Perspectives
Clin Ter 2016; 167 (5):e102-111. doi: 10.7417/CT.2016.1952

Health Data Entanglement and artificial intelligence-based analysis: a brand new methodology to improve the effectiveness of healthcare services

A. Capone[1,3], A. Cicchetti[2], F.S. Mennini[1,3], A. Marcellusi[3,4], G. Baio[5], G. Favato[1]

[1]*Institute for Leadership and Management in Health - Kingston University London, London, UK;* [2]*Department of Business Administration, Catholic University of Sacred Heart, Rome, Italy;* [3]*Economic Evaluation and HTA (CEIS- EEHTA) - IGF Department, Faculty of Economics, University of Rome «Tor Vergata», Rome, Italy;* [4]*Department of Demography, University of Rome «Sapienza», Rome, Italy;* [5]*Department of Statistical Science, University College London, London, UK*

Abstract

Healthcare expenses will be the most relevant policy issue for most governments in the EU and in the USA. This expenditure can be associated with two major key categories: demographic and economic drivers. Factors driving healthcare expenditure were rarely recognised, measured and comprehended. An improvement of health data generation and analysis is mandatory, and in order to tackle healthcare spending growth, it may be useful to design and implement an effective, advanced system to generate and analyse these data. A methodological approach relied upon the Health Data Entanglement (HDE) can be a suitable option. By definition, in the HDE a large amount of data sets having several sources are functionally interconnected and computed through learning machines that generate patterns of highly probable future health conditions of a population. Entanglement concept is borrowed from quantum physics and means that multiple particles (information) are linked together in a way such that the measurement of one particle's quantum state (individual health conditions and related economic requirements) determines the possible quantum states of other particles (population health forecasts to predict their impact). The value created by the HDE is based on the combined evaluation of clinical, economic and social effects generated by health interventions. To predict the future health conditions of a population, analyses of data are performed using self-learning AI, in which sequential decisions are based on Bayesian algorithmic probabilities. HDE and AI-based analysis can be adopted to improve the effectiveness of the health governance system in ways that also lead to better quality of care. ***Clin Ter 2016; 167(5):e102-111. doi: 10.7417/CT.2016.1952***

Key words: *Data entanglement, GDP, health governance, healthcare spending, self-learning artificial intelligence*

Introduction

Although many countries are currently trying to constrain or stabilize it, over the last two decades healthcare spending has been rising faster than the national GDP itself (1-3). Looking towards the future, according to a medium and long-term perspective, healthcare expense will be one of the most relevant policy issues for most governments in the EU and the USA (4). Although the growth has not been perfectly linear (indeed, economy and healthcare spending commonly grow at different rates), the share of the economy spent on healthcare in the USA (5) has quadrupled in 2009 compared to 1929. During the recession in 2008, the GDP growth was depressed concededly, but healthcare spending growth was not nearly as dampened. The concerns raised by projections are not simply academic: they have a real impact upon the wellbeing of the citizens and hence, upon the economy of nations.

In this article two critical issues are investigated: the growth of healthcare spending (including possible influencing factors); and also an advanced system to evaluate and tackle healthcare spending growth, which is actually a brand new methodology designed to improve the effectiveness of national health services. Hence, this paper is neither a systematic literature review nor an assessment of new studies. Rather, the aim of the current speculative research is to report the authors' opinions about basic principles and advanced future procedures that will help public health authorities make proper decisions and achieve national health goals.

The growth of healthcare spending

Despite the inevitable uncertainty surrounding cause-effect relationships, the public healthcare expenditure is associated with two major key categories: demographic and non-demographic (economic) drivers (Fig. 1) (4). Demographic drivers are directly correlated with the dynamic variations of population structure (basically the ageing process of a population). Over the next years, in all countries a continuous increase in age will be observed, leading to unprecedented shares of the population being 70-80 years and older. The greater the life expectations, the higher the demand for healthcare for a population, especially in elderly people with probable cognitive impairment and other chronic disabilities (6). This will have a further impact on social security as well as social and healthcare spending. However,

Correspondence: Dr. Alessandro Capone, Institute for Leadership and Management in Health, Kingston University London, Kingston Hill, KT2 7LB Kingston-upon-Thames, UK. Tel: +39347 - 5158247; Fax: +44 (0) 2084175026. E-mail: a.capone@kingston.ac.uk

ISSN 0009-9074

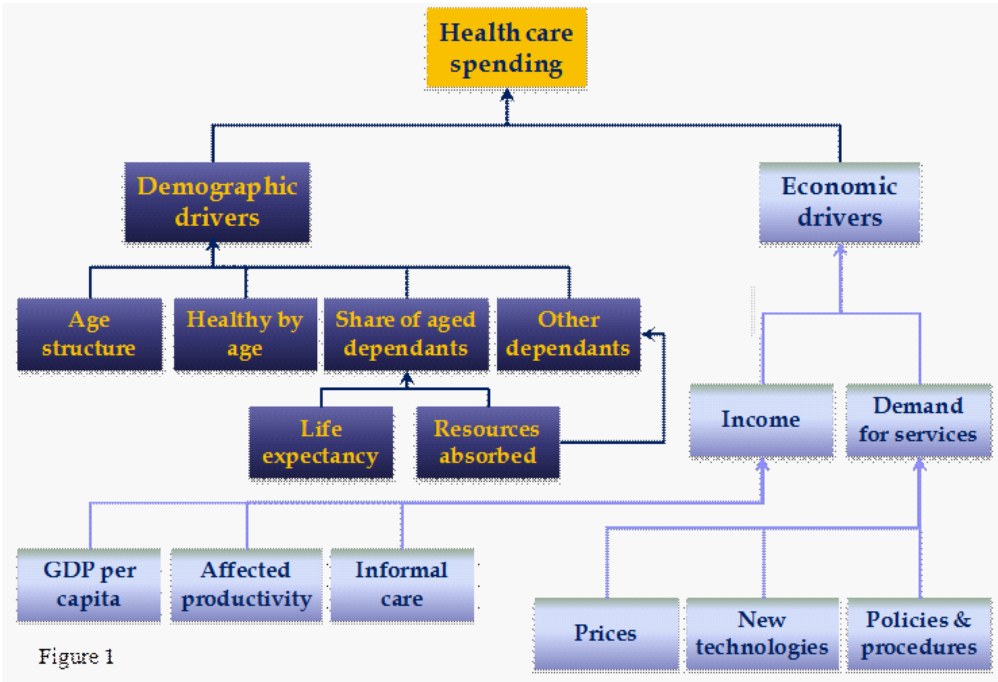

Fig. 1. Determinants of healthcare spending. Modified from the version by De La Maisonneuve et al. (4).

the ageing of the population cannot explain the whole story. The relationship between health expense and income is still a controversial issue (4). Indeed, independently of the precise income elasticity, the combined effect of demographic variations and income do not provide a comprehensive explanation for the health spending growth.

Some obvious factors can be mentioned: social preferences, public health standards, subsidy systems, taxation systems. Although the equation describing this association is not conclusive, some basic factors such as technological advancement, health policy decisions (i.e. access requirements and procedures) and prices are the most likely predictors for explaining the relationship between healthcare spending and income (4). Moreover, considering the deterioration in today's economic environment, it's worth mentioning the potential relationship between determinants of contribution to GDP and consumption of economic resources due to healthcare demand (Fig. 2). While economic determinants

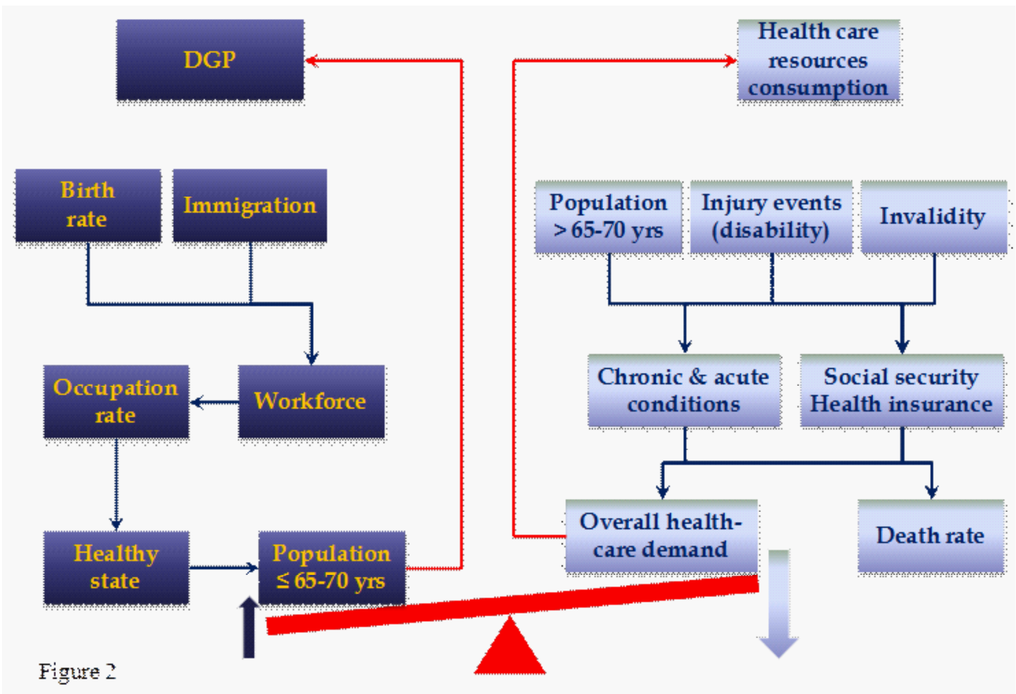

Fig. 2. Unbalanced power relationship between GDP contributors and economic resources absorbed by healthcare demand.

may produce different GDP results with a remarkable level of uncertainty, healthcare needs seem to be more easily predictable in consuming these resources (when suitable data exist, of course). Clear and comprehensive disclosures of uncertainties concerning GDP growth rate and, most importantly, the collection and analysis of that data driving an expected future risk to consume an increasing amount of healthcare resources are critical in this situation. The wise use of healthcare resources becomes a significant issue: in the future the GPD growth rate will be much less likely to result in higher living standards (quality of life or quality of services delivered). The burden of supporting an ageing society and the expanding cost of healthcare will take out a large proportion of workers' productivity improvements for years to come.

Health data and the economic assessment of health interventions

The availability of new highly targeted drugs and advanced health technologies produced insurmountable improvements in terms of health units gained worldwide, with a consequent average increase of life expectancy (7). However, the majority of advances in medical technology comes at higher prices compared with the current ones (8). As a result, new costs must be added to the healthcare funding for previously unavailable diagnostic procedures and treatments. So, in this turbulent economic time, health authorities and other bodies are increasingly requiring economic and clinical evidences as cost analysis and cost containment are becoming a central issue in each healthcare system. Nowadays, economic evaluations are routinely used in many countries to assess new health technologies and make decisions on pricing and reimbursement. The overall economic assessment should lead to a powerful enhancement in the quality of healthcare services delivered.

The cost-effectiveness and the economic sustainability of health interventions are commonly measured by means of parameters such as the cost per QALY gained, whose values should be included in a range of predetermined thresholds. QALY is used by major regulatory agencies (i.e. FDA, EMA, NICE) as an integral part of health technology assessment methods and procedures (9). Unfortunately, social costs are not systematically taken into account by regulatory agencies, even during the assessment of reimbursement procedure for new costly health technologies. Naturally, it can be argued that since utilities attributed to the recovery or preservation of a healthy state are closely related to a specific population (10, 11), values of QALY gained can be different between countries and hence international comparisons are not always performed or actually usable. However, the collection of data having a great impact like indirect costs (i.e. social security spending and the productivity lost due to absenteeism) cannot be neglected, especially in this moment. First of all, the reduction of absenteeism at the workplace can improve or recover the productivity and the contribution to the GDP growth rate. Second, a decline of presenteeism (a hidden measure of lost productivity due to workers' being on the job but not fully performing because of illness or other medical conditions), can generate an optimization of production and/or service costs. As a consequence, the quality of workforce's life and above all the industry's competitiveness profile can be substantially enhanced. An effective strategic advantage in a sluggish economy, especially bearing in mind that presenteeism may actually be a much costlier problem than absenteeism (approximately 3-10 times higher) (12, 13). Moreover, it's worth mentioning that presenteeism is quite common in tough economic times, when employees may be extremely afraid of losing their jobs.

At present time, there are several sources of health data: observational studies, claims databases, registries, PRO and other medical record linkage systems which can show different reliability and controversial degree of credibility from the regulatory agencies' perspective (14). Despite plenty of health data, a majority of them commonly show some caveats (15, 16) that may affect the credibility of some health assessment procedures (Table 1).

Health data entanglement and artificial intelligence-based analysis

In several health databases important pieces of information are still missing. As a result, available surrogate information cannot effectively inform probabilistic models used for better strategic decision making. The uncertainty associated with these data may undermine the effectiveness of any decision in public health governance. An improvement of health data generation and analysis is mandatory. Even in countries with a remarkable GDP growth rate like the United Arab Emirates, a recent study revealed that several health issues deserve to be accurately addressed (17). As an emerging suggestion, authors reported that reliable and valid longitudinal data must be considered as essential for planning population-based health programmes (17).

The increase of healthcare budgets in response to growing demand without its assessment and governance is no longer a viable option. Despite its magnitude in the overall economy, the key elements that drive healthcare expenditure have been rarely recognized, measured and comprehended. Therefore, to tackle healthcare spending growth, the design and implementation of an effective, advanced system to generate and analyze real and qualified health data may be advised. This system should allow the achievement of the following aims: a) measure the economic effectiveness of the current health policy procedures; b) estimate the value of health governance decisions and calculate the proper health funding; c) identify priorities and guide the implementation of corrective policies in order to improve the overall health state of a population with an economically sustainable use of resources. To this purpose, the methodological approach relied upon the HDE seems to be a suitable option. This approach is based on a multifaceted mix of technology, statistics, computer science, and medical knowledge. By definition, in the HDE a large amount of data sets having several sources are functionally interconnected and computed through learning machines that generate patterns of highly probable future health conditions of a population.

Table 1. Relevant limits of health data (15, 16).

Domain	Issue
Diagnosis	Diagnostic procedures that have not the recommended validation of medical associations
	Limited diagnostic information for a proper patient inclusion
	Healthcare needs
	Identification of appropriate funding
Completeness	Restrictions on access to drugs
	Intermediate outcomes
	Real measure of effectiveness
	Treatments delivered outside the considered system
	Health system structure
	Insurance coverage
	Referral patterns
Risk factors	Genetic susceptibility (genetic predisposition to diseases)
	Lifestyle habits and behavioural factors
	Environmental influences
Accuracy	Coding (i.e. classification of diseases)
	Subpopulation
	Sample size
Costs	Direct costs are not systematically gathered
	Management costs due to side effects are often not accounted for
	Indirect and social costs are sporadically collected
	Costing data to perform a complete budget impact assessment are often missing
Timeframe	Period of data collection
	Different follow-ups

Entanglement concept is borrowed from quantum physics and refers to the presence of correlations (i.e. inseparable interconnections) between observable physical quantities (in this case, wide health variables that are not necessarily resident or generated in the same place). Briefly, quantum entanglement means that multiple particles (or information) are linked together in a way such that the measurement of one particle's quantum state (in our more modest context, individual health conditions and related economic requirements) determines the possible quantum states of other particles (i.e. population health forecasts exploited to predict the overall economic impact and drive policy decisions). The value created by the HDE is based on a wider and more combined evaluation of clinical, economic and social effects generated by health interventions. Such an assessment is performed according to a predictive analytics approach (several interconnected variables can be considered). A comprehensive appraisal of clinical value, economic value, social value and system of governance value of health interventions is carried out per each beneficiary and patient subpopulations (Fig. 3). As a result of these specific analyses, public health decisions are data-driven and resources can be effectively used. Data can be generated, collected and stored by the means of several technological systems [including innova-

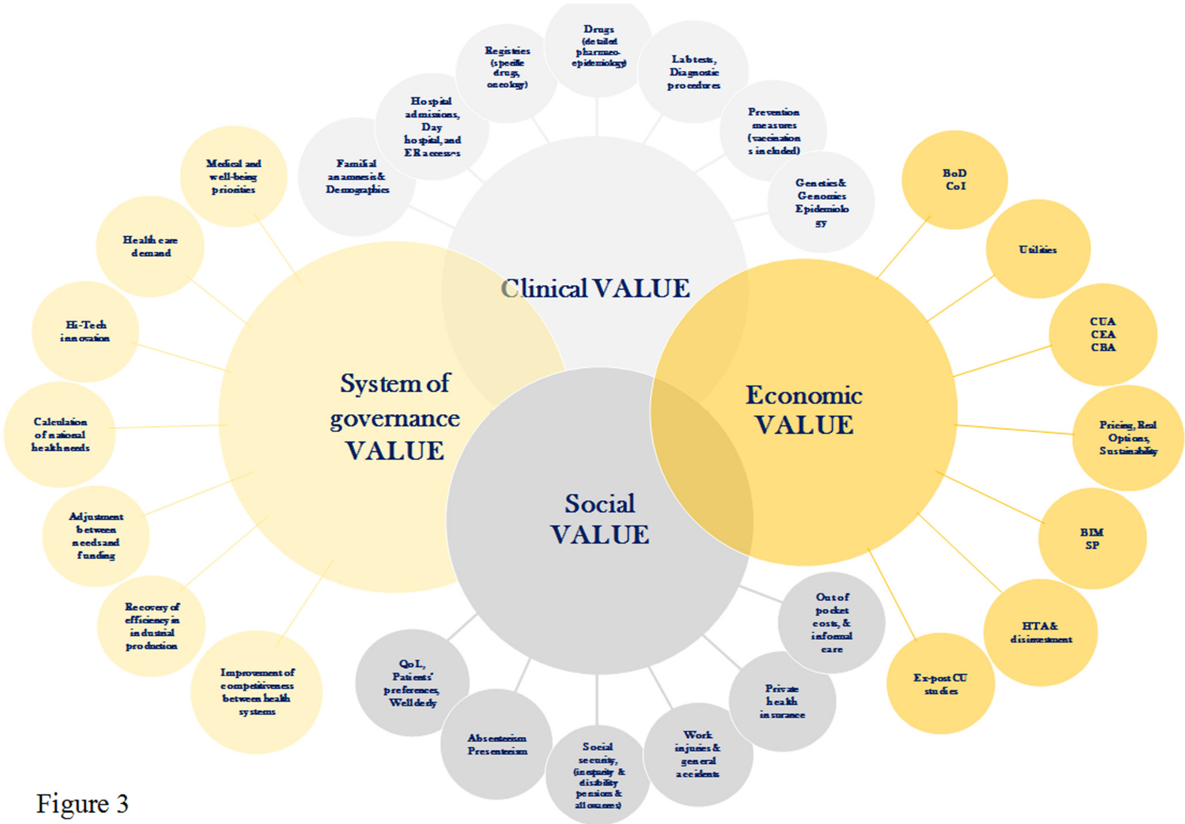

Figure 3

Fig. 3. Source of data entangled in the HDE. HTA - Health technology Assessment; Wellderly - Wellness of elderly people

tive smart devices (i.e. watches and phones) advanced web technologies, electronic cards, remote patient management such as telemedicine and telecare, or biosensors]. The remote patient monitoring is already more than a promising cost-effective tool to help deliver sweeping improvements in healthcare (18). Indeed, the innovative healthcare services delivered this way is supposed to reduce disease management programme costs while preserving, but it is much more likely to enhance, the quality of care provided. Beyond the fact that the entanglement approach provides a unique interconnection of qualified data to support health decisions, it is worth noting that analysis of data is performed using «self-learning AI», in which sequential decisions are based on Bayesian algorithmic probabilities. The approach is based on a direct approximation of a cybernetic system called an AIXI[1] agent. AIXI can be seen as a mathematical definition of AI. According to M. Hutter (19), AIXI is the most intelligent unbiased agent possible, used to formally solve a number of problem classes, including sequence prediction, strategic games, function minimization, reinforcement and supervised learning. Apparently, the major limit of the AIXI model is that it can be incomputable. To overcome this problem, Hutter developed a modified algorithm AIXI*tl*, which is still effectively more intelligent than any other time t and space l bounded agent (20). Whether U $(q, \alpha_1\alpha_2 \ldots\ldots \alpha_n)$ denotes the output of a universal machine U, informed by programme q and input $(\alpha_1\alpha_2 \ldots\ldots \alpha_n)$, $m \in N$ is a finite look ahead at the horizon, and $l(q)$ is in the amount of bits in the programme, once the cybernetic agent has picked the actions $(\alpha_1\alpha_2 \ldots\ldots \alpha_{t-1})$ and received the sequence of observations-reward pairs $o_1 r_1 o_2 r_2 \ldots\ldots o_{t-1} r_{t-1}$ from the environment, AIXI is equal to (1) (21):

$$a_t^* = \arg\max_{a_t} \sum_{o_t r_t} \ldots \max_{a_{t+m}} \sum_{o_{t+m} r_{t+m}} [r_t + \cdots r_{t+m}] \sum_{q:U(q,a_1 \ldots a_{t+m}) = o_1 r_1 \ldots o_{t+m} r_{t+m}} 2^{-l(q)}$$

Therefore, according to equation (1), AIXI considers the amount of the total rewards over all possible futures up m to steps ahead, weighs each of them by the complexity of programmes consistently with data collected in the past, and then picks the action that maximises expected future rewards. Assuming that the agent does not initially know the true environment (healthcare data must be progressively collected), it is suggested the development of models whose predictive performance improves as the agent gains experience. A way to provide such a model is to take into account the Bayesian perspective. According to Veness et al. (21), instead of committing to any single fixed environment model, the cybernetic agent uses a mixture of environment models. This implies a model class (a class of possible environments), the designation of an initial weight to each possible environment (the prior), and finally the updating of the weight for each model (computing the posterior) each time more experience is obtained. So, given a countable model class $M := \{\rho_1, \rho_2, \ldots\}$ and a prior weight $w_o^{\rho} > 0$ for each $\rho \in M$ such that, $\Sigma_{\rho \in M} w_o^{\rho} = 1$ the mixture environment model is given by (2):

$$\xi(x_{1:n}|a_{1:n}) := \sum_{\rho \in M} w_0^{\rho} \, \rho(x_{1:n}|a_{1:n})$$

[1]AIXI or super artificial intelligence is basically a simple concept. It combines a search strategy over all possible futures of an agent (a math operator using observations, rewards and actions) with a weight that favours thriftiness (in the sense of the complexity of a programme reaching this future, its sequence of actions and observations).

Hence, considering equation (1), the universal Bayesian agent can be formally expressed as:

$$a_t^* = \arg\max_{a_t} \sum_{o_t r_t} \dots \max_{a_{t+m}} \sum_{o_{t+m} r_{t+m}} [r_t + \cdots r_{t+m}] \sum_{\rho \in M} 2^{K(\rho)} \rho(x_{1:t+m}|a_{1:t+m})$$

where $\rho(x_{1:t+m}|a_{1:t+m})$ is the probability of observing $x_1, x_2 \cdots x_{t+m}$ given actions $a_1, a_2 \cdots a_{t+m}$; class M_U consists of all enumerable chronological semimeasures (19), which includes all computable ρ; and *K* (ρ) denotes the Kolmogorov complexity (22) of ρ with respect to *U*.

The direct AIXI approximation is a Bayesian optimality notion for universal reinforcement learning agents in unknown or partially known environments (19). Reinforcement learning is a general and influential paradigm for agents that learn from experience (23). According to J.S. Albus (24) the AI can be defined as « . . . the ability of a system to act appropriately in an uncertain environment, where appropriate action is that which increases the probability of success, and success is the achievement of behavioural subgoals that support the system's ultimate goal.» Therefore, to find out proper corrective actions and achieve a better health governance system, a Bayesian approach and Monte Carlo simulations must be included as a part of an AI self-learning cycle (Fig. 4) (21). The AIXI agent interacts with the environment (i.e. healthcare setting) in cycles. In each cycle, the agent executes an action and in turn receives observations and rewards. The only information available to the agent is the history of previous interactions (analysis of data previously generated and stored). The basic reinforcement learning problem is to build up an agent that collects as much reward as possible (expected positive futures) from a partial known environment over time (21). In other words, a sequence of phases called *learning* (in which future observations and rewards are predicted on the basis of past experiences), and *planning* (where the best future actions are determined and possibly implemented). The use of AI in health governance decisions is certainly a compelling application, and some university research centres are implementing this brilliant and effective application in healthcare setting (25, 26).

There is a further issue that deserves to be analyzed: why this approach should be adopted. The HDE and AI-based assessment provide an unprecedented world of information. First of all, this approach allows one to produce scenarios and forecasts concerning the effect of policy decisions on the health conditions of a population definitively and more accurately than anything else at the present; this is because it may consider the effect of many more variables (accurately measured) at the same time. No other system can claim to produce similar results. Large databases can provide interesting answers and uncover several potential factors influencing outcomes in the disease management of a given condition. However, there are also many biases associated with large databases and as a consequence a lot of uncertainty can be seen. This is basically due to the fact that one person's data usually are in different places (15). Additional key players in generating biases are factors

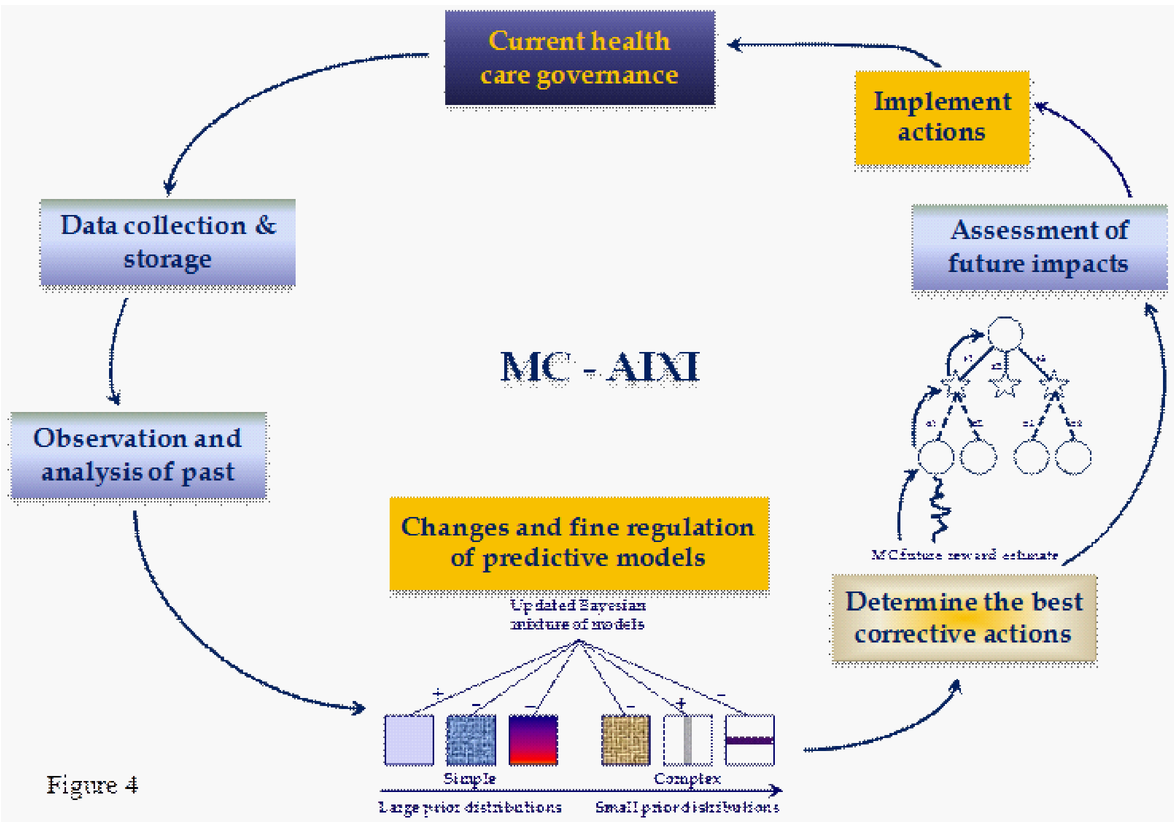

Fig. 4. Self-learning cycle of artificial intelligence. MC = Monte Carlo simulation. AIXI = artificial intelligence agent. Modified from Veness J. et al. [21].

driving the collection, coding, and preservation of the data; the extensive customization of different systems that collect similar data; the fragmentation of the healthcare delivery system and its records; and finally privacy and proprietary considerations (15). When we have to deal with huge amounts of information that are not specifically coupled to each other and/or not univocally related with individual patients, one can only piece together some pattern and on this basis make relatively reliable predictions of the future. That's why we wonder whether medicine can learn anything from the current experiences of Amazon or Netflix. They can make interesting customer profiles, improve strategic business decisions, allowing faster decision-making processes, and ultimately gain a competitive advantage. However, big databases (especially those relative to health) offer several technical and technological challenges. Due to its massive volume (from terabytes to exabytes), complexity and unstructured nature, big databases require advanced data storage, extremely high computational performance, as well as analysis and visualization techniques (27). Indeed, there still remain concerns that the statistical results for this observational data analysis may not be revealing valid causal pathways (28). In order to analyse big databases, some authors believe that familiar statistical approaches cannot be applied and therefore, analysts should explore new methods and techniques (29). So, at the moment, how is it possible to make proper health policy decisions and allocate economic resources to face the future health conditions of a population using big health databases?

The HDE provides an in-depth portrait of the national health condition performing a unique combination of macrodata (economic and social assessment) with micro or nanodata (clinical analysis). Actually, health-related data typically consist of information generated by various entities such as healthcare providers, healthcare encounters, social security, insurance companies and so on. Hence, the core of information commonly shows a set of attributes. All health-related data sources must be unequivocally associated and harmonized by bringing different data sets together. If there is a unique identifier (achieved by the means of smart devices, web technologies, health insurance electronic cards, biosensors or other) available for virtually all the individuals, then it is possible to have qualified data to perform a perfect data entanglement. When data elements are entangled, they can be considered as inferentially equivalent, and can thus be compared and used meaningfully in statistical analysis or even better processed with AI. The entanglement procedure therefore enables data integration; an approach that should be implemented for different scopes, including the necessity to obtain an economic assessment of resources used or to allocate the appropriate healthcare funding, and finally to investigate relatively rare events (i.e. orphan diseases). As a first step, the combination of macrodata is achieved considering individual and subpopulation data, then the AI predicts the key features of future conditions, health needs, funding, and proper policy resolutions to maintain and improve the health status of the population. With the HDE, fine-grained targeted data (namely nanodata - multiple daily measurements of blood pressure or glucose blood test are common examples) can be collected, allowing the understanding of individual patient needs (as well as patient population needs) and make appropriate decisions in healthcare management. This also gives the chance to use these data to inform comprehensive probabilistic models. Indeed, the full integration of fine-grained data and the use of the AI allow the identification of the most appropriate, effective and specific solutions pertaining to the given context. These health governance solutions take into account the evolution of health conditions, distinctively predicted in that population. The historical health data gathered with the HDE are more complete compared with the currently available systems. This is also true when a potential comparison with a large set of health data (the so called big databases) is considered. Indeed, the nature itself of the healthcare industry contributes to limit the usefulness and potential feasibility of big databases. The sharing of data among different functions and players is crucial, and creates challenges as well; so, important pieces of information often remain overlooked due to a lack of procedures in organizations for the full integration of data.

Outputs and solutions that come from the HDE implementation can be grouped in major clusters as reported in table 2 (see also figure 3). Estimates produced with this new system are different and have nothing to do with findings derived from health economic methods. The level of uncertainty surrounding outputs (futures) is expected to be extremely low or virtually removed. When machine learning is considered, the environment is commonly simulated with Markov decision process (sometimes with Bayesian

Table 2. The HDE outputs and solutions

Cluster	Output
Actual economic and clinical assessment	Multi-value based assessment of health interventions
	Health technology assessment
	Healthcare needs
	Identification of appropriate funding
Predictive simulators	Budget impact models
	Spending predictor models
	Disease models
	Epidemiological models
Self-learning data analyses	Computational profiling of national and regional health expenses
	Predictive behavioural prescriptive analysis
	Predictive treatment approach analysis
	Genetic-genomic profiling of disease risk and management (genomics)
Health governance models	Comprehensive effects of current and future health policy decisions including demographic dynamics of population (ageing)

networks) since reinforcement learning algorithms use dynamic techniques. The most relevant difference between the conventional techniques and the HDE (reinforcement learning algorithms or AIXI) is that the latter does not need knowledge and can always be utilized, even when other methods (such as the above mentioned) requiring exact or formally correct information (often missing) become infeasible (30). Naturally, some techniques such as cost effectiveness or cost utility analysis will continue to have their role in economic assessment; especially, when a new health intervention must be evaluated. Initially, conventional methodologies of economic assessment and HDE would be considered complementary since some outputs of HDE can be used to inform probabilistic economic models. Since the generation of individual health data is a time-dependent process, while the HDE is in its early stage, some data will be drawn from different sources having their own diversity, complexity and timeframe. A sort of supplementary approach dedicated to assimilating knowledge from biomedical data. Over time, the more HDE is used, the more specific and well-structured health information will be generated and available for self-learning AI analyses.

At the present, as far as we know, no other system should be capable of indicating with a comparable accuracy how to establish priorities (ex-ante assessment) and consequences (ex-post assessment) of health policy decisions. The HDE and the AI-based assessment are a unique tool because they are used to protect the health of the population (considering the demographic dynamics and epidemiology of the most important diseases), and the wellness of citizens (probably through a contribution in the achievement of the expected GDP growth).

Health data protection

Since the collection of sensitive health data and personal information are captured, the implementation of all requirements to assure data protection is absolutely mandatory and it must be taken into account in advance. Naturally, the patient's right to personal data privacy and the possibility to perform scientific research should be considered and balanced. Anonymisation of patient records, informed patient consent and data aggregating can be considered like fundamental rules to respect the patient's basic right to privacy. In recent years, privacy and security procedures have been evolving to keep up with the remarkable changes induced by the technological evolution. In the USA, health authorities modified some privacy and security rules associated with the electronic transmission of health information introducing the Health Information Technology for Economic and Clinical Health Act in 2009 (31).

In 2012, the EU released an updated version of the European General Data Protection Regulation (32) in which privacy and security requirements have been significantly strengthened compared to previous documents published in 1995 (33). Indeed, in order to produce innovations that are important for public health and useful in practice, research has to comply with generally accepted ethical, legal and administrative principles. The HDE is based on an advanced method evaluating large amounts of patient information. Therefore, its implementation would obviously need to also comply with data protection legislation, particularly in relation to access to and the safekeeping of the data. However, an additional peculiarity of the HDE approach is that the data storage system can be entangled as well. From an information technology perspective, data entanglement was initially suggested as a mechanism for increasing censorship resistance in document-storage systems (34, 35). Now entanglement is used to protect the data from an untrusted storage provider that might be tempted to damage or destroy the data through negligence or malice (36). When two files are entangled they are somehow linked together, and entanglement method provides a reliable settlement between strong robustness, security, pragmatism, and efficiency.

Conclusions

The burden of healthcare costs will continue to grow unless and until the efficiency and efficacy of healthcare systems will be accomplished. It is critical that policymakers take action to restrain the rising costs of healthcare today. Obviously, we have to bear in mind that the general aim is a continued public health improvement that will also help eliminate health gaps, as the relative position of vulnerable population groups gets better. HDE and AI-based analyses can be adopted to improve the effectiveness of health governance systems in ways that also lead to better quality of care.

Acknowledgements

The authors wish to acknowledge the support and relevant contributions of Professor Riccardo Torlone (Department of Informatics and Automation, University «Roma 3», Rome, Italy) who was also involved in the peer review of this article. Moreover, we would like to thank Nathan Spruit for his critical reading of this paper and valuable comments for improving the manuscript.

Conflict of interest

The authors declare that no competing interests exist. A. Capone, the corresponding author of this manuscript, confirms that the conflict of interest statements have been approved by all co-authors.

Abbreviations

AI - Artificial Intelligence
EMA - European Medicines Agency
EU - European Union
FDA - Food and Drug Administration
GDP - Gross Domestic Product
HDE - Health Data Entanglement
NICE - National Institute for health and Care Excellence

PRO - Patient Reported Outcomes
QALY - Quality-Adjusted Life-Years
USA - United States of America

References

1. Organisation for Economic Co-operation and Development (OECD). Economic outlook, analysis and forecasts. Economic Outlook. http://www.oecd.org/eco/outlook/economicoutlook.htm. Last accessed on June 2015
2. Organisation for Economic Co-operation and Development (OECD). Health Statistics 2013. http://www.oecd-ilibrary.org/social-issues-migration-health/data/oecd-health-statistics/system-of-health-accounts-health-expenditure-by-function_data-00349-en;jsessionid=5u1q6oooskltj4.x-oecd-live-01?isPartOf=/content/datacollection/health-data-en. Last accessed on June 2015
3. Organisation for Economic Co-operation and Development (OECD). Health at a Glance 2013. OECD Indicators, OECD Publishing. http://dx.doi.org/10.1787/health_glance-2013-en. Last accessed on June 2015
4. De La Maisonneuve C. and Oliveira Martins J. A projection method for public health and long-term care expenditures. Economics Department Working Papers No. 1048, OECD 2013. Full text available for download from the OECD web site: http://www.oecd.org/eco/growth/Health%20FINAL.pdf. Last accessed on June 2015
5. Conover C. American Health Economy Illustrated. Medical Industry Leadership Institute Open Education Hub. Collection structure revised: June 21, 2014. Ever-growing health share of economy: page 7. Full text available for download from the web site: https://hub.mili.csom.umn.edu/content/col10021/1.5/pdf. Last accessed on June 2015
6. Organisation for Economic Co-operation and Development (OECD). OECD work on health 2013-2014. Full text available for download from the OECD web site: http://www.oecd.org/health/health-brochure.pdf. Last accessed on June 2015
7. Olshansky SJ, Goldman DP, Zheng Y, Rowe JW. Aging in America in the twenty-first century: demographic forecasts from the MacArthur Foundation Research Network on an Aging Society. Milbank Quarterly 2009; 87(4): 842-862. doi: 10.1111/j.1468-0009.2009.00581.x.
8. Sorenson C, Drummond M, Burns LR. Evolving reimbursement and pricing policies for devices in Europe and the United States should encourage greater value. Health Aff 2013; 32(4): 788-796. doi: 10.1377/hlthaff.2012.1210
9. National Institute for Health and Clinical Excellence (NICE). Guide to the methods of technology appraisal (reference N0515). Full text available for download from the web site: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/191504/NICE_guide_to_the_methods_of_technology_appraisal.pdf. Last accessed on July 2015
10. Arnesen T, Trommald M. Roughly right or precisely wrong? Systematic review of quality-of-life weights elicited with the time trade-off method. J Health Serv Res Policy 2004; 9: 43-50
11. Mennini FS, Panatto D, Marcellusi A, et al. Time Trade-Off procedure for measuring health utilities loss with Human Papillomavirus–induced diseases: A multicenter, retrospective, observational pilot study in Italy. Clinical Therapeutics 2011; 33: 1084-95
12. Stewart WF, Ricci JA, Chee E, et al. Cost of lost productive work time among US workers with depression. JAMA 2003; 289(23): 3135-44
13. Hemp P. Presenteeism: at work - but out of it. Harv Bus Rev 2004; 82(10): 49-58
14. Garrison LP, Neumann PJ, Erickson P, et al. Using real-world data for coverage and payment decisions: The ISPOR Real-World Data. Task Force report. Value Health 2007; 10: 326-35. Full text available for download from the ISPOR web site: http://www.ispor.org/workpaper/RWD_TF/ISPOR-RealWorldDataTaskForceReport.pdf Last accessed on July 2015
15. Curtis LH, Brown J, Platt R. Four health data networks illustrate the potential for a shared national multipurpose big-data network. Health Affairs 2014; 33(7):1178-86
16. Hersh WR, Weiner MG, Embi PJ, et al. Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care 2013; 51(8 0 3): S30-S37 doi:10.1097/MLR.0b013e31829b1dbd
17. Loney T, Aw TC, Handysides DG, et al. An analysis of the health status of the United Arab Emirates: the «Big 4» public health issues. Glob Health Action 2013, 6: 20100. http://dx.doi.org/10.3402/gha.v6i0.20100
18. Dharmar M, Kuppermann N, Romano PS, et al. Telemedicine consultations and medication errors in rural emergency departments. Pediatrics 2013; 132: (6)1090-7
19. Hutter M. Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability. Springer 2005
20. Hutter M. A gentle introduction to the universal algorithmic agent AIXI. Technical Report IDSIA-01-03, Manno-Lugano, Switzerland, 2003. Full text available for download from the ResearchGate's web site: http://www.researchgate.net/publication/228851452_A_gentle_introduction_to_the_universal_algorithmic_agent_AIXI. Last accessed on July 2015
21. Veness J, Ng KS, Hutter M, et al. AIXI Approximation. Journal of Artificial Intelligence Research 2011; 40:95-142
22. Li M, and Vit´anyi P. An introduction to Kolmogorov complexity and its applications (Third edition). Springer 2008
23. Sutton RS, and Barto AG. Reinforcement Learning: An Introduction. MIT Press 1998
24. Albus JS. Outline for a theory of intelligence. IEEE Trans. Systems, Man and Cybernetics 1991; 21(3): 473-509
25. McCormick TH, Rudin C, Madigan D. Predicting Medical Conditions with Bayesian Hierarchical Rule Modeling. Proceedings of the 2011 INFORMS Data Mining and Health Informatics (DM-HI) Workshop. Full text available for download from: http://web.mit.edu/rudin/www/INFORMS2011DataMiningAndHealthInformaticsWorkshopProceedings.pdf. Last accessed on July 2015
26. Tulabandhula T, Rudin C. On Combining machine learning with decision making. Machine Learning (ECML-PKDD journal track), Online First, June, 2014. Full text available for download from: http://arxiv.org/pdf/1104.5061v2.pdf. Last accessed on July 2015
27. De Mauro A, Greco M, Grimaldi M. What is Big Data? A consensual definition and a review of key research topics. AIP Conference Proceedings 2015; 1644: 97-104
28. Atkinson G. Big data - What is it and what use is it? Journal of Ambulatory Care Management 2014; 37(3):196-8
29. George G, Haas MR, and Pentland A. Big Data and management. Academy of Management Journal 2014; 57(2): 321-6

30. McGeachie MJ, Chang HH, Weiss ST. CGBayesNets: Conditional Gaussian Bayesian Network Learning and Inference with Mixed Discrete and Continuous Data. PLoS Comput Biol 2014; 10(6): e1003676. doi: 10.1371/journal.pcbi.1003676
31. U.S. Department of Health & Human Services. Health Information Privacy. The Health Information Technology for Economic and Clinical Health (HITECH) Act. Washington, D.C. February 17, 2009. Full text available for download from the website of Department of Health & Human Services: http://www.hhs.gov/ocr/privacy/hipaa/administrative/enforcementrule/enfifr.pdf. Last accessed on July 2015
32. European Commission's proposal for a General Data Protection Regulation. EPF Position Statement, December 2012. Full text available for download from the web site: http://www.datasaveslives.eu/media/1110/general-data-protection-regulation-position-statement-final.pdf. Last accessed on August 2015
33. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Full text available for download from the web site: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML. Last accessed on August 2015
34. Mazieres D, and Waldman M. Tangler: A censorship-resistant publishing system based on document entanglements. In Proceedings of the 8th ACM Conference on Computer and Communications Security 2001; 126-35
35. Stubblefield A, Wallach DS. Dagster: Censorship-resistant publishing without replication. Technical Report TR01-380, Rice University 2001
36. Aspnes J, Feigenbaum J, Yampolskiy A, et al. Towards a theory of data entanglement. Theor Comput Sci 2007; 389: 26-43