ORIGINAL RESEARCH

The role of bioinformatics in developing clinical practice guidelines: principles and opportunities

Hisham M.F. Sherif

Cardiac Surgery, Christiana Hospital, Biomedical Engineering, University of Delaware, Newark, Delaware, USA

Correspondence: Hisham M.F. Sherif. Address: 4755 Ogletown-Stanton Road, Suite 1E50, Newark, Delaware, USA. E-mail: HSherif@Christianacare.org

Abstract

Clinical practice guidelines were developed by the professional organizations to provide standards of safe, effective, appropriate, equitable, affordable diagnostic and therapeutic care to promote the best yield at the least cost; to the patient, payers, health care professionals as well as to the healthcare system. One major objective of these guidelines is to prevent serious and/or harmful events in the course of disease. From a process analysis standpoint, such events are considered "errors" and should be investigated in accordance with the guidelines and tools for accident investigation. These tools include Process Mapping, Root-Cause Analysis, Influence Diagrams and Predictive Modeling methodology. The underlying principle is that all biologic environments are complex environments, involving a multitude of individual processes following multiple pathways and under the influence of numerous factors at multiple levels; as opposed to a single-axis main process line of the Ishikawa diagram. Process mapping mandates a thorough understanding of the multi-linear, inter-dependent and overlapping pathways involved in the genesis and maintenance of disease. Through this approach, the impact of various factors on the end-result can be more adequately assessed. More robust utilization of registry data is essential for comprehensive and complete data collection and organization. A probability-based predictive model may be best suited to incorporate all relevant factors, including others recognized in the future.

Key words

Bioinformatics, Decision-making, Practice guidelines

1 Background

Among the provisions of the Affordable Care Act of 2010 is an increased emphasis on improving the Quality of healthcare. This legislation provides recommendations for efforts towards improving:

- The safety of clinical decisions, both diagnostic and interventional; by improving outcomes in terms of mortality and morbidity and reducing medical errors.

- The effectiveness of the healthcare system; by reducing redundancy and waste and optimizing the allocation of resources.

- The appropriateness of healthcare decisions; by improving the indications and the selection of diagnostic and therapeutic measures.

- The equitability and accessibility of healthcare delivery to different segments of the population across different geographic and socioeconomic strata.

Hence, the "Health Information Technology for Economic and Clinical Health (HITECH) Act" has been introduced with the aim of providing support and incentives to promoting healthcare information technology solutions geared towards achieving the above-mentioned goals of improving the quality of healthcare. The objective is to foster the implementation of the bioinformatics tools and applications in improving healthcare decision-making, on the individual clinician level as well as on the regional or national healthcare system level.

2 Primo non nocere (Do No Harm)

The governing organizations for different areas of medical practice (such as Colleges, Boards and Societies) have had a long tradition in maintaining the professional standards that ensure the safety, effectiveness and appropriateness of clinical decisions through issuing Position Statements and Practice Guidelines. The ultimate objective of these regulations is to protect the general public from harm, which may result from inappropriate, ineffective or unnecessary diagnostic and/or therapeutic measures [1]. Although these guidelines are primarily designed for use by the medical professionals at different levels, they also have the added benefit of better informing the patients and members of the general public as well [2-4].

Recently, there has been a wider adoption of the concept of shared decision making; which has led to the increased focus on empowerment of the patients and their families, and promoting their engagement in a more active role in the clinical decision-making process. This is based on the fact that the patients are the critical stakeholders in the healthcare team and have the most vested interests in their well-being and Quality of Life. Therefore, the need to develop comprehensive, common-sense, practical and (most importantly) evidence-based guidelines becomes obvious.

Ideally, such guidelines should be based on reliable, reproducible high-level, scientific evidence derived from large-scale prospective, randomized controlled clinical trials. However, several obstacles to conducting large-scale prospective, randomized controlled trials remain, including: Insufficient number of subjects, insufficient funding, impracticality of study design and ethical considerations. Therefore, a significant part of the current practice guidelines is still based on intermediate-level scientific evidence derived from small clinical trials, case reports or "expert opinion". This is further complicated by the exponential growth in the volume and accessibility of scientific literature relevant to medical practice; driven by a robust and growing collaboration among different scientific disciplines such as genomics, proteomics, imaging technology, simulation and computer modeling. Unfortunately, the persistent "silo mentality" precluding effective communication among such traditionally separate disciplines has generated a widening gap between the scientific evidence and professional regulations. This disconnect has caused the guidelines to significantly lag behind the current scientific knowledge, to the point of being outdated or even presenting undue risk.

Since ancient times, the dictum of "First Do No Harm" remains the guiding principle for all practice guidelines. The ultimate objective of such regulations is to eliminate or significantly reduce the risk of death or disability resulting from disease, injury or medical practice.

Risk (R) is defined as the probability (p) of an adverse event (E), which can be expressed as follows: $R = p(E)$

Therefore, truly effective and reliable guidelines are – in fact – optimally constructed predictive models that provide an accurate, reproducible and consistent estimate of the probability of adverse and/or unexpected events in the course of disease, injury or medical practice. Such events are usually referred to as "errors".

3 Errors and processes

Simply put, an Error is "something that should not happen" [5]. In other words, it is an event that deviates from the normal or expected occurrence at a specific point within the context of a specific process in a specific environment. For example, putting the wrong number of candy pieces in a box is an error of the assembly process – specific to the packaging environment. In the aviation environment, an airplane crash is an error of the process involving the balance between lift and thrust; while aortic dissection is an error of the structural integrity process – specific to the aortic wall environment.

Each environment is composed of a number of different elements. The interaction of these elements with each other and with the environment describes a number of Processes. Each process is a sequence of events along a specific Process Line or Process Axis. Each process has a defined starting point, which can be termed "Baseline Definitions". The sequence of Events along the process line changes these definitions towards one or more end points or Results.

Along the process line, each Event is brought about by the effect of a Cause. The cause is usually an element of the environment, exerting its influence on the process line at a specific juncture or "node", representing the specific time it interacts with other factors. Each cause has its own specific effect on the process line, the significance of which is termed its "Weight". Depending on the number of Causes, their corresponding Weights, the direction in which the influence is applied relative to the process line, the balance of forces (influences) and the possible end results, the trajectory of the process line can remain the same or shift toward one or another end results.

In the candy packaging process, the baseline conditions are the empty box. The process line moves toward the result of a full box. This is effected by three successive events: Putting candy pieces in the box, counting the number of pieces then closing the box. Each event occurring in the correct way at the correct time leads to a successful result: A box filled with the right number of candy.

In analyzing any process, the graphical representation of its details describes a central axis intersected by the different vectors of the influence of all different causes. This well-established diagram [6] (The fish-bone, herring-bone or Ishikawa diagram) (see Figure 1) has considered one of the Seven Basic Quality Tools. Owing to its comprehensiveness and ease of use, it has been a standard tool in investigating accidents and errors in numerous environments, including industry, manufacture, management and healthcare.
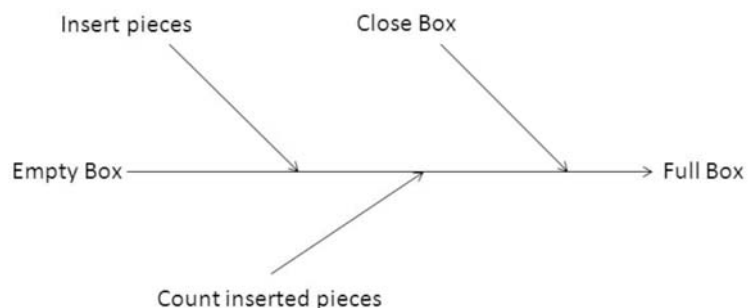
Figure 1. Candy packaging process

The effect of each factor can change, precipitating a change in the process as a whole. If the assembly line worker does not correctly count the pieces at the second "node", the result would be a half-full box, which is an Error. On the other hand, if he or she forgets to put the candy pieces inside the box, another Error results in an empty box (see Figure 2). By incorporating the trajectories of different results within the same process, this can be termed a "multi-axial Ishikawa diagram".

Another example of a process is taken from aviation: In level flight, there is a balance of the four basic forces (vectors) acting on the airplane. Lift opposes Weight, while aerodynamic drag is counteracted by Thrust from the engines. The net result is an airplane that maintains altitude because of the balance between lift and weight, while moving forward since the force vector for thrust is greater than that of drag.
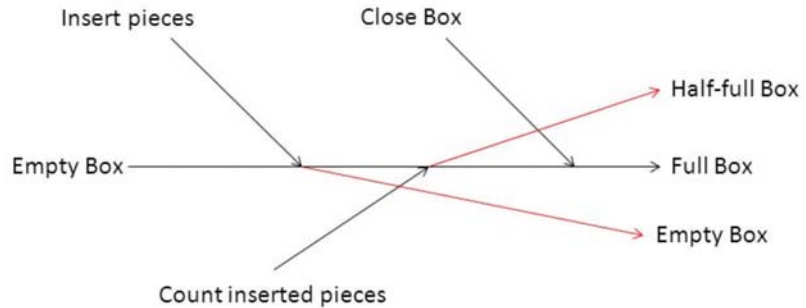
Figure 2. Candy packaging
process errors

However, the process of commercial flight to reach Point B from Point A involves a more intricate interaction between these basic factors. Hence, a process line for such a journey depicts the following segments: Takeoff, climb, level flight, descent and landing. Each segment involves a different balance of forces: Takeoff and climb require extra thrust and lift, while in descent and landing these forces are decreased, allowing weight and drag to gently decrease speed and altitude.

A fundamental concept here is that each of the basic factors (*e.g.*, lift, thrust, *etc.*) represents an entirely independent process. For instance, lift is dependent on the wing surface area, wing cross-section, angle of attack, the surface area of flaps, airplane speed and the air pressure. Thrust is a function of engine RPMs, air density, compression of air intake, fuel ignition, speed of exhaust leaving the nozzle, nozzle geometry, *etc*.

Examined from a different angle, the probability of level flight ($p = 1.0$ or 100%) depends on the combined probabilities of thrust, drag, weight and lift. According to the Bayesian Theory, the probability of each of these factors is dependent on the probability of multiple other factors. Thus, the probability of the End Result can be represented as the apex of a Decision Tree (see Figure 3).
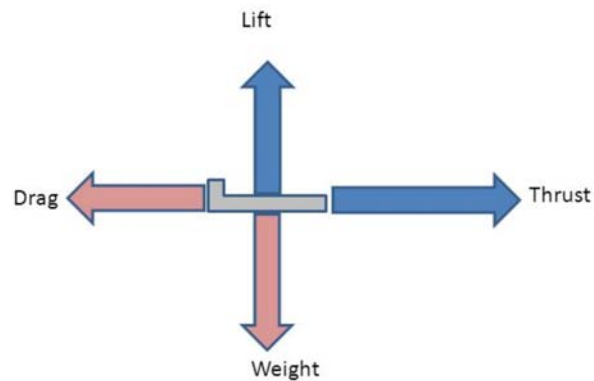

Figure 3. Balance of forces in level flight

4 Data categorization and processing

Another way of graphically depicting these relationships is as follows: A central process is dependent on a number of basic influencing processes. Each of these processes, in turn, is influenced by a number of factors. In addition, several of the outer factors may influence each other as well. The resulting diagram describes a Bayesian Network configuration (see Figure 4).

In this expanded influence diagram, the Central Process (commercial flight) has the following possible End Results: On-time arrival, Delay, Crash, Cancellation. It is directly influenced by a number of other processes (Determinants). Each Determinant process is, in turn, influenced by the Modifier set of variables. These factors can influence the Central Process and/or the Determinants. For example, the absence of a wing in the aircraft directly influences the lift, but it also directly affects the basic definition of the aircraft, thereby precluding the flight in the first place. Other factors with recognized but unclearly defined influence are called the Associated factors. In this example they may include low-pressure areas,

turbulence, torrential rain, *etc*. According to the NASA's Threat and Error model of accidents [7], Modifiers and Associated factors are "threats" that cause small but significant alterations in the course of process line by creating an "Unintended State" where the influence patterns of the variables are arranged differently from the optimal or intended design, ultimately resulting in a significant shift toward a deviation from the expected; *i.e.*, an Error.
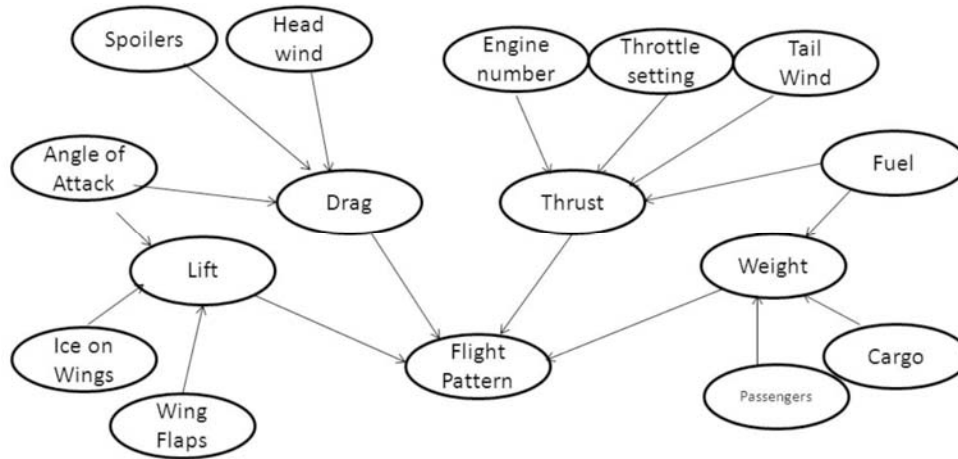
Figure 4. A Bayesian network model for flight

Similarly, most biologic environments involve a balance of forces acting on the same environment, organ or structure. For example, the Central Process of delivery of a baby has a trajectory with the following possible End Results: Successful Delivery, Stillbirth and Post-maturity (failure of delivery of a live fetus). This process is dependent on the following Determinants: Presence of a Pregnancy, Length of Pregnancy, Hormonal support of pregnancy, Uterine contractile function, and Anatomy of the birth canal. Modifiers for this process include: Female gender, ovulatory function, ovarian hormones, pituitary hormones, sexual activity, reproductive tract anatomy, cervical mucus, sperm number and function, *etc*. In this example, the absence or disturbance of female gender (as in some genetic disorders) can directly affect the central process, while simultaneously affecting other Modifiers such as ovulation and ovarian hormones. Associated factors may include alcohol intake, physical exercise, herbal and prescription medications, *etc*. This is the reason why – in developing meaningful, reliable and practical practice guidelines – all data relevant to a specific condition or disease should be collected in comprehensive detail as possible. By incorporating extensive representative data for all influencing factors, the predictive properties of the model are enhanced, increasing its ability to capture rare or atypical situations.

Establishing the highest possible order of certainty is an essential property of any decision-making algorithm or tool. Under ideal circumstances, the variable A always has a certain value B. In other words, the probability of A = B is 1.0 or 100%. This assumption offers a solid foundation for rule-based decision-making, which is well-established in computer programming. Simply put, rule-based decision-making assigns a specific action [X] based on the fulfillment of a specific assumption (*i.e.*, rule): IF [A] THEN [X]

Most of the current clinical guidelines for diagnostic, therapeutic or even resource allocation and payment are based on the identification of one or more combined rules. Surgery is recommended for aortic valve area < 0.6 cm^2; transcatheter aortic valve replacement is recommended for Society of Thoracic Surgeons Risk Score > 14; A repeat echocardiography within 30 days of the previous study may not be reimbursed; punitive financial measures are instituted against a hospital if there is a certain number of cases of ventilator-associated pneumonia, and many such "rules" are currently in force.

Rule-based decision making, however, has the following limitations:

- It assumes direct causative association;

- It requires and assumes a high degree of certainty;

- It assumes a singular causative relationship;

- It works best with variables of a discrete value;

- It is based on a Gaussian distribution;

- It cannot adequately account for confounding factors, continuous variables or borderline values;

- It assumes uniformity of substrate and uniformity of behavioral patterns in the environment.

By contrast, virtually all environments and processes in the healthcare field are highly complex, with numerous factors exerting their influences in a highly intricate pattern (which remains unknown or only partially understood in the majority of cases) with a high degree or inter-dependence and high level of uncertainty in both the value of variables (continuous or discrete) as well as the number and impact of possible End Results. In addition, such environments and processes are almost universally high-risk: to the patient, healthcare team or system.

Current guidelines were developed using evidence derived from standard statistical analytical methodologies. Such methods have significant limitations [8] of their predictive properties. These limitations include:

- Measurement errors, when there is inconsistency in measurements or assigning values to variables, particularly predictor values;

- Overfitting, or including too many variables in one predictive model; Omitted variable bias, which fails to adequately describe the association between variables rather that causative relationship;

- Multicollinearity, when two or more variables are highly correlated with one another. This is because standard statistical methods assume independent influence directions and does not adequately account for multiple interdependent patterns among variables;

- Reverse causation, which assumes a predetermined (*i.e.*, highly certain) causative relationship between risk factor and outcome factor;

- Selection effect, which stems of the observed sample not adequately representing the population of interest;

- The exponential increase in operational complexity relative to the increase in domain size;

- The assumption of a bell-shaped Normal (Gaussian) distribution, which ignores infrequent or rarely encountered processes or small-size populations whose mean and median fall well below the observed measures of central tendency.

Biologic environments and processes usually demonstrate a balance between two Determinant processes, each dependent on a number of other Modifier and Associated processes and factors. Therefore, the Central Process is actually a function of two functions, each – in turn – is a function of a multitude of other functions. Subsequently, the probability of an Event relative to the Central Process becomes a function of functions, correlating the probability of each factor influencing the process line pathway [9-12].

Hence, predicting the probability of an event (E) can be expressed as:

$$p(E) = p(C1) \cdot p(C2) \cdot p(C3)\ldots\cdot p(Cn) \tag{1}$$

$$p(E) = p\{C_1, C_2, \ldots C_n\} \tag{2}$$

Since the objective of guidelines is to identify and stratify the probability of high-risk events in such complex, high-risk environments, it becomes imperative to utilize a predictive model [13, 14] which has the following properties:

- Possibilistic decision-making property: To accommodate all the observed and suspected factors

- Probabilistic decision-making property: To adequately correlate the different probabilities of each event relative to the incorporated factors

- Prognostic decision-making property: To assign a hierarchy of the impact of predicted events on each of the possible End Results

- Provision for the Utility of any specific course of action based on the predicted event, given the effect of such recommended action on the result of the Central process.

Thus, a probability-based decision making tool should be designed to incorporate all known and suspected factors that do or may influence the course of a condition or a disease, assigning higher "weight" to those factors which have a more significant impact on this course toward one of more possible end results. In assigning such weight, the factors that contribute to the more serious and threatening result are assigned a higher weight, in accordance with the hierarchy of the possible end results. Also based on this hierarchy, the recommended course of action should correspond to these results.

For example, the central process of coronary atherosclerosis is a balance between normal proliferative and reparative intramural growth within the vascular wall, and the pro-apoptotic, pro-atherothrombotic process. In constructing a predictive model for the risk of myocardial infarction, all the factors influencing these two basic processes must be taken into account, with those having a more pronounced effect on the course of either process given a higher weight. Meanwhile, the possible outcomes (*e.g.*, massive myocardial infarction, stable angina, normal vascular growth) are arranged in a descending order of significance relative to their impact on the disease process, survival, Quality of Life, healthcare system burden, *etc*. An appropriate course of action is therefore recommended based on the probability of each end result.

5 Conclusions

Decision making in healthcare continues to present substantial challenges to the practitioners, patients as well as regulatory and administrative agencies. Inherently a very complex environment encompassing an almost infinite number of equally complex and high-risk environments, the field of healthcare defies attempts at simplistic rule-based decision-making. However, there is a rapidly growing need for robust, dynamic, real-life, adaptable and expandable bioinformatics solutions for decision-making support in clinical situations. Several applications have already been developed and are in use for isolated areas in biomedical sciences, such as gene mapping, drug design and development or optimization of specific drug therapy [15-18]. Although proven useful, such applications have a limited impact, if any, on the clinical practice, mainly because of a paucity of applications designed for the clinical, bedside practice areas. Most existing decision tree applications do not provide an automated clinical decision making or predictive function. Instead, the user usually supplements them with the patient-specific data and ultimately is the one to interpret the guideline content. Therefore, the healthcare community has remained reluctant to embrace computer-generated or interpretable guidelines. Reasons for this lack of adoption include a general distrust of automated systems in such high-risk situations, the perception of their incomplete consideration of all significant factors, the perceived loss of autonomy of practitioners and the lack of flexibility to fit different clinical environments.

The timing is fortuitous: As a beneficial result of the recent healthcare legislation, and partly driven by the Accountable Care Organizations and Managed Care Organzations, there has been a rapid growth in healthcare informatics and the availability of large data sets. This offers a unique opportunity for the development of simple yet powerful computer-based and computer-interpretable decision making tools as comprehensive, evidence-based clinical guidelines. Of particular

importance in this context is the promotion of their acceptance and use by real-life clinicians, and therefore the crucial requirement for these tools to mimic the conventional diagnostic and management methodology, and not be perceived as "cold, mechanical or inflexible" tools.

Probability based decision-making applications rooted in the Bayesian principle [19-22] incorporate a combined probabilistic and prognostic approach to predict the outcome of events in the disease process, the impact of interventions and their utility; in light of the preferences unique to decision-maker as they impact outcomes in terms of survival, morbidity, Quality of Life and other considerations; *e.g.*, patient wishes, available resources, financial impact, *etc*. Because of their flexibility, logical design approach, the ease of construction on simple, available platforms and the "familiar feel" during implementation, their ability to handle large volumes of data, they can be helpful tools for risk stratification, outcomes predictions and therefore the development of realistic, evidence-based clinical practice guidelines, designed to beneficially alter the course of disease processes and improve the overall quality of healthcare.

Competing interests

Acknowledgements

References

[1] Browne L. Regulation of Professions by the State. The Right to Regulate, Reasons Therefor, Methods in Use, and Attitude of Regulatory Bodies and the Courts, with Relation Thereto. Cal West Med. Aug 1935; 43(2): 119-123. PMid: 18743337.

[2] Woolf S. Practice Guidelines, a New Reality in Medicine II. Methods of Developing Guidelines. Arch Intern Med. 1992; 152(5): 946-952. http://dx.doi.org/10.1001/archinte.1992.00400170036007

[3] Grimshaw, J., Eccles, M., Russell, I. Developing clinically valid practice guidelines. Journal of Evaluation in Clinical Practice. 1995; 1: 37-48. http://dx.doi.org/10.1111/j.1365-2753.1995.tb00006.x

[4] Turner T, Misso M, Harris C, Green S. Development of evidence-based clinical practice guidelines (CPGs): comparing approaches. Implementation Science. 2008; 3: 45. http://dx.doi.org/10.1186/1748-5908-3-45

[5] Gerstein M. Flirting with disaster: Why accidents are rarely accidental. New York: Union Square Press. 2008.

[6] Ishikawa K. Guide to quality control. Kraus International Publications. White Plains, NY. 1982.

[7] Gunther D. Dealing with Emergency/Abnormal Situations with New Security Guidelines. NASA Ames Research Center Emergency and Abnormal Situations in Aviation Symposium. June 2013. Available from: http://humansystems.arc.nasa.gov/flightcognition/download/EAS_Symposium_Presentations/Security.pdf. Last accessed July 29th, 2014.

[8] Rizzo J, Chen J, Fang H, Ziangshin B, Eleftriades J. Statistical Challenges in Identifying Risk Factors for Aortic Disease. AORTA. April 2014; 2(2): 45-55. http://dx.doi.org/10.12945/j.aorta.2014.14-019

[9] North D. A tutorial introduction to decision theory. IEEE Transactions on systems science and cybernetics. Vol SSC-4, No 3. September 1968.

[10] Bayesian networks and decision making. University of Waterloo, Canada. Available from: http://pami.uwaterloo.ca/~basir/ECE457/week10.pdf. Last accessed July 29th, 2014.

[11] Cobb B, Shenoy P. Inference in hybrid Bayesian networks with mixtures of truncated exponentials. In: J. Vejnarova (ed.). Proceedings of the 6th workshop on uncertainty processing (WUPES-2003). 47-63. VSE-Oeconomica Publishers.

[12] Hommersom A, Lucas P. Generalising the Interaction Rules in Probabilistic Logic. Walsh, T. (ed.), Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence. 912-917. Menlo Park: AAAI Press.

[13] Peleg M. Computer-interpretable clinical guidelines: a methodological review. J Biomedical Informatics. 2013; 46: 744-763. PMid: 23806274. http://dx.doi.org/10.1016/j.jbi.2013.06.009

[14] Seroussi B, Bouaud J, Antoine E-C. ONCODOC: A successful experiment of computer-supported guideline development and implementation in the treatement of breast cancer. Artif Intell Med. 2001; 22(1): 43-64. http://dx.doi.org/10.1016/S0933-3657(00)00099-3

[15] Papageorgiou EI1, Roo JD, Huszka C, Colaert D. Formalization of treatment guidelines using Fuzzy Cognitive Maps and semantic web tools. J Biomed Inform. 2012 Feb; 45(1): 45-60. http://dx.doi.org/10.1016/j.jbi.2011.08.018

[16] John RI, Innocent PR. Modeling uncertainty in clinical diagnosis using fuzzy logic. IEEE Trans Syst Man Cybern, Part B. 2005; 35(6).

[17] Sadeghi S, Barzi A, Zarrin-Khameh N. Decision support system for medical triage. Stud Health Technol Inform. 2001; 81: 440-2. PMid: 11317786.

[18] Sadeghi S, Barzi A, Sadeghi N, King B. A Bayesian model for triage decision support. Int J Med Inform. 2006; 75: 403-11. http://dx.doi.org/10.1016/j.ijmedinf.2005.07.028

[19] Sherif H, *et al*. Design and Construction of a Computer-Based Logical System for Medical Diagnosis" In: Studies in Health Technology and Informatics. 2001; 81. James D. Westwood, et al, Editors.

[20] Sherif H. Is Practicing Medicine Virtually Impossible? Healthcare Informatics. August 2006; 23(8): 54-5. PMid: 16948337.

[21] Sherif HM, Rahman E, Mustafa N, Hosmane VR, Reddy V, Disabatino A, *et al*. Recovery After Resuscitation from Cardiac Arrest in ST Elevation Myocardial Infarction: A Computer-based Decision Support Tool. Stud Health Technol Inform. 2009; 142: 310-312. PMid: 19377175.

[22] Sherif HM, Johnson SS, Klair SA. Implementation of a Computer-Based Decision Support System for Outcomes Prediction and Clinical Triage: Initial Results of Two Pilot Studies. Journal of Cardiology and Therapeutics. 2013; 1: 71-77.