Review

Data Analytics to Support Policy Making for Noncommunicable Diseases: Scoping Review

Giorgos Dritsakis[1], PhD; Ioannis Gallos[1], PhD; Maria-Elisavet Psomiadi[2], PhD; Angelos Amditis[1], PhD; Dimitra Dionysiou[1], PhD

[1]Institute of Communication and Computer Systems, National Technical University of Athens, Athens, Greece

[2]Directorate of Operational Preparedness for Public Health Emergencies, Greek Ministry of Health, Athens, Greece

Corresponding Author:
Giorgos Dritsakis, PhD
Institute of Communication and Computer Systems
National Technical University of Athens
9, Iroon Politechniou Str. Zografou
Athens, 15773
Greece
Phone: 30 210772246
Email: giorgos.dritsakis@iccs.gr

Abstract

Background: There is an emerging need for evidence-based approaches harnessing large amounts of health care data and novel technologies (such as artificial intelligence) to optimize public health policy making.

Objective: The aim of this review was to explore the data analytics tools designed specifically for policy making in noncommunicable diseases (NCDs) and their implementation.

Methods: A scoping review was conducted after searching the PubMed and IEEE databases for articles published in the last 10 years.

Results: Nine articles that presented 7 data analytics tools designed to inform policy making for NCDs were reviewed. The tools incorporated descriptive and predictive analytics. Some tools were designed to include recommendations for decision support, but no pilot studies applying prescriptive analytics have been published. The tools were piloted with various conditions, with cancer being the least studied condition. Implementation of the tools included use cases, pilots, or evaluation workshops that involved policy makers. However, our findings demonstrate very limited real-world use of analytics by policy makers, which is in line with previous studies.

Conclusions: Despite the availability of tools designed for different purposes and conditions, data analytics is not widely used to support policy making for NCDs. However, the review demonstrates the value and potential use of data analytics to support policy making. Based on the findings, we make suggestions for researchers developing digital tools to support public health policy making. The findings will also serve as input for the European Union–funded research project ONCODIR developing a policy analytics dashboard for the prevention of colorectal cancer as part of an integrated platform.

Introduction

Noncommunicable Diseases as a Public Health Challenge

Noncommunicable diseases (NCDs), such as cardiovascular or chronic respiratory diseases, cancer, and diabetes, account for 74% of all deaths globally [1]. In the European Union (EU), NCDs are responsible for almost 80% of the disease burden and most premature deaths [2]. They affect quality of life and life expectancy, create numerous challenges for patients and their families, and have a large financial impact, costing EU economies more than 100 billion US$ annually. As many NCDs are age-related, their burdens are increasing partly due to the

prolonged lifespan of the population [3]. NCDs are strongly associated with a number of preventable factors, such as smoking, physical inactivity, harmful alcohol use, and unhealthy diet, and environmental factors, such as air, water, soil pollution, and chemical exposure. Interventions for controlling such risk factors and promoting health and well-being have the potential to reduce the prevalence of NCDs by as much as 70% [4]. To this end, the European Commission has launched an initiative to support effective policies and actions to reduce the burden of major NCDs and improve citizens' health and well-being [2].

The successful management of NCDs requires the integration of the best available scientific evidence into decision-making [5]. Effective use and reporting of data can guide the process and empower policy makers to better understand and act [6]. On the other hand, research findings that directly apply to the policy of interest, when available, are often inconsistent, out of date, or of poor quality. As a result, policy making is traditionally based on social context, political agendas, expert opinion, or the media, all of which are usually biased [7]. Currently, this traditional policy-making approach, which abides more to the notion of clinical health services delivery, is simultaneously challenged from various aspects as it appears unable to meet the novel needs of decision makers [8]. From population surveillance data and indicators to big data and time-tagged trends, policy makers' informed decisions nowadays predispose the effective deployment of technological advancements regardless of whether they concern NCD screening and treatment or managing emerging crises. The recent COVID-19 pandemic signaled the transition into the digital health era, where a newfound model of supporting policy makers' decision processes is needed [9].

Harnessing Digital Technologies for Public Health Policy Making

In the last 2 decades, the growth of health care data in quantity and complexity and the rapid advances in the fields of big data analytics (BDA) and artificial intelligence (AI) present an opportunity to transform conventional policy making into a data-driven process, utilizing various health-related data sources such as electronic health records (EHRs), public health databases, and social networking platforms [10-12]. The integration of routinely collected real-world data and use of advanced analytical techniques for improved decision-making is also referred to as "precision public health" as opposed to traditional public health [13].

Reliable data, combined with data analytics, play key roles at various stages of the public health policy-making process by aiding in understanding, priority setting, resource allocation optimization, identification of the optimal intervention, implementation, and evaluation [14]. This capability allows policy makers to base their decisions on empirical evidence by identifying patterns, trends, and correlations that might not be evident otherwise. Such informed decision-making is vital in areas like health care and can lead to more effective outcomes [10]. Specifically, data analytics may improve the delivery of public services, enabling governments to anticipate public health crises, optimize public health initiatives, increase adherence from the public, and enhance the overall responsiveness of

services [15]. This iterative process of policy evaluation and adjustment, supported by real-time data monitoring, ensures that strategies are continuously refined to achieve intended goals. Analytics can identify high-impact areas for efficient resource allocation, such as targeting aggressive policies where they are most needed (eg, the outbreak of a pandemic).

Data analytics approaches are commonly categorized into 3 broad types: descriptive, predictive, and prescriptive analytics. The 3 types answer different questions but use similar methodologies that can be applied individually or in combination, depending on the health policy objectives, data availability, and decision-making context. Descriptive analytics summarize past and present trends and patterns in the data to answer the question "What happened?" Predictive analytics use primarily historical data to create models that answer the question "What will happen?" Prescriptive analytics employ data-driven models and optimization algorithms to recommend the most effective actions, interventions, or allocation strategies trying to answer the question "What should I do?"

The use of data analytics in the context of policy making has also been referred to as "policy analytics" by some authors [16]. It has been suggested that policy analytics include data-driven tools that respond to a policy need and use a transparent development process [17]. Examples of policy analytics techniques are statistics, simulations, data mining, machine learning (ML), social networks, and geographic info [18].

Rationale and Aims

Currently, there is no comprehensive overview of the use of data analytics tools to support policy making for NCDs. Such an overview could guide stakeholders and organizations involved in policy making, engineers and companies managing such tools, data analysts, and researchers, and could highlight gaps and unmet needs in the area. The present review has been performed in the context of the Horizon Europe project ONCODIR [19]. The project seeks to identify risk factors associated with colorectal cancer (CRC) and will integrate multidisciplinary research methods and technologies to deliver evidence-based and personalized recommendations on CRC. The findings of this review will serve as an evidence base for the development of the policy analytics component of ONCODIR's integrated platform.

The primary aim of this scoping review is to investigate the landscape of data analytics tools and platforms designed to support evidence-based policy making for NCDs. The secondary aim is to explore the adoption of these tools by policy makers and the factors affecting this.

Methods

Design

A scoping review was selected given the aim of this study. It has been suggested that scoping reviews are more suitable than traditional systematic reviews and meta-analyses when focusing on rapidly evolving topics, as they allow more timely synthesis of evidence. The inclusive and broad nature of scoping reviews also makes them better at looking into the range of evidence on a particular topic [20]. We conducted our review according to

XSL•FO

RenderX

the Joanna Briggs Institute guidelines to ensure quality and reliability [21]. We followed the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) guidelines (Multimedia Appendix 1) and developed a search strategy based on the Population, Concept, and Context framework as follows [22]:

- Population: humans with NCDs
- Concept: digital data analytics tools
- Context: public health policy making

Based on this framework, in December 2023, we searched the PubMed database for publications written in English from 2013 to date. The search was repeated in July 2024. We searched for articles published in the last decade to explore the latest advances in the field. We searched for original research and review articles published in peer-reviewed journals and no other sources to ensure the highest quality of the studies.

A query was first developed in PubMed and was then adapted for IEEE Xplore: (("health"[Title/Abstract] OR "healthcare"[Title/Abstract] OR "clinical"[Title/Abstract] OR "medical"[Title/Abstract]) AND ("policy"[Title/Abstract] OR "policies"[Title/Abstract] OR "policymaking"[Title/Abstract] OR "policy-making"[Title/Abstract] OR "policy making"[Title/Abstract] OR "decision making"[Title/Abstract] OR "decision-making"[Title/Abstract] OR "decision maker"[Title/Abstract]) AND ("analytics"[Title/Abstract] OR "analytic"[Title/Abstract]) AND ("data"[Title/Abstract] OR "evidence"[Title/Abstract] OR "evidence-based"[Title/Abstract])).

Given that the concept of policy analytics and the associated terminology are not yet well defined, we chose broad terms in the first place to widen the results and assessed eligibility later case by case. For instance, we did not use a term for NCD as it appeared that usually the names of the individual conditions are used. Moreover, we preferred "health" as "public health" is not consistently used in the context of policy making for NCDs.

Study Selection

Title and abstract screening and full-text review were performed according to predefined eligibility criteria (Textbox 1). A snowballing approach was also used to identify any additional articles from the reference lists of the screened articles. We included studies describing data analytics tools or their use in detail. These had to be concrete tools or applications designed specifically with the aim to support policy makers as opposed to models, algorithms, or other theoretical frameworks, such as decision-analytic models or cost-effectiveness analyses. Analytics tools had to be designed for or applied to a specific NCD as these have been defined according to the World Health Organization (WHO) [23]. We excluded analytics not performed on specific conditions or not applied to health data (eg, health care management). We also excluded studies on views or perceptions about analytics.

From all included articles, we extracted data related to the disease studied, purpose of the study and tool, analytics used, implementation, and link with policy making. We performed a narrative, descriptive data synthesis of the techniques used and their implementation.

Textbox 1. Eligibility criteria for article selection.

Inclusion criteria

- Original research or review.
- Study on humans with noncommunicable diseases (NCDs).
- The study includes a concrete digital data analytics tool that can be used by a policy maker.
- The study includes a tool that is designed to support health policy making.
- The study includes analytics performed on health-related data.

Exclusion criteria

- Commentaries, viewpoints, protocols, perspectives, or other study types.
- Study on infectious diseases or no disease (eg, health management).
- The study describes theoretical models, frameworks, or statistical algorithms.
- The study includes tools for clinical decision support, patient benefit, or other purposes.
- The study includes analytics performed on the literature.

Results

Study Characteristics

The search yielded 1850 articles from the PubMed and IEEE Xplore databases (Figure 1). After article duplicates were removed, 1836 remained for title and abstract screening. A total of 59 articles were identified for full-text review. Five studies met our eligibility criteria and were included in the review. Two more studies were identified via snowballing of the reference

lists of screened articles. Furthermore, 2 articles were added that included additional applications of the identified tools, as one of the aims of the review was to explore implementation and uptake. In total, 9 articles were included in this review. Articles were published between 2017 and 2022. The articles described 7 data analytics tools or integrated platforms including a data analytics component. The studies described the tools in detail and included use case scenarios, pilot studies, workshops, and examples of implementation that illustrate how analytics

can be used to support policy making for NCDs. Four studies were conducted in the United States, 1 in Australia, and 1 in Canada. Three studies were conducted in various countries of Europe (ie, Greece, Germany, Slovenia, Spain, United Kingdom, Sweden, Finland, Northern Ireland, Republic of Ireland, and Denmark).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) flow diagram of the stages of the scoping review.
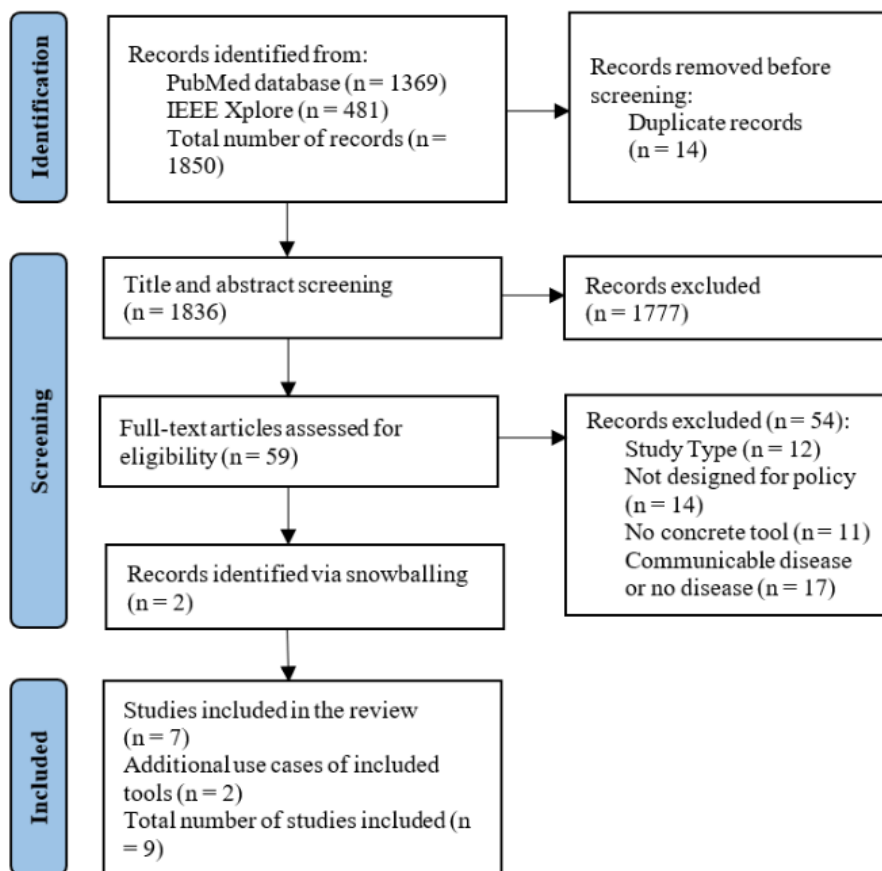
Tools Overview

The 7 tools reviewed are summarized in Table 1. Three integrated platforms (EVOTION, MIDAS, and CrowdHEALTH) were designed to support public health policy decisions for a range of conditions and include a data analytics component supporting both descriptive and predictive analytics [24-28]. Users can create policy models, define the way in which data should be analyzed in order to produce the evidence useful for public health policy making, and obtain analytical results of how this evidence may support or contradict various policy actions. The other 4 tools (PoPHQ, PoPHR, Social InfoButtons, and RiskScape) were designed for disease monitoring through the integration, descriptive analysis, and visualization of population health data from various sources [29-32].

Two of the 7 tools were designed to address specific NCDs. EVOTION was specifically designed for hearing loss and PoPHQ was designed for obesity prevention. The other platforms were applicable to and piloted with a wide range of conditions, including obesity in adults and children, cardiovascular conditions, chronic kidney disease, chronic respiratory conditions, cancer, mental health, asthma, and diabetes. Three platforms were piloted in obesity, physical activity, and nutrition. Two platforms were piloted with each of the following conditions: respiratory, cardiovascular, and diabetes. CrowdHEALTH was piloted with cancer [25]. Six of the 7 tools were designed for or used with NCDs only, while PoPHR was applicable to infectious diseases as well [30].

Table 1. Overview of the tools included in this review.

Tool	Type	Analytics supported	Implementation reported	Study
EVOTION	Public health policy decision support	Descriptive and predictive	Analytics applied, use case example, and workflow	Prasinos et al [27] and Saunders et al [28]
CrowdHEALTH	Public health policy decision support	Descriptive and predictive	Use case example, workflow, and use cases including health care stakeholders	Moutselos et al [26] and Mavrogiorgou et al [25]
MIDAS	Public health policy decision support	Descriptive and predictive	Pilots with policy makers	Shi et al [24]
PoPHQ	Health informatics tool	Descriptive	Design workshops where stakeholders identified user stories and use cases	Canfell et al [29]
RiskScape	Public health surveillance	Descriptive	Real-life implementation	Canfell et al [29] and Cocoros et al [32]
Social InfoButtons	Public health analytics	Descriptive	Use case scenarios	Ji et al [31]
PoPHR	Visualization of population health data	Descriptive	Description and functionalities with examples, and initial implementation ongoing	Shaban-Nejad et al [30]

Study Settings

Details of the use cases are provided in Table 2. The studies presented the 7 tools using different settings, including descriptions of the functionalities with examples, use case scenarios where authors presented a workflow, pilot studies with actual data from various sources, applications of analytics on data collected using the tool, design workshops with stakeholders, and real-life implementation.

CrowdHEALTH was piloted in 5 different countries and 6 use case scenarios with stakeholders from various health care organizations [25]. Relevant data from patients and healthy adults and children were used to address policy needs on various conditions. For example, the Slovenian pilot used data on physical activity to analyze the physical fitness and weight status of children, assess its development over time, and predict future levels. This provided a basis for the implementation of policies that link school and health data for early intervention monitoring and evaluation. Interestingly, this use case was based on a real-life policy described in the Slovenian National Program on Nutrition and Physical Activity for Health 2015-2025 [26]. MIDAS was piloted in 4 countries using social media, MEDLINE analytics, and news media data on a range of conditions [24]. Ji et al [31] used data on treatments and symptoms posted by patients on social media to evaluate the effectiveness of the Social InfoButtons platform. PoPHR was piloted with a randomly sampled, open cohort of 25% of the Montreal population in Canada [30]. PoPHQ is currently in the mock-up phase but has been designed to integrate anonymized electronic medical record data from Queensland, with a final total sample size of approximately 5 million [29]. No pilot data were reported for Risk Scape, but the Massachusetts Department of Public Health (MDPH) is currently using the platform to monitor conditions of interest using EHR data that are updated monthly from clinical practice groups that cover approximately 20% of the state population [32]. Prasinos et al [27] described the workflow of policy creation, selection, and execution of analytics using the EVOTION platform. Saunders et al [28] reported 3 applications of BDA techniques using a dataset synthesized from the EVOTION data repository.

Table 2. Details of the analytics included in this review.

Study and country	Noncommunicable disease	Analytics	Purpose	Data
Prasinos et al [27]				
Greece, United Kingdom, and Denmark	Hearing loss (HL)	• Basic statistics • Linear regression • Principal component analysis • Inferential statistics	• Investigate the impact of hearing interventions on quality of life (QOL)	Synthetically generated dataset from the EVOTION data repository
Saunders et al [28]				
Greece, United Kingdom, and Denmark	HL	• Predictive modeling • Regression analysis • Generalized linear mixed model • Correlations	• Estimate the risk of noise-induced HL • Predict hearing aid (HA) usage from changes to the sound environment • Examine the association between physical activity and HA usage	Synthetically generated dataset from the EVOTION data repository
Moutselos et al [26]				
Slovenia	Obesity	• Forecasting • Simulation • Causal analysis	• Effectiveness of various interventions for obesity prevention • Early detection of children with increased risk linked to poor physical fitness	Large-scale data from the national surveillance system on the physical and motor development of children
Mavrogiorgou et al [25]				
Spain	Obesity in adults	• Clinical pathway mining • Risk stratification	• Identification of overweight patients	Demographic information, hospitalization, emergency and hospital episodes, and morbidity
Sweden	Cardiovascular and chronic kidney disease	• Clinical pathway mining • Risk stratification • Causal analysis	• Monitor patients	Demographic, drug usage, and practitioners' consultation data
Slovenia	Fitness and obesity in childhood	• Clinical pathway mining • Risk stratification • Causal analysis	• Analyze physical fitness and weight status • Predict future levels of fitness and somatic development	Physical activity, sedentariness, sleep, heart rate, socioeconomic status, and parental physical activity
Greece	Chronic respiratory conditions	• Risk stratification	• Monitor disease progression and health care expenditure for improved chronic disease management of patients	Biosignals from pulse oximeters, blood pressure meters, glucometers, spirometers, weighing scales, and activity trackers
Germany	Nutrition and physical activities	• Clustering analysis • Correlation analysis	• Understand the influence of nutritional habits and physical activity on overall health and QOL	Physical and activity data provided by activity trackers
Greece	Cancer care	• Causal analysis	• Evaluate the impact of online coaching and medical education	Diagnosis, treatment, comorbidities, health behaviors, and side effects
Shi et al [24]				
Spain	Obesity	• Random forests/least absolute shrinkage and selection operator	• Identify the risk factors of childhood obesity	Controlled and open data including social media analysis, MEDLINE analytics, and news media analysis
Finland	Mental health	• Lexis diagram analysis • Descriptive analysis	• Aggregate, summarize, and visualize risk factors • Evaluate health, social, and educational status	Controlled and open data including social media analysis, MEDLINE analytics, and news media analysis

Study and country	Noncommunicable disease	Analytics	Purpose	Data
Northern Ireland	Social care for children	• Markov chain • Long short-term memory (LSTM) network	• Track patterns of behavior over time • Estimate the probability of transition between different types of care • Improve the protection of children	Controlled and open data including social media analysis, MEDLINE analytics, and news media analysis
Republic of Ireland	Diabetes	• Autoregressive integrated moving average (ARIMA) model	• Forecast the consumption of diabetic drugs	Controlled and open data including social media analysis, MEDLINE analytics, and news media analysis
Canfell et al [29]				
Australia	Obesity	• Comparison across age groups • Counts by facility • Counts and percentages • Stratification by suburb and facility	• Target interventions across the life course • Direct resources • Justify the problem • Compare obesity across regions	"Mock-up" without patient data[a]
Canfell et al [29] and Cocoros et al [32]				
United States	Diabetes, hypertension, asthma, and obesity	• Heat maps by zip code • Stratification by demographics and comorbidities • Time series analyses with trend statistics • Data aggregation and visualization	• Review, analyze, map, and trend aggregate data • Prevalence of selected conditions • Identify health disparities	Electronic health records (EHRs) from 3 clinical practice groups that cover approximately 20% of the state population (>1.2 million)
Ji et al [31]				
United States	Posttraumatic stress disorder and asthma	• Statistical • Geospatial • Temporal • Topic investigation • Association discovery • Recommendation discovery • Visualization	• Compute statistical aggregates • Explore data according to a geographic feature • Analyze trends over time • Explore correlations between treatments, side effects, symptoms, and conditions • Discover treatment recommendations • Integrate openly available health data	Openly available health data sources including SMN, Twitter, MedHelp, WebMD, CDC, and PubMed
Shaban-Nejad et al [30]				
Canada	Chronic diseases (eg, diabetes, hypertension, coronary heart disease, and stroke)	• Visualization • Stratification • Filtering • Statistical algorithms to detect changes in an indicator over time and space	• Explore and visualize available indicators • Create coherent portraits of population health and health system performance • Evaluate the effects of public health interventions	No actual data used in this study; initial implementation ongoing[b]

[a]Final total sample size is estimated to be approximately 5 million (anonymized electronic health record data from Queensland).

[b]Initial implementation with a randomly sampled, open cohort of 25% of the population of the Census Metropolitan Area of Montreal, Quebec; in the process of implementation in the entire population of the province of Quebec.

Data Analytics Applied

Table 2 presents the data analytics applied in detail, including the setting, purpose, data used, NCD studied, and country where the study took place. The choice of analytics is strongly linked to the policy need being addressed and the nature of the tool.

A wide range of analytical techniques were used, which can be summarized as follows.

Descriptive Analytics

All tools employed descriptive analytics. The first level was data ingestion, integration, cleaning, and preprocessing, such

as removal of duplicates and errors, imputation of missing data, handling of outliers, and standardization of data formats [24,25,30,31]. The next step was data exploration using basic descriptive statistics and inferential statistics, such as the identification of risk factors for a specific condition. Static or interactive visualizations were used, including scatter plots, heat maps, bar or box plots, and pie charts. Three of the tools employed more specialized techniques, including geospatial analytics or mapping (ie, exploration of data in a geographical area), temporal analytics (ie, tracking trends over time), and comparative analytics, which identifies differences between groups of measurements, such as disease prevalence across different age groups [29,30,32]. Other types of analyses used were clustering analysis (ie, grouping of objects based on measures of similarity) and correlation analysis to identify the strength of the linear association between variables. All tools included a visualization dashboard or user interface, although there were differences in the type or level of user interaction.

Predictive Analytics

Three out of the 7 tools employed predictive analytics. Methods included regression analysis and statistical modeling. Contemporary frameworks were also used for prediction and forecasting, including ML, deep learning, and simulation modeling [24,28]. Other types of analyses that were used for predictive purposes included risk stratification analysis; causal analysis, which models the behavior of the target variable of interest; and clinical pathway analysis, which models the process followed during treatment of a patient with respect to a particular condition [25,26].

Prescriptive Analytics

No study implemented a concrete prescriptive methodology. Three out of the 7 tools were decision support systems designed to make policy recommendations. The authors referred to the prescriptive capabilities of the tools and demonstrated the policy creation process [24,27,29]. However, none of them presented a related use case actually applying this type of analytics.

Tool Implementation

Out of the 7 platforms reviewed, 1 has been fully implemented in real life. RiskScape is used by MDPH to monitor conditions of interest using EHR data updated monthly from 3 clinical practice groups that cover approximately 20% of the state population [32]. It has a key role in demonstrating the need and burden for MDPH's applications for funding through the identification of inequitably burdened populations. The authors suggest that the platform unloads analytical burden from health departments, centralizes information in an efficient electronic environment, and offers clinical practices a holistic understanding of disease patterns and management practices. RiskScape is an open-source software and is publicly available [33]. Shaban-Nejad et al [30] reported that PoPHR had been initially implemented and was in the process of being fully implemented in Montreal, Quebec. According to the authors, the platform can be used by policy makers to improve decisions related to the planning, implementation, and evaluation of population health and health system interventions.

For the remaining 5 tools, the reviewed studies included analytics examples, use cases, or pilots. MIDAS was evaluated by policy makers in the pilot studies and successfully achieved all key progress indicators [24]. The platform received positive feedback on its capacity to integrate and analyze data. The pilots demonstrated how custom-tailored analytics produced knowledge and results that are actionable by public health policy makers and gave them insights for possible future interventions. Based on these results, the authors concluded that MIDAS is transferable, sustainable, and scalable across policies, data, and regions. Mavrogiorgou et al [25] reported how each use case of the CrowdHEALTH platform provided insights that can be used in policy making. Stakeholders from various sectors attended a workshop and provided feedback on the purpose, interface, and overall design of PoPHQ [29]. They identified various uses of the platform to create stories for 4 different end users: public health practitioners, systems planners, researchers, and generic users. PoPHQ is planned to be implemented in Queensland with a population of 5 million [29]. Saunders et al [28] reported that a policy maker could use EVOTION analytics as evidence to expand guidelines aimed at preventing noise-induced hearing loss or simulate hearing aid uptake and usage if urban planning organizations were to project an increase in everyday acoustic noise due to changed requirements for official noise prevention initiatives. Finally, Social InfoButtons can be used by governments for disease surveillance [31].

Discussion

Data Analytics for Public Health Policy Making

This scoping review was conducted to explore the data analytics tools designed for public health policy making for NCDs and their implementation. The review was motivated by the emerging need for approaches harnessing BDA, AI, and other novel technologies to improve public health policy making. It was also motivated by the EU-funded research project ONCODIR developing a policy analytics dashboard for the prevention of CRC as part of an integrated platform.

We presented 2 different types of tools enabling data analytics for policy making for NCDs: (1) tools designed for public health monitoring and surveillance that aggregate openly available data or data from electronic medical records and have mainly descriptive analytics and visualization functionalities and (2) integrated platforms designed for policy decision support with both descriptive and predictive analytics functionalities. Previously, Canfell et al [13] reviewed the use of real-world data for precision public health in NCDs and identified surveillance platforms integrating descriptive, comparative, and geospatial analytics. Our review, with its different scope, extends these findings and further demonstrates that predictive analytics can be used for the management of NCDs to inform policy decisions. A variety of ML techniques were used in the studies included in this review for forecasting. Moreover, classic statistical methods, such as logistic regression analysis, were adopted. ML techniques are increasingly used with large population health datasets to improve public health surveillance, disease prediction, and delivery of interventions [34]. On the other hand, prescriptive analytics provide actionable

recommendations to policy makers and can have a key role in precision public health. It must be noted here that even though 3 of the platforms included in this review were designed as policy decision support tools, no prescriptive analytics were actually applied in the pilot studies or use cases that generated policy recommendations. Instead, predictive models, risk estimation, and forecasting provided insights that aimed to support policy makers in making decisions. We could claim that the 3 tools reviewed, which employ predictive analytics, are more advanced toward supporting policy decisions compared to tools with descriptive analytics capabilities only. However, the actual usefulness of a tool for policy makers depends on the specific policy need and scenario.

NCDs Studied

Among the 4 major types of NCDs, cancer appears to be the one that is least studied in the context of data analytics for policy making as only the CrowdHEALTH platform was piloted with cancer. CrowdHEALTH used data from a web platform related to patients' diagnosis, treatment, comorbidities, health behaviors, and side effects to assess the impact of online coaching and medical education and predict future behaviors of cancer patients [25]. The type of cancer supported by the platform was not specified. According to the authors, given the absence of specific policies for the provision of medical information and online coaching and the increased patient support needs, such an approach may be useful for the improvement of resource allocation in the health care system among others. Other studies have explored protocols for mapping breast cancer registry data [35], the use of modeling to optimize cancer screening and predict catchment areas, and the use of AI for risk stratification of cancer patients [36-38]. However, none of these studies included tools designed to be used by policy makers. To the best of our knowledge, no policy-making platform with an analytics component has been designed for or used with cancer.

To address this gap, the EU-funded ONCODIR research project aims to develop an intelligent policy analytics dashboard as part of an integrated platform to support the primary prevention of CRC. The dashboard will incorporate retrospective data on CRC incidence, risk factors, and other relevant data as well as prospective data from a mobile app, and will enable descriptive and predictive analytics to provide insights to inform CRC prevention policies.

Use of Tools by Policy Makers

Our findings show very limited real-world use of analytics by policy makers. This is in line with previous studies showing limited implementation of digital tools for NCDs [13]. Only RiskScape is fully implemented and is also publicly available to use. Shaban-Nejad et al [30] reported that PoPHR had been initially implemented and was in the process of being fully implemented in Montreal, Quebec, but no more reports have been published since then [30]. In a subsequent study from 2020 not included in this review, it was reported that PoPHR was in the process of being deployed in Quebec for routine use by public health professionals [39]. In the same study, the authors reported their plans to extend the use of PoPHR to recommend interventions that are likely to be the most effective.

For most of the other tools' use cases, pilots or evaluation workshops were reported that involved policy makers. For instance, Saunders et al [28] reported some applications of analytics using the EVOTION platform with implications for policy makers [28]. In another study not included in this review, Dritsakis et al [40] reported a series of workshops where EVOTION was demonstrated to stakeholders in 4 countries and evaluated using a Strengths, Weaknesses, Opportunities, and Threats methodology [40]. The study highlighted the potential of the tool together with obstacles and risks that need to be addressed, such as the complicated mechanism of data collection and analysis and the lack of major analytic capabilities required for public health policy decision-making (eg, economic evaluation).

Overall, the 7 tools included in this review have been mostly designed as research prototypes in academic settings. Policy makers were involved in the development or use in some way, and all studies highlight the potential use of analytics to support policy making. However, there is very limited uptake or plans for use by policy makers reported in the reviewed studies. This has implications for the usefulness of the tools as most of them have not been tested in real-life settings. RiskScape (fully implemented) and PoPHR (initially implemented) are currently the most mature data analytics tools that can aid policy making according to the findings of this review. The fact that most tools were developed in the last 4 years could explain the poor uptake to some extent as the tools may not be implemented yet or the implementation may not have been published. Finally, it must be noted that besides RiskScape, it is unclear if the rest of the platforms are accessible and available for use.

Adoption of Digital Health Technologies

Many studies have explored factors influencing the adoption of digital health technologies by policy makers. Innovative solutions are incorporated into health policy functions at a slower pace than health transformation into a digital asset [41]. Lack of advanced infrastructure, low interoperability levels among critical actors, and bureaucracy pose barriers to the acceptance of new technologies that are frequently exacerbated by safety concerns. Other challenges are organizational fragmentation creating siloed data systems, difficulty in data sharing due to privacy and security issues, and concerns around data quality [42]. Moreover, systemic and organizational issues exist owing to the core characteristics of public health authorities as administrative bodies that lack regulatory frameworks and a data governance culture as a whole [43].

A very important obstacle from the perspective of end users is often the poor IT literacy and lack of digital skills by policy makers and public health professionals [44]. Another aspect is whether these tools actually meet the needs of end users. For example, reviews on the use of visual analytics in mental health care planning have shown that despite the availability of advanced visualization tools, such as geographical maps, the majority of experts use simple, familiar, and readily available visualizations, and a very small percentage of digital tools are actually used for policy and planning [45,46]. Despite a clear need to extract information from highly complex data, such barriers and concerns hinder the use of analytics as part of

decision-making and lead policy makers to use approaches that are most familiar to them and widely understood, and keep relying on expert opinion and intuition.

Poor uptake of evidence-based tools is also related to challenges inherent in the policy-making process. Moving evidence into practice will always require political engagement and therefore will be influenced by political agendas [28]. Moreover, there will always be urgent problems and limited funds that will require policy makers to use economics, statistics, and scientific skills to rapidly interpret evidence and provide solutions as the recent COVID-19 pandemic showed. The successful use of data-driven tools for responsive and accurate public health decisions for NCDs requires the optimization and reorganization of the public health sector and workflows [13]. Some priorities that have been reported include investment in modern data management infrastructure, development of strategic partnerships, need for AI transparency and reproducibility, and explicit consideration of equity and bias [47]. To effectively use new technologies and the large amount of health care data that are now becoming available to guide policy decisions, it is necessary to overcome the computational, algorithmic, and technological obstacles of an extremely heterogeneous data landscape, as well as a variety of legal, normative, governance, and policy limitations. Moreover, the transition into the digital health era requires a digital health background and key IT skills to ensure that policy makers have the capacity to make the most out of digital health innovations. A paradigm shift is required for policy makers at local, regional, and national levels to overcome cultural barriers that influence their acceptance and implementation of novel technologies [48].

Recommendations

Moving forward, we recommend that researchers developing advanced data analytics tools for policy making should engage with end users throughout the process, respond to their needs, and present results in a way that is easy to interpret. They should specify the type and purpose of analytics, how policy makers can use it, and what they can achieve with it. They should also secure the required resources to make the tool available and sustainable after it has been launched. Finally, developers must promote the implementation of digital health tools for decision-making as this could positively influence the IT literacy of policy makers. On the other hand, public health professionals should be encouraged by the potential of data analytics tools to inform evidence-based decisions and empower their digital health skills to be able to use the developed IT tools. Overall acceptance of novel technologies in domains and processes where current uptake is limited will facilitate the required paradigm shift.

To unlock the full potential of data analytics, authorities need to prioritize robust data infrastructure, interoperability, data integration, and the initiation of high-quality nationwide data registries (eg, cancer registries) to monitor public health. Training programs for data governance capacity building, hiring of specialists to key positions, and international collaborative efforts on major public health threats could help transition to more informed and responsive public health policy making. Finally, a strong legislative framework on data sharing that

focuses on maximizing utility while also preserving privacy could be a major accelerating factor in fostering a data-driven culture. Establishing trust between government authorities and individuals would eventually create a conducive environment that promotes data sharing for the benefit of the public health system as a whole.

Methodology

The scope of the review and the corresponding search strategy were intentionally broad to allow us to explore a relatively new concept that is not yet well defined. The term policy analytics, although present in the literature, was rarely used in the studies reviewed here. Instead, authors used separate terms to refer to (1) the analytics and tools designed (eg, data integration, aggregation, visualization, or analysis; BDA; platform; data-driven; evidence-based; and system) and (2) the use or purpose of the analytics (eg, policy [decision] making, [precision] public health, surveillance, monitoring, and population health). There are differences in the way the applications of tools are reported in the literature. For example, the term "use case" may refer to the actual use of the tool in a particular setting or scenario and examples of how the tool can be used. Various terms are also used to refer to conditions (eg, NCDs, public health, or the name of the individual condition) and end users (eg, policy makers, decision makers, and public health professionals). This highlights the difficulty to comprehensively review all such tools and their applications.

A quality assessment of the included studies was not performed in this review. Quality assessment is an optional step for scoping reviews. Even though it would provide a more robust evaluation of the evidence and help identify any potential biases, it was not considered crucial given the aims of the review (we were not interested in the actual data collected or the findings). However, the different types, aims, and scopes of the included studies have been clearly presented in the Study Characteristics subsection in the Results section.

Finally, we only searched 2 databases compared to other scoping reviews that included more sources of information. However, we selected 2 reliable databases that together cover both the medical and technological fields of literature and are therefore sufficient to explore the concept of data analytics tools and public health policies. The search strategy as a whole ensures the quality of the studies reviewed.

Conclusions

There are many digital tools incorporating descriptive and predictive analytics that can support policy making in different ways for a range of NCDs. However, the majority of these tools are not widely used by policy makers in real-life. We have discussed factors affecting the adoption of digital health technologies as a whole and have also made recommendations on how stakeholders involved in the development and use of data analytics tools for public health policy making can help increase adoption and maximize the impact of such tools in supporting the policy-making process. The effectiveness and actual usefulness of the tools reviewed in this study should be assessed on the basis of the specific policy needs, settings, populations, conditions, and data with which they were tested.

Acknowledgments

Authors' Contributions

GD made substantial contributions to the conception and design of the work throughout the manuscript preparation. IG contributed to the parts of the manuscript specific to data analytics. MEP contributed to the parts of the protocol specific to public health policy making. AA reviewed and approved the final manuscript. DD made substantial contributions to the conception and design of the work and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist.
[PDF File (Adobe PDF File), 84 KB-Multimedia Appendix 1]

References

1. Noncommunicable diseases. WHO. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2024-09-30]
2. Healthier together – EU non-communicable diseases initiative. EU Commission. URL: https://health.ec.europa.eu/non-communicable-diseases/healthier-together-eu-non-communicable-diseases-initiative_en [accessed 2024-09-30]
3. Kehler DS. Age-related disease burden as a measure of population ageing. The Lancet Public Health. Mar 2019;4(3):e123-e124. [doi: 10.1016/s2468-2667(19)30026-x]
4. REPORT on non-communicable diseases (NCDs). European Parliament. URL: https://www.europarl.europa.eu/doceo/document/A-9-2023-0366_EN.html [accessed 2024-09-30]
5. Implementation roadmap 2023–2030 for the Global action plan for the prevention and control of NCDs 2013–2030. WHO. URL: https://www.who.int/teams/noncommunicable-diseases/governance/roadmap [accessed 2024-09-30]
6. Data, analytics and delivery for impact in focus 2022: priorities for regions and countries. WHO. URL: https://iris.who.int/handle/10665/360468 [accessed 2024-09-30]
7. van de Goor I, Hämäläinen R, Syed A, Juel Lau C, Sandu P, Spitters H, et al. REPOPA consortium. Determinants of evidence use in public health policy making: Results from a study across six EU countries. Health Policy. Mar 2017;121(3):273-281. [FREE Full text] [doi: 10.1016/j.healthpol.2017.01.003] [Medline: 28139253]
8. Kostkova P. Grand challenges in digital health. Front Public Health. 2015;3:134. [FREE Full text] [doi: 10.3389/fpubh.2015.00134] [Medline: 26000272]
9. Kostkova P, Saigí-Rubió F, Eguia H, Borbolla D, Verschuuren M, Hamilton C, et al. Data and digital solutions to support surveillance strategies in the context of the COVID-19 pandemic. Front Digit Health. 2021;3:707902. [FREE Full text] [doi: 10.3389/fdgth.2021.707902] [Medline: 34713179]
10. Chao K, Sarker MNI, Ali I, Firdaus RR, Azman A, Shaed MM. Big data-driven public health policy making: Potential for the healthcare industry. Heliyon. Sep 2023;9(9):e19681. [FREE Full text] [doi: 10.1016/j.heliyon.2023.e19681] [Medline: 37809720]
11. Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. BMC Med Inform Decis Mak. Apr 10, 2021;21(1):125. [FREE Full text] [doi: 10.1186/s12911-021-01488-9] [Medline: 33836752]
12. Gomes MAS, Kovaleski JL, Pagani RN, da Silva VL, Pasquini TCDS. Transforming healthcare with big data analytics: technologies, techniques and prospects. J Med Eng Technol. Jan 19, 2023;47(1):1-11. [doi: 10.1080/03091902.2022.2096133] [Medline: 35852400]
13. Canfell OJ, Kodiyattu Z, Eakin E, Burton-Jones A, Wong I, Macaulay C, et al. Real-world data for precision public health of noncommunicable diseases: a scoping review. BMC Public Health. Nov 24, 2022;22(1):2166. [FREE Full text] [doi: 10.1186/s12889-022-14452-7] [Medline: 36434553]
14. Global Evidence Commission update 2024: Building momentum in strengthening domestic evidence-support systems, enhancing the global evidence architecture, and putting evidence at the centre of everyday life. McMaster Health Forum. URL: https://www.mcmasterforum.org/docs/default-source/evidence-commission/update-2024.pdf [accessed 2024-09-30]
15. Brownson RC, Chriqui JF, Stamatakis KA. Understanding evidence-based public health policy. Am J Public Health. Sep 2009;99(9):1576-1583. [doi: 10.2105/ajph.2008.156224]

16. Tsoukias A, Montibeller G, Lucertini G, Belton V. Policy analytics: an agenda for research and practice. EURO Journal on Decision Processes. Jun 2013;1(1-2):115-134. [doi: 10.1007/s40070-013-0008-3]

17. Meinard Y, Barreteau O, Boschet C, Daniell KA, Ferrand N, Girard S, et al. What is policy analytics? An exploration of 5 years of environmental management applications. Environ Manage. May 20, 2021;67(5):886-900. [FREE Full text] [doi: 10.1007/s00267-020-01408-z] [Medline: 33474617]

18. Gil-Garcia J, Pardo T, Luna-Reyes L. Policy Analytics: Definitions, Components, Methods, and Illustrative Examples. In: Gil-Garcia J, Pardo T, Luna-Reyes L, editors. Policy Analytics, Modelling, and Informatics. Public Administration and Information Technology, vol 25. Cham. Springer; 2018:1-16.

19. ONCODIR project. URL: https://www.oncodir.eu/ [accessed 2024-09-30]

20. Lewinter K, Hudson S, Kysh L, Lara M, Betz C, Espinoza J. Reconsidering reviews: the role of scoping reviews in digital medicine and pediatrics. NPJ Digit Med. Dec 10, 2020;3(1):158. [FREE Full text] [doi: 10.1038/s41746-020-00368-2] [Medline: 33303889]

21. Peters M, Godfrey C, Khalil H, McInerney P, Parker D, Soares C. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc. Sep 2015;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]

22. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Tunçalp, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and explanation. Ann Intern Med. Oct 02, 2018;169(7):467-473. [doi: 10.7326/m18-0850]

23. Global action plan for the prevention and control of noncommunicable diseases 2013-2020. WHO. URL: https://iris.who.int/handle/10665/94384 [accessed 2024-09-30]

24. Shi X, Nikolic G, Fischaber S, Black M, Rankin D, Epelde G, et al. System architecture of a European platform for health policy decision making: MIDAS. Front Public Health. Mar 31, 2022;10:838438. [FREE Full text] [doi: 10.3389/fpubh.2022.838438] [Medline: 35433572]

25. Mavrogiorgou A, Kiourtis A, Maglogiannis I, Kyriazis D, De Nigro A, Blanes-Selva V, et al. CrowdHEALTH: An e-Health Big Data Driven Platform towards Public Health Policies. In: Proceedings of the 6th International Conference on Information and Communication Technologies for Ageing Well and e-Health (ICT4AWE 2020). 2020. Presented at: 6th International Conference on Information and Communication Technologies for Ageing Well and e-Health; May 3-5, 2020:241-249; Prague, Czech Republic. [doi: 10.5220/0009388802410249]

26. Moutselos K, Maglogiannis I. Evidence-based public health policy models development and evaluation using big data analytics and web technologies. Med Arch. Feb 2020;74(1):47-53. [FREE Full text] [doi: 10.5455/medarh.2020.74.47-53] [Medline: 32317835]

27. Prasinos M, Basdekis I, Anisetti M, Spanoudakis G, Koutsouris D, Damiani E. A modelling framework for evidence-based public health policy making. IEEE J. Biomed. Health Inform. May 2022;26(5):2388-2399. [doi: 10.1109/jbhi.2022.3142503]

28. Saunders G, Christensen J, Gutenberg J, Pontoppidan N, Smith A, Spanoudakis G, et al. Application of big data to support evidence-based public health policy decision-making for hearing. Ear Hear. 2020;41(5):1057-1063. [FREE Full text] [doi: 10.1097/AUD.0000000000000850] [Medline: 31985536]

29. Canfell OJ, Davidson K, Woods L, Sullivan C, Cocoros NM, Klompas M, et al. Precision public health for non-communicable diseases: An emerging strategic roadmap and multinational use cases. Front Public Health. Apr 8, 2022;10:854525. [FREE Full text] [doi: 10.3389/fpubh.2022.854525] [Medline: 35462850]

30. Shaban-Nejad A, Lavigne M, Okhmatovskaia A, Buckeridge DL. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. Ann N Y Acad Sci. Jan 17, 2017;1387(1):44-53. [doi: 10.1111/nyas.13271] [Medline: 27750378]

31. Ji X, Chun SA, Cappellari P, Geller J. Linking and using social media data for enhancing public health analytics. Journal of Information Science. Feb 01, 2016;43(2):221-245. [doi: 10.1177/0165551515625029]

32. Cocoros NM, Kirby C, Zambarano B, Ochoa A, Eberhardt K, Rocchio, SB C, et al. RiskScape: A data visualization and aggregation platform for public health surveillance using routine electronic health record data. Am J Public Health. Feb 2021;111(2):269-276. [doi: 10.2105/ajph.2020.305963]

33. RiskScape. ESP. URL: https://www.esphealth.org/riskscape [accessed 2024-09-30]

34. Khoury M, Engelgau M, Chambers D, Mensah G. Beyond public health genomics: Can big data and predictive analytics deliver precision public health? Public Health Genomics. Jul 17, 2018;21(5-6):244-250. [FREE Full text] [doi: 10.1159/000501465] [Medline: 31315115]

35. Roberts E, Kumar B, Collins N, Guo L, King G, Wong M, et al. Guidelines for the mapping of cancer registry data: results from a breast cancer expert panel study. J Public Health Manag Pract. 2013;19(3):E1-E10. [doi: 10.1097/PHH.0b013e318268aef1] [Medline: 23364274]

36. Kim JJ, Tosteson AN, Zauber AG, Sprague BL, Stout NK, Alagoz O, et al. Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) consortium. Cancer models and real-world data: Better together. J Natl Cancer Inst. Feb 03, 2016;108(2):djv316. [FREE Full text] [doi: 10.1093/jnci/djv316] [Medline: 26538628]

37. Salmasi L, Capobianco E. Predictive assessment of cancer center catchment area from electronic health records. Front Public Health. Nov 16, 2017;5:303. [FREE Full text] [doi: 10.3389/fpubh.2017.00303] [Medline: 29201863]

38. Torrente M, Sousa PA, Hernández R, Blanco M, Calvo V, Collazo A, et al. An artificial intelligence-based tool for data analysis and prognosis in cancer patients: Results from the Clarify Study. Cancers (Basel). Aug 22, 2022;14(16):4041. [FREE Full text] [doi: 10.3390/cancers14164041] [Medline: 36011034]

39. Okhmatovskaia A, Buckeridge D. Intelligent tools for precision public health. Stud Health Technol Inform. Jun 16, 2020;270:858-863. [doi: 10.3233/SHTI200283] [Medline: 32570504]

40. Dritsakis G, Trenkova L, Śliwińska-Kowalska M, Brdarić D, Pontoppidan NH, Katrakazas P, et al. Public health policy-making for hearing loss: stakeholders' evaluation of a novel eHealth tool. Health Res Policy Syst. Oct 29, 2020;18(1):125. [FREE Full text] [doi: 10.1186/s12961-020-00637-2] [Medline: 33121491]

41. Stoumpos AI, Kitsios F, Talias MA. Digital transformation in healthcare: Technology acceptance and its applications. Int J Environ Res Public Health. Feb 15, 2023;20(4):3407. [FREE Full text] [doi: 10.3390/ijerph20043407] [Medline: 36834105]

42. Ingram E, Cooper S, Beardon S, Körner K, McDonald HI, Hogarth S, et al. Barriers and facilitators of use of analytics for strategic health and care decision-making: a qualitative study of senior health and care leaders' perspectives. BMJ Open. Feb 17, 2022;12(2):e055504. [FREE Full text] [doi: 10.1136/bmjopen-2021-055504] [Medline: 35177457]

43. Sørensen K, Levin-Zamir D, Duong T, Okan O, Brasil V, Nutbeam D. Building health literacy system capacity: a framework for health literate systems. Health Promot Int. Dec 13, 2021;36(Supplement_1):i13-i23. [FREE Full text] [doi: 10.1093/heapro/daab153] [Medline: 34897445]

44. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, et al. Mapping factors that affect the uptake of digital therapeutics within health systems: Scoping review. J Med Internet Res. Jul 25, 2023;25:e48000. [FREE Full text] [doi: 10.2196/48000] [Medline: 37490322]

45. Chung Y, Bagheri N, Salinas-Perez JA, Smurthwaite K, Walsh E, Furst M, et al. Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review. International Journal of Information Management. Feb 2020;50:17-27. [doi: 10.1016/j.ijinfomgt.2019.04.012]

46. Walsh EI, Chung Y, Cherbuin N, Salvador-Carulla L. Experts' perceptions on the use of visual analytics for complex mental healthcare planning: an exploratory study. BMC Med Res Methodol. May 07, 2020;20(1):110. [FREE Full text] [doi: 10.1186/s12874-020-00986-0] [Medline: 32380946]

47. Fisher S, Rosella LC. Priorities for successful use of artificial intelligence by public health organizations: a literature review. BMC Public Health. Nov 22, 2022;22(1):2146. [FREE Full text] [doi: 10.1186/s12889-022-14422-z] [Medline: 36419010]

48. Rajamani S, Hultman G, Bakker C, Melton GB. The role of organizational culture in health information technology implementations: A scoping review. Learn Health Syst. Jul 11, 2022;6(3):e10299. [FREE Full text] [doi: 10.1002/lrh2.10299] [Medline: 35860317]

Abbreviations

AI: artificial intelligence
BDA: big data analytics
CRC: colorectal cancer
EHR: electronic health record
EU: European Union
MDPH: Massachusetts Department of Public Health
ML: machine learning
NCD: noncommunicable disease