Data Analytics to support Policy-making for Non-Communicable Diseases: A Scoping Review

Giorgos Dritsakis, Ioannis Gallos, Maria-Elisavet Psomiadi, Angelos Amditis, Dimitra Dionysiou

Submitted to: Online Journal of Public Health Informatics
on: May 09, 2024

Table of Contents

Data Analytics to support Policy-making for Non-Communicable Diseases: A 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 GR
[2]Directorate of Operational Preparedness for Public Health Emergencies Greek Ministry of Health Athens GR

Corresponding Author:
Giorgos Dritsakis PhD
Institute of Communication and Computer Systems
National Technical University of Athens
9, Iroon Politechniou Str. Zografou
Athens
GR

Abstract

Background: There is an emerging need for evidence-based approaches harnessing large amounts of healthcare data and novel technologies (such as Artificial Intelligence, AI) to optimize public health policy-making.

Objective: The aim of the present study was to explore the data analytics tools designed specifically for policy-making in Non-Communicable Diseases (NCDs) and their implementation.

Methods: A scoping review was conducted after searching the PubMed database 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. Tools incorporate descriptive and predictive analytics. Some tools were designed to include recommendations for decision support but no pilot studies have been published that apply prescriptive analytics. The tools were piloted with a range of conditions with cancer being the condition least studied. Implementation of tools included use cases, pilots or evaluation workshops were reported that involved policy-makers. However, our findings demonstrate very limited real-world use of analytics by policy-makers, in line with previous studies.

Conclusions: Despite the availability of tools designed for different purposes and conditions, data analytics are 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 a number of suggestions for researchers developing digital tools to support public health policy-making. The findings will also serve as input for the EU-funded research project ONCODIR developing a policy analytics dashboard for the prevention of colorectal cancer as part of an integrated platform.

(JMIR Preprints 09/05/2024:59906)
DOI: https://doi.org/10.2196/preprints.59906

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?
✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Data Analytics to support Policy-making for Non-Communicable Diseases: A Scoping Review

Giorgos Dritsakis[1*], PhD; Ioannis K. Gallos[1], PhD; Maria-Elisavet Psomiadi[2], RN, MSc, PhD(c); Angelos Amditis[1], PhD; and 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

*Correspondence: giorgos.dritsakis@iccs.gr, dimitra.dionysiou@iccs.gr

Abstract

Background: There is an emerging need for evidence-based approaches harnessing large amounts of healthcare data and novel technologies (such as Artificial Intelligence, AI) to optimize public health policy-making.

Objective: The aim of the present study was to explore the data analytics tools designed specifically for policy-making in Non-Communicable Diseases (NCDs) and their implementation.

Methods: A scoping review was conducted after searching the PubMed database 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. Tools incorporate descriptive and predictive analytics. Some tools were designed to include recommendations for decision support but no pilot studies have been published that apply prescriptive analytics. The tools were piloted with a range of conditions with cancer being the condition least studied. Implementation of tools included use cases, pilots or evaluation workshops were reported that involved policy-makers. However, our findings demonstrate very limited real-world use of analytics by policy-makers, in line with previous studies.

Conclusions: Despite the availability of tools designed for different purposes and conditions, data analytics are 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 a number of suggestions for researchers developing digital tools to support public health policy-making. The findings will also serve

as input for the EU-funded research project ONCODIR developing a policy analytics dashboard for the prevention of colorectal cancer as part of an integrated platform.

Keywords: policy-making; public health; non-communicable diseases; data analytics; digital tools

Introduction

Non-communicable diseases as a public health challenge

Non-Communicable Diseases (NCDs), such as cardiovascular and chronic respiratory diseases, cancer or 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 the majority of premature deaths [2]. They affect quality of life and life expectancy, create numerous challenges for patients and their families and have a big financial impact, costing EU economies more than €100 billion annually. As many NCDs are age-related, their burden is 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, 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 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 NCDs' 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 two decades, the growth of healthcare data in quantity and complexity as well as rapid advances in the field 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, 11, 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 [29].

Data analytics plays a key role at various stages of the public health policy-making process by aiding understanding, priority setting, resource allocation optimization, identification of the optimal intervention, implementation and evaluation [13]. Data analytics approaches are commonly categorized into three broad types: descriptive, predictive and prescriptive analytics. The three 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 summarizes past and present trends and patterns in the data to answer the question "What happened?". Predictive analytics uses primarily historical data to create models that answer the question "What will happen". Prescriptive analytics employs data-driven models and optimization algorithms to recommend the most effective actions, interventions, or allocation strategies trying to answer the question of "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 [14]. It has been suggested that policy analytics include data-driven tools that respond to a policy need and use a transparent development process

[15]. Examples of policy analytics techniques are statistics, simulations, data mining, machine learning, social network and geographic info [16].

Rationale & 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 also highlight gaps and unmet needs in the area. The present review has been performed in the context of the Horizon Europe project ONCODIR [42]. 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 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. A secondary aim is to explore the adoption of these tools by policy-makers and the factors affecting this.

Methods

Design

A scoping review was most appropriate as the aim of the study was to explore a relatively new concept. We conducted our review according to the Joanna Briggs Institute guidelines to ensure quality and reliability [17]. We followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews (PRISMA-ScR) and developed a search strategy based on the Population, Concept and Context framework as below [18]:

- 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. We searched for articles published in the last decade to explore the latest advances in the field.

The PubMed query we used was as follows:

(("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. Also, 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 (**Table 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 and/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 WHO [19]. We excluded analytics not performed on specific conditions or not applied on health data (e.g. healthcare 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.

Results

Study characteristics

The search yielded 1369 articles from the PubMed database (Figure 1). After article duplicates were removed, 1346 remained for title and abstract screening. A total of 31 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 reference lists of screened articles. Also 2 more 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 the present 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 describe the tools in detail and include use case scenarios, pilot studies, workshops, examples of implementation that illustrate how the analytics can be used to support policy-making for NCDs. Four studies were conducted in the USA, one in Australia and one in Canada. Three studies were conducted in various countries of Europe, i.e. Greece, Germany, Slovenia, Spain, UK, Sweden, Finland, Northern Ireland, Republic of Ireland and Denmark.

Tools overview

The 7 tools reviewed are summarized in **Table 2**. Three integrated platforms (EVOTION, MIDAS, 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 [22-26]. Users can create policy models, define the way in which data should be analyzed in order to produce the evidence useful for public health policymaking and obtain analytical results of how this evidence may support or contradict various policy actions. The other 4 tools (PoPHQ, PoPHR, Social InfoButtons, RiskScape) were designed for disease monitoring through the integration, descriptive analysis and visualization of population health data from various sources [20, 21, 27, 28].

Two of the 7 tools were designed to address specific NCDs. EVOTION was specifically designed for Hearing Loss (HL) and PopHQ for obesity prevention. The rest of the 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 [23]. Six of the 7 tools were designed for or used with NCDs only while PoPHR was applicable to infectious diseases too [21].

Study settings

Details of the use cases are in **Table 3**. The studies present the 7 tools using different settings including descriptions of the functionalities with examples, use case scenarios where authors present a workflow, pilot studies with actual data from various sources, applications of analytics on data collected using the tool, design workshops with stakeholders or real-life implementation.

CrowdHEALTH was piloted in 5 different countries and 6 use case scenarios with stakeholders from various healthcare organizations [23]. 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, 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 interventions 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 [24]. MIDAS was piloted in 4 countries using Social Media, MEDLINE analytics and News Media data on a range of conditions [22]. Ji et al. (2017) used data on treatments and symptoms posted by patients on Social Media to evaluate the effectiveness of the Social InfoButtons platform [27]. PopHR was piloted with a randomly sampled, open cohort of 25% of the Montreal population in Canada [21]. PopHQ is currently in mock-up phase but has been designed to integrate anonymized electronic medical record data from Queensland with a final total sample size of ∼5 million [20]. No pilot data was 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 [28]. Prasinos et al. (2020) described the workflow of policy creation, selection and execution of analytics using the EVOTION platform [25]. Saunders et al. (2020) reported three applications of

BDA techniques using a dataset synthesized from the EVOTION data repository [26].

Data analytics applied

Table 3 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 they address 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 [21, 22, 23, 27]. The next step was data exploration using basic descriptive statistics and inferential statistics such as identification of risk factors for a specific condition. Static or interactive visualizations were used including scatter plots, heatmaps, bar and box plots or pie charts. Three of the tools employed more specialized techniques including geospatial analytics or mapping (i.e. exploration of data in a geographical area), temporal analytics (i.e. tracking trends over time) and comparative analytics, which identifies differences between groups of measurements such as disease prevalence across different age groups [20, 21, 28]. Other types of analyses used were clustering analysis (i.e. 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 visualisation 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 modelling. Contemporary frameworks were also used for prediction and forecasting, including machine learning, deep learning and simulation modelling [22, 26]. Other types of analyses that were used for predictive purposes is risk stratification analysis, causal analysis that models the behavior of the target valuable of interest and clinical pathway analysis, which models the process followed during treatment of a patient in respect to a particular condition [23, 24].

Prescriptive Analytics

No study implemented a concrete prescriptive methodology. Three out of 7 tools were decision support systems designed to make policy recommendations. Authors referred to the prescriptive capabilities of the tools and demonstrated the policy creation process [20, 22, 25]. However, none of them presented a related use case actually applying this type of analytics.

Tools implementation

Out of the seven platforms reviewed, one is 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 [28]. It has a key role in demonstrating need and burden for MDPH's applications for funding through the identification of inequitably burdened populations. The authors suggest that the platform unloads analytic 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 [43]. Shaban-Nejad et al. (2017) reported that PoPHR had been initially implemented and was in the process of being fully implemented in Montreal, Quebec [21]. 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 include analytics examples, use cases or pilots. MIDAS was evaluated by policy-makers in the pilot studies and successfully achieved all Key Progress Indicators [22]. 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. (2020) reported how each use case of the CrowdHEALTH platform provided insight that can be used in policy-making [23]. Stakeholders from various sectors attended a workshop and provided feedback on the purpose, interface and overall design of PopHQ [20]. They identified various uses of the platform to create stories for 4 different end-users: public health practitioner, systems planner, researcher, generic user. PoPHQ is planned to be implemented in Queensland with a population of 5 million [20]. Saunders et al. (2020) reported that a policy-maker could use the EVOTION analytics as evidence to expand guidelines aimed at preventing noise-induced HL or simulate Hearing Aid (HA) 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 [26]. Finally, Social Infobuttons can be used by governments for disease surveillance [27].

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 demonstrated two different types of tools enabling data analytics for policy-making for NCDs – (a) 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 (b) integrated platforms designed for policy decision support with both descriptive and predictive analytics functionalities. Previously, Canfell et al. 2022 reviewed the use of real-world data for precision public health in NCDs and identified surveillance platforms integrating descriptive, comparative and geospatial analytics [29]. 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 Machine Learning (ML) techniques were used in studies included present review for forecasting as well as classic statistical methods such as logistic regression analysis. ML techniques are increasingly used with large population health datasets to improve public health surveillance, disease prediction and delivery of interventions [45]. 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 three 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.

NCDs studied

From the 4 major types of NCDs cancer seems 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 [23]. 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 towards the improvement of resource allocation in the healthcare system among others. Other studies have explored protocols for mapping breast cancer registry data [30], use of modelling to optimize cancer screening and predict catchment areas and use of AI for risk stratification of cancer patients [31-33]. 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 demonstrate very limited real-world use of analytics by policy-makers. This finding is in line with previous studies showing limited implementation of digital tools for NCDs [29]. Only RiskScape is fully implemented and is also publicly available to use. Shaban-Nejad et al. (2017) reported that PoPHR had been initially implemented and was in the process of being fully implemented in Montreal, Quebec but no more reports were published since then [21]. In a subsequent study from 2020 not included in this review, it was reported that PoPHR was in the process of being deployed in the Quebec for routine use by public health professionals [44]. In the same study, 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 the EVOTION platform, Saunders et al 2020 reported some applications of analytics with implications for policy-makers [26]. In another study not included in this review Dritsakis et al. (2020) reported a series of workshops where EVOTION was demonstrated to stakeholders in four countries and evaluated using a Strengths Weaknesses Opportunities and Threats methodology [34]. The study highlighted the huge 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 (e.g. 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. 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

A number of studies have explored factors influencing the adoption of digital health technologies by policy-makers. Innovative solutions are incorporated in health policy functions at a slower pace than health transforms into a digital asset [35]. Lack of advanced infrastructure, low interoperability levels among critical actors, and bureaucracy pose a barrier 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 [36]. Systemic and organizational issues exist as well due to the core characteristics of public health authorities as administrative bodies that lack regulatory frameworks and a data governance culture as a whole [38].

A very important obstacle from the end users' perspective is often the poor IT literacy and lack of digital skills by policy-makers and public health professionals [37]. Another aspect is whether these tools actually meet the needs of the end-users. For example, reviews on the use of visual analytics in mental healthcare planning have showed 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 [39, 40]. 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 utilize 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 [26]. Also, 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 an optimization and reorganization of the public health sector and workflows [29]. Some priorities that have been reported are the investment in modern data management infrastructure, the development of strategic partnerships, the need for AI transparency and reproducibility and the explicit consideration of equity and bias [41]. To effectively utilize new technologies and the big amounts of healthcare data that is 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. Also, 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.

Terminology

This review was set out to explore the data analytics tools designed to support policy-making for NCDs. 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 it is present in the literature, was rarely used in the studies reviewed here. Instead, authors use separate terms to refer to: (a) the analytics and tools designed (e.g. data integration / aggregation / visualization / analysis, big data analytics, platform, data-driven, evidence-based, system) and (b) to the use/purpose of the analytics [e.g. policy (decision) making, (precision) public health, surveillance, monitoring, population health). There are differences in the way the applications of the tools are reported in the literature. For example, the term ‘use case’ may refer to actual use of the tool in a particular setting or scenarios and examples of how the tool can be used. A variety of terms is also used to refer to the conditions (e.g. NCD, public health or the name of the individual condition) and the end users (e.g. policy-makers, decision-makers, public health professionals). This highlights the difficulty to comprehensively review all such tools and their applications.

Conclusions

Although a number of digital tools are available to inform policy-making for NCDs incorporating descriptive and predictive analytics, the majority of them are not widely used in real-life. Moving forward, we recommend that researchers involved in the development and use of data analytics tools for policy-making:

1. Engage with policy-makers throughout the process, respond to their needs and present results in a way that is easy to interpret.
2. Empower the digital health skills of policy-makers to ensure they can utilize the developed IT tools.
3. Specify the type and purpose of analytics, how policy-makers can use it and what they can achieve with it.
4. Secure the required resources to make the tool available and sustainable after it has been developed.

Acknowledgments

GD made substantial contributions to the conception and design of the work throughout the manuscript preparation. IG contributed to the part of the manuscript specific to data analytics. MP 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. The research has been funded by the European Union's Horizon Europe research and innovation program under grant agreement No. 101104777 (ONCODIR).

Conflicts of interest

None declared.

Abbreviations

NCD: Non-Communicable Disease

BDA: Big Data Analytics

AI: Artificial Intelligence

CRC: Colo-Rectal Cancer

PRISMA-ScR: Preferred Reporting Items for Systematic reviews and Meta-Analyses Extension for Scoping Reviews

HL: Hearing Loss

HA: Hearing aid

MDPH: Massachusetts Department of Public Health

References

1. https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases

2. EU Commission, 2022. Healthier together EU noncommunicable diseases initiative

3. Kehler, D.S., 2019. Age-related disease burden as a measure of population ageing. *The Lancet Public Health*, *4*(3), pp.e123-e124.

4. European Parliament. REPORT on non-communicable diseases (NCDs). 2023/2075(INI). Committee on the Environment, Public Health and Food Safety.

5. World Health Organization, 2021. Development of an Implementation Roadmap 2023–2030 for the Global Action Plan for the Prevention and Control of NCDS 2013–2030. *WHO: Geneva, Switzerland*.

6. Data, analytics and delivery for impact in focus 2022: priorities for regions and countries. Geneva: World Health Organization; 2022. Licence: CC BY-NC-SA 3.0 IGO.

7. van de Goor, I., Hämäläinen, R.M., Syed, A., Lau, C.J., Sandu, P., Spitters, H., Karlsson, L.E., Dulf, D., Valente, A., Castellani, T. and Aro, A.R., 2017. Determinants of evidence use in public health policy making: results from a study across six EU countries. *Health policy*, *121*(3), pp.273-281.

8. Kostkova P. (2015). Grand challenges in digital health. *Frontiers in public health*, *3*, 134. https://doi.org/10.3389/fpubh.2015.00134

9. Kostkova, P., Saigí-Rubió, F., Eguia, H., Borbolla, D., Verschuuren, M., Hamilton, C., Azzopardi-Muscat, N., & Novillo-Ortiz, D. (2021). Data and Digital Solutions to Support Surveillance Strategies in the Context of the COVID-19 Pandemic. *Frontiers in digital health*, *3*, 707902. https://doi.org/10.3389/fdgth.2021.707902

10. Chao, K., Sarker, M.N.I., Ali, I., Firdaus, R.R., Azman, A. and Shaed, M.M., 2023. Big data-driven public health policy making: Potential for the healthcare industry. *Heliyon*, *9*(9).

11. Secinaro, S., Calandra, D., Secinaro, A., Muthurangu, V. and Biancone, P., 2021. The role of artificial intelligence in healthcare: a structured literature review. *BMC medical informatics and decision making*, *21*, pp.1-23.

12. Myller Augusto Santos Gomes, João Luiz Kovaleski, Regina Negri Pagani, Vander Luiz da Silva & Tatiana Cabreira de Severo Pasquini (2023) Transforming healthcare with big data analytics: technologies, techniques and prospects, Journal of Medical Engineering & Technology, 47:1, 1-11

13. Global Commission on Evidence to Address Societal Challenges. 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. Hamilton: McMaster Health Forum, 2024

14. Tsoukias, A., Montibeller, G., Lucertini, G. and Belton, V., 2013. Policy analytics: an agenda for research and practice. *EURO Journal on Decision Processes*, *1*(1-2), pp.115-134.

15. Meinard, Y., Barreteau, O., Boschet, C., Daniell, K.A., Ferrand, N., Girard, S., Guillaume, J.H., Hassenforder, E., Lord, M., Merad, M. and Nabavi, E., 2021. What is policy analytics? An exploration of 5 years of environmental management applications. *Environmental Management*, *67*, pp.886-900.

16. Gil-Garcia, J.R., Pardo, T.A. and Luna-Reyes, L.F., 2018. Policy analytics, modelling, and informatics. *Policy Analytics: Definitions, Components, Methods, and Illustrative Examples*, pp.1-16.

17. Peters, M.D., Godfrey, C.M., Khalil, H., McInerney, P., Parker, D. and Soares, C.B., 2015. Guidance for conducting systematic scoping reviews. *JBI Evidence Implementation*, *13*(3), pp.141-146.

18. Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D., Horsley, T., Weeks, L. and Hempel, S., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Annals of internal medicine*, *169*(7), pp.467-473.

19. World Health Organization, 2013. *Global action plan for the prevention and control of noncommunicable diseases 2013-2020*. World Health Organization.

20. Canfell, O.J., Davidson, K., Woods, L., Sullivan, C., Cocoros, N.M., Klompas, M., Zambarano, B., Eakin, E., Littlewood, R. and Burton-Jones, A., 2022. Precision public health for non-communicable diseases: an emerging strategic roadmap and multinational use cases. *Frontiers in Public Health*, *10*, p.508.

21. Shaban-Nejad, A., Lavigne, M., Okhmatovskaia, A. and Buckeridge, D.L., 2017. PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data. *Annals of the New York Academy of Sciences*, *1387*(1), pp.44-53.
22. Shi, X., Nikolic, G., Fischaber, S., Black, M., Rankin, D., Epelde, G., Beristain, A., Alvarez, R., Arrue, M., Pita Costa, J. and Grobelnik, M., 2022. System architecture of a European platform for health policy decision making: MIDAS. *Frontiers in public health*, *10*, p.838438.
23. Mavrogiorgou, A., Kiourtis, A., Maglogiannis, I., Kyriazis, D., De Nigro, A., Blanes-Selva, V., García-Gómez, J.M., Menychtas, A., Soric, M., Jurak, G. and Lustrek, M., 2020. CrowdHEALTH: An e-Health Big Data Driven Platform towards Public Health Policies. In *ICT4AWE* (pp. 241-249).
24. Moutselos, K. and Maglogiannis, I., 2020. Evidence-based public health policy models development and evaluation using big data analytics and web technologies. *Medical Archives*, *74*(1), p.47.
25. Prasinos, M., Basdekis, I., Anisetti, M., Spanoudakis, G., Koutsouris, D. and Damiani, E., 2022. A modelling framework for evidence-based public health policy making. *IEEE Journal of Biomedical and Health Informatics*, *26*(5), pp.2388-2399.
26. Saunders, G.H., Christensen, J.H., Gutenberg, J., Pontoppidan, N.H., Smith, A., Spanoudakis, G. and Bamiou, D.E., 2020. Application of big data to support evidence-based public health policy decision-making for hearing. *Ear and hearing*, *41*(5), p.1057.
27. Ji, X., Chun, S.A., Cappellari, P. and Geller, J., 2017. Linking and using social media data for enhancing public health analytics. *Journal of Information Science*, *43*(2), pp.221-245.
28. Cocoros, N.M., Kirby, C., Zambarano, B., Ochoa, A., Eberhardt, K., Rocchio, SB, C., Ursprung, W.S., Nielsen, V.M., Durham, N.N., Menchaca, J.T. and Josephson, M., 2021. RiskScape: a data visualization and aggregation platform for public health surveillance using routine electronic health record data. *American Journal of Public Health*, *111*(2), pp.269-276.
29. Canfell, O.J., Kodiyattu, Z., Eakin, E., Burton-Jones, A., Wong, I., Macaulay, C. and Sullivan, C., 2022. Real-world data for precision public health of noncommunicable diseases: a scoping review. *BMC Public*

Health, *22*(1), p.2166.
30. Roberts, E.M., Kumar, B., Collins, N., Guo, L., King, G., Wong, M., Barlow, J., Bichler, J., Cady, L., Engel, C. and Garrett, D., 2013. Guidelines for the Mapping of Cancer Registry Data. *Journal of Public Health Management and Practice*, *19*(3), pp.E1-E10.
31. Kim, J.J., Tosteson, A.N., Zauber, A.G., Sprague, B.L., Stout, N.K., Alagoz, O., Trentham-Dietz, A., Armstrong, K., Pruitt, S.L., Rutter, C.M. and Population-based Research Optimizing Screening through Personalized Regimens (PROSPR) consortium, 2016. Cancer models and real-world data: better together. *Journal Of The National Cancer Institute*, *108*(2), p.djv316.
32. Salmasi, L. and Capobianco, E., 2017. Predictive assessment of cancer center catchment area from electronic health records. *Frontiers in Public Health*, *5*, p.303.
33. Torrente, M., Sousa, P.A., Hernández, R., Blanco, M., Calvo, V., Collazo, A., Guerreiro, G.R., Núñez, B., Pimentao, J., Sánchez, J.C. and Campos, M., 2022. An artificial intelligence-based tool for data analysis and prognosis in cancer patients: Results from the Clarify study. *Cancers*, *14*(16), p.4041.
34. Dritsakis, G., Trenkova, L., Śliwińska-Kowalska, M., Brdarić, D., Pontoppidan, N.H., Katrakazas, P. and Bamiou, D.E., 2020. Public health policy-making for hearing loss: stakeholders' evaluation of a novel eHealth tool. *Health research policy and systems*, *18*, pp.1-12.
35. Stoumpos, A. I., Kitsios, F., & Talias, M. A. (2023). Digital Transformation in Healthcare: Technology Acceptance and Its Applications. *International journal of environmental research and public health*, *20*(4), 3407. https://doi.org/10.3390/ijerph20043407
36. Ingram, E., Cooper, S., Beardon, S., Körner, K., McDonald, H.I., Hogarth, S., Gomes, M. and Sheringham, J., 2022. 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*, *12*(2), p.e055504.
37. van Kessel, R., Roman-Urrestarazu, A., Anderson, M., Kyriopoulos, I., Field, S., Monti, G., Reed, S.D., Pavlova, M., Wharton, G. and Mossialos, E., 2023. Mapping factors that affect the uptake of digital therapeutics within health systems: scoping review. *Journal of Medical Internet Research*, *25*(1), p.e48000.

38. Sørensen, K., Levin-Zamir, D., Duong, T. V., Okan, O., Brasil, V. V., & Nutbeam, D. (2021). Building health literacy system capacity: a framework for health literate systems. *Health promotion international*, *36*(Supplement_1), i13–i23. https://doi.org/10.1093/heapro/daab153

39. Chung, Y., Bagheri, N., Salinas-Perez, J.A., Smurthwaite, K., Walsh, E., Furst, M., Rosenberg, S. and Salvador-Carulla, L., 2020. Role of visual analytics in supporting mental healthcare systems research and policy: A systematic scoping review. *International Journal of Information Management*, *50*, pp.17-27.

40. Walsh, E.I., Chung, Y., Cherbuin, N. and Salvador-Carulla, L., 2020. Experts' perceptions on the use of visual analytics for complex mental healthcare planning: an exploratory study. *BMC medical research methodology*, *20*, pp.1-9.

41. Fisher, S. and Rosella, L.C., 2022. Priorities for successful use of artificial intelligence by public health organizations: a literature review. *BMC Public Health*, *22*(1), p.2146.

42. https://www.oncodir.eu/

43. https://www.esphealth.org/riskscape

44. Okhmatovskaia, A. and Buckeridge, D.L., 2020. Intelligent Tools for Precision Public Health. In *Digital Personalized Health and Medicine* (pp. 858-863). IOS Press.

45. Khoury, M.J., Engelgau, M., Chambers, D.A. and Mensah, G.A., 2019. Beyond public health genomics: can big data and predictive analytics deliver precision public health?. *Public health genomics*, *21*(5-6), pp.244-250.

Figures and Tables

Figure 1. PRISMA flow diagram of the stages of the scoping review.

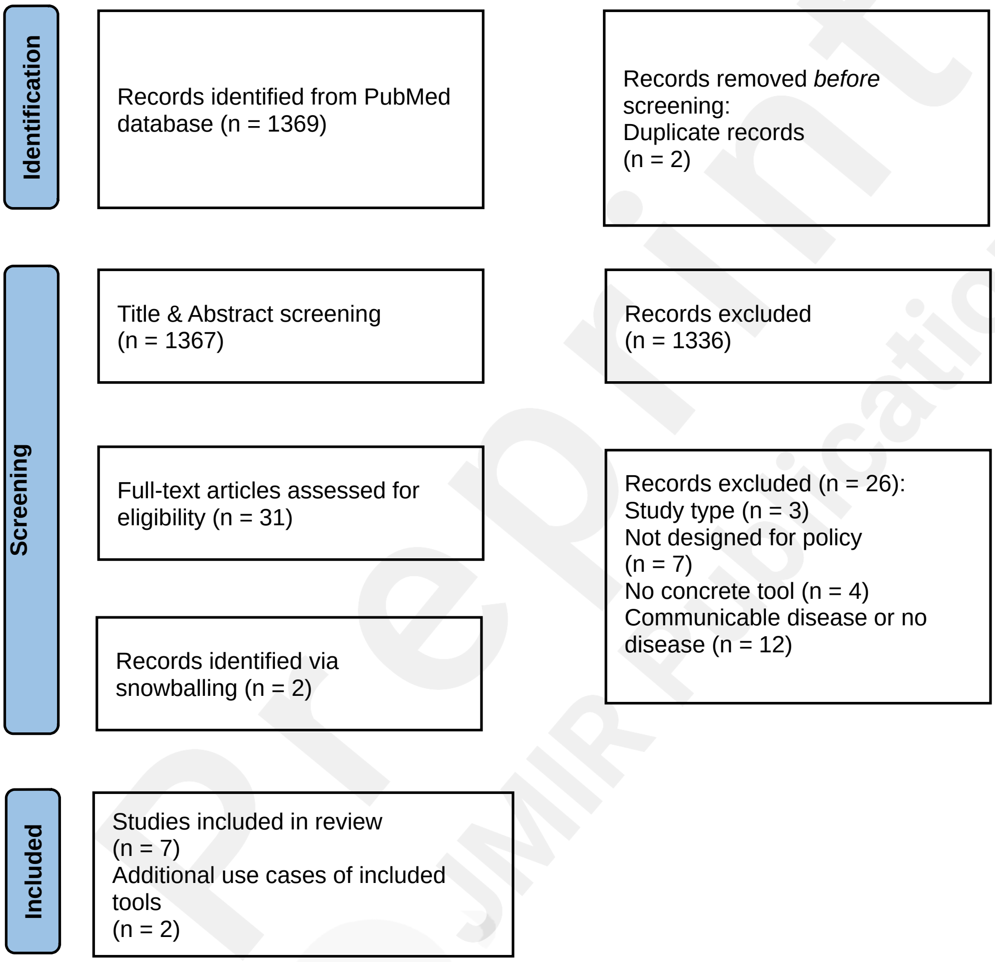

Table 1. Eligibility criteria for article selection.

Inclusion criteria	Exclusion criteria
Original research or review	Commentaries, viewpoints, protocols, perspectives or other study types

Studies on humans with a Non-Communicable Disease (NCD)	Articles on infectious diseases or no disease (e.g. health management)
The study includes a concrete digital data analytics tool that can be used by a policy-maker	The study describes theoretical models or frameworks or statistical algorithms
The study includes a tool that is designed to support health policy-making	The study includes tools for clinical decision support, patient benefit or other purposes
The study includes analytics performed on health-related data	The study includes analytics performed on the literature

Table 2. Overview of the tools included in the present review.

Tool	Type	Analytics supported	Implementation reported	Study
EVOTION	Public Health Policy Decision Support	Descriptive Predictive	Analytics applied Use case scenarios	Prasinos et al (2022) Saunders et al (2020)
CrowdHEALTH		Descriptive Predictive	Use cases	Moutselos et al (2020) Mavrogiorgou et al (2020)
MIDAS		Descriptive Predictive	Pilots	Shi et al (2022)
PoPHQ	Health informatics tool	Descriptive	Design workshops Use Cases	Canfell et al (2022)
RiskScape	Public health surveillance	Descriptive	Real-life implementation	Canfell et al (2022) Cocoros et al (2021)
Social InfoButtons	Public health analytics	Descriptive	Use case scenarios	Ji et al (2017)
PopHR	Visualization of population health data	Descriptive	Initial implementation	Shaban-Nejad et al (2017)

Table 3. Details of the analytics applied in the present review.

Study (Tool)	Setting	Country	NCD	Analytics applied	Purpose	Data
Prasinos et al. 2022 (EVOTION)	Use case example & workflow	Data collected: Greece, UK, Denmark	Hearing Loss (HL)	Basic Statistics Linear Regression Principal component Analysis Inferential Statistics	To investigate the impact of hearing health interventions on Quality of Life (QoL)	Synthetically generated dataset made by 10M of records from the EVOTION data repository
Saunders et al. 2020 (EVOTION)	Applications of big data analytics	As above	As above	Predictive modelling	To estimate the risk of noise-induced HL by modelling the combined impact of factors	As above
				Regression analysis Generalized linear mixed model	To predict Hearing Aid (HA) usage from changes to the sound environment	
				Correlations	To examine the association between physical activity and HA usage	
Moutselos et al. 2020 (CrowdHEALTH)	Use case example & workflow	Slovenia	Obesity	Forecasting Simulation Causal analysis	Effectiveness of various interventions on obesity prevention in schools Early detection of children with increased risk linked to poor physical fitness	Large scale data from Slovenia's national surveillance system on physical and motor development of children.
Mavrogiorgou et al. 2020	Use cases including	Spain	Obesity in adults	Clinical pathway mining	Identification of overweight patients	Demographic information, hospitalization, emergency

Study (Tool)	Setting	Country	NCD	Analytics applied	Purpose	Data
(CrowdHEALTH)	healthcare stakeholders			Risk stratification		and hospital at home episodes, morbidity
		Sweden	Cardiovascular Chronic kidney disease	Clinical pathway mining Risk stratification Causal analysis	Monitoring 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 of children Predict future levels of fitness and somatic development	Physical activity, sedentariness, sleep, heart-rate, socio-economic status and parental physical activity
		Greece	Chronic respiratory conditions	Risk stratification	Monitor disease progression and healthcare expenditure for improved chronic disease management of patients	Bio-signals from pulse oximeters, blood pressure meters, glucometers, spirometers, weighing scales and activity trackers
		Germany	Nutrition and physical activities	Clustering analysis Correlation analysis	Understand 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	Data on diagnosis, treatment, comorbidities, health behaviors and side-effects
Shi et al. 2022 (MIDAS)	Pilots with policy-makers	Spain	Obesity	RandomForests / Least Absolute	Identify risk factors of childhood obesity	Controlled and open data including:

Study (Tool)	Setting	Country	NCD	Analytics applied	Purpose	Data
				Shrinkage and Selection Operator		● Social Media Analysis ● MEDLINE analytics ● News Media Analysis
		Finland	Mental health	Lexis diagram analysis Descriptive analysis	Aggregate, summarize and visualize risk factors Evaluate health, social and education status	
		Northern Ireland	Social care for children	Markov chain LSTM Network	Track patterns of behavior over time Estimate probability of transition between different types of care Improve children protection	
		Republic of Ireland	Diabetes	ARIMA	Forecast the consumption of diabetic drugs	
Canfell et al. 2022 (PopHQ)	Design workshop where stakeholders identified user stories	Australia	Obesity	Comparison of obesity across age groups	Target interventions across the life course	Currently 'mock-up' without using patient data. The final total sample size is estimated to be ~5 million. Vision: to use anonymized Electronic Health Records (EHR) data from Queensland
				Obesity counts by facility	Direct resources	
				Counts and percentages of obesity	Justify the problem	
				Stratification by suburb and facility	Compare obesity across regions	
Canfell et al. 2022	Real-life	USA	Diabetes,	Heat maps by zip	Automated analysis of EHR.	EHR data updated monthly

Study (Tool)	Setting	Country	NCD	Analytics applied	Purpose	Data
Cocoros et al. 2021 (RiskScape)	implementation		hypertension, asthma, obesity	code Stratifications 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 Identification of potential health disparities	from 3 clinical practice groups that cover ~20% of the state population (>1.2 million).
Ji et al. 2017 (Social InfoButtons)	Use case scenarios	USA	PTSD, asthma	Statistical Geospatial Temporal Topic investigation Association discovery Recommendation discovery Visualization	Compute statistical aggregates, e.g. number of patients suffering from a condition Explore data according to a geographic feature, e.g. concentration of health conditions in a geographical area. Analyze trends over time Explore correlations between treatments, side effects, symptoms and conditions. Discover treatment recommendations given the	Openly available health data sources including: SMN, Twitter, MedHelp, WebMD, CDC and PubMed.

Study (Tool)	Setting	Country	NCD	Analytics applied	Purpose	Data
					symptoms Integrate openly available health data and make them easily understandable by physicians, healthcare staff and patients.	
Shaban-Nejad et al. 2017 (PopHR)	Description and functionalities with examples	Canada	Chronic diseases, such as diabetes, hypertension, coronary heart disease, and stroke	Visualizations: maps, bar charts, data tables, time series, scatter plot Stratification by age, sex and region Filtering Statistical algorithms to detect changes in an indicator over time and space	Exploration and visualization of available indicators for a defined population Create coherent portraits of population health and health system performance Evaluate the effects of public health interventions or other policies and programs	No data used in the study. Initial implementation: Randomly sampled, open cohort of 25% of the population of the Census Metropolitan Area of Montreal, Quebec. In the process of implementing in the entire population of the province of Quebec.