Journal of Clinical Medicine and Research Vol 2 Issue 5 2025

Original Research Article

Open Access

An Exploration of Public Health Risk Early Warning Mechanisms Based on Big Data Technology

Cheng-De Bai*

Disease Control and Prevention Center of Ningdong Energy and Chemical Industry Base, Ningxia Hui Autonomous Region, 750411, China

***Correspondence to:** Cheng-De Bai, Disease Control and Prevention Center of Ningdong Energy and Chemical Industry Base, Ningxia Hui Autonomous Region, 750411, China, E-mail: 1282329158@qq.com

Abstract: With the acceleration of globalization and urbanization, public health events such as emerging infectious diseases have occurred frequently, posing serious threats to human health and social stability. Traditional public health monitoring and early warning systems have shown clear limitations in terms of the breadth, depth, timeliness of data acquisition, as well as the integration and analysis of multi-source heterogeneous data. Leveraging its characteristics of massive volume, high velocity, and diversity, big data technology provides critical technical support for building a new generation of intelligent public health risk early warning mechanisms. This paper systematically explores such mechanisms by first analyzing the dilemmas faced by traditional early warning systems, and then elaborating on the core logic, technical architecture, and data ecosystem of big data–empowered public health early warning. Innovatively, energy and chemical enterprises are introduced as a representative case to demonstrate how internal early warning systems constructed through big data can, in turn, enhance regional public safety.

Keywords: Big data; Public health; Risk early warning; Intelligent monitoring; Energy and chemical enterprises; Data integration

Introduction

Since the beginning of the 21st century, major public health emergencies such as SARS and COVID-19 have occurred repeatedly, warning that humanity has entered a highly interconnected era of risk. These events not only cause substantial losses of life and property but also severely disrupt global economic and social order. Establishing an efficient public health risk early warning mechanism has thus become a global consensus and a key governance priority. Traditional early warning systems mainly rely on statutory disease reporting systems and other structured data sources. When confronted with complex threats such as unknown pathogens, these systems suffer from inherent shortcomings, including delayed responses, fragmented information, and passive reaction modes. The core challenges lie in single-source data dependence, static analytical methods, and rigid warning thresholds. The new generation of information technology revolution offers fresh solutions to these challenges. By integrating multidimensional and heterogeneous data streams, big data technology enables the identification of risk patterns and facilitates a fundamental shift from "post-event response" to "pre-event prediction" in public health risk management.

1. Dilemmas and Challenges of Traditional Public Health Risk Early Warning Systems

Traditional public health risk early warning systems are largely modeled on the *World Health Organization's International Health Regulations (2005)*, with medical institutions serving as core nodes in statutory infectious disease reporting networks. While these systems have historically played an important role in controlling the spread of many known infectious diseases, their limitations become increasingly evident when confronted with novel and complex crises:

(1) Data silos and fragmented information: Early warning data are dispersed across multiple government departments and private institutions. Administrative barriers and the lack of unified standards hinder effective data sharing and integration. For example, in cases of poisoning caused by chemical plant leaks, early signals may be scattered across emergency call records (e.g., 120 emergency services), social media platforms, and other sources. Without timely correlation and integration, effective early warning becomes difficult to achieve [1].

(2) Delayed response: The linear process of "diagnosis–reporting–analysis–warning" inherently leads to time lags. From the onset of individual illness to the issuance of an early warning, several days or even weeks may elapse. For rapidly spreading respiratory infectious diseases, outbreaks may have already expanded widely by the time a warning is issued.

(3) Perception blind spots and insufficient coverage: Traditional systems rely heavily on proactive reporting by medical institutions, resulting in monitoring blind spots for individuals who do not seek medical care, have mild symptoms, or are located in remote areas. Moreover, these systems focus primarily on known pathogens and exhibit limited capability in issuing early warnings for health risks caused by non-infectious factors.

(4) Static analytical models: Early warning thresholds are typically established based on historical averages and lack adaptive adjustment to complex and dynamic variables. This often leads to "false positives" during periods of high population mobility and "missed warnings" when risks gradually accumulate.

2. Mechanism Construction of Big Data–Empowered Public Health Risk Early Warning

Big data technology does not merely increase the volume of available data; rather, it reconstructs the entire public health risk early warning process through new data-driven thinking and advanced technical approaches. Its core mechanism can be summarized into four interconnected stages: multi-source perception, intelligent integration, dynamic early warning, and coordinated response.

2.1 Holistic Perception of Multi-Source Heterogeneous Data

The data ecosystem underpinning modern public health risk early warning constitutes a complex mega-system. It has long surpassed the boundaries of traditional medical data and evolved into a multidimensional perception network that integrates online and offline information, physical and digital signals, as well as individual- and environment-level data. The core data sources of this network include, but are not limited to: medical and health data represented by electronic health records, electronic medical records, medical insurance settlement data, and laboratory testing results; behavioral and public opinion data represented by search engine query logs, social media text, online consultation platform records, and pharmaceutical retail data, which are highly sensitive in reflecting public health concerns and early self-reported symptoms; environmental and spatial data represented by air quality monitoring, water quality surveillance, meteorological data, and geographic information system (GIS) data, which are used to assess environmental health risks; as well as population and socioeconomic data represented by census statistics, mobile signaling data, transportation checkpoint records, and enterprise attendance data [2]. Notably, operational data from specific industries—such as safety production monitoring and pollutant emission data from energy and chemical enterprises—have also become indispensable sensory nodes for risk perception. Through multiple technical approaches, including application programming interfaces (APIs), web crawlers, and Internet of Things (IoT) sensors, real-time or near–real-time acquisition of the above data sources can be achieved.

2.2 Data Integration and Intelligent Analysis Architecture

The raw data collected from diverse sources are inherently disordered and must undergo a series of preprocessing steps—such as data cleaning, standardization, anonymization, and integration—before they can be transformed into valuable analytical assets. On this basis, a layered intelligent analysis architecture functions as the "brain" of the entire early warning system. At the foundational level, data lakes or data warehouses serve as centralized storage hubs for massive volumes of raw and processed data, accommodating structured, semi-structured, and unstructured data formats. Above this lies a powerful computing engine layer, which relies on distributed computing frameworks such as Hadoop and Spark to efficiently handle batch and stream processing tasks, thereby ensuring analytical timeliness. The core of the architecture is the intelligent analytics layer, where a variety of advanced algorithms converge. Natural language processing (NLP) techniques enable in-depth analysis of textual data from social media and news reports, extracting health-related keywords, sentiment tendencies, and potential events. Machine learning and deep learning models act as predictive engines: whether forecasting influenza trends using long short-term memory (LSTM) networks or simulating disease transmission pathways through graph neural networks, these models demonstrate strong analytical and inferential capabilities. Spatiotemporal data analysis, combined with GIS technologies, supports hotspot detection, clustering analysis, and diffusion modeling of risk events, thereby assigning precise geographic attributes to early warnings. Finally, the outcomes of these complex analyses are presented through the visualization and decision-support layer in intuitive forms such as interactive dashboards, heat maps, and risk maps. This enables decision-makers to gain clear situational awareness and actionable insights, completing a closed loop from data to knowledge and ultimately to informed decision-making.

2.3 Dynamic and Tiered Early Warning Models

Building upon the aforementioned advanced intelligent analytics capabilities, the new generation of early warning models has fundamentally broken free from the constraints of traditional static thresholds, demonstrating unprecedented flexibility and precision. The primary feature of these models lies in their dynamic threshold mechanisms. Rather than relying on fixed numerical trigger points, early warnings are adaptively adjusted according to multiple real-time factors, such as population mobility patterns, changes in meteorological conditions, and fluctuations in historical baseline levels. This ensures that warning sensitivity remains consistently aligned with the prevailing risk environment. Secondly, the models adopt a multi-indicator fusion strategy, moving beyond sole dependence on confirmed case counts, which are inherently lagging indicators. Instead, they comprehensively integrate multiple weak yet highly correlated early signals, including abnormal surges in symptom-related search indices, sudden increases in the retail sales of specific pharmaceuticals, and abrupt changes in environmental pollutant concentrations. Through methods such as weighted averaging or ensemble learning, these signals are synthesized into a composite risk score. Finally, the models enable fine-grained tiered responses. Based on the magnitude of the composite risk score, the system automatically activates different levels of warning status (e.g., blue, yellow, orange, and red) and precisely disseminates corresponding warning information and recommended actions to designated responsible entities, such as community grid managers, regional hospitals, centers for disease control and prevention, and government emergency command offices. This targeted dissemination ensures the timeliness and effectiveness of response measures [3]. Through this series of mechanisms, public health risk early warning is transformed from a passive, delayed, and coarse approach into a proactive, agile, and precise closed-loop system.

3. Case Study: Practices of Energy and Chemical Enterprises in Public Health Risk Early Warning

The energy and chemical industry has long been recognized as a key sector for safety production and environmental risk prevention, owing to the extensive involvement of flammable, explosive, toxic, and hazardous substances in its production processes.

Vol 2 Issue 5 2025

Once leakage, explosion, or pollution incidents occur, they can easily lead to large-scale acute or chronic health damage, constituting typical regional public health risks. Therefore, the application of big data technologies to internal health management and external risk early warning in this industry holds strong demonstrative value.

3.1 Risk Characteristics and Data Foundations

Public health risks associated with energy and chemical enterprises mainly arise from two types of events:

(1) Sudden accidents: Such as toxic gas leaks (e.g., chlorine or ammonia) caused by storage tank rupture, which may result in mass poisoning among residents in downwind areas within a short period of time.

(2) Long-term chronic exposure: Such as sustained excessive emissions of wastewater and exhaust gases, which may lead to abnormally high incidence rates of cancer, respiratory diseases, and other health conditions in surrounding communities.

To address these risks, large-scale energy and chemical enterprises have typically deployed relatively comprehensive data acquisition systems. First, DCS/SCADA systems continuously monitor key operational parameters of production equipment, including pressure, temperature, flow rate, and liquid level. Second, continuous emission monitoring systems (CEMS) conduct 24-hour monitoring of air quality and water quality at plant boundaries and surrounding sensitive sites. Third, personnel positioning and health management systems record employees' locations and physiological indicators—such as heart rate and body temperature—through smart wearables and access control systems, and integrate these data with annual occupational health examination records. Fourth, emergency command and communication systems document the entire process of incident reporting and emergency response. Collectively, these systems generate massive volumes of high-frequency, spatiotemporally precise operational data, laying a solid foundation for the development of enterprise-specific public health risk early warning models.

3.2 Construction and Application of Big Data–Driven Early Warning Mechanisms

Taking a large domestic petrochemical industrial base as an example, the big data–based public health risk early warning platform it has established mainly consists of the following functional modules.

3.2.1 Internal Employee Health Risk Early Warning

The platform integrates employees' real-time physiological data (collected via wearable devices), job-related exposure risk levels (derived from chemical Material Safety Data Sheet, MSDS, databases), and abnormal indicators identified in historical health examination records. Machine learning models are employed to dynamically assess the health risks of individual employees. For instance, when an employee working in a benzene operation area exhibits a continuous decline in white blood cell counts over three consecutive days as detected by wearable devices, the system automatically issues a yellow-level warning, prompting the occupational health management department to conduct targeted monitoring and intervention. In this way, the prevention and control threshold for occupational diseases is shifted forward.

3.2.2 External Environmental Health Risk Early Warning

The platform is connected to hundreds of environmental monitoring micro-stations distributed across the industrial site and within a 5-kilometer surrounding area, and it integrates forecast data from meteorological authorities, including wind speed, wind direction, and atmospheric stability. Once an abnormal surge in the concentration of characteristic pollutants (e.g., volatile organic compounds, VOCs) is detected at any monitoring point, the system immediately activates dispersion simulation models (such as AERMOD) to predict pollutant migration pathways and impact ranges over the subsequent 1–6 hours. Simultaneously, the platform captures data from social media and local online forums in surrounding communities to analyze whether there is a concentration of discussions related to unusual odors or physical discomfort. If the simulation results show a high degree of consistency with public sentiment signals, the system automatically issues a joint early warning to local government emergency management offices, health authorities, and affected communities, and recommends the initiation of evacuation or protective measures.

3.2.3 Accident Scenario Simulation and Emergency Response Optimization

Under routine conditions, the platform continuously conducts "digital twin" simulations based on historical accident data and modeling techniques to evaluate the effectiveness of different emergency response plans. When an actual incident occurs, the platform can rapidly retrieve the optimal response plan and dynamically adjust the deployment of rescue forces and the allocation of medical resources based on real-time feedback from the site, such as aerial drone imagery and the positioning data of rescue personnel. This approach minimizes casualties to the greatest extent possible.

3.3 Implications of the Case Study

This case demonstrates that energy and chemical enterprises are not only potential sources of public health risks but can also serve as intelligent "sentinels" within regional public health security networks. By integrating internal safety production data with external public health demands, such enterprises can effectively safeguard the health rights of employees and surrounding residents, while simultaneously providing precise and reliable micro-level data support for government-led macro-level early warning systems. This integration achieves a win–win outcome between corporate social responsibility and public safety governance.

4. Optimization Pathways and Future Prospects

(1) Establishing law-based and standardized data-sharing mechanisms: At the national level, dedicated regulations on public health data sharing should be promulgated to clearly define the rights and responsibilities of all stakeholders. A privacy-preserving computing framework—such as federated learning and secure multi-party computation—should be established to ensure that data are "usable but not visible," thereby maximizing data value while protecting personal privacy.

(2) Enhancing model transparency and human–machine collaboration: Priority should be given to the development of explainable artificial intelligence (XAI) technologies to ensure that the logic of early warning models is transparent and their results are trustworthy. At the same time, expert review mechanisms should be established to integrate AI-driven intelligent analysis with the experiential judgment of domain experts, forming a "human-in-the-loop" decision-making model[4].

(3) Strengthening interdisciplinary talent cultivation and international cooperation: Interdisciplinary programs should be established in higher education institutions to cultivate a new generation of professionals proficient in both public health and data science. Meanwhile, active participation in global public health data-sharing networks—such as the World Health Organization's Global Outbreak Alert and Response Network (GOARN)—is essential to jointly address cross-border public health threats.

Conclusion

Public health security is closely linked to national security. In the face of complex and evolving health risks, constructing a modern risk early warning mechanism driven by big data technology is an inevitable choice. Through theoretical analysis, technical exploration, and case studies, this paper demonstrates that big data technologies can substantially enhance the foresight, precision, and coordination of public health risk early warning. The case of energy and chemical enterprises further shows that high-risk industries can also safeguard public safety through technological innovation. Future early warning systems will evolve into intelligent networks that integrate data from governments, enterprises, and other stakeholders, enabling both the detection of subtle health signals and the anticipation of threats posed by abnormal events.

References

[1] Zhang Yaofeng, Geng Zhilin, He Ruibo. Construction of a risk prevention and control system for public health emergencies driven by big data and intelligent decision-making[J]. Chinese Journal of Public Health Management, 2022, 38(04): 439–443.

[2] Wang Chao, Xu Wendong, Shi Ruyi, et al. Big data–driven public health risk monitoring: Theoretical framework and practical reflections (in English)[J]. Journal of Information Resources Management, 2022, 12(03): 118–136.

[3] Wang Chao. Research on big data–driven public health risk governance[D]. Lanzhou University, 2020. DOI:10.27204/d.cnki.glzhu.2020.003426.

[4] Wang Bing, Bai Minghao. Risk information perception and analytical models for public health emergencies in the big data environment[J]. Journal of Intelligence, 2021, 40(11): 176–181.