ABSTRACT

Introduction: NCDs (non-communicable diseases) are considered an important social issue and a financial burden to the health care systems in the EU which can be decreased if cost-effective policies are implemented, along with proactive interventions. The CrowdHEALTH project recognizes that NCD poses a burden for the healthcare sector and society and aims at focusing on NCDs’ public health policies. Aim: The aim of this paper is to present the concept of Public Health Policy (PHP), elaborate on the state-of-the-art of PHPs development, and propose a first approach to the modeling and evaluation of PHPs used in a toolkit that is going to support decision making, the Policy Development Toolkit (PDT).

Methods: The policy creation module is a part of the PDT aiming to integrate the results of the rest of the health analytics and policy components. It is the module that selects, filters, and aggregates all relevant information to help policy-makers with the decision making process. The policies creation component is connected to the visualization component to provide the final users with data visualization on different PHPs, including outcomes from data-driven models, such as risk stratification, clinical pathways mining, forecasting or causal analysis models, outcomes from cost-benefit analysis, and suggestions and recommendations from the results of different measured KPIs, using data from the Holistic Health Records (HHRs).

Results: In the context of CrowdHEALTH project, PHP can be defined as the decisions taken for actions by those responsible in the public sector that covers a set of actions or inactions that affect a group of public and private actors of the health care system. In the CrowdHEALTH project, the Policy Development Toolkit works as the main interface between the final users and the whole system in the CrowdHEALTH platform. The three components related to policy creation are: (i) the policy modeling component, (ii) the population identification component and (iii) the policy evaluation component. In policy evaluation, KPIs are used as measurable indicators to help prevent ambiguity problems in the interpretation of the model and the structure.

Conclusions: This initial Policy creation component design might be modified during the project life circle according to the concept complexity.

Keywords: Policy Making, Public Health Policy, Policy Creation.

1. INTRODUCTION

About 36 million deaths, or 63% of the 57 million deaths that happened on a global scale in 2008, were attributed to NCDs. Cardiovascular diseases were the first cause (48% of NCDs), cancers followed with 21%, chronic respiratory diseases with 12% and diabetes with 3.5%. According to the WHO, the total annual number of deaths from NCDs will rise to 55 million by 2030 if no significant change occurs (1). NCDs also make up for the healthy life years lost based on the Disability Adjusted Life Years (DALY) (2) while NCDs are considered an important social and financial burden to the health care systems and are a barrier to development, especially related to the aging population in developed countries and the EU member states (3),(4). NCDs burden can be decreased if cost-effective policies are implemented, along with proactive interventions and regular monitoring of NCDs. Enabling health systems to respond...
promptly and effectively to the health-care needs of people with NCDs can reduce premature deaths. Public policies regarding risk factors such as tobacco use, unhealthy diet, physical inactivity, and the harmful use of alcohol can be implemented in several sectors (4). The CrowdHEALTH project recognizes that NCD poses a burden to the healthcare sector and society and aims at focusing on NCD’s public health policies. Health policy refers to decisions, plans, and actions that are implemented to achieve specific health care goals within a society. A health policy defines a vision for the future that can be achieved by establishing targets and points of short and medium-term reference. Also, it sets priorities, outlines the expected roles of different groups, builds consensus, and raises people’s awareness (5). The aim of EU public health policies is to ameliorate human health, prevent disease and support change of Europe’s health systems (6). Health Policy may be difficult to define as the definition is based on the experience of who defines it. However, following the general definition of (7, 8) Public Health Policy (PHP) is the decisions taken by those responsible in the public sector that covers a set of actions or inactions that affect a group of public and private actors of the health care system to achieve specific health care goals. PHP takes into account the specific context and characteristics of the region where it has to be implemented affecting many different actors that have to be considered during its design. When a PHP is developed, a proper evaluation is mandatory to assess whether the actions or inactions are serving the defined goals. This process is monitored using Key Performance Indicators (KPIs) to evaluate whether the proposed goals are achieved and to consider the appropriateness of the selected indicators.

2. AIM
The aim of this paper is to present the concept of Public Health Policy (PHP), elaborate on the state-of-the-art of PHPs development, and propose a first approach to the modeling and evaluation of PHPs using a toolkit that is going to support decision making, the Policy Development Toolkit (PDT).

3. METHODS
The Policy Development Toolkit works as the main interface between the final users and the whole system in the CrowdHEALTH platform. The three components related to policy creation are: (i) the policy modeling component, (ii) the population identification component and (iii) the policy evaluation component. The policy creation module is a part of the PDT aiming to integrate the results of the rest of the health analytics and policy components. It is the module that selects, filters, and aggregates all relevant information to help policymakers with the decision making process. Hence, this module, as well as the whole PDT, is regarded as a decision support system for policy-makers. PDT does not only create new PHPs, but it can also be used to improve already existing PHPs. The policies creation component is connected to the visualization component to provide the final users with data visualization of different PHPs, including outcomes from data-driven models, such as risk stratification, clinical pathways mining, forecasting or causal analysis models, outcomes from cost-benefit analysis, and suggestions and recommendations from the results of different measured KPIs, using data from the Holistic Health Records (HHRs). The policy modeling component aims at providing a formal structure to a PHP. The structure is focused mainly on the key performance indicator of an existing or a new particular goal that is related to a PHP. A KPI is a clear measurable indicator with a clear definition and a mathematical formula, which helps prevent ambiguity problems in the interpretation of the model and the structure. However, since a KPI is related to specific goals of a PHP, the relation will be kept explicit in the model of the health policy. Also, particular data-driven models, specific data and sets of active and passive actors related to the PHP and the defined KPI can be made explicit in the formulation of the formal model of the PHPs, to relate possible predictions to the KPIs and hypothesize evolutions of the indicators for the improvement and development of PHPs. The policy evaluation component is devoted to the assessment of the different health policies that are under consideration in the PDT using a policy model that has to be defined in the previous component. The evaluation of the policies depends on the core concept of KPIs that defines and bases the formal policy models. Furthermore, each KPI will be associated with certain Organization for Economic Co-operation and Development (OECD) evaluation tiers (12) and the performance metrics defined, to clarify the role of each KPI inside a PHP. Besides, the policy evaluation component is responsible for obtaining the actual value of the KPI, based on its mathematical definition and the available data from the HHRs. Thus, it is very important to define clear and measurable KPIs and their associated mathematical formulas, as well as the population that is evaluated by the KPIs. The population identification component aims basically at providing a proper set of stored and available data from the HHRs that identify the proper population to measure the success of the PHP. This component can be seen as a system that applies filters to the actual data and gathers the results as an evaluation set to send it to the policy evaluation component. The population identification component is of utmost importance to clearly define the population for which the PHP is meant, before defining the different components of the PDT architecture. It is important because the policy formal model and structure depends on the understanding of health policies and its main features.

4. RESULTS
4.1. Policy modeling
Based on the definition of PHPs, the use of ontologies for policy modeling appears to be a rather complex task. The intrinsic complexity of PHPs and the extrinsic influences and dependencies on uncontrollable factors makes the design and development of PHPs rather un-
Figure 1. Conceptual structure of policy model

4.1. KPI-based PHP

The main goal of the population identification component is to provide a data set to evaluate the proposed KPI within the policy evaluation component, as well as for identifying data from the HHR population that can be used for running a data-driven predictive model. Hence, this component will serve two different but analog purposes: population identification for policy evaluation, and population identification for running predictive models. Identification of the proper population for policy evaluation must be based on a set of inclusion and exclusion criteria, which have to be included in metadata information. This metadata is data that provides information about the KPIs used to evaluate the policies and information on the predictive models, regardless of their nature, i.e. risk stratification, pathway mining, causal analysis, or forecasting. The metadata will include information for identifying the proper population through the use of filters to select the subjects under study. The data of the CrowdHEALTH project will be stored in the CrowdHEALTH data store, which in the end will be a relational data-
Table 1. Example of KPI and PHP conceptual structure

| ITEM | Description |
|------|-------------|
| KPI  | Rate of adults identified as obese. The prevalence of obesity in adults is known to be around 15% of the population. However, there is only around 1.5% of diagnoses of obesity reported using ICD-9 codes, which implies that obesity conditions is not being taken into account when reporting and storing secondary diagnosis of the patients. The goal of this KPI would be to report an increase in the prevalence of the obesity condition in the codification of the morbidities of each patient, promoting thus a systematic detection of this condition. |
| FORMULA | Numerator: Number of adult population properly identified as obese (ICD9ICD10) Denominator: Number of adult population |
| DATA | The data for computing this KPI and report its evolution should be obtained from the databases of the Health Department La Fe including primary and secondary care information about morbidity. |
| STAKEHOLDERS | People interested in the evolution of this KPI who may be reported are hospital managers, the head of the Endocrinology Service, and Experts on Public Health, among others. |
| GOAL | To sensitize healthcare professionals to boost the systematic detection of overweight and obesity in the population. |
| PHP | Strategy for physical activity, nutrition and prevention of obesity |
| ACTORS | People that may be affected by this PHP are citizens that are obese but they have not been correctly diagnosed and whose diagnosis has not been correctly stored. Also, endocrinologists may be affected as their work may be improved by this systematic detection of obesity. |
| MODELS | A model can be used to help detect people that are likely to be in the obesity profile. This is the risk stratification model proposed by UPV in Deliverable 5.2 based on a semi-supervised learning approach. |
| DATA FOR MODELS | The data to be used for the models will be determined during the development and validation of the semi-supervised model. |

| ITEM | Description |
|------|-------------|
| KPI  | Health-related fitness index. The estimation of fitness based on motor tests 600 m run, bent arm hang, and sit up, which gives the estimation of health-related fitness in percentages according to age and gender. |
| FORMULA | Numerator: Aggregated data from the health-related fitness index Denominator: 1 |
| DATA | The data for computing this KPI and report its evolution is gathered every April with field testing in all (about 600) primary and secondary schools in Bulgaria and imported directly in SLOCC database by school administrators. The percentages (according to sex and age) of fitness data are computed within SLOCC web application. |
| STAKEHOLDERS | People interested in the evolution of this KPI and who may be reported are Experts on Public Health, researchers, children’s parents. |
| GOAL | To promote good physical activity habits and physical fitness. |
| ITEM | Description |
| PHP | Nutrition and Physical Activity for health. |
| ACTORS | People that may be affected by this PHP are mainly children. |
| MODELS | A forecasting model can be used to help in the prediction of future evolution of health-related fitness index. |
| DATA FOR MODELS | The data to be used for the models will be determined during the development and validation of the forecasting model. |

Table 2. Example of another KPI and PHP conceptual structure

Variables that have to be selected for each subject under study. Hence, if a KPI or a predictive model needs the “gender” of the individuals, then this could be included in the select clause within the SQL query. Hence, there is a need to design a schema that will provide all necessary information within the metadata, to identify the population, as explained in the previous paragraphs. This metadata information can be developed using an XML schema that should be able to encapsulate all the needed attributes and sub-elements to represent the necessary filters and variables in detail, its parameters, and its structure. The XML Schema Definition (XSD) for population identification contains definitions for fields that are used as filters. It specifies the types and value ranges and the valid values for filtering in a where clause within an SQL query. The XSD is shown in Table 3. There are two main types of data to be considered when filtering SQL queries: numerical data and...
categorical data. Hence, there are two basic elements called NumericalDataField and CategoricalDataField that must be unique from other names in the PopulationIdentification element. Each element identifies each main type, respectively. Both elements share a set of common attributes like name, displayName, and varType. The name is a string that is used to identify the variable name. The displayName is an optional attribute that may be used by applications to refer to that field. However, the XML document only considers as significant the value of the name. If displayName is not given, then it takes the value of the name as default. The varType is a restricted type that allows using the following values: numerical, categorical, and dates. The NumericalDataType requires a minimum and a maximum value to be defined to provide enough information for an SQL query to be correctly defined. On the other hand, the CategoricalDataType requires a sequence of attributes with valid values for the appropriate type. The value number of fields is the number of fields that are described in the XSD document. This number can be otherwise defined. This sort of contextual information is not considered at present but could be an improvement for future versions. The filter for including citizens who have a diagnosis of overweight or obesity is identified through the variable “age”, which is a numerical data field. Hence, there are two basic elements called NumericalDataField and CategoricalDataField that must be unique from other names in the PopulationIdentification element. Each element identifies each main type, respectively. Both elements share a set of common attributes like name, displayName, and varType. The name is a string that is used to identify the variable name. The displayName is an optional attribute that may be used by applications to refer to that field. However, the XML document only considers as significant the value of the name. If displayName is not given, then it takes the value of the name as default. The varType is a restricted type that allows using the following values: numerical, categorical, and dates. The NumericalDataType requires a minimum and a maximum value to be defined to provide enough information for an SQL query to be correctly defined. On the other hand, the CategoricalDataType requires a sequence of attributes with valid values for the appropriate type. The value number of fields is the number of fields that are described in the XSD document. This number can be added for consistency checks. Based on the XSD, it is possible to define an XML document that will serve as a basis for identifying a particular population for evaluating a policy through its defined KPIs. At the same time, this document can be used for identifying a population to execute the data-driven models developed for each PHP. An example of an XML document for population identification is explained next. It focuses on collecting adult citizens who have been previously diagnosed with as overweight or obese patients based on ICD–9–CM codes (Table 4). The adult population is identified through the variable “age”, which is a numerical data type, with the minimum age being “18” and with a maximum value chosen to be large enough. It is clear that depending on the country, the adult age values of the “age” variable, which is 18 and 120, respectively, are transformed into “age >= 18 AND age <= 120”. Regarding the categorical data fields, the schema requires at least one valid value, but it can include more than one. Each valid value will be transformed into a list of valid values that will be part of an “IN” clause. Hence, the categorical data field for identifying the population in Table 4 is transformed into the following SQL expression “IN (‘278’, ‘278.0’,...
To summarize and generalize, on one hand, a numerical data field with the name “name”, with a minimum value “min” and a maximum value “max” will become part of the where clause of the SQL query with the form “name >= min AND name <= max”. On the other hand, a categorical data field with the name “name”, and a sequence of filter valid values “value1”, “value2”, ... will become part of the where clause with the form “name IN (‘value1’, ‘value2’, ...)”.

4.3. Policy evaluation

PHPs were described and modeled through a fundamental conceptual schema, in which the KPIs were at the core of each PHP model. KPI-centered models are expected to allow both developers and policy-makers to define, support and evaluate PHPs. The fact that KPIs are used as the core of PHPs is to facilitate the process of policy evaluation, enabling at the same time the possibility to visualize past and present results for comparison purposes, but also to forecast results (in case a forecasting model is developed for the proper KPI). Also, KPIs are invaluable tools that contribute to the process of health system performance monitoring. However, for KPIs to be effective, they need to have clear definitions to ensure that the data collected is consistent, reliable and, when possible, sharing definitions with official institutions like the WHO, OECD, EU, and other national and regional institutions. Hence, the evaluation of the PHP depends on proper definitions of the different KPIs related to that PHP. The definition of the KPI should include a well-formed formal mathematical formula without ambiguity, with proper term definitions and inclusion and exclusion criteria to ensure their validity and reliability. Valid KPIs measure what they are intended to measure, whereas reliable KPIs produce consistent results, regardless of who performs the measurement. The aim of the formal definition is to support the Population Identification component to identify the proper dataset to evaluate the KPI and then to compute the values of the PHP. The Policy Evaluation component is devoted to computing and assessing the policies through the KPIs. This component requires two inputs to compute its outcome and provide it to the Policy Creation component, where the variables are used for computing the value of the KPI based on its formula. The final result may be compared later with the results of data sources from different regions in Europe. For example, the current results from HULAFE datasets are that less than 3% of the capita of obesity follow the aforementioned formulation (10).

As a first approach, each factor of the KPI formula will be transformed manually into an SQL query to collect the corresponding information. Later versions may provide a more flexible approach to enable a semi-automatic interpretation of the numerator and denominator of the indicators. For instance, the KPI of the example tries to estimate the rate of the adult population that has been identified as obese. This is translated into a mathematical formula in Figure 4. To compute this formula, both factors need to be calculated as aggregated information from the CrowdHEALTH data store. The first factor, the numerator, may be gathered using an SQL expression such as:

Whereas the denominator may be gathered using:

The final result may be compared later with the result of the expectations from public health surveys, and also with the results of data sources from different regions in Europe. For example, the current results from HULAFE datasets are that less than 3% of the capita of the Health Department is correctly diagnosed as obese.
public health surveys that state that the obese adult population represents around 14% of the population. This result would indicate a failure in the goal of promoting a systematic detection of obesity in the adult population. A closer value of the indicator to the expectations would point out a better outcome of public health policies regarding systematic detection of obesity. Thus, this KPI would be able to assess if the systematic detection of obesity is successful or not.

5. CONCLUSION
The present work defines what a Public Health Policy (PHP) is in the context of the CrowdHEALTH project (12–17). It also proposes state-of-the-art best practices of PHPs in general and in NCDs particularly. The paper also presents the technologies and methodologies based on ontologies and semantic reasoning that can be used for modeling PHPs providing examples for further understanding. A structure for policy modeling based on Key Performance Indicators as the core of the structure is proposed since KPIs enable direct measurement of parts of the PHP. Finally, methodologies that can be developed for the three different components communication: policy modeling, policy evaluation and population identification within the policy creation component are presented. This is an initial design that might be modified during the project life cycle according to the conceptual complexity and the available information managed by the platform.

- Acknowledgements: CrowdHEALTH project is co-funded by the Horizon 2020 Programme of the European Commission Grant Agreement number: 727560 – Collective wisdom driving public health policies.
- Author’s contribution: Each author gave substantial contribution in acquisition, analysis and data interpretation. Each author had a part in preparing article for drafting and revising it critically for important in intellectual content. Each author gave final approval of the version to be published and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.
- Conflict of interest: not declared.
- Financial support: nil.

REFERENCES
1. World Health Organization. Global action plan for the prevention and control of noncommunicable diseases 2013–2020. 2013. Available at: https://apps.who.int/iris/handle/10665/94384
2. World Health Organization. Preventing chronic diseases: a vital investment: WHO global report. 2005. Available at: https://apps.who.int/iris/handle/10665/43314
3. World Health Organization. Global status report on non-comunicable diseases 2010. 2011. Available at: https://www.who.int/nmh/publications/ncd_report2010/en/
4. Bloom DE, Cafiero ET, Jané-Llopis E, Abrahams-Gessel S, Bloom LR, Fathima S, et al. The Global Economic Burden of Non-communicable Diseases. Harvard School of Public Health (2011)
5. World Health Organization. Health Policy. Available at: http://www.who.int/topics/health_policy/en/
6. European Commission. EU health policies. Available from: https://ec.europa.eu/health/policies/policy_en
7. Buse K, Mays N, Walt G, Gilson L. Making Health Policy (Understanding Public Health). London Open University Press2012; 236.
8. Smith PC, Mossialos E, Papanicolas I, Leatherman S. Performance Measurement for Health System Improvement. Experiences, Challenges and Prospects. Smith PC Cambridge University Press. 2009.
9. Agencia Española de seguridadalimentaria y nutrición. Estrategia NAOS: estrategia para la nutrición, actividadfísica y prevención de la obesidad. Ministra de Sanidad y Consumo. 2005.
10. Agencia Española de seguridadalimentaria y nutrición. Evaluación y seguimiento de la EstrategiaNAOS: conjunto mínimo de indicadores. Ministerio de sanidad, política social e igualdadAgencia Española de seguridadalimentaria y nutrición. 2011.
11. Ministry of Health. Resolucija o nacionalneprogramu o prehrani in telesnejavnosti za zdravje 2015–2025 (ReNPPT–D2), RepublikeSlovenije. 2015. Available at: https://www.uradni-list.si/glasilo–uradni–list–rs/vsebina/2015–01–2441 sop=2015–01–2441
12. Perakis K, Miltiadou D, De Negro A, Torelli F, Montandon L, Mantas J, et al. Data Sources and Gateways: Design and Open Specification. Acta Inform Med. 2019 Dec; 27(5): 341–347. doi: 10.5455/aim.2019.27.341-347.
13. Wajid U, Orton C, Magdalinou A, Mantas J, Montandon L. Generating and Knowledge Framework: Design and Open Specification. Acta Inform Med. 2019 Dec; 27(5): 362–368. doi: 10.5455/aim.2019.27.362-368.
14. Kyriazis D, Autexier S, Boniface M, Engen V, Jimenez-Peris R, Jordan B, et al. The CrowdHEALTH Project and the Holistic Health Records: Collective Wisdom Driving Public Health Policies. Acta Inform Med. 2019 Dec; 27(5): 369–373. doi: 10.5455/aim.2019.27.369-373.
15. Magdalinou A, Mantas J, Montandon L, Weber P, Gallos P. Disseminating research Outputs. The CrowdHEALTH Project. Acta Inform Med. 2019 Dec; 27(5): 348–355. doi: 10.5455/aim.2019.27.348-355.
16. Malliaros S, Xénakis C, Moldovan G, Mantas J, Magdalinou A, Montandon L. The Intergrated Holistic Security and Privacy Framework Deployed in CrowdHEALTH Project. Acta Inform Med. 2019 Dec; 27(5): 333–340. doi: 10.5455/aim.2019.27.333-340.
17. Arah O, Westert G, Hurst J, Klazinga N A conceptual framework for the OECD Health Care Quality Indicators Project. Int J Qual Heal Care. 2006;18(Suppl 1),5–13.