

Epidemiology and Health Informatics

A narrative and analytical overview of population health evidence, digital systems, and responsible data use

Prof Ir Rudy C. Tarumingkeng, PhD

www.rudycr.com/cv.pdf

IBM-ASMI 2026

[Academic Article](#)

Abstract (Abstrak)

Epidemiology and health informatics have converged into a strategic field for understanding population health and improving decisions. Epidemiology contributes concepts of distribution, determinants, risk, causation, surveillance, and evaluation, while health informatics contributes data architecture, interoperability, analytics, information systems, and digital tools for action. This article develops a comprehensive narrative account of their relationship across the health-data life cycle. It examines electronic health records, public-health surveillance, population-health management, causal inference, artificial intelligence, natural-language processing, genomics, One Health, privacy, ethics, interoperability, implementation, evaluation, and governance, with attention to Indonesia and other low- and middle-income settings. The central argument is that digital capacity does not automatically create valid evidence or equitable health improvement. Value depends on fit-for-purpose data, explicit denominators, transparent provenance, rigorous epidemiological design, human-centered implementation, protection of privacy, and continuous evaluation after deployment. Health informatics is therefore best understood not as a technical adjunct to medicine, but as a socio-technical infrastructure for epidemiological intelligence, learning health systems, and collective well-being.

Keywords

Epidemiology; health informatics; digital health; public-health surveillance; electronic health records; artificial intelligence; interoperability; health equity; causal inference; Indonesia.

Article Orientation

This article is a conceptual and practice-oriented synthesis. It does not present a new empirical dataset or replace formal clinical guidance. Examples clarify the relationship between epidemiological reasoning, information systems, and public decisions. The references combine peer-reviewed scholarship with authoritative institutional sources.

Conceptual Architecture

Epidemiological question	Informatics capability	Decision use
What is happening, where, and to whom?	Surveillance, registries, dashboards, geographic and temporal data	Detect priorities and allocate attention
Why is it happening?	Linked data, phenotyping, causal analysis, contextual metadata	Design prevention and evaluate determinants
What may happen next?	Prediction, forecasting, anomaly detection, scenario models	Prepare services and target early action
Did the intervention work fairly?	Outcome measurement, feedback, subgroup monitoring, reproducible evaluation	Learn, adapt, and protect equity

1. Introduction

Epidemiology and health informatics have become inseparable fields in the organization of modern health knowledge. Epidemiology asks how health-related events are distributed across populations, what factors shape those distributions, and which interventions can prevent illness or reduce its consequences. Health informatics asks how health information is represented, captured, exchanged, interpreted, protected, and converted into action. One field supplies the population logic and questions of evidence; the other supplies many of the infrastructures, computational methods, and organizational arrangements through which contemporary answers are produced. Their convergence is therefore more than the replacement of paper with screens. It is a reconfiguration of how societies observe health, define risk, coordinate services, and learn from interventions.

The practical significance of this convergence is visible in both infectious disease and chronic disease. A surveillance officer may combine laboratory confirmations, emergency visits, pharmacy dispensing, school absenteeism, mortality records, community reports, and environmental measurements to assess whether a health threat is changing. A hospital may use an electronic health record to identify patients at risk of deterioration, evaluate readmissions, or examine differences in treatment. A ministry may connect an immunization registry with civil registration, geography, and primary-care records to locate communities that are not being reached. Each activity depends on epidemiological concepts such as case definition, denominator, time at risk, confounding, and bias, while also depending on informatics concepts such as interoperability, provenance, identity management, metadata, and user-centered design.

Digital transformation creates opportunity, but it does not abolish the classical problems of population health. More records do not automatically produce valid estimates. A dashboard may display thousands of observations while omitting people who lack access to care. An algorithm may predict risk accurately in the

institution where it was trained but fail in a rural clinic with different disease prevalence, documentation practices, or laboratory capacity. A social-media signal may appear early because a vocal population is overrepresented online, not because incidence has actually risen. Informatics can accelerate observation and response, yet it can also accelerate error, exclusion, surveillance overreach, and false confidence.

This article develops a narrative and analytical account of the relationship between epidemiology and health informatics. It examines epidemiological reasoning, the health-data life cycle, surveillance, electronic health records, population health management, causal inference, artificial intelligence, natural-language processing, genomics, One Health, privacy, interoperability, implementation, evaluation, governance, and applications in low- and middle-income countries. Its central argument is that health informatics should be treated as a socio-technical public-health capability. Its value is realized only when data are fit for purpose, methods are epidemiologically defensible, systems fit real workflows, governance protects rights, and decisions return benefits to the populations whose data made those decisions possible.

2. Epidemiological Reasoning in a Data-Intensive Era

Epidemiology is fundamentally a discipline of comparison. It compares disease occurrence between places, periods, exposures, groups, and interventions. Incidence describes the emergence of new events in a population at risk, whereas prevalence describes the proportion or burden existing at a particular time. Risk ratios, rate ratios, odds ratios, hazard ratios, attributable fractions, and standardized mortality ratios answer different questions and require different denominators. Informatics systems often place these measures next to one another in a dashboard, but visual proximity can conceal conceptual differences. A well-designed health-information system must preserve the meaning of each measure rather than presenting numbers as interchangeable indicators of health.

The denominator is especially important when administrative and clinical data are used for epidemiological purposes. A hospital database is not a census of disease in the community; it is a record of people who reached that institution and were documented there. A laboratory database may overrepresent people who were tested, while a claims database may reflect reimbursement rules as much as clinical need. Electronic data therefore require an explicit description of the source population, eligibility criteria, observation window, missingness, and pathways through which individuals entered or left the data system. These features determine whether a result can be generalized and whether an apparent change reflects disease, access, testing, coding, or recording behavior.

Modern informatics also makes repeated analyses easy, increasing the risk of undisclosed analytic flexibility. Investigators can test many definitions, predictors, subgroups, lag structures, and model specifications before selecting a result that looks persuasive. Reproducible epidemiology responds by specifying protocols, preserving code, documenting data transformations, and distinguishing exploratory from confirmatory analyses. A data pipeline should make it possible to trace a published estimate back to source records, inclusion rules, versioned code, and the analytic dataset used to produce it. Reproducibility is not only a research virtue. It is a public-health safety requirement when estimates influence policy or allocation of scarce resources.

The information environment also changes the speed at which evidence is produced. During an outbreak, a preliminary signal may be valuable even when uncertainty is substantial. During long-term chronic-disease research, slower and more carefully validated measurement may be preferable. Epidemiologists and informaticians must therefore match the evidence standard to the decision. Early-warning systems may prioritize sensitivity and rapid investigation; official burden estimates may require representative sampling and calibration; clinical decision support may require prospective evaluation and monitoring for harm. The

right question is not whether a dataset is large or real time, but whether its design and limitations are appropriate for the decision at hand.

3. The Health-Data Life Cycle

Health information begins as an event in the world: a person experiences symptoms, seeks care, receives a test, takes a medicine, or dies. It becomes data through observation and documentation. A clinician selects a diagnosis code, a laboratory instrument produces a result, a community health worker records a visit, or a sensor produces a measurement. Informatics studies the structures that make those observations computable and usable. Epidemiology then asks whether the recorded event corresponds to the phenomenon of interest and how it should be interpreted relative to a population. The distinction between event, observation, record, and inference should remain visible throughout the data life cycle.

Data quality is multidimensional. Completeness asks whether expected fields or events are present. Timeliness concerns the delay between occurrence, recording, transmission, and availability. Validity asks whether a value represents what it claims to represent. Consistency concerns agreement across fields, systems, or time. Accuracy is often difficult to establish because the true value may not be directly observable. A diagnosis code can be accurate for billing but imprecise for research; a free-text note can contain rich clinical detail but be difficult to classify consistently. Quality assessment should therefore be tied to use. The same data may be adequate for service planning and inadequate for estimating incidence.

The life cycle includes ingestion, cleaning, linkage, transformation, analysis, communication, use, retention, and disposal. Each stage can introduce selection or measurement bias. Record linkage may fail more often for people with common names, unstable addresses, or incomplete identifiers. Deduplication may remove legitimate repeated episodes. Imputation may smooth away meaningful heterogeneity. Aggregation may protect privacy while hiding small-area disparities. A credible informatics program documents these decisions as part of the evidence, not as invisible technical housekeeping. Data provenance should include who collected the data, under which protocol, using which instrument, and with what changes over time.

A useful principle is to separate data minimization from data impoverishment. Collecting every possible variable can increase privacy risks, storage costs, and analytic confusion. Collecting too little can prevent adjustment for confounding, obscure inequities, or make results impossible to interpret. The goal is purposeful sufficiency: the minimum information needed to answer a defined question responsibly, together with metadata that preserve meaning and limitations. In practice, this means designing information systems around use cases, population needs, and decision pathways rather than around the technical convenience of accumulating unstructured data.

4. Public-Health Surveillance and Early Warning

Public-health surveillance is the continuous, systematic collection, analysis, interpretation, and dissemination of health data for action. Informatics changes surveillance by connecting sources that traditionally operated in separate administrative domains. Laboratory information systems, case notifications, hospital admissions, emergency calls, pharmacy sales, veterinary reports, environmental sensors, school attendance, and community-health applications can contribute complementary signals. The purpose of integration is not to create a single perfect dataset. It is to assemble timely and interpretable evidence while preserving knowledge of how each source was generated.

Traditional indicator-based surveillance generally begins with a case definition and reporting process. It is strong when definitions are stable, reporting coverage is known, and legal or professional obligations support completeness. Event-based surveillance searches for signals that may not yet satisfy a formal case

definition, including unusual clusters, rumors, media reports, or community alerts. Syndromic surveillance uses symptoms, encounters, or proxy measures to detect changes before laboratory confirmation is available. Each approach has a different relationship between sensitivity, specificity, timeliness, and workload. Informatics platforms should allow these streams to be compared rather than treating one as a universal replacement for the others.

Digital epidemiology extends surveillance into online and networked environments. Search queries, social-media posts, web reports, mobility patterns, and digital contact networks can reveal concerns, behavior, or possible disease activity. These data may be valuable because they are fast and geographically detailed, but they are socially patterned. Internet access, language, platform use, media attention, and public trust affect who generates a signal. A sudden increase in online discussion may reflect a news story or political controversy rather than a change in incidence. Signals should therefore trigger investigation, not be interpreted as self-validating measurements of disease.

A mature early-warning system includes an operational loop: detect, verify, assess, communicate, respond, and learn. Detection without verification generates alert fatigue. Verification without a response pathway produces information theater. Communication must explain uncertainty and avoid unnecessary stigma. After action, the system should record whether the signal was useful, how quickly it was investigated, what resources were consumed, and what changes are needed. Informatics makes this feedback loop possible when alerts, decisions, and outcomes are captured in a form that can be reviewed.

5. Electronic Health Records as Epidemiological Resources

Electronic health records are primarily designed to support individual care, documentation, coordination, and billing, but they also provide an important source for epidemiological research. Their strength is longitudinal detail: encounters, diagnoses, prescriptions, laboratory results, measurements, referrals, and clinical narratives may be linked at the patient level. They can support rapid cohort construction, pragmatic outcome measurement, and monitoring of care processes. Their weakness is that the data are shaped by clinical workflow. What is not needed for care may never be recorded, and what is recorded may reflect reimbursement, institutional policy, or copied-forward text rather than a new clinical observation.

Phenotyping is the process of translating a clinical concept into an operational rule that identifies people or events in data. A simple phenotype may use a diagnosis code; a more complex phenotype may combine codes, medications, laboratory thresholds, imaging results, temporal relationships, and exclusion criteria. The rule should be clinically credible and empirically validated. If a study defines diabetes using a single code, it may misclassify rule-out diagnoses. If it requires repeated measures, it may improve specificity but exclude people with limited care access. Phenotype performance can change across institutions because coding practices and populations differ.

EHR-based studies require attention to immortal time, informative observation, confounding by indication, and missing-not-at-random measurements. A patient must remain observable long enough to be classified in certain exposure groups, which can create artificial survival advantages. Treatment is often given because a patient is at higher risk, so a crude comparison may make an effective treatment appear harmful. Laboratory tests are ordered selectively, meaning missing results can contain information about clinical judgment. Epidemiological design and causal reasoning must therefore guide the analysis from the beginning, not be added after a predictive model has been fitted.

Responsible EHR use depends on partnership with clinicians, patients, information officers, and public-health agencies. Data users should understand the workflow that generated the record, while system designers should understand how epidemiological questions use time, denominators, and uncertainty. Governance should specify permitted uses, access controls, audit trails, retention, and processes for

correcting errors. When EHR data are returned to clinical teams as quality indicators or decision support, the feedback should be evaluated for unintended consequences such as alert fatigue, inequitable targeting, or additional documentation burden.

6. Population Health Management and Health Equity

Population-health management translates information into coordinated action for defined groups. It may involve identifying people at high risk of complications, closing gaps in immunization, improving hypertension control, reducing avoidable admissions, or coordinating services for older adults with multiple conditions. Informatics supports registries, risk stratification, outreach workflows, referral tracking, and outcome measurement. Epidemiology contributes the population definition, risk-factor logic, comparison groups, and methods for assessing whether interventions improve outcomes rather than simply increasing activity.

Equity requires more than adding demographic fields to a dashboard. It requires examining who is represented, who is missing, who is classified incorrectly, and who bears the costs of an intervention. A risk model may underpredict need among people with limited access to care because it interprets fewer recorded diagnoses as lower risk. A digital appointment system may improve convenience for connected households while creating a new barrier for people without reliable internet or literacy in the system language. Equity-sensitive informatics evaluates performance across sex, age, ethnicity, disability, geography, socioeconomic position, migration status, and other locally relevant dimensions while avoiding unnecessary collection of sensitive attributes.

Small-area analysis can reveal variation hidden by national or regional averages. Geographic-information systems allow analysts to map service locations, travel time, environmental exposures, and outcomes. However, maps can stigmatize communities when data are displayed without context or when small counts make individuals identifiable. Ecological associations should not be treated as individual causal effects. Community participation is essential when interpreting place-based data, particularly in Indigenous and marginalized communities whose knowledge may not be captured by administrative categories. A technically precise map can still be socially misleading if it ignores local history and power.

Population-health programs should measure intermediate and final outcomes. Counting outreach calls or completed referrals is useful, but it does not establish improved health. Evaluation should examine uptake, adherence, clinical outcomes, patient experience, access, costs, and distribution of benefits. Interrupted time series, difference-in-differences, stepped-wedge designs, and pragmatic trials can help determine whether observed improvement exceeds what would have occurred without the program. Informatics supplies repeated measurements; epidemiology supplies the counterfactual discipline needed to interpret them.

7. Causal Inference and Real-World Evidence

Health informatics has expanded the availability of real-world data, but causal inference remains difficult. A predictive association answers whether information helps forecast an outcome. A causal question asks what would happen if an exposure or intervention were changed. These are not equivalent. A variable can improve prediction without being a useful intervention target, and a strong causal factor may have little predictive value when it is common across the population. Informatics projects should state explicitly whether their purpose is description, prediction, explanation, evaluation, or decision support.

Directed acyclic graphs and potential-outcome frameworks help clarify assumptions. Analysts must identify the exposure, outcome, time zero, eligible population, treatment strategies, follow-up period, censoring

rules, and variables that precede treatment. Adjusting for a mediator can block part of the effect being estimated; adjusting for a collider can create a spurious association. Automated feature selection does not resolve these conceptual problems. A large EHR can contain thousands of variables, but statistical abundance is not a substitute for a defensible causal design.

Target-trial emulation is a practical bridge between epidemiology and routinely collected data. Investigators specify the hypothetical randomized trial they would like to conduct and then emulate its eligibility, treatment assignment, follow-up, outcome, and analysis using observational records. This approach makes hidden choices visible and can reduce common biases. It does not transform observational data into randomized evidence; unmeasured confounding, measurement error, treatment switching, and missing follow-up remain important. Its value lies in disciplined specification and transparent sensitivity analysis.

Real-world evidence becomes more credible when multiple data sources are triangulated. Clinical records can be combined with registries, laboratory data, claims, surveys, mortality records, and patient-generated information. Agreement across sources is informative, but disagreement can also reveal systematic bias. A registry may capture clinically confirmed disease while claims show service use, and the difference may identify access barriers. Informatics should preserve source-specific meanings so that integration does not erase the contrasts needed for interpretation.

8. Artificial Intelligence and Machine Learning in Health

Artificial intelligence and machine learning are used across the health-information landscape. Applications include image interpretation, natural-language processing, risk prediction, anomaly detection, resource forecasting, recommender systems, and conversational interfaces. In epidemiology, models can help classify cases, estimate nowcasts, detect clusters, predict hospital demand, and integrate high-dimensional exposures. In clinical care, they may support diagnosis, triage, treatment selection, and monitoring. Their usefulness depends on the quality of labels, representativeness of data, calibration, interpretability, workflow fit, and consequences of error.

Model performance is often summarized with discrimination metrics such as area under the receiver-operating-characteristic curve, precision-recall measures, sensitivity, specificity, and F1 score. These are not enough for implementation. Calibration asks whether predicted probabilities correspond to observed frequencies. Decision-curve analysis examines whether using the model yields more benefit than treating nobody or everybody under specified thresholds. External validation tests portability across time, place, institution, and population. A model that performs well in a retrospective dataset may fail prospectively when clinicians change behavior in response to its predictions.

Bias can enter through the target, sample, measurement process, labels, missingness, and deployment context. An algorithm trained to predict health-care expenditure may learn patterns of spending rather than need, thereby disadvantaging groups that historically received less care. A natural-language model may perform better on notes written in the dominant language or by well-resourced hospitals. Fairness cannot be reduced to one universal metric because different definitions of equal performance can conflict. Responsible development requires stakeholder participation, subgroup analysis, documentation, monitoring, and a process for contesting or correcting decisions.

Human oversight should be designed, not assumed. Clinicians may over-trust an automated recommendation, ignore a useful warning, or experience alert fatigue when a system generates too many low-value prompts. A decision aid should communicate intended use, relevant uncertainty, data recency, and conditions under which it should not be used. High-stakes systems need rollback procedures, incident reporting, model versioning, and post-deployment surveillance. The proper objective is not to replace

professional judgment but to improve its quality, timeliness, and consistency while preserving accountability.

9. Natural-Language Processing and Unstructured Data

A large proportion of health information is unstructured: clinical notes, discharge summaries, radiology reports, referral letters, laboratory comments, call-center transcripts, and patient messages. Natural-language processing can identify symptoms, exposures, social conditions, medication changes, disease severity, and patient concerns that are not captured in structured fields. It can support phenotyping, cohort discovery, information retrieval, summarization, translation, and surveillance. The challenge is that language is context-dependent. Negation, uncertainty, temporality, abbreviations, copied text, and local conventions can reverse the apparent meaning of a phrase.

Clinical text also contains multiple voices and purposes. A note may mix a patient's report, a clinician's observation, a differential diagnosis, a historical condition, and a plan for future testing. A machine that counts disease terms without recognizing these distinctions can produce substantial misclassification. Annotation schemes should specify the unit of analysis, concept boundaries, assertion status, experiencer, and temporal relation. Validation should include representative notes and clinically meaningful error review, not only aggregate accuracy against an incomplete reference standard.

Large language models introduce new capabilities and new risks. They can summarize records, generate draft communications, answer questions, or assist coding, but they may produce plausible statements unsupported by the source record. In epidemiology, a fluent synthesis can obscure uncertainty and make it difficult to identify which evidence was actually used. Systems should retrieve and cite source data where appropriate, distinguish extraction from inference, preserve audit logs, and prevent sensitive information from being exposed to unauthorized services. Evaluation should test factuality, omission, subgroup performance, and effects on human decisions.

For countries with multiple languages and limited annotated datasets, participatory development is especially important. Local health workers and clinicians can identify expressions that general models miss, while communities can explain culturally specific terms and concerns. Transfer learning may reduce data requirements, but it does not remove the need for local validation. A modest, transparent model trained on local language and workflow may be more useful than a larger model whose errors are difficult to detect.

10. Genomics, Precision Public Health, and One Health

Genomic information adds a molecular dimension to epidemiology. Pathogen sequencing can identify transmission clusters, track mutations, distinguish reintroduction from local spread, and support antimicrobial-resistance surveillance. Human genomics can inform disease risk, pharmacogenomics, and research into biological mechanisms. When linked with clinical, geographic, and environmental data, genomic systems can contribute to precision public health by targeting interventions according to heterogeneous risk rather than assuming that populations are uniform.

The power of genomic epidemiology depends on sampling and metadata. A sequence without collection date, location, host information, or linkage to a case may be difficult to interpret. Overrepresentation of samples from large hospitals or well-resourced regions can distort the apparent evolution and geography of a pathogen. Rapid sequencing pipelines need clear sampling frames, quality control, standardized metadata, and transparent reporting of missingness. The same principle applies to human genomic research: analytic sophistication cannot compensate for biased recruitment or inadequate consent.

One Health extends information architecture across human, animal, and environmental domains. Zoonotic emergence, antimicrobial resistance, food safety, water quality, wildlife movement, and climate-sensitive diseases cannot be understood through a human clinical system alone. Interoperability in One Health is not merely a technical problem because sectors use different definitions, mandates, privacy expectations, and professional cultures. Governance must clarify when data can be shared, who is responsible for verification, and how communities and animal owners are protected from unjust economic or social consequences.

Genomic and One Health systems illustrate the double-edged character of advanced informatics. They can shorten the interval between detection and intervention, yet they can also enable intrusive surveillance or stigmatize a location, species, or group. Public trust requires proportionality, transparency, independent oversight, and demonstrable public benefit. The scientific value of a dataset does not by itself justify indefinite retention or unrestricted secondary use.

11. Privacy, Security, Ethics, and Public Trust

Health data are intimate because they concern bodies, relationships, behavior, reproduction, disability, mental health, and mortality. Privacy is not merely a technical property achieved by removing names. Re-identification can occur through combinations of dates, geography, rare conditions, genetic information, and external datasets. Security concerns confidentiality, integrity, and availability: data should not be disclosed improperly, altered without detection, or made inaccessible when needed for care. Ethical governance must address all three dimensions and recognize that harms can affect individuals, families, communities, and groups.

Consent is important but cannot carry the entire governance burden. In routine care, people may have limited ability to negotiate data practices. In public-health emergencies, legal authority may permit data use without individual consent, but necessity and proportionality still matter. Secondary research may use broad consent, dynamic consent, waivers, or other models depending on risk and jurisdiction. Whatever the model, people should receive understandable information about purposes, safeguards, retention, sharing, and their options. Community engagement is particularly important when collective interests and group harms are possible.

De-identification, pseudonymization, encryption, role-based access, secure enclaves, audit logs, and differential privacy can reduce risk, but each has limitations. Pseudonymization preserves linkage and therefore does not eliminate identifiability. Aggregation can hide small populations or weaken usefulness. Differential privacy introduces controlled noise and requires management of privacy budgets. Technical measures should be combined with organizational controls: access committees, sanctions, training, incident response, independent review, and clear accountability for vendors and partners.

Trust is built through reciprocity. Communities are more likely to support data use when they can see benefits, understand decisions, and participate in governance. Results should be returned in accessible forms, and systems should not extract information from populations without improving services or sharing knowledge back. Public trust can be damaged by a single undisclosed breach, discriminatory deployment, or unexplained policy change. Ethical health informatics therefore treats trust as an operational outcome that must be measured, maintained, and repaired when harmed.

12. Interoperability, Standards, and Data Architecture

Interoperability is the ability of systems to exchange, interpret, and use information consistently. Technical interoperability allows messages to move. Syntactic interoperability ensures that data follow a shared structure. Semantic interoperability preserves meaning across systems. Organizational interoperability

aligns policies, workflows, responsibilities, and incentives. A country may have excellent network connectivity and still lack useful interoperability if a laboratory result cannot be interpreted in the receiving system or if institutions cannot agree on who may act on it.

Standards such as HL7 FHIR, openEHR, DICOM, ICD, SNOMED CT, LOINC, and ISO-related frameworks provide important building blocks, but implementation requires governance. Terminologies must be maintained, mappings must be tested, and local concepts must be represented without being erased by imported categories. A diagnosis code used for reimbursement may not capture the clinical detail required for surveillance. A national minimum dataset should be designed around priority use cases and reviewed as care, technology, and public-health needs evolve.

Data architecture should support both operational and analytical use. Transactional systems are optimized for current care and workflow reliability; analytical platforms are optimized for longitudinal integration, quality checks, and reproducible analysis. Extracting data from an operational system can create burden and risk if it is not governed. Lake, warehouse, and federated architectures each involve trade-offs in cost, control, latency, and privacy. The correct architecture depends on the decisions to be supported, the capacity of institutions, and the need for local versus centralized stewardship.

Interoperability should be evaluated by outcomes rather than by the number of connected interfaces. Does a referral arrive with the information needed for action? Can a public-health team identify cases without repeatedly requesting spreadsheets? Can a patient access and correct a record? Can researchers reproduce a cohort definition across sites? These questions connect technical performance to health outcomes, professional work, and public value.

13. Implementation, Human Factors, and Organizational Change

Health-information systems fail as often from organizational misfit as from software defects. Implementation changes roles, documentation practices, professional autonomy, accountability, and the timing of work. A system that is technically sound may be rejected if it increases duplicate entry, slows consultations, or conflicts with local responsibilities. Conversely, an apparently simple tool may succeed when it reduces friction, fits existing routines, and makes benefits visible to users. Implementation should begin with workflow observation and stakeholder analysis rather than with procurement specifications alone.

Human-factors design examines how people perceive information, make decisions, coordinate under pressure, and recover from error. Interfaces should present the right information at the right time, support recognition rather than memory, and make critical states visible without overwhelming users. Alerts should be prioritized and evaluated for burden. Data entry should minimize unnecessary clicks and allow correction. In low-connectivity settings, offline capability, power constraints, device sharing, and maintenance must be treated as design requirements rather than exceptional conditions.

Change management is continuous. Training is necessary but insufficient because competence develops through supported practice. Local champions can translate system goals into credible operational benefits. Feedback channels help identify defects and unintended effects. Implementation metrics should combine adoption, fidelity, usability, workload, data quality, equity, and outcomes. The question is not simply whether users log in, but whether the system improves the work and the health decision it was intended to support.

Leadership matters because information systems cross boundaries. A hospital information officer, public-health director, clinical leader, data-protection officer, and community representative may each hold part of the authority required for change. Governance forums should clarify decision rights and escalation paths.

Procurement contracts should address data ownership, portability, service continuity, security updates, algorithmic transparency, and exit arrangements. Organizational learning is strongest when failures are treated as evidence for redesign rather than occasions for blame concealment.

14. Health Informatics in Low- and Middle-Income Countries

The Global South contains diverse health systems, from highly connected urban hospitals to community programs operating with intermittent power and paper records. Informatics strategies must begin with public-health priorities and existing capabilities, not with an assumption that every setting should reproduce the infrastructure of high-income countries. Digital tools can support maternal and child health, immunization, tuberculosis control, malaria surveillance, medicine supply chains, disaster response, and chronic-disease management, but only when they are integrated with supervision, financing, training, and referral capacity.

Fragmentation is a common challenge. Projects may be funded by different partners, use incompatible platforms, and report overlapping indicators to separate dashboards. This increases workload and can weaken national stewardship. Countries benefit from enterprise architecture, shared standards, open interfaces, common registries, and a clear division between national public goods and locally adaptable applications. The goal is not uniformity for its own sake; local innovation should remain possible while core identities, terminologies, reporting rules, and privacy protections remain coherent.

Digital inequity operates at several levels. Communities may lack connectivity, devices, affordable data, electricity, language support, or confidence in digital services. Health workers may have limited time and inadequate technical support. Institutions may not have data engineers, security officers, or procurement capacity. A national strategy should therefore invest in people and institutions as much as in software. Capacity includes epidemiological reasoning, data governance, maintenance, cybersecurity, evaluation, and the ability to negotiate with vendors.

Indonesia illustrates the need for a layered approach. A large archipelago contains different disease profiles, administrative capacities, geographies, and patterns of access. National platforms can support standards and aggregation, while provinces, districts, facilities, and community workers need tools adapted to their workflows. Interoperable immunization, laboratory, civil-registration, and primary-care information can strengthen surveillance, but the system should also preserve offline pathways and community-level verification. The essential measure of success is whether information reaches a decision-maker in time and whether the resulting action improves health fairly.

15. Evaluation, Reproducibility, and Evidence Quality

Evaluation asks whether an informatics intervention produces value under real conditions. Different questions require different designs. Usability testing can identify interface problems before deployment. A diagnostic-accuracy study can estimate sensitivity and specificity. A cluster randomized trial can assess effectiveness when individual randomization is impractical. An interrupted time series can examine changes after a policy or system introduction. A process evaluation can explain why an intervention worked in one site and failed in another. No single method is sufficient for the full socio-technical question.

Digital-health studies should report the data-generating context as carefully as they report the algorithm or intervention. Important details include setting, population, workflow, staffing, device and connectivity conditions, training, implementation date, comparator, missing data, and changes during follow-up. For machine learning, reporting should include data splitting, preprocessing, feature definitions, hyperparameter selection, calibration, external validation, subgroup performance, and intended use. For

surveillance, reporting should include case definitions, reporting coverage, timeliness, alert thresholds, verification, and false-alert burden.

Reproducibility has technical and institutional dimensions. Code and documentation support computational reproduction, but data access may be limited by privacy, contracts, or national law. In such cases, researchers can provide synthetic data, executable analysis plans, containerized workflows, aggregate validation results, or secure review environments. Version control is essential because source systems, code lists, and model parameters change. A published result should state which version of the data and software was used and how future users can determine whether an update is comparable.

Evidence quality also depends on publication and deployment incentives. Rapid innovation may reward novelty over careful evaluation, while operational programs may have little incentive to publish negative results. Registries, living evaluations, independent audits, and post-market monitoring can reduce this imbalance. Policymakers should ask not only whether a system can work, but under what conditions, for whom, at what cost, and with what risks.

16. Governance and Policy for Integrated Health Information

Governance is the arrangement through which values, authority, accountability, and resources shape information practice. It includes law and regulation, but also institutional policy, professional standards, technical rules, contracts, and community oversight. Effective governance distinguishes data stewardship from data ownership. A health ministry may be responsible for a national registry without claiming an unlimited right to every use of the data. Patients and communities may have legitimate interests even when records are held by providers or public agencies.

Policy should define purpose limitation, proportionality, transparency, security, access, correction, retention, and redress. These principles must be operationalized in ordinary workflows. Who approves a new linkage? Who reviews an algorithm before deployment? Who informs affected people after a breach? Who can suspend a model when performance changes? Who pays for maintaining a registry after a project ends? Unanswered questions create accountability gaps in which harms can persist while each institution assumes another is responsible.

National strategies should align public health, clinical care, research, social protection, digital government, and cybersecurity. A common architecture can reduce duplication, but centralization should not become an excuse for unnecessary data accumulation or exclusion of local expertise. Independent oversight is useful for high-risk systems, especially those involving sensitive populations, automated decisions, genetic data, or cross-border exchange. Procurement policy should require interoperability, auditability, security, accessibility, and transparent evaluation rather than purchasing only visible features.

Policy should also support learning. Rules that are too vague fail to protect people; rules that are too rigid can prevent beneficial research and emergency response. Regulatory sandboxes, staged deployment, proportional review, and clear sunset clauses can allow experimentation while limiting exposure. The aim is accountable innovation: systems may evolve, but their purposes, risks, evidence, and lines of responsibility remain visible.

17. Future Directions and Strategic Priorities

The next phase of epidemiology and health informatics will be shaped by convergence. Cloud and edge computing will distribute analytic capability. Interoperable records will make longitudinal and cross-institutional studies more feasible. Wearables and home devices will expand measurement beyond formal encounters. Genomics, environmental monitoring, and climate data will support more integrated models of

risk. Artificial intelligence will assist both routine operations and complex scientific reasoning. These developments should be judged by whether they improve prevention, care, resilience, equity, and democratic accountability.

One priority is the development of learning health systems that connect care, analysis, intervention, and feedback. Every clinical encounter should not be treated as a research experiment, but systems can be designed so that routine data help identify gaps, test improvements, and return lessons to practice. This requires governance that respects autonomy and privacy, common data models, embedded evaluation, and clinicians who can interpret evidence. Learning must include patient-reported outcomes and experience, not only measurements convenient for institutions.

A second priority is trustworthy artificial intelligence. Trustworthy systems will need robust external validation, transparent documentation, monitoring for drift, human-centered interfaces, and pathways for appeal. Model cards, data sheets, and audit reports can make assumptions visible, but documentation must be connected to institutional action. Organizations should be able to retire a system when its risks exceed its benefits. Education should prepare health professionals to question algorithms constructively rather than accepting or rejecting them as a matter of ideology.

A third priority is resilient public infrastructure. Health informatics cannot depend entirely on temporary projects, proprietary lock-in, or a small number of specialists. Governments and institutions need sustainable financing, open standards, local technical capacity, reliable connectivity, cyber resilience, and data institutions that can operate during crisis. Resilience also includes social legitimacy. Communities must see that information systems serve them, that errors can be corrected, and that innovation will not be used to reduce people to scores without context.

18. Methodological Toolkit for Epidemiology and Informatics

An integrated project should begin with a question map rather than a software map. The team should specify the population, health event, exposure or intervention, comparison, time horizon, and decision that may follow. A descriptive question may require a representative denominator and standardized measurement. A predictive question may require a timely stream and repeated validation. A causal question requires explicit assumptions about treatment, confounding, and follow-up. A service-improvement question may require workflow observation and mixed methods. Making this distinction early prevents the common error of selecting a fashionable method before defining what the evidence must accomplish.

Descriptive epidemiology remains essential in a digital environment. Analysts need to calculate counts, proportions, rates, trends, distributions, and uncertainty before applying complex models. Stratification can reveal effect modification and inequity that a single average obscures. Age standardization may be needed when populations differ in structure. Time series should account for reporting delays, seasonality, and changes in case definition. Informatics systems should make these routine checks easy, visible, and reproducible. A dashboard that begins with a sophisticated forecast but cannot show the underlying numerator, denominator, and missingness is difficult to trust.

Statistical modeling should be selected according to the data-generating process. Logistic regression, Poisson or negative-binomial models, survival analysis, multilevel models, time-series methods, spatial models, and machine learning each make different assumptions and answer different questions. Model complexity can be justified when it improves calibration, handles structure, or supports a decision, but complexity should not be used to disguise weak measurement. Analysts should assess sensitivity to alternative case definitions, missing-data strategies, lag periods, geographic boundaries, and unmeasured confounding. Sensitivity analysis is not an optional appendix; it is a way to show how much a conclusion depends on assumptions.

Mixed methods are particularly important because health information systems are embedded in organizations and communities. Quantitative records may show that an intervention was adopted, while interviews reveal that staff are entering data retrospectively to satisfy reporting requirements. A model may suggest that a community has low service need, while ethnographic work shows that people avoid a facility because of cost, distance, stigma, or distrust. Participatory methods can improve variable definitions and help interpret unexpected results. The strongest projects connect quantitative patterns with the lived processes that generated them.

Geographic information and spatial epidemiology add another layer of reasoning. Maps can support resource allocation, identify clusters, estimate travel time, and relate outcomes to environmental exposures. However, spatial scale affects results, and boundaries may reflect administrative convenience rather than lived geography. Analysts should consider the modifiable areal unit problem, spatial autocorrelation, boundary changes, and uncertainty in geocoding. Exact locations of vulnerable households should not be displayed simply because the software permits it. Spatial analysis should protect confidentiality while preserving enough resolution for a meaningful decision.

Implementation science provides a bridge from evidence to sustained practice. Frameworks may examine intervention characteristics, inner and outer settings, individual capability, implementation processes, and maintenance. The point is not to add theoretical labels to a project but to identify mechanisms. Did a system succeed because it reduced workload, because leadership made use mandatory, because staff received feedback, or because it was adapted to local language? Implementation measures should distinguish reach, adoption, fidelity, adaptation, cost, and sustainability. Without this distinction, a project may report success because a small pilot worked under unusually favorable conditions.

19. Education, Professional Roles, and Workforce Capacity

The convergence of epidemiology and informatics creates new educational requirements. Health professionals need enough data literacy to interpret denominators, bias, uncertainty, calibration, and causal claims. Informaticians need enough public-health understanding to recognize how case definitions, surveillance coverage, and social determinants shape data. Managers need to connect technical investments with workflow, governance, budgets, and outcomes. Community leaders need accessible explanations of what data are collected, why they matter, and how people can participate in oversight. A workforce strategy should therefore include multiple levels rather than training only a small group of specialists.

Undergraduate and professional curricula can use case-based learning. Students might follow a respiratory-outbreak signal from community report to laboratory confirmation, compare possible denominators, critique a dashboard, and design a communication plan. Another exercise might ask students to evaluate a readmission model across two hospitals with different documentation practices. Such cases teach that technical decisions and ethical decisions are intertwined. Learners should practice reading data dictionaries, questioning missingness, describing data provenance, interpreting model uncertainty, and communicating limitations to non-specialist decision-makers.

The emerging workforce is interdisciplinary. Epidemiologists, clinicians, statisticians, software engineers, data stewards, cybersecurity specialists, geographers, social scientists, ethicists, and community representatives each contribute different forms of expertise. Team effectiveness depends on translation across professional languages. A clinical expert may describe a syndrome that has no direct code; an engineer may identify a scalable architecture; a community worker may explain why a field is systematically missing. Institutions should reward this collaborative work because it is often less visible than an individual technical product but more important for safe deployment.

Professional development must continue after procurement. Health workers need support when definitions, interfaces, or reporting rules change. Managers need skills in vendor governance, contract review, and evaluation. Data protection officers need to understand the operational realities of care and surveillance. Researchers need training in reproducible computational practice and responsible artificial intelligence. Local mentorship networks, open educational resources, public-sector fellowships, and university partnerships can help spread capacity beyond capital cities and large institutions.

Ethical competence should be practical rather than ceremonial. Teams should be able to identify who could be harmed by a dataset, model, linkage, or public display. They should ask whether the benefits are distributed fairly, whether people can contest an error, and whether the system creates obligations that users cannot realistically meet. Students and professionals should learn to recognize automation bias, privacy leakage, discriminatory proxies, and the tendency to treat recorded need as actual need. In this way, education becomes part of the safety architecture of health informatics.

20. Data Stewardship, Sustainability, and Public Value

Data stewardship is the practical discipline of caring for information across its life cycle. It includes defining data owners and custodians, maintaining dictionaries, approving access, monitoring quality, preserving provenance, and deciding when retention is no longer justified. Stewardship is sometimes treated as administrative overhead, but it is a scientific function. If a code list changes without documentation, a trend may be misread. If a facility is removed from a reporting network, apparent incidence may fall. If a model is retrained without a record of the new data, its performance cannot be compared with the previous version. Stewardship makes change interpretable.

Sustainability has financial, technical, organizational, and political dimensions. A pilot may succeed because external funding pays for devices, connectivity, training, and specialist support. When the grant ends, maintenance may stop and staff may return to parallel paper processes. Sustainable programs budget for hosting, replacement cycles, cybersecurity, licensing, user support, data-quality review, and independent evaluation. They also define who can operate the system during leadership changes, emergencies, or vendor withdrawal. A technically impressive prototype is not a sustainable public-health asset until these responsibilities are visible.

Open standards and modular architecture can reduce dependence on a single vendor, but openness must be implemented carefully. Public documentation, portable data formats, application programming interfaces, and transparent licensing can support competition and local adaptation. At the same time, institutions must protect sensitive code, credentials, and security details. Procurement should avoid both extremes: closed systems that make exit impossible and poorly governed openness that exposes data or creates inconsistent versions. The strategic objective is controlled flexibility.

Public value is broader than efficiency. A system may save staff time, but it may also improve the timeliness of outbreak response, make an underserved population visible, reduce duplication, support patient autonomy, or strengthen accountability. Evaluation should therefore include social and institutional outcomes. Who gained access? Which decisions became faster or more accurate? Which burdens shifted to patients or frontline staff? Did the system reduce uncertainty or merely move it into a less visible place? These questions help leaders distinguish genuine transformation from digital decoration.

Data stewardship also requires attention to historical records. Retrospective datasets may contain classifications and practices that reflected earlier social assumptions. Linking such data to current systems can reproduce old inequalities while giving them a modern appearance. Analysts should document changes in diagnostic criteria, geography, service availability, and population structure. Where possible, affected

communities should help interpret historical categories. Ethical use of legacy data recognizes that preservation of information does not preserve the circumstances under which it was produced.

Emergency preparedness tests every part of the information system. During a crisis, reporting volume rises, definitions change, staff are reassigned, and decisions must be made before data are complete. Systems should support graceful degradation, offline workflows, manual correction, source versioning, and clear communication of provisional estimates. Exercises can reveal whether an alert reaches the right person, whether a laboratory result is visible to the response team, and whether a public dashboard can be updated without creating contradictory messages. Preparedness is built before the emergency, not during it.

Finally, public value depends on institutional humility. No model, registry, dashboard, or surveillance stream captures the whole of health. Data are partial traces of lived experience. Good institutions make this limitation explicit, invite correction, and avoid presenting computational outputs as neutral facts detached from choices. Humility does not mean refusing technology; it means using technology with disciplined confidence. The most trustworthy systems are those that are powerful enough to help and transparent enough to be questioned.

21. Applied Case Vignettes

Consider a district that notices a rise in acute respiratory complaints. The first signal comes from primary-care visits, but the district also examines laboratory turnaround time, pharmacy sales, school absenteeism, meteorological conditions, and hospital oxygen use. Each source has a different reporting delay and population coverage. The informatics platform displays them in a common timeline while retaining source labels and uncertainty. Epidemiologists review whether the increase can be explained by a coding change, a clinic relocation, seasonal smoke, or genuine transmission. The system is useful because it connects signals without pretending that they are identical. The response may include targeted testing, risk communication, and support for affected facilities.

A second case involves hypertension management. A primary-care registry identifies people with repeated high blood-pressure readings but no recent follow-up. The outreach team uses the registry to invite patients for review, yet the program also checks whether people in remote communities are underrepresented because they have fewer recorded visits. The evaluation compares control rates, missed appointments, travel burden, and patient experience before and after implementation. If control improves only among people who can attend the clinic, the program has increased activity without solving the access problem. Data become valuable when they reveal this distributional result.

A third case concerns an artificial-intelligence model for hospital deterioration. The model is trained on historical records and performs well in the development hospital. Before deployment elsewhere, the team examines calibration, missingness, nursing documentation, disease prevalence, and workflow differences. A silent pilot assesses whether clinicians understand the alerts and whether the model changes response time. Governance defines who can override the recommendation, how errors are reported, and when the model must be suspended. The case shows that safe deployment is an organizational process, not the final step of software development.

These vignettes share a common lesson. Effective informatics connects data to a defined decision, makes uncertainty visible, evaluates consequences across groups, and creates a feedback loop. The same technical platform can support good or poor practice depending on the epidemiological question, the surrounding institutions, and the values that guide use.

22. Boundaries and Limitations

This narrative has several boundaries. It treats epidemiology and health informatics as connected fields, but each contains specialized methods that cannot be fully covered in one article. Infectious-disease modeling, environmental epidemiology, clinical informatics, biomedical imaging, health economics, and implementation science each have their own technical literatures. The purpose here is to establish a coherent map of relationships and practical questions, not to provide a complete methods textbook.

The discussion also emphasizes responsible use rather than assuming that every new technology is beneficial. Some digital interventions may improve access, while others may intensify exclusion or shift work to already burdened staff. The effect of a system depends on the setting, the people who use it, the people represented in its data, and the institutional choices surrounding it. Readers should therefore treat the principles in this article as a basis for local inquiry and evaluation rather than as universal claims about a particular product or policy.

Evidence is also uneven across countries and populations. Much published research comes from well-resourced health systems with more complete digital records, stable connectivity, and specialist staff. Applying those findings in Indonesia or other diverse settings requires attention to transportability, local disease patterns, language, governance, workforce, and informal care. Local evidence is not merely a replication of global evidence; it is necessary for understanding how information systems interact with local institutions and communities.

Finally, the pace of technical change means that tools, standards, regulations, and risks will continue to evolve. Durable principles are therefore more important than allegiance to one platform: clarify the question, preserve meaning, measure quality, protect people, evaluate consequences, and maintain accountability. These principles allow institutions to adopt useful innovations while retaining the capacity to revise or reject them when evidence changes.

The article also assumes that information should remain connected to action. An accurate estimate that never reaches a decision-maker has limited public-health value, while a rapid signal that triggers an inappropriate intervention can cause harm. This is why communication, leadership, and institutional learning appear throughout the discussion. Evidence must be translated into choices, choices must be documented, and outcomes must return to the data system so that future decisions become better informed.

A final practical implication is that institutions should begin with a small number of high-value use cases and build outward. A reliable immunization registry, a useful laboratory exchange, or a well-evaluated chronic-disease registry can create more public value than many disconnected applications. Each successful use case can strengthen standards, skills, trust, and governance. Incremental progress is not a retreat from innovation; it is a way to make digital transformation durable and accountable.

In teaching and management, this principle can be expressed as a simple discipline: every data element should have a defined meaning, every indicator should have a defined denominator, every model should have a defined use, and every automated decision should have a defined human responsibility. Such discipline helps organizations resist the temptation to equate data volume with intelligence. It also creates a common language through which professors, managers, clinicians, engineers, public officials, and communities can discuss what a health-information system is doing and whether it is serving its purpose.

That common language is itself an institutional asset because it makes collaboration, accountability, and learning possible across professional and administrative boundaries.

It helps transform isolated technical projects into coherent public-health capabilities that can be evaluated, improved, and sustained over time.

23. Conclusion

Epidemiology and health informatics form a shared discipline of evidence and action. Epidemiology supplies the population concepts needed to interpret health events, compare risks, evaluate interventions, and protect against misleading conclusions. Health informatics supplies data structures, computational methods, interoperable systems, and organizational processes needed to observe complex health realities at scale. Their integration can strengthen surveillance, clinical care, research, planning, and policy, but only when digital capability is joined to epidemiological judgment.

The central challenge is not data scarcity alone. It is the alignment of data with purpose. A record is valuable when its meaning, quality, provenance, and limitations are understood. An algorithm is valuable when it improves a decision in a real workflow and when its errors are detectable and addressable. A national platform is valuable when it helps people receive better services, not merely when it produces a polished dashboard. A privacy safeguard is valuable when it protects dignity while allowing proportionate public benefit.

For Indonesia and other countries navigating rapid digital change, the strategic task is to build systems that are interoperable but locally responsive, innovative but accountable, analytic but humane. Investment should combine infrastructure with epidemiological education, informatics competence, governance, cybersecurity, and community participation. Health information should circulate to support prevention and care while remaining subject to legitimate safeguards. Evaluation should examine not only efficiency and accuracy but also equity, trust, workload, sustainability, and health outcomes.

The future of public health will not be determined by the quantity of data collected or the sophistication of software alone. It will be determined by the quality of questions asked, the integrity of methods used, the fairness of systems deployed, and the willingness of institutions to learn openly from evidence. When these conditions are met, health informatics becomes more than a technical service: it becomes an infrastructure for epidemiological intelligence and collective well-being.

Glossary (Glossarium)

Algorithmic bias. Systematic differences in model performance or decisions that disadvantage particular groups, often arising from data, labels, design, or deployment context.

Causal inference. Methods for estimating the effect of an exposure or intervention rather than merely describing or predicting an association.

Electronic health record. A longitudinal digital record of health information generated across clinical encounters and related services.

Epidemiology. The study of the distribution and determinants of health-related states or events in specified populations and the application of that study to control health problems.

Health informatics. The interdisciplinary study and practice of using information, data, and computational systems to support health and health care.

Interoperability. The ability of systems to exchange, interpret, and use information in a dependable and meaningful way.

Phenotype. An operational definition used to identify a clinical or epidemiological condition in a dataset.

Population-health management. Organized use of data and coordinated services to improve outcomes for a defined population.

Public-health surveillance. Continuous, systematic collection, analysis, interpretation, and dissemination of health data for action.

Syndromic surveillance. Surveillance based on symptoms, encounters, or proxy indicators that may precede confirmed diagnoses.

Target-trial emulation. A design strategy that specifies a hypothetical randomized trial and emulates its components using observational data.

Data provenance. Documentation of where data originated, how they were collected, transformed, linked, and used.

Digital epidemiology. Epidemiological practice that uses data generated outside traditional health systems, including online and networked sources.

Calibration. The agreement between predicted probabilities and the observed frequency of outcomes.

One Health. An integrated approach recognizing connections among human, animal, plant, and environmental health.

References (Daftar Pustaka)

- American Medical Informatics Association. (2024). Informatics: Research and practice. <https://amia.org>
- Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *JAMA*, 319(13), 1317–1318. <https://doi.org/10.1001/jama.2017.18391>
- Buntin, M. B., Burke, M. F., Hoaglin, M. C., & Blumenthal, D. (2011). The benefits of health information technology: A review of the recent literature shows predominantly positive results. *Health Affairs*, 30(3), 464–471. <https://doi.org/10.1377/hlthaff.2011.0178>
- Brownstein, J. S., Freifeld, C. C., & Madoff, L. C. (2009). Digital disease detection—Harnessing the Web for public health surveillance. *New England Journal of Medicine*, 360(21), 2153–2155. <https://doi.org/10.1056/NEJMp0900702>
- Centers for Disease Control and Prevention. (2024). Public health surveillance and public health informatics resources. <https://www.cdc.gov>
- Delbanco, T., Walker, J., Bell, S. K., Darer, J. D., Elmore, J. G., Farag, N., Feldman, H. J., Mejilla, R., Ngo-Metzger, Q., Ralston, J. D., Ross, S. E., Trivedi, N., Vodicka, E., & Leveille, S. G. (2012). Inviting patients to read their doctors' notes: A quasi-experimental study and a look ahead. *Annals of Internal Medicine*, 157(7), 461–470. <https://doi.org/10.7326/0003-4819-157-7-201210020-00002>
- Eysenbach, G. (2009). Infodemiology and infoveillance: Tracking online health information and cyberbehavior for public health. *American Journal of Preventive Medicine*, 36(5 Suppl.), S154–S158. <https://doi.org/10.1016/j.amepre.2008.10.022>
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014. <https://doi.org/10.1038/nature07634>
- Greenhalgh, T., Wherton, J., Papoutsis, C., Lynch, J., Hughes, G., A'Court, C., Hinder, S., Fahy, N., Procter, R., & Shaw, S. (2017). Beyond adoption: A new framework for theorizing and evaluating nonadoption, abandonment, scale-up, spread, and sustainability of health and care technologies. *Journal of Medical Internet Research*, 19(11), e367. <https://doi.org/10.2196/jmir.8775>
- Hernán, M. A., & Robins, J. M. (2020). Causal inference: What if. Chapman & Hall/CRC. <https://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/>
- Kamel Boulos, M. N., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/SARS-CoV-2 epidemic and associated events around the world. *International Journal of Health Geographics*, 19, 8. <https://doi.org/10.1186/s12942-020-00202-8>

- Krumholz, H. M. (2014). Big data and new knowledge in medicine: The thinking, training, and tools needed for a learning health system. *Health Affairs*, 33(7), 1163–1170. <https://doi.org/10.1377/hlthaff.2014.0053>
- Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343(6176), 1203–1205. <https://doi.org/10.1126/science.1248506>
- Morley, J., Machado, C. C. V., Burr, C., Cows, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A systematic review. *Social Science & Medicine*, 260, 113172. <https://doi.org/10.1016/j.socscimed.2020.113172>
- Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—Big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216–1219. <https://doi.org/10.1056/NEJMp1606181>
- Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, 380(14), 1347–1358. <https://doi.org/10.1056/NEJMr1814259>
- Rajkomar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E., Irvine, J., Le, Q., Litsch, K., & Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *npj Digital Medicine*, 1, 18. <https://doi.org/10.1038/s41746-018-0029-1>
- Rumbold, J. M. M., & Pierscione, B. K. (2017). The effect of the general data protection regulation on medical research. *Journal of Medical Internet Research*, 19(2), e47. <https://doi.org/10.2196/jmir.7108>
- Salathé, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., Campbell, E., Cattuto, C., Khandelwal, S., Mabry, P. L., & Vespignani, A. (2012). Digital epidemiology. *PLoS Computational Biology*, 8(7), e1002616. <https://doi.org/10.1371/journal.pcbi.1002616>
- Shortliffe, E. H., & Cimino, J. J. (Eds.). (2021). *Biomedical informatics: Computer applications in health care and biomedicine* (6th ed.). Springer. <https://doi.org/10.1007/978-3-030-58721-5>
- Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine*, 25, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>
- Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, 15(11), e1002689. <https://doi.org/10.1371/journal.pmed.1002689>
- Wiens, J., Saria, S., Sendak, M., Ghassemi, M., Liu, V. X., Doshi-Velez, F., Jung, K., Heller, K., Kale, D., Saeed, M., Ossorio, P. N., Thadaney-Israni, S., & Goldenberg, A. (2019). Do no harm: A roadmap for responsible machine learning for health care. *Nature Medicine*, 25, 1337–1340. <https://doi.org/10.1038/s41591-019-0548-6>
- World Health Organization. (2021). *Global strategy on digital health 2020–2025*. World Health Organization. <https://www.who.int/publications/i/item/9789240020924>
- World Health Organization. (2024). *Digital health*. <https://www.who.int/health-topics/digital-health>