Review Article

DIGITAL HEALTH

DIGITAL HEALTH
Volume 10: 1–20

DOI: 10.1177/20552076241256877
journals.sagepub.com/home/dhj
Sage

Precision public health, the key for future outbreak management: A scoping review

Ellappa Ghanthan Rajendran[1], Farizah Mohd Hairi[1], Rama Krishna Supramaniam[1] and Tengku Amatullah Madeehah T Mohd[2]

Abstract

Background: Precision Public Health (PPH) is a newly emerging field in public health medicine. The application of various types of data allows PPH to deliver more tailored interventions to a specific population within a specific timeframe. However, the application of PPH possesses several challenges and limitations that need to be addressed.

Objective: We aim to provide evidence of the various use of PPH in outbreak management, the types of data that could be used in PPH application, and the limitations and barriers in the application of the PPH approach.

Methods and analysis: Articles were searched in PubMed, Web of Science, and Science Direct. Our selection of articles was based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) for Scoping Review guidelines. The outcome of the evidence assessment was presented in narrative format instead of quantitative.

Results: A total of 27 articles were included in the scoping review. Most of the articles (74.1%) focused on PPH applications in performing disease surveillance and signal detection. Furthermore, the data type mostly used in the studies was surveillance (51.9%), environment (44.4), and Internet query data. Most of the articles emphasized data quality and availability (81.5%) as the main barriers in PPH applications followed by data integration and interoperability (29.6%).

Conclusions: PPH applications in outbreak management utilize a wide range of data sources and analytical techniques to enhance disease surveillance, investigation, modeling, and prediction. By leveraging these tools and approaches, PPH contributes to more effective and efficient outbreak management, ultimately reducing the burden of infectious diseases on populations. The limitation and challenges in the application of PPH approaches in outbreak management emphasize the need to strengthen the surveillance systems, promote data sharing and collaboration among relevant stakeholders, and standardize data collection methods while upholding privacy and ethical principles.

Keywords

Big data, communicable disease, data-driven, outbreak, precision public health

Submission date: 18 August 2023; Acceptance date: 7 May 2024

Introduction

Background

Precision public health (PPH) is a new area of public health that supports the growth of precision medicine (PM) by leveraging improvements in the latest technology and information gained from Big Data (BD) to deliver more tailored public health initiatives in the communities.[1] In short, PPH is defined as an approach to implementing the right intervention

[1]Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
[2]Faculty of Medicine and Health Science, University of Islamic Science Malaysia, Kuala Lumpur, Malaysia

Corresponding author:
Farizah Mohd Hairi, Department of Social and Preventive Medicine, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia.
Email: farizah@ummc.edu.my

for the right population at the right time.[1] According to experts, PPH is capable of improving the ability to prevent disease, promote health, and reduce health disparities in populations by applying emerging methods and technologies for measuring disease, pathogens, exposures, behaviors, and susceptibility in populations; and developing policies and targeted implementation programs to improve health.[2]

This application involves the use of single or multiple data sources with large volumes of data known as BD[3,4] and utilizes methods such as machine learning[5–7] and artificial intelligence.[8,9] The combination of data sources such as omics data, clinical data, social data, patient-generated health data, environmental data, and demographic data can provide a complete perspective of a patient or community.[8–10] The application of PPH and BD would be able to improve the accuracy in select fields of public health which includes disease surveillance, risk prediction, targeted intervention, and understanding of disease.[11] This advancement in health-related technology was further catalyzed by the recent COVID-19 pandemic. Various studies have shown the significant impact of PPH approaches in managing the COVID-19 pandemic, especially in understanding the novel disease,[12] clinical management,[13,14] disease spread and burden,[15–17] population behavior,[18–20] and the effectiveness of strategies implemented.[21–23] However, the application of PPH possesses several challenges and limitations. First, there is limited knowledge and theory on PPH which thus limits our understanding of it.[24] Secondly, the availability of data sources and resources to cater to PPH applications. Current technology limits the extent to which the large volume of data could be utilized in public health fields, especially in middle- and low-income countries.[25] Furthermore, various ethical and privacy issues should be addressed such as data security, individual right to access personal information, and accessibility of vast amounts of data by various stakeholders.[2]

In this scoping review, we aimed to explore the literature on the PPH approaches in outbreak management. Outbreak management is defined as coordinated and systematic approaches taken to control and mitigate the spread of infectious/communicable disease or other harmful agents within a specific population or geographical region.[26] Hence, this review can provide insights into the various types of health-related data which could be utilized in various stages of outbreak management. The findings may be used as a policy guideline for the implementation of PPH at various stages of healthcare systems, especially in outbreak management.

Methods

Review questions

1. How does the PPH approach complement the conventional outbreak management practice in responding to infectious disease outbreaks?
2. How could various types of nonconventional health-related data complement outbreak management?
3. What are the barriers and challenges of implementing the PPH approach in outbreak management?

Eligibility criteria

A preliminary search was performed in major databases such as PubMed, Web of Science (WoS), the Cochrane Database of Systemic Review, and Joanna Briggs Institute (JBI) Evidence Synthesis to identify any new or ongoing reviews on the study topic and no reviews or studies were found.

The present review used the PCC (Population, Concept, and Context) framework proposed by the JBI. The JBI scoping review is a widely used reviewing method that incorporated other known frameworks proposed by Arksey and O'Malley, and enhancement proposed by Levac, Colquhon, and O'Brien.[27] The eligibility criteria were justified based on the article type, language, population, concept, and context (Table 1).

PRISMA extension for Scoping Reviews (PRISMA-ScR)

In our study, we employed the PRISMA-ScR framework as a guiding tool to ensure transparency and methodological rigor in the conduct and reporting of our scoping review. By adhering to PRISMA-ScR guidelines, we systematically identified, selected, and synthesized relevant literature pertaining to our research question, following a predefined protocol.

The adherence to the PRISMA-ScR checklist was robust, with 20 out of the 22 relevant criteria fully met in the conduct of this scoping review (Appendix 1). This dedication to adherence underscores our commitment to transparency and methodological rigor in reporting the review process and findings. While two criteria specific to systematic reviews were not applicable to our scoping review methodology, the overall compliance with PRISMA-ScR guidelines enhances the credibility and reproducibility of our review.

In our study, although the research protocol was developed and followed throughout the conduct of the scoping review, it was not formally published due to financial constraint. Despite the protocol not being published, rigorous adherence to a predefined protocol was maintained throughout the study process to ensure consistency and transparency in the conduct of the scoping review.

Search strategy

The search strategy aimed to locate published studies. The search was conducted in PubMed, WoS, and Science Direct databases using the keywords "precision public health" and "communicable disease" (Table 2). The search strategy was

Table 1. Article eligibility criteria.

Criteria	Inclusion	Exclusion
Article type	All types of experimental studies, observational studies, and text or opinion papers show potential evidence of the PPH application in outbreak management.	Individual cases or a small series of cases without broader generalizability or implications for outbreak management. Studies that do not show potential PPH application in outbreak management. Not published during the study period.
Study period	Studies conducted between 1st January 2017 and 1st January 2023.	Studies conducted before 1st January 2017 and after 1st January 2023.
Language	Published in the English language.	Not published in the English language.
Concept	PPH applications including Big Data or health-related data-driven applications.	Theoretical concepts or model development without outcome measurements.
Context	Communicable diseases or outbreak management, such as surveillance, resource allocations, and preventive and control measures.	Any other field of public health (non-communicable diseases, health education, etc.)

Table 2. Search strategy.

Database	Timeframe	Search terms	Studies (N = 1059), n
PubMed	From 1st January 2017 to 1st January 2023	((“precision public health"[Title/Abstract] OR “big data"[Title/Abstract] OR “data driven"[Title/Abstract]) AND (“communicable disease"[Title/Abstract] OR “outbreak"[Title/Abstract]))	333
Web of Science	From 1st January 2017 to 1st January 2023	(TS = (“precision public health” OR “big data” OR “data driven”)) AND TS = (“communicable disease” OR “outbreak”)	626
Science Direct	From 1st January 2017 to 1st January 2023	((“precision public health” OR “big data” OR “data driven”) AND (“communicable disease” OR “outbreak”))	100

conducted based on index terms and adapted for each database and information source. For instance, the PPH is closely related to BD and data-driven analytics while outbreak management, as defined previously, involves the management of communicable diseases. Hence, these keywords were included in the search terms. The PPH topic was not explored using Medical Subject Headings (MeSH) terms due to the highly specific and novel nature of the topic. No appropriate MeSH terms are available to accurately capture the concept. Hence, relying solely on keywords search yielded more relevant results. The reference list of the selected studies was screened as an additional source of evidence. We included publications in the English language in this study. We retrieved studies conducted between the 1st of January 2017 and the 1st of January 2023. This limitation was relevant to ensure that the review captures the most recent and relevant literature on PPH. Limiting the search to recent years helps to focus on current trends, developments, advancements in the PPH-related field.

Selection of studies

Following the search, all identified citations were grouped and added to the EndNote version X9 year 2018 (by Clarivate Analytics, PA, USA) and duplicates were removed. Two review authors (EGR and RKS) independently screened the eligibility of the studies. Studies with titles and abstracts that did not fit the eligibility criteria were excluded from the study. Following that, the retrieval and assessment of the full text were performed, and only relevant studies were kept using predefined inclusion and exclusion criteria. Any disagreement between the review authors was resolved by further discussion with the third review author (FMH) who acts as an arbiter. Studies were included if they describe the PPH approach, including BD application in communicable disease or outbreak management.

Data extraction

Data extraction from the studies was performed independently by two of the review authors (EGR and RKS) using a data extraction tool developed by the reviewers. The researchers first identified the key variables or pieces of information that needed to be extracted from each study to address the research question and objectives of the scoping review. These variables were selected based on their relevance to the topic under investigation and the specific aims of the review. For instance, in terms of types of data, the researchers established inclusion criteria based on the types of data sources relevant to the research question and objectives. The inclusion criteria were determined based on the scope of the review and the types of data sources commonly used in PPH applications.

Once the relevant variables were identified, the researchers developed a structured data extraction tool to systematically capture these variables from each included study. The tool was designed to ensure consistency and completeness in data extraction across all studies.

The data extracted included specific details as follows:

1. Author(s) name.
2. Publication year.
3. Type of study, including cross-sectional, cohort, experimental, and review.
4. Disease studied includes specific infectious/communicable disease such as dengue and influenza or combination of diseases.
5. Type(s) of data source, namely traditional health-related data (for example, medical records, genetic data, and surveillance data) and non-conventional data (for example meteorology data, zoonotic data, and Internet data).
6. Type(s) of PPH application, explaining the approach in disease management, for instance, predicting risk, allocation of resources, and signal detection.
7. Findings, the outcomes of the study.
8. Type(s) of limitations/barriers the researcher identified during the study was conducted.

The summary of the articles is presented in Table 3.

Data analysis

The key information from the selected studies was systematically extracted, including data as mentioned previously. This step ensured that relevant data were captured in a structured manner for further analysis.

The extracted data were then analyzed thematically to identify common themes, patterns, and trends across the literature. This involved organizing the findings into meaningful categories or concepts that captured the essence of the research findings. With regards to types of PPH application, we constructed a table based on the review by Dolley et al. on the role of BD in PPH as the use of BD is vital in implementing the precision of public health strategies.[11] These roles include performing disease surveillance and signal detection, predicting risk, targeting interventions, and understanding disease.

Moreover, rather than relying on quantitative measures of evidence, the synthesized findings were presented in a narrative format. This approach allowed for a qualitative interpretation of evidence, highlighting key themes, relationships, and areas of consensus or contention within the articles.

The synthesized findings were interpreted in the context of the research question and objectives of the scoping review. Any implications, limitations, or gaps identified in the literature were discussed, providing insights into the state of knowledge on the topic and guiding future research directions.

Patient consent

In the context this scoping review, patient consent was not required as the synthesis of existing literature does not involve direct interaction with human subjects or their personal information.

Results

Search outcome

The systematic search process (Figure 1) involved a comprehensive search of relevant literature sources, including PubMed, WoS, and Science Direct databases, to identify articles addressing the research question. The initial search yielded a total of 1058 articles, with PubMed contributing 332 articles, WoS contributing 626 articles, and Science Direct contributing 100 articles. Following the removal of duplicates, 714 unique articles remained for further evaluation.

The screening phase commenced with the assessment of article titles to determine their suitability for inclusion in the scoping review. During this process, 647 articles were excluded due to a variety of reasons diverging from the research questions. Among these exclusions were studies focusing on non-communicable diseases such as cancer, Parkinson's disease, and diabetes mellitus, which did not directly address infectious/communicable disease outbreaks or the application of PPH approaches in outbreak management. Additionally, articles centered on the development of decision-making algorithms or artificial intelligence modeling were excluded as they were not directly related to PPH strategies or outbreak management. Furthermore, studies

Table 3. Article summary.

No.	Author(s)	Publication Year	Country of Origin	Type of Study	Disease Studied	Types of Data	Type of PPH Application	Findings	Types of Limitations/Barriers
1	Jung et al.	2019	Korea	Cohort	Infectious Disease	1. Surveillance/Epidemiological Data 2. Personal Health-related Data	Performing disease surveillance and signal detection	Insurance claim data complement the low-reporting CDC data to improve surveillance.	1. Data quality and availability 2. Health disparities and inequity
2	Moller et al.	2018	Denmark	Review	Food-borne disease	1. Personal Social-related data.	1. Performing disease surveillance and signal detection 2. Targeting treatment or intervention	1. Generating hypotheses 2. Analytical epidemiology 3. Tracking back of source/vehicle of the outbreak 4. Supporting existing evidence/ hypothesis 5. Targeted intervention of the exposed of the outbreak investigations.	1. Data quality and availability
3	Chen et al.	2019	China	Cohort	Avian Influenza A (H7N9)	1. Internet Query Data 2. Social Media Data	Performing disease surveillance and signal detection	1. Social media index related with H7N9 consistently negative lag values. 2. Search engine index on H7N9 increased 4 weeks prior to occurrence.	1. Data quality and availability 2. Data integration and interoperability 3. Privacy concerns
4	Ippoliti et al.	2019	Italy	Cohort	Bluetongue vectors and West Nile diseases	1. Environmental/Geographical Data	Predicting risk	The data can be used to monitor the environmental/climate changes that could facilitate VBD outbreak and function as early warning system.	1. Data quality and availability 2. Privacy concerns
5	Du et al.	2020	Haiti	Experimental	Cholera	1. Clinical Data 2. Surveillance/Epidemiological Data	1. Performing disease surveillance and signal detection 2. Targeting treatment or intervention	1. Antibiotic intervention is more effective when budget is abundant. 2. ORT is more cost-effective intervention when the infected population is large.	1. Data quality and availability 2. Data integration and interoperability
6	Betancor et al.	2021	Germany	Cohort	Conjunctivitis	1. Internet Query Data	Performing disease surveillance and signal detection	Correlation identified between search data volume and density of patients presented with conjunctivitis.	1. Data quality and availability 2. Infrastructure and technical capabilities
7	Huang et al.	2020	China	Experimental	COVID-19	1. Surveillance/Epidemiological Data 2. Clinical Data 3. Internet Query Data	Performing disease surveillance and signal detection	1. Online medical consultation data has better predicting power than OMS. 2. OMA failed to predict trends of COVID-19.	1. Data quality and availability 2. Data integration and interoperability
8	Nei L et al.	2020	China	Experimental	COVID-19	1. Global Positioning Data	1. Targeting treatment or intervention 2. Resource allocation and optimization	The data source able to show the effectiveness of public interventions by measuring the number of people presented at specific place during specific time.	1. Public acceptance and trust 2. Health disparities and inequity
9	Zhu et al.	2022	Japan	Cross sectional	COVID-19	1. Environmental/Geographical Data	Performing disease surveillance and signal detection	Wastewater surveillance can be a predictor to implement targeted intervention and provide accurate outcome when the is a lag in conventional reporting system.	1. Data quality and availability 2. Infrastructure and technical capabilities 3. Cost and resource allocation

(continued)

Table 3. Continued.

No.	Author(s)	Publication Year	Country of Origin	Type of Study	Disease Studied	Types of Data	Type of PPH Application	Findings	Types of Limitations/Barriers
10	Razavi et al.	2021	Iran	Cross sectional	COVID-19	1. Environmental/Geographical Data	Predicting risk	1. Risk is higher at public transport stations and pharmacies. 2. Risk maps could help to make arrangements in high-risk areas to reduce outbreak.	1. Data quality and availability 2. Infrastructure and technical capabilities
11	Minetto et al.	2020	USA	Experimental	COVID-19	1. Global Positioning Data	1. Targeting treatment or intervention 2. Resource allocation and optimization	The framework able to automatically detect different types of vehicles from satellite images indicating the economic activity. This help decision makers in spotting situation that demand immediate action.	1. Data quality and availability 2. Cost and resource allocation
12	Xiao et al.	2021	China	Cohort	COVID-19	1. Environmental/Geographical Data 2. Surveillance/Epidemiological Data	Understanding disease	The median incubation period of the city varies and the longest when the meteorological temperature maintained at 10 to 15 degrees Celsius. The association facilitates the prediction of the transmission and evolution cycle of COVID-19 outbreak.	1. Data quality and availability
13	Jimenez et al.	2020	Spain	Cohort	COVID-19	1. Internet Query Data 2. Surveillance/Epidemiological Data	Performing disease surveillance and signal detection	1. Google Trends correlated with daily incidence of confirmed COVID-19 cases, hospitalization, ICU admission and death. 2. Google Trend data correlated with daily incidence of COVID-19 with an 11-day time lag.	1. Data quality and availability 2. Data integration and interoperability 3. Privacy concerns
14	Ma et al.	2022	USA	Cohort	COVID-19	1. Internet Query Data 2. Surveillance/Epidemiological Data	Performing disease surveillance and signal detection	The framework was capable to predict the COVID-19 death incident 1 to 4 weeks ahead.	1. Privacy concerns 2. Public acceptance and trust
15	Ali et al.	2020	Indonesia	Experimental	Dengue	1. Surveillance/Epidemiological Data 2. Environmental/Geographical Data	Performing disease surveillance and signal detection	1. Local weather data is able to create an early warning system of DHF cases in the locality. 2. Able to control DHF outbreak and low budget allocation required.	1. Data quality and availability 2. Data integration and interoperability
16	Sanchez et al.	2022	Brazil	Experimental	Dengue	1. Surveillance/Epidemiological Data 2. Entomology/Zoology Data	Performing disease surveillance and signal detection	By using EDI time series as a predictor, the best performance was obtained for 6 to 4 weeks before the target week for dengue incidence.	1. Data quality and availability 2. Data integration and interoperability
17	Li et al.	2022	Brazil	Experimental	Dengue	1. Environmental/Geographical Data 2. Surveillance/Epidemiological Data	Performing disease surveillance and signal detection	LTSM time series 5 weeks ahead had the best performance to predict the dengue cases.	1. Data quality and availability 2. Data integration and interoperability
18	Kerdprasop et al.	2019	Thailand	Experimental	Dengue	1. Environmental/Geographical Data 2. Surveillance/Epidemiological Data	Performing disease surveillance and signal detection	Chi-squared automatic interaction detection (CHAID) model is the best forecasting model for dengue incidence in the chosen 4 different provinces in Thailand.	Not mentioned.

(continued)

Table 3. Continued.

No.	Author(s)	Publication Year	Country of Origin	Type of Study	Disease Studied	Types of Data	Type of PPH Application	Findings	Types of Limitations/Barriers
19	Sylvestre et al.	2022	France	Review	Dengue	1. Environmental/Geographical Data 2. Surveillance/Epidemiological Data 3. Clinical Data 4. Internet Query Data 5. Social Media Data	Performing disease surveillance and signal detection	Combination of novel Big Data streams with machine-learning methods for dengue surveillance and prediction showed a promising improvement in the accuracy of the outcome compared to conventional data source alone.	1. Data integration and interoperability
20	Huang et al.	2017	China	Experimental	HFMD	1. Environmental/Geographical Data 2. Surveillance/Epidemiological Data 3. Internet Query Data	Performing disease surveillance and signal detection	A model that includes cases, query data, and meteorological data provided best and effective estimating ability which involve the past history of cases, climate changes, and people's reaction and response.	1. Data quality and availability
21	Yoon et al.	2022	Korea	Experimental	Highly pathogenic avian influenza (HPAI)	1. Global Positioning Data 2. Surveillance/Epidemiological Data 3. Environmental/Geographical Data	Predicting risk	1. Livestock vehicle movement is strongly associated with the HPAI virus spread during an outbreak. 2. Targeted control measure implementation could be conducted with limited resources.	Not mentioned.
22	Almoodi et al.	2021	Malaysia	Review	Infectious Disease	1. Social Media Data	Performing disease surveillance and signal detection	1. Motivation related to disease mitigation: Sentiment analysis has significant role in monitoring, discovery, news sharing and policies making aspects. 2. Motivation related to data analysis: Public data sentiment capable of preventing an outbreak, visualizing transmission, tracking trends which leads to more efficient strategy implementations. 3. Theoretical contribution: Provide the insight on the capabilities of social media data in providing potential benefits in studies. 4. Practical contribution: Helps to determine highly important information that can be used for sentiment analysis and data mining and the method of data extraction.	1. Data quality and availability
23	Alessa et al.	2018	USA	Review	Influenza		Performing disease surveillance and signal detection	1. Most studies used Pearson correlation and root Mean Squared Error method. 2. SNEFT yields a very high correlation coefficient with the use of ground truth. 3. Eliminating repeated tweet and repost within a week by same user improve the outcome.	1. Data quality and availability 2. Health disparities and inequity

(continued)

Table 3. Continued.

No.	Author(s)	Publication Year	Country of Origin	Type of Study	Disease Studied	Types of Data	Type of PPH Application	Findings	Types of Limitations/Barriers
								4. Distinction between the infection and awareness shows high correlation with CDC data using Pearson correlation.	
24	Walsh et al.	2022	India	Cohort	Japanese Encephalitis	1. Environmental/Geographical Data 2. Clinical Data 3. Personal Social-related Data 4. Entomology/Zoology Data	1. Performing disease surveillance and signal detection 2. Predicting risk	1. The risk increases sharply as the species richness increased (0–4 species) and peaked at 5 species and decreased sharply. 2. Area with highest level richness were associated with low risk of JEV outbreaks.	1. Data quality and availability 2. Infrastructure and technical capabilities
25	Zhang et al.	2019	China	Experimental	Pertussis	1. Environmental/Geographical Data 2. Internet Query Data 3. Personal Social-related Data	Performing disease surveillance and signal detection	1. Strong positive association observed at 2-weeks lag for BI, 1-week lag for temperature and 2-weeks lag for rainfall. 2. Use of search index and climate data for monitoring can assist governing body to implement cost-effective control activity with a longer time window.	1. Data quality and availability
26	Cummings et al.	2018	Uganda	Cohort	SARI	1. Surveillance/Epidemiological Data 2. Genomic Data	1. Predicting risk 2. Understanding disease	1. Combination of spatiotemporal analytics with a viral oligonucleotide probe capture and high-throughput sequencing platform able to detect and characterize zoonotic and vaccine preventable viral pathogens within specific subpopulations in Uganda. 2. Applied in real-time to routinely collected surveillance data, may enable targeted deployment of vaccination resources, trigger investigations of unusual clusters, facilitate infection control measures, and direct surveillance activities to "hotspots" from which outbreaks are most likely to emerge.	1. Data quality and availability 2. Infrastructure and technical capabilities
27	Davis et al.	2018	USA	Experimental	West Nile Virus	1. Surveillance/Epidemiological Data 2. Environmental/Geographical Data 3. Entomology/Zoology Data	Performing disease surveillance and signal detection	1. The risk of human disease was highest after periods of high temperature and high negative vapor deficit, but only when mosquito pools in the early season showed a rapid spread of the virus. 2. Atmospheric water vapor was a more accurate predictor of WNWV cases compared to precipitation, soil moisture, and evapotranspiration.	1. Data quality and availability 2. Health disparities and inequity

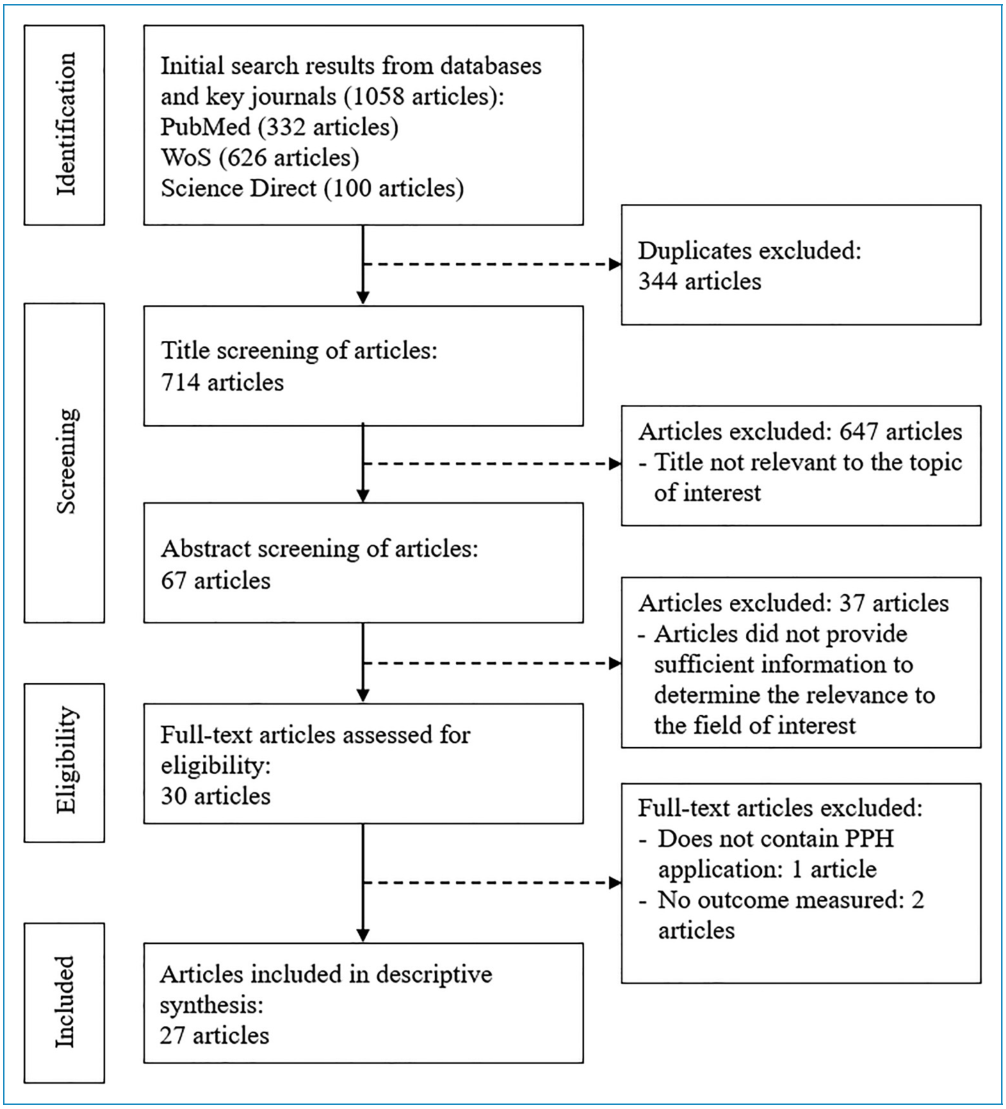

Figure 1. Flowchart of preferred reporting items for systematic reviews and meta.

related to BD application in non-health-related sectors such as agriculture and economy were deemed outside the scope of the review, which specifically targeted health-related data and outbreak management. Finally, case studies lacking generalizability or potential impact of PPH approaches on outbreak management aspects were also excluded, indicating a lack of relevance to the research questions. These exclusions were made to ensure that the articles included in the scoping review were aligned with the defined research objectives, focusing specifically on the role of PPH in outbreak management.

Subsequently, abstract screening was conducted on the remaining 67 articles. After careful evaluation, an additional 37 articles were excluded due to no information on potential PPH application in infectious/communicable or outbreak management were stated (exclusion criteria in Table 1). The remaining 30 articles underwent a thorough full-text assessment to ascertain their eligibility for inclusion in the scoping review. Among the full-text articles, one was excluded as it did not pertain to the specific application of PPH, while two articles were excluded due to the absence of outcome measures related to the research question. This meticulous selection process resulted in a final set of 27 articles that met the eligibility criteria and were deemed appropriate for inclusion in the descriptive synthesis of the scoping review.

Study demographics

Table 4 shows the demographic of studies included in the review. Most of the articles were published in the year 2020 ($n = 7$, 25.9%) followed by the year 2020 ($n = 6$, 22.2%) and the year 2019 ($n = 5$, 18.5%). Meanwhile, 12 out of 27 articles (44.4%) were experimental study designs and followed by cohort ($n = 9$, 33.3%), review ($n = 4$, 14.8%), and cross-sectional ($n = 2$, 7.4%). We found that most of the selected articles originated from China ($n = 6$, 22.2%) followed by the United States of America ($n = 4$, 14.8%), while Brazil and Korea had two articles (7.4%) each. Based on the type of disease studied in the articles, COVID-19 accounted for the highest number ($n = 8$, 296%) reflecting the significant impact of the ongoing pandemic on public health research. Meanwhile, diseases such as dengue ($n = 5$, 18.5%), influenza ($n = 3$, 11.1%), infectious diseases ($n = 2$, 3.7%), and West Nile virus outbreak ($n = 2$, 3.7%), also received attention in the literature.

Principal findings

PPH applications. The reviewed articles were classified based on the five themes of PPH application in outbreak management as mentioned in methods chapter. On top of that, we found that two articles were focused on resource allocation and optimization which was included in the theme of PPH application. Some of the reviewed articles contained more than one type of PPH application. Most of the review articles were focused on the PPH approach in performing disease surveillance and signal detection ($n = 20$, 74.1) followed by predicting risk ($n = 5$, 18.5%), targeting interventions ($n = 4$, 14.8%), understanding the disease ($n = 2$, 7.4%), and resource allocation and optimization ($n = 2$, 7.4%).

Data sources in the PPH application. From the article analysis, we found that various data types were used in implementing PPH in outbreak management, either as a single data type or in combination. The widely used data in PPH application was surveillance or epidemiological data (51.9%) such as national infectious disease surveillance data,[28–30] local region dengue incidence data,[31,32] history of disease occurrence data,[33] and COVID-19 morbidity and mortality rate.[34] Meanwhile, environmental or geographical data (44.4%) were also often used in studies such as daily local weather data which includes mean humidity and temperature,[31,35] wastewater surveillance for the COVID-19 virus,[36] and urban topography data.[37] Similarly, Internet queries and social media data (37.0% and 22.2%, respectively) such as Google search trends or Baidu search indexes on specific diseases,[30,33,38–42] Weibo posting index,[38] and Twitter trends.[18,41] Additionally, personal health-related data (3.7%) such as personal insurance claim data,[28] and personal social-related data (11.1%) such as consumer purchase data,[43] human footprint data,[44] and school calendar patterns[42] were utilized as well. We also found that genomic data[44,45] and various clinical data (25.9%) such as patient medical records,[29,41] laboratory data,[44] and online medical appointment or consultation data[30] were included in some of the studies. Another important data type that is crucial in PPH implementation is global positioning data which could be retrieved such as mobile positioning terminals,[46] street or satellite images,[47] and global positioning systems (GPS) coordinates data.[48] Lastly, entomology or zoology data were also utilized in outbreak management such as vector egg density index in the dengue outbreak,[49] bird species surveillance in the Japanese Encephalitis outbreak,[44] and mosquito infection growth rate in the West Nile virus outbreak.[50]

Gaps and challenges in PPH application. Our synthesis revealed several noteworthy limitations and barriers associated with PPH application in outbreak management. The quality and availability of data emerged as the most prevalent limitation mentioned in the reviewed articles ($n = 22$, 81.5%), followed by data integration and interoperability ($n = 8$, 29.6%), infrastructure and technical capabilities ($n = 5$, 18.5%), privacy concerns ($n = 4$, 14.8%), and health disparity and inequity ($n = 4$, 14.8%). Meanwhile, cost and resource allocation and public acceptance and trust were the least mentioned in the articles ($n = 2$, 7.4%, respectively). On the other hand, in two of the articles, the authors did not mention the limitations or barriers encountered during the study period.[32,48]

Discussion

Principal findings

To date, several articles have been published regarding the role of PPH and BD in the public health field.[51–55] However, to our knowledge, this is the first review focusing specifically on the PPH application in outbreak management. The results from this scoping review will aim to address the application of PPH in outbreak management. We aim to describe the types of PPH approaches applied in outbreak management as well as the limitations and challenges encountered throughout the implementations.

PPH applications in outbreak management

Disease surveillance involves systematic collection, analysis, and interpretation of health-related data to determine early signals of potential disease outbreaks. PPH leverages various sources of data to improve the accuracy and timeliness of disease surveillance and monitoring by including not only currently available conventional disease surveillance systems, but also integrating other non-conventional data sources such as social media and Internet query data,[18,30,33,34,38–42] and environmental data.[31,35–37,56] For instance, Chen et al. (2019) found that the number of Internet query data in search engines and social media increased as early as four weeks before the epidemic occurrence.[38] Similarly, Li et al. (2022) developed a dengue prediction framework by integrating multiple environmental

Table 4. Study demographics.

Characteristic	n	%
Year		
2017	1	3.7
2018	4	14.8
2019	5	18.5
2020	6	22.2
2021	4	14.8
2022	7	25.9
Study design		
Cohort	9	33.3
Review	4	14.8
Experimental	12	44.4
Cross-sectional	2	7.4
Country of origin		
Brazil	2	7.4
China	6	22.2
Denmark	1	3.7
France	1	3.7
Germany	1	3.7
Haiti	1	3.7
India	1	3.7
Indonesia	1	3.7
Iran	1	3.7
Italy	1	3.7
Japan	1	3.7
Korea	2	7.4
Malaysia	1	3.7
Spain	1	3.7

(continued)

Table 4. Continued.

Characteristic	n	%
Thailand	1	3.7
Uganda	1	3.7
USA	4	14.8
Disease studied		
Cholera	1	3.7
Conjunctivitis	1	3.7
COVID-19	8	29.6
Dengue	5	18.5
Food-borne diseases	1	3.7
HFMD	1	3.7
Infectious diseases	2	7.4
Influenza	3	11.1
Japanese Encephalitis	1	3.7
Pertussis	1	3.7
SARI	1	3.7
West Nile virus	2	7.4
Types of data used		
Surveillance/epidemiological data	14	51.9
Personal health-related data	1	3.7
Personal social-related data	3	11.1
Internet query data	10	37.0
Social media data	6	22.2
Environmental/geographical data	12	44.4
Clinical data	7	25.9
Genomic data	2	7.4
Global positioning data	4	14.8
Entomology/zoology data	5	18.5

(continued)

Table 4. Continued.

Characteristic	n	%
Type of PPH application		
Performing disease surveillance and signal detection	20	74.1
Predicting risk	5	18.5
Targeting treatment or intervention	4	14.8
Understanding disease	2	7.4
Resource allocation and optimization	2	7.4
Limitations/barriers		
Data quality and availability	22	81.5
Privacy concerns	4	14.8
Data integration and interoperability	8	29.6
Infrastructure and technical capabilities	5	18.5
Health disparities and inequity	4	14.8
Cost and resource allocation	2	7.4
Public acceptance and trust	2	7.4
Not mentioned	2	7.4

data and Brazil's national historical dengue data which accurately forecast the weekly dengue cases five weeks ahead.[56] The ability of PPH approaches to forecast an outbreak earlier than the actual occurrence provides adequate time to establish effective prevention and control measure strategies to reduce or nullify the possible impact due to outbreak occurrence.

The risk prediction potential of the PPH approaches involves the adaptation of predictive analytics models to identify specific populations or regions with a higher risk of disease outbreak occurrences. These predictive models integrate various risk factors, such as biological data,[44] socio-demographic data,[44] environmental alterations,[35,37] and population behavior patterns[48] to generate specific precise risk scores or assessments. For instance, Yoon et al. (2022) conducted a study to determine the risk of highly pathogenic avian influenza (HPAI) in Korea using deep learning computation. The authors utilized livestock surveillance program data, local meteorological reports, and the movement of livestock vehicles in the interior of the poultry farm to predict the risk of HPAI virus spread. They found that livestock vehicle movements were strongly associated with the HPAI virus spread during an outbreak.[48] Meanwhile, in Italy, Ippoliti et al. (2019) attempted to develop a risk map of West Nile Disease (WND) by adopting a data-driven spatial clustering approach using topographical, environmental, and climate factors.[35] They found the Bluetongue vector was strongly associated with high mean temperature, dry environments, flattened and sunny surfaces, and low vegetation. Based on this information, they were able to develop a much more precise risk map that could facilitate decision-makers to prioritize preventive measures and resource allocations more effectively.

The fundamental application of PPH is the right intervention for the right population at the right time.[2] Therefore, targeted intervention is an essential component of PPH to optimize the outcomes of the intervention. This method is tailored based on individual or population-level characteristics which involve more than one type of data. For instance, a review published by Moller et al. (2018) showed that various types of consumer purchase data, such as loyalty card data, bank card data, and history of transactions were able to facilitate outbreak management.[43] More tailored intervention could be provided to specific exposed populations based on the outbreak investigation. On top of that, the review also reported that consumer purchase data were used in generating hypotheses, analytical epidemiology, tracing back the source of an outbreak, and supporting existing evidence or hypothesis.

The other key application of PPH in outbreak management is the capability to optimize the allocation of limited healthcare resources, such as manpower, medical supplies, and infrastructures. Via utilization of data-driven simulations and models, PPH endeavors to identify resource gaps, anticipate future resource needs and optimize the resources allocation to regions or populations with the highest disease burden or vulnerability. The evidence of such advantages was mentioned by Minetto et al. (2020) in a study conducted in the United States of America to develop a framework to map the flow of vehicles over the area of interest with the combination of publicly available data sources, such as satellite images and OpenStreetMap, and machine learning-based vehicle detection method.[47] The author reported that the proposed framework was able to provide stable and accurate vehicle counts in most regions of interest. This information helps the decision makers to COVID-19 control strategies based on the different region of interest groups and their expected behavior. On the other hand, the adoption of PPH also allows for more reliability and accuracy in measuring the effectiveness of an intervention or strategy. For instance, during the COVID-19 pandemic, China implemented five phases of public gathering striction strategies to control COVID-19 transmission.[46] Nei et al. (2020) conducted a study to evaluate the impact of these strategies using mobile terminal positioning data as the source of information. The application of mobile terminal positioning data in the difference-in-difference (DID) model provided information on the number of people present at a specific time duration at designated points of interest. The author was able to show the effectiveness of public gathering restrictions by measuring the number of people presented at specific places during a specific time.

The concept of PPH application of leveraging data and advanced analytics allows us to gain insights into the underlying causes, mechanisms, and dynamics of diseases. Integration of various types of data from multiple sources, such as genomic data, clinical data, social data, environmental data, and social determinants of health allows PPH to provide a much more accurate understanding of disease etiology and progression.[1,11] The recent COVID-19 pandemic is a perfect example of how quickly we were able to understand the nature of the disease. A study was conducted in two different cities in China to understand the nature of COVID-19 using environmental and national surveillance data.[57] The author found that the incubation period of COVID-19 disease varies, and the longest incubation period was seen if maintained mean temperature at 10 to 15 degrees Celsius (50 to 59 degrees Fahrenheit). This evidence of association played an important role in the prediction of the transmission and evolution cycle of the COVID-19 outbreak in China. Meanwhile, a study was conducted in Uganda to identify clusters of unexplained, influenza-negative severe acute respiratory infections (SARI) by applying spatiotemporal analysis using SARI surveillance data and genomic data.[45] The analysis identified nine statistically significant clusters accounting for 10.4% of unexplained cases. The method, applied in realtime to routinely collected surveillance data, able to mitigate the outbreak of preventable viral disease by targeted vaccination strategy, capable to signal outbreak investigation of unusual clusters of pathogens and resource allocations to higher risk areas (hotspots).

Despite having several proof-of-concept applications of PPH in outbreak management, there are still many unexplored aspects in this approach. As tailored or data-driven approaches had been well established in other sectors such as finance, economy, and education, there is potential of adopting these applications into healthcare system particularly in outbreak management. For instance, the utilization of data-centric strategies within health marketing such as tailored communication and engagement, seamlessly measuring success, and better insight of the target population.[58–60] These strategies facilitate the customization of messaging and communication directed towards distinct demographic segments, yielding substantial advantages.[58] Moreover, data analytics offers the capacity to furnish immediate and dynamic insights into the performance of a campaign.[60] This real-time understanding empowers agile adjustments and refinements to endeavors and stratagems. On top of that, utilizing data encompassing demographic attributes, geographical disposition, as well as health-related behaviors and attitudes, affords the capacity to formulate campaigns of heightened efficacy and discerningly allocate focus to paramount health imperatives.[58,59]

Furthermore, incorporating robotic technology to enhance the effectiveness of public health strategies, especially outbreak management will be the next groundbreaking approach. This unexplored technology in public health have shown tremendous benefits in clinical practices.[61–63] Some possible advantages of robotic applications in outbreak management includes enhancing data collection,[64,65] smarter prediction of outbreak progression,[64,65] act as a health educator,[64,66] remote checkups using biosensors,[65,67] automated logistic and delivery systems,[64] as well as surveillance and preventive strategies in highly hazardous locations.[67] However, these possible applications warrant further research and concept development in public health discipline.

Types of data in PPH applications

The PPH approaches are dependent on the large volume of information from conventional and non-conventional data sources. One of the important health-related data sources that are readily available is molecular profile data.

This data are information regarding the features of molecular profiles, such as genes, proteins, chemical compounds, carbohydrates, and metabolites which help us to provide us a better understanding of relationships between molecules of an organism. The research of such interactions is critical in the discipline of PPH.[68] This research can provide numerous discoveries, such as determining whether there are distinct variants of a disorder and identifying characteristics that may contribute to the effectiveness of a possible treatment.[45,69] In a disease outbreak, the use of such data can provide accurate information regarding the different variants that exist within the same species of pathogen which could provide a better understanding of the pathogenesis and virulence of the pathogen. This information is vital to establish effective intervention and prevention measures during the outbreak.[29,70]

Another health-related data that is vital in PPH approaches is clinical data. Clinical data includes diagnostic procedures such as Computerized Tomography scans, electrocardiograms, and x-rays. In addition, certain forms of clinical data may be acquired and controlled entirely by patients through the use of health equipment such as digital scales, wearables, and digital sphygmomanometers.[11] However, such clinical data resources and applications are not routinely supervised or recorded in healthcare institutions or patients' Electronic Health Records (EHR).[10] Clinical data may be obtained from various sources, including EHRs, administrative data, claims data, patient registries, questionnaires, and clinical studies data.[71] Such data could provide valuable information in the surveillance of population at risk,[29,45,72] determining disease burden[28] and forecasting the occurrence of future disease outbreak.[41]

Meanwhile, social data, nonconventional data is simply defined as information made publicly available on social networking sites, such as a user's location, network of friends, and language.[73] Extending this approach, the use and interpretation of specific facts are inextricably linked to social relationships such as behavioral data. This would include social networking sites like Facebook, Twitter, and Google (for example, browsing history and Google Trends), as well as data on social engagements. These data are capable to provide additional information on disease surveillance[38,74] and complement conventional data to provide more accurate information.[40] For example, Deiner et al. (2016) examined the incidence of diagnosed conjunctivitis using Twitter and Google search data.[74] They found a strong correlation between seasonal conjunctivitis and Google search data and a moderate correlation with Twitter data. However, caution should be exercised when utilizing social data since many of these study findings may not be reproducible.[10]

On the other hand, environmental data is associated with information obtained about the environment in which individuals and communities live, such as water quality,[36] climate,[35,57] landscape suitability,[44] global positioning satellite,[48] geographical information system.[35,56] Several studies have shown evidence of the importance of incorporating environmental data into outbreak surveillance and management help to improve the accuracy of predicting the risk of a disease outbreak,[44,48] targeted surveillance[35] and forecast of a disease outbreak[56] which led to early preparedness and interventions. The types of data that are applicable in the PPH approach are limitless and depend widely on the availability and accessibility of such data sources that have the possibility to further enhance the effectiveness of outbreak management.

Tackling the gaps and challenges in PPH approach in outbreak management

The most prevalent limitation was data quality and availability. This finding underscores the significant challenge posed by poor quality or inadequate data, fragmented data, and non-standardized data collection methodologies. These factors lead to ineffective implementation of PPH strategies in outbreak management. For instance, the limitation on the information on COVID-19 surveillance availability in the public domain to identify contact and infected data leads to delays in outbreak investigation and prevention strategies.[57] Data quality has been a major concern in utilizing social media data whereby the noisy nature of data could reduce the accuracy of analytic outcomes.[18]

Secondly, due to the complexities associated with integrating data from diverse sources, such as national healthcare systems, public health surveillance systems, and meteorological monitoring systems. Variations in data formats, compatibility between different data systems, and restriction in seamless data exchange mechanisms contribute to the challenge. For instance, developing novel BD streams commonly uses public domain data sources such as disease surveillance data, Internet queries, and environmental data.[52] However, some data such as genomic data and entomological data are underutilized most likely due to limited access to designated personnel only.[41] Many studies attempted to integrate various data sources into single-stream BD. However, due to the different data systems and the absence of seamless data exchange mechanisms, the variation of data type in the BD stream is limited.[29,34,56]

Meanwhile, insufficient infrastructure and outdated technological resources contribute to the limited implementation of PPH in outbreak management. Moreover, limited access to necessary tools and expertise further constrains the applicability of PPH in outbreak management, such as COVID-19 forecast whereby manual collection of wastewater samples may lead to short-term heterogeneity of viral RNA abundance.[36] Additionally, resource constraints, such as inadequate budget allocation and accessibility to the latest technologies can compromise the ability to predict an outbreak incidence and disease burden. For instance, the estimation of conjunctivitis incidence using Internet query data is more saturated in urban regions with adequate Internet access than the rural communities.[39]

Privacy concern is one of the main barriers, especially upon involving personal data such as personal health records, mobile surveillance data, social media data, and genomic data. Issues about the collection, use, and sharing of sensitive health and social data, such as insurance claim data[28] and consumer purchase data[43] can encounter public resistance and legal challenges. Therefore, policymakers should prioritize data security and discretion regardless of the data type, sources, storage, and accessibility.[75]

On the other hand, public health agencies face challenges in addressing health disparities and ensuring equitable access to healthcare services, and implementing interventions across diverse population groups.[45] These challenges are often seen in low- and middle-income countries where the margin of health disparities and inequity is very high compared to high-income countries.[45] Furthermore, within a country setting, inequity is seen between urban and rural regions, especially in infrastructure and health service availability and limited information on public health surveillance.[76]

The cost incurred to possess the latest technologies limits the advancement in public health systems, especially in low- and middle-income countries.[45,47] Similarly, lesser resources, including trained personnel, competing priorities, and allocation discrepancies impede the successful implementation of PPH approaches in outbreak management.[36]

Lastly, public acceptance and trust emphasize the importance of garnering public acceptance and establishing trust in PPH approaches. As mentioned by Nei et al. (2020), there is a possible risk of overestimation of the effect of the study, in this case, the effectiveness of public intervention using mobile positioning data, because the acceptance of the public towards the evaluation process was not considered.[46] On the other hand, Ma et al. (2022) were concerned that the search index data is highly sensitive to media coverage leading to instability that could propagate into the prediction of an outbreak.[40]

Addressing barriers and challenges in PPH necessitates a multifaceted and comprehensive approach encompassing scientific, technological, ethical, and social strategies. Data collection and integration obstacles, such as the lack of high-quality and interoperable data sources, can be mitigated by investing in data infrastructure, fostering data sharing and collaboration, and implementing advanced integration techniques.[77,78] Meanwhile, in terms of privacy and ethical concerns associated with sensitive health data, a robust data governance framework is essential, emphasizing data anonymization, informed consent, and encryption for secure storage.[79,80] Similarly, navigating regulatory and policy uncertainties involves advocating for adaptive regulations that keep pace with innovation while safeguarding individual rights.[81] Effective communication and education strategies should be tailored for various stakeholders, and success stories should be highlighted to underscore the value of precision interventions.[82]

PPH approaches can exacerbate health disparities if not implemented equitably. Therefore, overcoming health disparities that might arise from precision approaches requires inclusivity, diversity in data collection, and interventions that consider socioeconomic factors.[83] Meanwhile, health literacy challenges such as, relying on individuals' understanding of their health data, which may require a high level of health literacy, can be addressed by developing user-friendly tools and educational resources that simplify complex health data interpretation.[84] On top of that, cultural beliefs and societal norms can influence individuals' willingness to engage with precision health initiatives. It is necessary to tailor strategies that align with community preferences and values, engaging community leaders and influencers to promote acceptance and adoption.[85,86]

Meanwhile, to manage the rapid pace of technological advancements, collaboration between researchers, technologists, and public health professionals is crucial to ensure responsible integration of emerging technologies like AI and genomics.[87,88] Moreover, rigorous evaluation methods and collaborations between researchers, clinicians, and data scientists are vital to validate the impact of tailored interventions. On top of that, it is challenging to implement PPH as it can be resource-intensive, especially in resource-limited settings. Hence, managing resource allocation demands prioritization of cost-effective interventions, potentially through public-private partnerships.[89,90]

In essence, conquering the barriers and challenges of PPH necessitates a harmonious blend of these strategies, facilitated by collaboration between diverse stakeholders, including researchers, public health experts, policymakers, and communities. This collective effort is pivotal in unlocking the full potential of precision interventions to enhance outbreak management and population health.

Strengths and limitations

Our scoping review possesses several limitations, Firstly, we limited our interest to the studies that contained PPH application outcomes in outbreak management. Therefore, studies involving predictive model development and application of artificial intelligence or deep learning framework without any outcome measurements were not included in our scoping review. Secondly, we only included published English language medium articles. There are possibilities of articles with successful and effective implementation of the PPH approach that were not published or in different languages. Furthermore, we focused on publications published in the past six years only to retrieve studies related to current practice, hence missing out on the publications before the timeline. We may also have limited access to more publications due to the search term that we developed. Lastly, the review does not develop and promote practical recommendations as the methodology did not involve methodological evaluation and evidence rating levels. However, this review could help to identify the advantages and gaps in the PPH approaches in outbreak management and the need for future systematic reviews.

Conclusion

In conclusion, PPH applications in outbreak management utilize a wide range of data sources and analytical techniques to enhance disease surveillance, investigation, modeling, and prediction. Through the integration of various types of data, such as genomic data, personal data, Internet query data, environmental data, and global positioning data, authorities can gain valuable insights into disease occurrence, spread, and impact. Modeling and prediction techniques, including metapopulation models, risk mapping algorithms, and machine learning frameworks, enable proactive interventions, targeted resource allocation, and evidence-based decision-making. Additionally, sentiment analysis of social media data provides valuable insights into public perceptions and sentiments, enabling authorities to address concerns, combat misinformation, and foster public engagement. By leveraging these tools and approaches, PPH contributes to more effective and efficient outbreak management, ultimately reducing the burden of infectious diseases on populations. Meanwhile, the limitations and challenges in the application of PPH approaches in outbreak management emphasize the need to strengthen the surveillance systems, promote data sharing and collaboration among relevant stakeholders, and standardize data collection methods while upholding privacy and ethical principles. By enhancing data quality and availability, decision- and policymakers can make informed decisions, implement a targeted intervention, and effectively control and mitigate outbreaks using limited and effective resource allocation strategies.

Acknowledgments: We would like to thank the Director General of Health Malaysia for allowing us to publish this scoping review protocol. We would like to express our gratitude to the Department of Social and Preventive Medicine, University of Malaya for its support and guidance in the development of this protocol.

Contribution list

- Conception of the study—EGR, FMH, RKS, TAM
- Data collection—EGR, RKS
- Data analysis and interpretation—EGR, RKS, TAM
- Drafting the article—EGR
- Critical revision of the article—EGR, FMH, RKS, TAM
- Final approval of the version to be published—FMH, RKS, TAM

Data availability: All data relevant to the present study will be provided.

Declaration of conflicting interests: The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethics and dissemination: This scoping review is part of the research entitled "Measuring Awareness on Precision Public Health among Healthcare Providers", registered with the National Medical Research Register (NMRR), registration ID: NMRR-21-932-59827 (IIR). The findings of the scoping review will be presented at National and International Conferences, Continuous Medical Education, and will be disseminated to relevant stakeholders within the Ministry of Health Malaysia.

Funding: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research is funded by a University of Malaya Impact Oriented Interdisciplinary Research Grant (IIRG007A-19FNW).

Gurantor: FMH

ORCID iD: Ellappa Ghanthan Rajendran https://orcid.org/0000-0003-2666-9645

References

1. Rasmussen SA, Khoury MJ and Del Rio C. Precision public health as a key tool in the COVID-19 response. *JAMA* 2020; 324: 933.
2. Khoury MJ, Iademarco MF and Riley WT. Precision public health for the era of precision medicine. *Am J Prev Med* 2016; 50: 398–401.
3. Haafza LA, Awan MJ, Abid A, et al. Big data COVID-19 systematic literature review: Pandemic crisis. *Electronics (Basel)* 2021; 10: 1–21.
4. Hamid S, Bawany NZ, Sodhro AH, et al. A Systematic review and IoMT based big data framework for COVID-19 prevention and detection. *Electronics (Basel)* 2022; 11: 1–21.
5. Harvey D, Valkenburg W and Amara A. Predicting malaria epidemics in Burkina Faso with machine learning. *Plos One* 2021; 16: 1–16.
6. Hasan A, Putri ERM, Susanto H, et al. Data-driven modeling and forecasting of COVID-19 outbreak for public policy making. *ISA Trans* 2022; 124: 135–143.
7. Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016; 316: 2402.
8. Gunasekeran DV, Tseng RMWW, Tham Y-C, et al. Applications of digital health for public health responses to

COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. *npj Dig Med* 2021; 4: 1–6.
9. Jamshidi MB, Lalbakhsh A, Talla J, et al. Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. *IEEE Access* 2020; 8: 109581–95.
10. Velmovitsky PE, Bevilacqua T, Alencar P, et al. Convergence of precision medicine and public health into precision public health: toward a big data perspective. *Front Public Health* 2021; 9: 561873.
11. Dolley S. Big data's role in precision public health. *Front Public Health* 2018; 6: 68.
12. Aghaali M, Kolifarhood G, Nikbakht R, et al. Estimation of the serial interval and basic reproduction number of COVID-19 in Qom, Iran, and three other countries: a data-driven analysis in the early phase of the outbreak. *Transbound Emerg Dis* 2020; 67: 2860–2868.
13. Alwashmi MF. The use of digital health in the detection and management of COVID-19. *Int J Environ Res Public Health* 2020; 17: 1–7.
14. Awan MJ, Bilal MH, Yasin A, et al. Detection of COVID-19 in chest X-ray images: A Big data enabled deep learning approach. *Int J Environ Res Public Health* 2021; 18: 1–16.
15. Bhardwaj R and Bangia A. Data driven estimation of novel COVID-19 transmission risks through hybrid soft-computing techniques. *Chaos Solitons & Fractals* 2020; 140: 1–16.
16. Buonanno P and Puca M. Using newspaper obituaries to "nowcast" daily mortality: evidence from the Italian COVID-19 hot-spots. *Health Policy* 2021; 125: 535–540.
17. Chakraborty T and Ghosh I. Real-time forecasts and risk assessment of novel coronavirus (COVID-19) cases: a data-driven analysis. *Chaos Solitons & Fractals* 2020; 135: 1–10.
18. Alamoodi AH, Zaidan BB, Zaidan AA, et al. Sentiment analysis and its applications in fighting COVID-19 and infectious diseases: a systematic review. *Expert Syst Appl* 2021; 167: 114155.
19. Chong M and Park HW. COVID-19 in the Twitterverse, from epidemic to pandemic: information-sharing behavior and twitter as an information carrier. *Scientometrics* 2021; 126: 6479–6503.
20. Dai BT, Tan SY, Chen SR, et al. Measuring the impact of COVID-19 on China's population migration with mobile phone data. *Acta Phys Sin* 2021; 70: 068903-1–068903-10.
21. Aleta A, Hu QT, Ye JC, et al. A data-driven assessment of early travel restrictions related to the spreading of the novel COVID-19 within mainland China. *Chaos Solitons & Fractals* 2020; 139: 1–6.
22. Aleta A and Moreno Y. Evaluation of the potential incidence of COVID-19 and effectiveness of containment measures in Spain: a data-driven approach. *BMC Med* 2020; 18: 1–12.
23. Cavallaro C, Bujari A, Foschini L, et al. Measuring the impact of COVID-19 restrictions on mobility: a real case study from Italy. *J Commun Netw* 2021; 23: 340–349.
24. Kee F and Taylor-Robinson D. Scientific challenges for precision public health. *J Epidemiol Community Health* 2020; 74: 311–314.
25. Sahay S. Big data and public health: challenges and opportunities for low and middle income countries. *J Communications of the Association for Information Systems* 2016; 39: 20.
26. Raslan O. Outbreak Management. *IFIC Basic Concepts of Infection Control* 2011: 57: 57–59.
27. Peters MDJ, Godfrey C, McInerney P, et al. Scoping reviews. *Joanna Briggs Institute Reviewer's Manual* 2017; 2015: 1–24.
28. Jung J, Im JH, Ko YJ, et al. Complementing conventional infectious disease surveillance with national health insurance claims data in the Republic of Korea. *Sci Rep* 2019; 9: 1–9.
29. Du M, Sai A and Kong N. A data-driven optimization approach for multi-period resource allocation in cholera outbreak control. *Eur J Oper Res* 2021; 291: 1106–1116.
30. Huang WS, Cao BL, Yang G, et al. Turn to the internet first? Using online medical behavioral data to forecast COVID-19 epidemic trend. *Inf Process Manag* 2021; 58: 1–17.
31. Ali K, Ma'rufi I, Wiranto , et al. Acm. Variability of local weather as early warning for dengue hemorrhagic fever outbreak in Indonesia. Proceedings of 2020 10th international conference on bioscience, biochemistry and bioinformatics (ICBBB 2020) 2020. p. 129–134.
32. Kerdprasop K, Kerdprasop N, Chansilp K, et al. The use of spaceborne and oceanic sensors to model dengue incidence in the outbreak surveillance system. Computational science and its applications – ICCSA 2019, PT I: 19TH international conference, Saint Petersburg, Russia, JULY 1–4, 2019, PROCEEDINGS, PT I2019. p. 447–460.
33. Huang DC and Wang JF. Monitoring hand, foot and mouth disease by combining search engine query data and meteorological factors. *Sci Total Environ* 2018; 612: 1293–1299.
34. Jimenez AJ, Estevez-Reboredo RM, Santed MA, et al. COVID-19 symptom-related Google searches and local COVID-19 incidence in Spain: Correlational study. *J Med Internet Res* 2020; 22: 1–11.
35. Ippoliti C, Candeloro L, Gilbert M, et al. Defining ecological regions in Italy based on a multivariate clustering approach: A first step towards a targeted vector borne disease surveillance. *Plos One* 2019; 14: 1–21.
36. Zhu YF, Oishi W, Maruo C, et al. COVID-19 case prediction via wastewater surveillance in a low-prevalence urban community: a modeling approach. *J Water Health* 2022; 20: 459–470.
37. Razavi-Termeh SV, Sadeghi-Niaraki A, Farhangi F, et al. COVID-19 risk mapping with considering socio-economic criteria using machine learning algorithms. *Int J Environ Res Public Health* 2021; 18: 1–20.
38. Chen Y, Zhang YZ, Xu ZW, et al. Avian influenza A (H7N9) and related internet search query data in China. *Sci Rep* 2019; 9: 1–9.
39. Kammrath Betancor P, Tizek L, Zink A, et al. Estimating the incidence of conjunctivitis by comparing the frequency of google search terms with clinical data: retrospective study. *JMIR Public Health Surveill* 2021; 7: e22645.
40. Ma SM and Yang SH. COVID-19 forecasts using Internet search information in the United States. *Sci Rep* 2022; 12: 1–16.
41. Sylvestre E, Joachim C, Cecilia-Joseph E, et al. Data-driven methods for dengue prediction and surveillance using real-world and Big Data: A systematic review. *PLoS Negl Trop Dis* 2022; 16: 1–22.

42. Zhang YZ, Bambrick H, Mengersen K, et al. Using big data to predict pertussis infections in Jinan city, China: a time series analysis. *Int J Biometeorol* 2020; 64: 95–104.
43. Møller FT, Mølbak K and Ethelberg S. Analysis of consumer food purchase data used for outbreak investigations, a review. Euro surveillance: bulletin Europeen sur les maladies transmissibles. *Eur Commun Disease Bull* 2018; 23: 1–9.
44. Walsh MG, Pattanaik A, Vyas N, et al. A biogeographical description of the wild waterbird species associated with high-risk landscapes of Japanese encephalitis virus in India. *Transbound Emerg Dis* 2022; 69: E3015–E3E23.
45. Cummings MJ, Tokarz R, Bakamutumaho B, et al. Precision surveillance for viral respiratory pathogens: virome capture sequencing for the detection and genomic characterization of severe acute respiratory infection in Uganda. *Clin Infect Dis* 2019; 68: 1118–1125.
46. Nie L, Guo X, Yi CQ, et al. Analyzing the effects of public interventions on reducing public gatherings in China during the COVID-19 epidemic via mobile terminals positioning data. *Math Biosci Eng* 2020; 17: 4875–4890.
47. Minetto R, Segundo MP, Rotich G, et al. Measuring human and economic activity from satellite imagery to support city-scale decision-making during COVID-19 pandemic. *IEEE Trans On Big Data* 2021; 7: 56–68.
48. Yoon H, Lee I, Kang H, et al. Big data-based risk assessment of poultry farms during the 2020/2021 highly pathogenic avian influenza epidemic in Korea. *Plos One* 2022; 17: 1–10.
49. Sanchez-Gendriz I, de Souza GF, de Andrade IGM, et al. Data-driven computational intelligence applied to dengue outbreak forecasting: a case study at the scale of the city of natal, RN-Brazil. *Sci Rep* 2022; 12: 6550.
50. Davis JK, Vincent GP, Hildreth MB, et al. Improving the prediction of arbovirus outbreaks: a comparison of climate-driven models for west Nile virus in an endemic region of the United States. *Acta Trop* 2018; 185: 242–250.
51. Balicer RD, Luengo-Oroz M, Cohen-Stavi C, et al. Using big data for non-communicable disease surveillance. *Lancet Diabetes & Endocrinology* 2018; 6: 595–598.
52. Barrett MA, Humblet O, Hiatt RA, et al. Big data and disease prevention: from quantified self to quantified communities. *Big Data* 2013; 1: 168–175.
53. Bempong NE, De Castaneda RR, Schutte S, et al. Precision Global Health - The case of Ebola: a scoping review. *J Glob Health* 2019; 9: 1–12.
54. Bide P and Padalkar A, IEEE Survey on diabetes Mellitus and incorporation of big data, machine learning and IoT to mitigate it. 2020 6TH International Conference on Advanced Computing And Communication Systems (ICACCS) 2020. p. 1–10.
55. Bilkey GA, Burns BL, Coles EP, et al. Optimizing precision medicine for public health. *Front Public Health* 2019; 7: 42.
56. Li ZC, Gurgel H, Xu L, et al. Improving dengue forecasts by using geospatial big data analysis in google earth engine and the historical dengue information-aided long short term memory modeling. *Biology-Basel* 2022; 11: 1–14.
57. Xiao ZP, Guo WB, Luo ZQ, et al. Examining geographical disparities in the incubation period of the COVID-19 infected cases in Shenzhen and Hefei, China. *Environ Health Prev Med* 2021; 26: 1–10.
58. Cheng T, Horbay B, Nocos R, et al. The role of tailored public health messaging to young adults during COVID-19: "There's a lot of ambiguity around what it means to be safe". *Plos One* 2021; 16: e0258121.
59. Grier S and Bryant CA. Social marketing in public health. *Annu Rev Public Health* 2005; 26: 319–339.
60. Jose Ramon S. Using data sciences in digital marketing: framework, methods, and performance metrics. *J of Innov & Knowl* 2021; 6: 92–102.
61. Antonella M, Cinzia, Federico M, et al. Robotics in clinical and developmental psychology. *Comprehen Clin Psychol* 2022; 10: 121–140.
62. Kim J. Use of robots as a creative approach in healthcare ICT. *Healthc Inform Res* 2018; 24: 155.
63. Marcos-Pablos S and García-Peñalvo FJ. More than surgical tools: a systematic review of robots as didactic tools for the education of professionals in health sciences. *Adv in Health Sci Educ* 2022; 27: 1139–1176.
64. Sarker S, Jamal L, Ahmed SF, et al. Robotics and artificial intelligence in healthcare during COVID-19 pandemic: a systematic review. *Rob Auton Syst* 2021; 146: 103902.
65. Zhao Z, Ma Y, Mushtaq A, et al. Applications of robotics, artificial intelligence, and digital technologies during COVID-19: a review. *Disaster Med Public Health Prep* 2022; 16: 1634–1644.
66. Wei C-W, Kao H-Y, Wu W-H, et al. The influence of robot-assisted learning system on health literacy and learning perception. *Int J Environ Res Public Health* 2021; 18: 11053.
67. Trevelyan JP, Kang S-C and Hamel WR. Robotics in hazardous applications. In: Siciliano B and Khatib O, editors. *Springer handbook of robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2008. pp. 1101–1126.
68. Mougin F, Auber D, Bourqui R, et al. Visualizing omics and clinical data: which challenges for dealing with their variety? *Methods* 2018; 132: 3–18.
69. Savage N. Calculating disease. *Nature* 2017; 550: S115–SS7.
70. Marziano V, Poletti P, Trentini F, et al. Parental vaccination to reduce measles immunity gaps in Italy. *ELIFE* 2019; 8: 1–34.
71. Rich J. Library Guides: Data Resources in the Health Sciences: Clinical Data 2021 [updated 13 December 2021; cited 2022 9 January]. Available from: https://guides.lib.uw.edu/hsl/data/findclin.
72. Ben Yahia N, Kandara MD and BenSaoud NB. Integrating models and fusing data in a deep ensemble learning method for predicting epidemic diseases outbreak. *Big Data Research* 2022; 27: 1–9.
73. Frankenfield J. Social Data Investopedia2020 [updated 28 August 2020; cited 2022 9 January]. Available from: https://www.investopedia.com/terms/s/social-data.asp.
74. Deiner MS, Lietman TM, McLeod SD, et al. Surveillance tools emerging from search engines and social media data for determining eye disease patterns. *JAMA Ophthalmol* 2016; 134: 1024.
75. McGraw D and Mandl KD. Privacy protections to encourage use of health-relevant digital data in a learning health system. *npj Dig Med* 2021; 4: 1–11.
76. Alessa A and Faezipour M. A review of influenza detection and prediction through social networking sites. *Theoret Biol And Med Modell* 2018; 15: 1–27.

77. Reddy CK and Aggarwal CC. *Healthcare data analytics*. New York: CRC Press, 2015.
78. Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019; 6: 1–25.
79. Saltz JS and Dewar N. Data science ethical considerations: a systematic literature review and proposed project framework. *Ethics Inf Technol* 2019; 21: 197–208.
80. Barocas S and Boyd D. Engaging the ethics of data science in practice. *Commun ACM* 2017; 60: 23–25.
81. Valle-Cruz D, Criado JI, Sandoval-Almazán R, et al. Assessing the public policy-cycle framework in the age of artificial intelligence: from agenda-setting to policy evaluation. *Gov Inf Q* 2020; 37: 101509.
82. Hanh TN. *The art of communicating*. New York: Random House, 2013.
83. Nelson A. Unequal treatment: confronting racial and ethnic disparities in health care. *J Natl Med Assoc* 2002; 94: 666–668.
84. Dewalt DA. Health literacy from A to Z: practical ways to communicate your health message. *Prev Chronic Dis* 2005; 2: A28 eCollection 2005 Apr.
85. Meyer E. The culture map: Breaking through the invisible boundaries of global business. *Public Affairs* 2014; 1: 15–31.
86. Thomas DC. Cultural intelligence: People skills for global business: ReadHowYouWant. com, 2008.
87. Negnevitsky M. *Artificial intelligence: a guide to intelligent systems*. Canada: Pearson Education, 2005.
88. Xu M, David JM and Kim SH. The fourth industrial revolution: opportunities and challenges. *Int J of Finan Res* 2018; 9: 90.
89. Aktaş E, Ülengin F and Önsel Şahin Ş. A decision support system to improve the efficiency of resource allocation in healthcare management. *Socioecon Plann Sci* 2007; 41: 130–146.
90. Guindo LA, Wagner M, Baltussen R, et al. From efficacy to equity: literature review of decision criteria for resource allocation and healthcare decision-making. *Cost Eff Resour Alloc* 2012; 10: 9.