

Decision Intelligence in Public Health – DIONE

Jacob Stolk¹ and **Simone Nyon¹**

¹ *Dione Complex Systems, Gold Coast, Queensland, Australia*
Email: jacob@dionecomplexsystems.com

Abstract: Public health surveillance is the ongoing systematic collection, analysis, interpretation, and dissemination of health data for the planning, implementation, and evaluation of public health action. To achieve effective public health interventions, it is pivotal to analyse and interpret the vast amounts of data collected by surveillance systems to enable good understanding of all factors having an impact on health.

For example, we can consider child protection, which is an important public health issue. Often relatively extensive data exist on families in official statistics, research reports, social services reports, school and medical records, etc. However, these data are dispersed and hard if not impossible to relate and compare. This leads to numerous interventions that are conducted without adequate knowledge of target families that are to benefit from these interventions. Inadequate knowledge also leads to lack of intervention where it is needed, such as undetected cases of child abuse. In many cases the basic data that are needed for intervention decisions exist, but are not available to decision makers due to inadequate communication and lack of data integration, analysis and interpretation.

Chronic condition management is another area where extensive disparate data exist from statistics and various health services and intervention agencies. In this area too there are numerous organizations offering services and a great need to better coordinate these services to achieve better outcomes for patients and also to reduce soaring costs of the healthcare system.

We propose DIONE (Decision Intelligence for Organizations in Networked Environments), a decision support system that uses a complex systems approach to offer a means to integrate and maximally exploit *all* available data to optimize intervention decisions. It will transform dispersed data into relevant information for decision makers. The proposed system has its place on top of existing research and data collection, rather than duplicating any existing research, and will optimize utilization of existing data and research results.

DIONE will enable more focussed health and social service interventions to encourage and help people to manage life stages and events, in view of achieving the best possible health and behaviour outcomes. The system will use data mining and analysis techniques such as predictive analytics, text analytics, data mining, machine learning, and Bayesian statistics to integrate and exploit all available quantitative and qualitative data from heterogeneous data sources, such as statistics, social services, schools, medical services, etc., to identify risk factors and people most at risk. This will enable more focussed health and behaviour interventions. The problem with risk factors for health outcomes is that they are in general hard or impossible to observe directly. The proposed data mining techniques will be calibrated with synthetic data, obtained from agent-based simulation or micro-simulation models of the populations to be studied. For a neural network, for example, this amounts to training it with the synthetic data. The trained neural network can then be applied to real observable data to predict the risk factors of interest for the population. Knowledge of these risk factors will constitute a significant contribution to decision making, for example on allocation of resources to specific areas or people with certain characteristics.

Keywords: *Public health, decision support system, complex systems, multi-agent simulation, Bayesian networks*

1. INTRODUCTION

Public health surveillance is the ongoing systematic collection, analysis, interpretation, and dissemination of health data for the planning, implementation, and evaluation of public health action (Thacker & Berkelman 1988; Choi 2012). To achieve effective public health interventions, it is pivotal to analyse and interpret the vast amounts of data collected by surveillance systems to enable good understanding of all factors having an impact on health.

For example, we can consider child protection, which is an important public health issue in regard to the long term health issues of victims of child abuse and neglect. Often relatively extensive data exist on families in official statistics, research reports, social services reports, school and medical records, etc. However, these data are dispersed and hard if not impossible to relate and compare (Coates 2015). This leads to numerous interventions that are conducted without adequate knowledge of target families that are to benefit from these interventions. Inadequate knowledge also leads to lack of intervention where it is needed, such as undetected cases of child abuse. Better knowledge could also make it possible to flag down perpetrators or people who are likely to become perpetrators. In many cases the basic data that are needed for intervention decisions exist, but are not available to decision makers due to inadequate communication and lack of data integration, analysis and interpretation, as well as poor, inconsistent and conflicting inter-agency collaboration (Coates 2015; Taplin 2017).

Chronic condition management is another area where extensive disparate data exist from statistics and various health services and intervention agencies. In this area too, where there are numerous organizations offering services, there is a great need to better coordinate these services to achieve better outcomes for patients and also to reduce soaring costs of the healthcare system (Norris et al. 2002).

In these and other areas of public health data integration is important to take into account all available information for making the best possible policy decisions, but is hard to achieve in practice. It is difficult to match data on the same persons or phenomena from different sources. There can also be uncertainty about the reliability of data, data can be missing, and there can be frequent changes in data and data definitions, many different data formats. Much relevant information is only available in unstructured formats, such as case reports or clinical notes.

As there are complex relationships between factors influencing health, it is often not enough to do just a statistical analysis on the relationship between just a few factors. In addition, different levels of factors influencing health have to be taken into account, for example: country, region, population, locality, environment, group, family, individual, socio-economic factors and mental health.

To address all these considerations, there is a need for a multi-level methodology that can handle the complexity of the phenomena that are studied and take into account uncertainty in relationships and data.

2. THEORY AND METHODOLOGY

This section aims to highlight theoretical ideas that are useful for the development of a decision support system addressing the issues discussed in the introduction. It is not an exhaustive literature review, but introduces theoretical ideas and illustrates their application with some examples.

2.1. Complex Systems, Multi-Agent Systems and Hierarchy: Emergent Models

Complex systems can be modelled using multi-agent simulation. Evolutionary algorithms, and in particular genetic programming, are general-purpose algorithms to solve a variety of optimisation problems. In Stolk (2005) ideas from multi-agent simulation and from evolutionary algorithms are combined in a methodology to discover emergent macro-level regularities or patterns in simulations of complex systems. These macro-level regularities are models of the behaviour of macro-level agents, in other words emergent models. Therefore, the methodology is called the *Emergent Models* methodology. It defines ways to derive macro-level behaviour from micro-level properties and behaviour, and discovers models at the macro-level implied by those that describe the micro-level. This methodology can also be applied to the inverse problem of discovering micro-level behaviour of the composing entities of a complex system from data on its macro-level properties and behaviour. The Emergent Models methodology is a powerful methodology to study complex systems in general and systems in public health in particular.

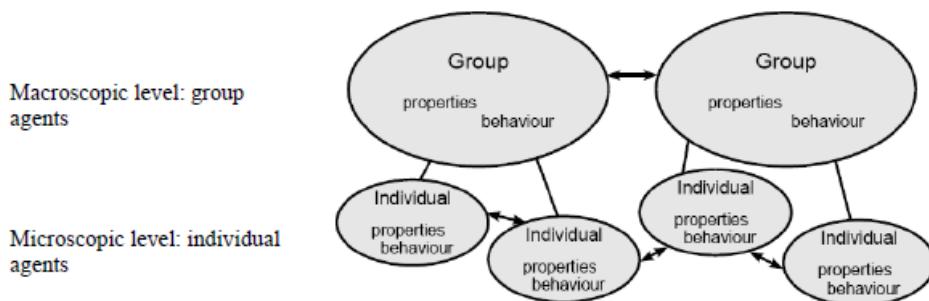


Figure 1. Levels in a multi-agent simulation.

Organizations are good examples of complex systems suitable for simulation using multi-agent approaches as described, for example, by Ferber (1999) and Wooldridge (2002). First, in an organization relatively autonomous entities (organizational units, individuals) with their own behaviour respond to environmental stimuli, as well as satisfying goals. Second, organizational units and individuals in an organization interact with each other and with processes in their environment. Agents can simulate this by appropriate communication capabilities. Third, organizations are structured hierarchically. Micro-level entities act together to constitute macro-level entities. Macro-level properties and behaviour are derived from micro-level properties and behaviour.

Thus, emergent models of group behaviour describe important phenomena in organizations, as they directly address the fundamental problem of emergence: how to derive properties and behaviour at the group level, or macro-level, from those at the individual level, or micro-level? Emergent Models is a systematic method to derive macro-level properties and behaviour and to obtain models on the level of group agents in computer simulations, as illustrated by Figure 1. Alternatively, the methodology can be applied to derive individual or subsystem characteristics, given desired whole system behaviour, in a way analogous to the derivation of genetic networks from whole organism behaviour described by Stolk (2005). In a decision support system desired behaviour is to be defined by users of the system: the system will assist in making optimal decisions to achieve that behaviour.

2.2. Complex Systems and Uncertainty: Flood Incident Management

An example of using emergent models for organizational systems is a proposal described by Stolk (2009) to model the Flood Incident Management (FIM) system of the British Environment Agency. The FIM system is a complex system. Complex systems can be modelled by multi-agent simulation. Therefore, the proposed methodology includes modelling FIM using multi-agent simulation. Risk is an important aspect of FIM. In the context of artificial intelligence much research has been done on Bayesian networks as a method to analyse uncertainty and risk, for example by Pearl (2009). Bayesian networks contain nodes that represent causes and effects with associated conditional probabilities. Bayesian decision networks are an extension of Bayesian networks particularly suited to decision making in circumstances of uncertainty. In addition to cause and effect nodes, they contain nodes that represent decision variables and nodes that represent objective variables. Decisions are evaluated by inference algorithms. Therefore, the proposed methodology includes the use of Bayesian decision networks to model risk and uncertainty. In the proposed methodology multi-agent simulation and Bayesian decision networks are integrated in a coherent conceptual framework for risk assessment in FIM.

The Environment Agency's Flood Incident Management (FIM) model describes processes at several levels: the FIM end-to-end process and activity diagrams at levels 1, 2 and 3. Part of the FIM model defines processes related to flood warning and response. A simplified representation of these processes is given in Figure 2. We see that FIM processes are described as interacting components at several levels of detail. The top-level end-to-end process is refined in successively more detailed sub-processes.

The Environment Agency's FIM Model can serve as a basis for a multi-agent model of the FIM process. Loosely, every box in a diagram of the model corresponds to an agent. The behaviour of an agent consists of: observing its environment; processing information obtained from observations; acting on its environment.

Bayesian decision networks are used in this multi-agent system to model probabilistic decision making by each agent. Actions by agents lead to results with associated probabilities. Probabilistic nodes of the diagram represent aspects of agent behaviour, including uncertainties in observations (for example, reliability of observed or communicated information), information processing (for example, cognitive mechanisms, reasoning errors, or faulty computing equipment) and actions (for example, faulty communication equipment, or physical obstruction). Decision nodes represent decisions made by the agents. The impact of uncertainty in the decisions by each agent on the behaviour of the whole system can be analysed by running multi-agent simulations.

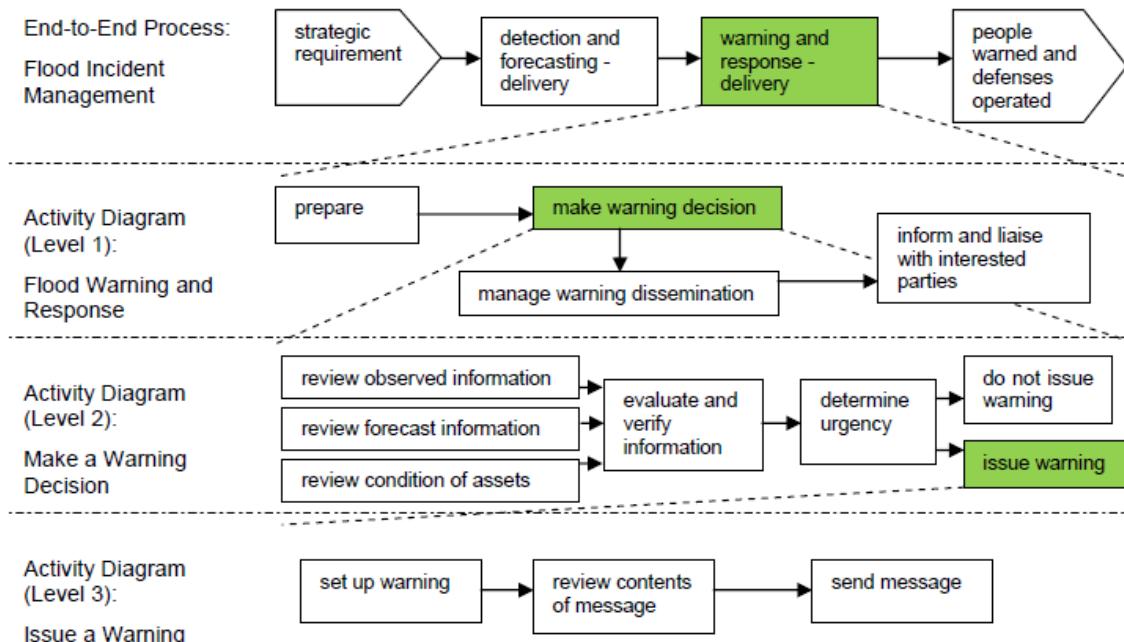


Figure 2. Flood incident management processes.

Flood incident management is a specific area where risk and uncertainty at multiple levels are very important in making decisions. Public health is another area where relationships between causal factors and health outcomes have a probabilistic nature and exist at different levels. Probabilistic and multi-level considerations need to be taken into account in modelling and decision making. The following example illustrates this.

2.3. Multi-level Bayesian Networks for the Analysis of Hierarchical Health Care Data

A method for multi-level modelling and learning combined with Bayesian network modelling with application to analysis of hierarchical health care data has been proposed by Lappenschaar et al. (2013). They call this method *multi-level Bayesian networks* and use it to analyse interactions and probabilistic dependencies between multiple diseases, using patient data obtained from multiple sources, such as family medical practices. Models studied are multi-level regression models and multi-level Bayesian networks for predicting either diabetes mellitus or heart failure using data grouped by practice, where the urbanity of the practice is modelled as higher level variable. It is assumed that urbanity is independent of the disease, given the practice.

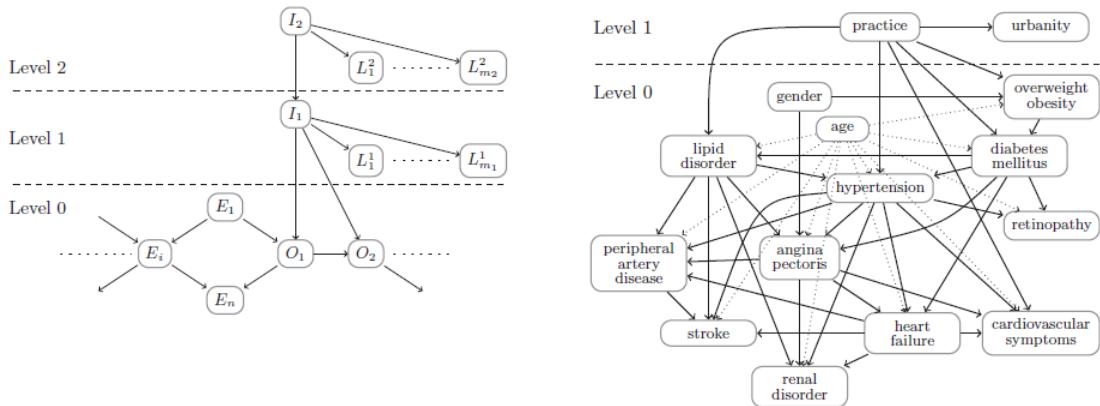


Figure 3. Multilevel Bayesian networks (Lappenschaar et al. 2013).

To build the network structure, Lappenschaar et al. (2013) make use of two approaches: model the structure manually based on existing medical knowledge; or learn the structure from data. For structure learning the constraints imposed by the multi-level Bayesian network can be captured by blacklisting and whitelisting edges, which can be incorporated into a wide range of structure learning algorithms. Blacklisting an edge between two variables means that the edge will never be included in the network structure learned by a structure learning algorithm, and whitelisting an edge means that it must always be included in the learned network structure. For example, the necessary edges between I_1 and all variables $O_i \in O$ are whitelisted, whereas edges from a lower level to a higher level are all blacklisted.

Multi-level Bayesian networks allow modelling of interactions between variables at the same level, such as different diseases or diseases and symptoms, and interactions between variables at different levels, such as influence of environmental factors on diseases (urban or rural situation of practice in the above example).

3. A DECISION INTELLIGENCE SYSTEM FOR PUBLIC HEALTH – DIONE

Building on the ideas in section 2, it is proposed to develop a software environment enabling the construction of complex systems models of public health issues, using a combination of multi-agent simulation and Bayesian networks at multiple levels of analysis. The multi-agent simulation environment will incorporate Bayesian network analysis and in particular the influence diagram extension of Bayesian networks (allowing decision support as a method suitable for a whole system approach to risk).

In influence diagrams qualitative information is reflected in probabilities. Relationships between objects, agents or functions are reflected in a network of nodes and relevance arrows. It is assumed that management decisions can alter the probabilities of component performance descriptors and that cost information of decisions can be included.

This section summarily describes functional requirements of an environment that can serve as a support tool for decision making in view of optimising decisions on public health issues.

DIONE is meant to be a flexible system to allow researchers to model their research problem and use data mining algorithms to extract meaningful information from data to solve their research problem and assist in making optimal decisions based on evidence from data. Relevant variables for the medical conditions to be studied and relevant data are defined by the researcher using the system. The researcher can also experiment with various machine learning algorithms to make inferences from the data to learn models that best describe the causal links with the medical condition of interest and possible interventions that can be effective.

3.1. Functional requirements related to system components

Understand, assess and model the relevant factors and their relationships in a public health issue:

- define all different system components, including possible causal factors and possible interventions
- enter data on uncertainty of system components as estimated conditional probabilities of factors given values of causal factors

3.2. Functional requirements related to the whole system

Understand and model the integration of key factors within a complex public health system:

- define how all system components are interrelated
- simulate system behaviour using definitions of system components and their interrelationships, as well as data on uncertainty

3.3. Functional requirements related to management decisions about the system

Understand, assess and model possible interventions to improve public health:

- understand risk factors and their relation to health outcomes
- define various options for possible interventions
- estimate costs and health improvement for the different options
- simulate the system with the different options
- evaluate which option is the best in terms of overall public health improvement in relation to cost

Enable different user categories to use the decision making tool with information available at their level. Data security and confidentiality are important and access is to be controlled dependent on authorization levels if different users.

4. DISCUSSION AND CONCLUSIONS

For public health interventions a practical approach to modelling of complex systems and uncertainty is required to model real systems. To realise such a practical approach it is recommended to:

- use multi-agent simulation to model in an intuitive way the systems involved; as seen in section 2, multi-agent simulation is particularly suitable for simulation of complex systems, as it offers a natural way to describe system components and their interactions and can be used effectively to simulate such systems in order to study their behaviour; a public health system can also be described in a very natural way as a system of interacting agents; in public health agents are decision makers acting in an environment of health conditions and causal effects on these conditions, information on which is obtained by analysing data on relevant exogenous and decision variables; therefore, multi-agent simulation can be used to gain insight in such a system;
- combine multi-agent simulation with Bayesian networks to capture risk and uncertainty; probability estimates can incorporate in a natural way expert judgements; as seen in section 2, Bayesian networks are a particularly relevant formalism for describing probability and risk aspects of a public health system, because they offer an intuitive way to define probabilities of different outcomes on the level of individual system components; these probabilities need not to be objectively defined, but can be based on the judgment of experts with practical knowledge of the system; and they offer a method to integrate component-level knowledge in a whole system level outcome.

We propose DIONE (Decision Intelligence for Organizations in Networked Environments), a decision support system that will offer a means to integrate and maximally exploit all available data to optimize intervention decisions. It will transform dispersed data into relevant information for decision makers. The proposed system has its place on top of existing research and data collection, rather than duplicating any existing research, and will optimize utilization of existing data and research results. Specific data to be

analysed are to be defined by users of the system, at least initially; it can also be imagined at a later stage that a researcher just defines a research problem and the system would use AI techniques to locate relevant data.

DIONE will enable more focussed health and social service interventions to encourage and help people to manage life stages and events, in view of achieving the best possible health and behaviour outcomes. The system will use data mining and analysis techniques such as predictive analytics, text analytics, data mining, machine learning and Bayesian statistics to integrate and exploit all available quantitative and qualitative data from heterogeneous data sources, such as statistics, social services, schools, medical services, etc., to identify risk factors and people most at risk. The system will provide an intuitive user interface to make it straightforward to define variables of interest, data to be analysed and data mining techniques to be used.

This will enable more focussed health and behaviour interventions. The problem with risk factors for health outcomes is that they are in general hard or impossible to observe directly. The proposed data mining techniques will be calibrated with synthetic data, obtained from agent-based simulation or micro-simulation models of the populations to be studied. For a neural network, for example, this amounts to training it with the synthetic data. The trained neural network can then be applied to real observable data to predict the risk factors of interest for the population. Knowledge of these risk factors will constitute a significant contribution to decision making, for example on allocation of resources to specific areas or people with certain characteristics.

REFERENCES

- Coates D. (2015). Working with families with parental mental health and/or drug and alcohol issues where there are child protection concerns: inter-agency collaboration. *Child & Family Social Work*, 22, 1-10.
- Choi B.C.K. (2012). The past, present, and future of public health surveillance. *Scientifica*, 2012, Article ID 875253, Hindawi Publishing Corporation, <http://dx.doi.org/10.6064/2012/875253>.
- Ferber J. (1999). *Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence*. Addison-Wesley, Harlow, UK.
- Harding N.J. (2012). *Application of Bayesian Networks to Problems within Obesity Epidemiology*. PhD Thesis, University of Manchester, Manchester, UK.
- Lappenschaar M., Hommersom A., Lucas P.J.F., Lagro J. and Visscher S. (2013). Multilevel Bayesian networks for the analysis of hierarchical health care data. *Artificial Intelligence in Medicine*, 57, 171–183.
- Norris L.S., Nichols P.J., Caspersen C.J., Glasgow R.E., Engelgau M.M., Jack L., Isham G., Snyder S.R., Carande-Kulis V.G., Garfield S., Briss P., McCulloch D. and the Task Force on Community Preventive Services (2002). The effectiveness of disease and case management for people with diabetes: A systematic review. *American Journal of Preventive Medicine*, 22, 15-38.
- Pearl J. (2009). *Causality: Models, Reasoning, and Inference*. 2nd ed., Cambridge University Press, New York, USA.
- Stolk J. (2009). *Complex systems simulation for risk assessment in flood incident management*. 18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation, Cairns, Australia 13-17 July 2009, <http://mssanz.org.au/modsim09>.
- Stolk H.J. (2005). *Emergent Models in Hierarchical and Distributed Simulation of Complex Systems*. PhD Thesis, University of Queensland, Brisbane, Australia.
- Taplin S. (2017). Prenatal reporting to child protection: Characteristics and service responses in one Australian jurisdiction. *Child Abuse & Neglect*, 65, 68-76.
- Thacker S.B. and Berkman R.L. (1988). Public health surveillance in the United States. *Epidemiologic Reviews*, 10, 164–190.
- Wooldridge M.J. (2002). *An Introduction to Multiagent Systems*. John Wiley & Sons, Chichester, UK.