BL, Vachuska K, Subramanian SV, Sampson RJ. Neighborhood socioeconomic inequality based on everyday mobility predicts COVID-19 infection in San Francisco, Seattle, and Wisconsin. Sci Adv. 2022;8(7):eabl3825. doi: 10.1126/sciadv.abl3825
61. Gauvin L, Tizzoni M, Piaggesi S, Young A, Adler N, Verhulst S, et al. Gender gaps in urban mobility. Humanit Soc Sci Commun. 2020;7(1):1–13.
62. Cantor JH, McBain RK, Pera MF, Bravata DM, Whaley CM. Who is (and is not) receiving telemedicine care during the COVID-19 pandemic. Am J Prev Med. 2021. Sep;61(3):434–438. doi: 10.1016/j.amepre.2021.01.030
63. Lian W, Wen L, Zhou Q, Zhu W, Duan W, Xiao X, et al., "Digital health technologies respond to the COVID-19 pandemic in a tertiary hospital in China: development and usability study," J Med Internet Res. 2020;22(11):e24505. JMIR Publications, Toronto, Canada. doi: 10.2196/24505
64. Kim EJ, Moretti ME, Kimathi AM, Chan SY, Wootton R. "Use of provider-to-provider telemedicine in Kenya during the COVID-19 pandemic," Front Public Health. 2022;10:1028999. Frontiers Media SA. doi: 10.3389/fpubh.2022.1028999
65. Ganjali R, Eslami S, Samimi T, Sargolzaei M, Firouraghi N, MohammadEbrahimi S, et al. Clinical informatics solutions in COVID-19 pandemic: Scoping literature review. Inform Med Unlocked. 2022;100929. doi: 10.1016/j.imu.2022.100929
66. Rambaud K, van Woerden S, Palumbo L, Salvi C, Smallwood C, Rockenschaub G, et al. "Building a Chatbot in a Pandemic," J Med Internet Res. 2023;25:e42960. JMIR Publications, Toronto, Canada. doi: 10.2196/42960
67. Salerno J, Coughlin SS, Goodman KW, Hlaing WM. Current ethical and social issues in epidemiology. Ann Epidemiol. 2023;80:37–42. doi: 10.1016/j.annepidem.2023.02.001
68. Zhao Y, He X, Feng Z, Bost S, Prosperi M, Wu Y, et al. Biases in using social media data for public health surveillance: A scoping review. Int J Med Inform. 2022;104804. doi: 10.1016/j.ijmedinf.2022.104804
69. Williams S. Data action: Using data for public good. MIT Press. 2022.
70. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. "Dissecting racial bias in an algorithm used to manage the health of populations,". Science. 2019;366(6464):447–453. American Association

for the Advancement of Science. doi: 10.1126/science.aax2342
71. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. “Potential biases in machine learning algorithms using electronic health record data,” JAMA Intern Med. 2018;178(11):1544–1547. American Medical Association. doi: 10.1001/jamainternmed.2018.3763
72. Kramer ADI, Guillory JE, Hancock JT. "Experimental evidence of massive-scale emotional contagion through social networks." Proc Natl Acad Sci U S A. 2014;111(24):8788–8790. doi: 10.1073/pnas.1320040111
73. Segura A. “Epidemics and epidemiology: back to the future,” Gac Sanit. 2023;37:102277. doi: 10.1016/j.gaceta.2022.102277
74. Ferretti A, Vayena E. In the shadow of privacy: Overlooked ethical concerns in COVID-19 digital epidemiology. Epidemics. 2022;41:100652. doi: 10.1016/j.epidem.2022.100652
75. Kostkova P. Disease surveillance data sharing for public health: the next ethical frontiers. Life Sci Soc Policy. 2018;14(1):1–5.
76. Vela MB, Erondu AI, Smith NA, Peek ME, Woodruff JN, Chin MH. “Eliminating explicit and implicit biases in health care: evidence and research needs,” Annu Rev Public Health. 2022;43(1):477–501. Annual Reviews. doi: 10.1146/annurev-publhealth-052620-103528
77. Bower JK, Patel S, Rudy JE, Felix AS. “Addressing bias in electronic health record-based surveillance of cardiovascular disease risk: finding the signal through the noise,” Curr Epidemiol Rep. 2017(4):346–352. Springer. doi: 10.1007/s40471-017-0130-z
78. Chiolero A, Santschi V, Paccaud F. “Public health surveillance with electronic medical records: at risk of surveillance bias and overdiagnosis,” Eur J Public Health. 2013;23(3):350–351. Oxford University Press. doi: 10.1093/eurpub/ckt044
79. Hicks B, Kaye JA, Azoulay L, Kristensen KB, Habel LA, Pottegård A. “The application of lag times in cancer pharmacoepidemiology: a narrative review,” Ann Epidemiol. 2023;84:25–32. Elsevier.
80. Xu J, Xiao Y, Wang WH, Ning Y, Shenkman EA, Bian J, et al. “Algorithmic fairness in computational medicine,” *EBioMedicine*. 2022:84. doi: 10.1016/j.ebiom.2022.104250
81. Tversky A, Kahneman D. “Availability: A heuristic for judging frequency and probability,” Cogn Psychol. 1973;5(2):207–232.
82. Shaw RJ, Harron KL, Pescarini JM, Pinto Junior EP, Allik M, Siroky AN, et al. “Biases arising from linked administrative data for epidemiological research: a conceptual framework from registration to analyses,” Eur J Epidemiol. 2022;37(12):1215–1224. Springer. doi: 10.1007/s10654-022-00934-w
83. Lewin A, Brondeel R, Benmarhnia T, Thomas F, Chaix B. “Attrition bias related to missing outcome data: a longitudinal simulation study,” Epidemiology. 2018;29(1):87–95. LWW. doi: 10.1097/EDE.0000000000000755
84. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. BMJ Evid Based Med. 2018;23(1):21–22. Royal Society of Medicine. doi: 10.1136/ebmed-2017-110883
85. Lipsitch M, Tchetgen ET, Cohen T. “Negative controls: a tool for detecting confounding and bias in observational studies,” Epidemiology. 2010;21(3):383–388. LWW. doi: 10.1097/EDE.0b013e3181d61eeb
86. Stockham N, Washington P, Chrisman B, Paskov K, Jung JY, Wall DP. “Causal modeling to mitigate selection bias and unmeasured confounding in internet-based epidemiology of COVID-19: model development and validation,” JMIR Public Health Surveill. 2022;8(7):e31306. doi: 10.2196/3130
87. Engelmann L. “Digital epidemiology, deep phenotyping and the enduring fantasy of pathological omniscience,” Big Data Soc. 2022;9(1):20539517211066451. SAGE Publications Sage UK: London, England.
88. Flores L, Kim S, Young SD, “Addressing bias in artificial intelligence for public health surveillance,” J Med Ethics. 2024;50(3):190–194. doi: 10.1136/jme-2022-108875
89. European Commission. (2024). European Health Data Space. Retrieved from https://health.ec.europa.eu/ehealth-digital-health-and-care/european-health-data-space-en.
90. European Centre for Disease Prevention and Control (ECDC). Long-term surveillance framework 2021–2027. Stockholm: ECDC; 2023. Available from: https://www.ecdc.europa.eu/sites/default/files/documents/long-term-surveillance-framework-2021-2027.pdf.
91. Andermann A, Pang T, Newton JN, Davis A, Panisset U. Evidence for Health II: Overcoming barriers to using evidence in policy and practice. Health Res Policy Syst. 2016;14:1–7. Springer.
92. Topol EJ. “Medical forecasting.” Science. 2024;384(6698):eadp7977. American Association for the Advancement of Science. doi: 10.1126/science.adp7977
93. Narayanan A, Kapoor S. *AI Snake Oil*: *What Artificial Intelligence Can Do*, *What It Can’t*, *and How to Tell the Difference*. Princeton University Press; 2024.
94. .Tan YR, Agrawal A, Matsoso MP, Katz R, Davis SLM, Winkler AS, et al. A call for citizen science in pandemic preparedness and response: beyond data collection. BMJ Glob Health.

2022;7(6):e009389. doi: 10.1136/bmjgh-2022-009389
95. .Chan AT, Brownstein JS. Putting the public back in public health—surveying symptoms of Covid-19. N Engl J Med. 2020;383(7):e45. doi: 10.1056/NEJMp2016259
96. Marley G, Dako-Gyeke P, Nepal P, Rajgopal R, Koko E, Chen E, et al., "Collective intelligence–based participatory COVID-19 surveillance in Accra, Ghana: pilot mixed methods study," JMIR Infodemiology. 2024;4(1):e50125. doi: 10.2196/50125
97. .Briand SC, Cinelli M, Nguyen T, Lewis R, Prybylski D, Valensise CM, et al. Infodemics: A new challenge for public health. Cell 2021;184(25):6010–6014. doi: 10.1016/j.cell.2021.10.031
98. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn YY, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad Sci U S A. 2020;117(21):11220–11222. doi: 10.1073/pnas.2005335117
99. Chafetz H, Zahuranec AJ, Marcucci S, Davletov B, Verhulst S. The# Data4COVID19 Review: Assessing the Use of Non-Traditional Data During A Pandemic Crisis. SSRN. 2022;4273229.
100. European Centre for Disease Prevention and Control. RespiCast. Available from: https://respicast.ecdc.europa.eu
101. European Centre for Disease Prevention and Control. EpiPulse: European surveillance portal for infectious diseases. Available from: https://www.ecdc.europa.eu/en/publications-data/epipulse-european-surveillance-portal-infectious-diseases
102. WHO. Regional strategy for health security and emergencies 2022–2030: report of the Secretariat. 2022
103. Cohen J. 'Cycles of panic and neglect': Head of Pandemic Prevention Institute explains its early death. Science. 2022. Available from: https://www.science.org/content/article/cycles-panic-and-neglect-head-pandemic-prevention-institute-explains-its-early-death
104.