Capturing Scientific Knowledge on Medical Risk Factors

Allan Third[1], Eleni Kaldoudi[2], George Gkotsis[3], Stefanos Roumeliotis[2], Kalliopi Pafili[2], John Domingue[1]

[1]Knowledge Media Institute
Open University
Milton Keynes, UK
+44 (0) 1908 659708/655014
{allan.third,
john.domingue}@open.ac.uk

[2]Democritus University of Thrace
Dragana
Alexandroupolis, Greece
+30 2551030329
kaldoudi@med.duth.gr,
{st_roumeliotis,
kpafili}@hotmail.com}

[3] King's College London
Biomedical Research Centre Nucleus
London, UK
+44 (0) 20 3228 8538
george.gkotsis@kcl.ac.uk

ABSTRACT

In this paper, we describe a model for representing scientific knowledge of risk factors in medicine in an explicit format which enables its use for automated reasoning. The resulting model supports linking the conclusions of up-to-date clinical research with data relating to individual patients. This model, which we have implemented as an ontology-based system using Linked Data, enables the capture of risk factor knowledge and serves as a translational research tool to apply that knowledge to assist with patient treatment, lifestyle, and education. Knowledge captured using this model can be disseminated for other intelligent systems to use for a variety of purposes, for example, to explore the state of the available medical knowledge.

Categories and Subject Descriptors

H.2.8 Database applications – scientific databases

I.2.1 Applications and Expert Systems – Medicine and Science

I.2.4 Knowledge Representation Formalisms and methods

General Terms

Measurement, Design, Experimentation, Standardization.

Keywords

Health, comorbidities, risk, scientific modelling, knowledge capture, semantics, ontology, Linked Data.

1. INTRODUCTION

An important task in medicine is the assessment of risk. This depends on scientific knowledge derived by rigorous clinical studies regarding the (quantified) factors affecting clinical changes. Existing risk prediction tools typically only cover a very limited range of patient states, and the scientific knowledge informing the predictions is hardcoded into the tool. This makes them limited in application, particularly for patients with comorbidities (multiple co-occurring conditions), and rapidly out of date. An explicit representation of this knowledge, covering a wide (and, more importantly, expandable) range of risks and outcomes, would enable more sophisticated and maintainable risk prediction, prevention and management.

Conference'10, Month 1–2, 2010, City, State, Country.

We present here a model for representing the scientific knowledge of risk factors in just such a way. Existing work on modelling risks in clinical research focuses on the process and detail of clinical studies performed to identify and quantify risk factors. The work presented here focuses instead on the *output* of such clinical studies and proposes a generic model for the concept of health risk factor.

Representations of scientific knowledge regarding health risk factors will be useful in clinical decision support systems and personalised care services. Additionally, intelligent systems can analyse knowledge regarding a particular medical subdomain in order to identify gaps or interesting areas in currently available research. This work was carried out in the context of the EU funded FP7-ICT project CARRE (Grant no. 611140), which seeks to provide personalised decision support to patients in managing comorbidities associated with cardiorenal disease.

2. PREVIOUS WORK

A number of models have been proposed for capturing various aspects of clinical research at various levels of granularity. As there is no other model addressing the concept of risk factor, to the best of our knowledge, we compare related work addressing similar concepts and level of abstraction. In particular, we consider the Ontology-Based eXtensible data model (OBX, [1]), the models maintained by the Clinical Data Interchange Standards Consortium (CDISC, [2]), and the Ontology of Clinical Research (OCRe, [3]).

OBX was designed for the Immunology Database and Analysis Project (Immport, [4]) to be a generic model for representing the results of clinical research, including data from case report forms down to molecular data from specimens. It represents significant details of experimental design, modelling them at a high enough level to make it possible to capture many different kinds of research output. The aim of OBX is to promote data reuse. However, it does not focus on the broader scientific knowledge which can be *learned* from this data after suitable statistical analysis. A "Finding" in OBX is designed to represent the outcome of a particular assay or test, such as a blood sample analysis or a patient assessment, not a population-level risk factor.

The CDISC standards, and OCRe, both take a more top-down approach to the modelling of clinical research, although from different perspectives. CDISC are responsible for standards relating to the data required for case report forms and the requirements and concepts in clinical trial structures necessary for regulatory reporting to the US Federal Drug Administration (FDA). In general, these models are focused on data interchange formats and on the conceptual modelling of proposed and ongoing clinical trials, in order to support the vital but complex

reporting needed for regulatory purposes. OCRe is also a formal model of the clinical trials process, but from a much more scientific perspective than the CDISC standards. OCRe models the design and analysis of a clinical research study in great detail and accommodates the wide range of possible types of study. The aim is to represent the scientific knowledge *about* studies in order to promote reuse, accountability and transparency in the scientific process in medicine. For example, OCRe is used by the Human Studies Database Project (HSDB, [5]) to support federated querying of data regarding human studies across a range of institutions holding such data. One of the use cases of OCRe is to support search and aggregation of trial data in order at least to semi-automate literature research to conduct systematic reviews and meta-analysis.

A general theme of all of these models is the focus on the knowledge about the science, to promote different kinds of good working practice in research and to support research-related tasks. That is, existing models aim to support the process of generating new scientific knowledge in medicine (i.e. new medical evidence). What we do not see are models intended to support the *application* of scientific results in practice. Treatment guidelines or lifestyle recommendations for patients typically come from the output of the systematic review process, and constitute the vast amount of medical evidence on which clinical practice is largely based. The model presented here is intended to remedy that omission. This work is thus complementary to the existing modelling work, and it would be fruitful to work further on the possible ways to link these models, to promote the full "bench-to-bedside" pathway.

Existing algorithms for risk prediction for, e.g., cardiovascular risk, include the Framingham equation [6], the Joint British Societies (JBS) formula [7] and the ASSIGN score [8]. These only take account of a limited set of risk factors and possible outcomes, as these have been produced by specific clinical studies – thus can be limited in application. For example, the ASSIGN score is specialised for Scottish populations, and, while Framingham includes diabetes as a risk factor, it is omitted from the JBS formula (diabetic patients are always high-risk). More fundamentally, each of these hardcode the scientific knowledge about risk *into* the prediction formula itself, thus requiring new versions to be created to accommodate new scientific knowledge. By concentrating on the explicit modelling of this knowledge, it is possible to separate the knowledge from the reasoning task to be carried out with it, allowing new and updated knowledge to be added easily and in a transparent fashion.

3. RISK FACTORS

The following subsections present the requirements for modelling the concept of risk factor in medicine, how this is quantified and estimated via evidence from research studies.

3.1 Risk factors in medicine

In medicine, risk is the probability of a negative outcome on the health of a population of subjects. The agents responsible for that risk are called risk factors when they aggravate a situation and are used to predict (up to a degree) the occurrence of a condition or deterioration of a patient's health dividing the population into high and low risk groups [9]. In general, risk factors can be:

- **Environmental:** Chemical, physical, mechanical, biological and psychosocial elements that constitute risk factors to public health.
- **Demographic:** Empirical findings have pointed out that age, sex, race, location, and religion all affect public health.
- **Genetic:** Any predisposition to conditions and habits hardcoded in the human genome.
- **Behavioral/Lifestyle-related** Human behaviors that are marked as "risky" and have proven to cause deterioration or provide added risk, such as smoking, overeating, unprotected sex, excessive alcohol consumption, drug abuse and a sedentary lifestyle.
- **Biomedical:** These include clinical diagnoses such as diabetes, and states such as pregnancy, present in a patient that can influence his/her health by creating or affecting other conditions.

Each of the above elements, severally or in combination, may be descriptive of a particular patient or population, and may predispose such a patient/member of that population to develop further such conditions.

Extending work on general risk analysis [10],[11], we can present a risk factor as a triplet, which includes the *source of the risk*, the *outcome* and an expression of their *association*. The source of the risk is an agent (an event, a condition, a disorder or any other factor) that is shown via empirical studies to be associated with a consequence, that is, the outcome. The outcome itself is a negative health condition or disorder. Most often the outcome itself is found to be a source of another risk factor.

Thus in the general case the source and the outcome can both be treated as health related conditions (including disorders). In this work, we collectively refer to both the source and the outcome as *risk elements.* Not all elements can occur in both roles. In particular, "fixed markers" such as date of birth, genetics or ethnic origin cannot be modified or affected by anything, and so cannot be the outcome of a risk [12].

The association between the source and the outcome is a complex construct which describes the type of relation, the likelihood of an outcome to occur, and the initial conditions under which such likelihood can be estimated. The relation between the source and the outcome may not always be proven causation. Following the Unified Medical Modelling System's (UMLS) Semantic Network [13], associations between a risk factor and the associated condition may include:

- **issue_in:** the risk factor is a point of discussion for a condition;
- **affects:** the risk factor produces a direct effect on the condition;
- **causes:** the risk factor brings about the condition; and
- **complicates**: the risk factor causes another (risk) factor to become more complex (recursive).

3.2 Risk probability

The existence of a risk factor is not a determinant of consequence but the degree of its influence can be statistically calculated. The way to measure the likelihood requires a certain quantitative biomarker and observational studies that statistically calculate a probability. Different study designs and analyses can generate different types of probability measures [14].

A commonly used risk measure is the Relative Risk or Risk ratio (RR), which is the ratio of the probability of an event occurring

(for example, developing a disease) in an exposed group to the probability of the event occurring in a non-exposed group.

Another metric of relative risk is the Hazard Ratio (HR) (e.g. [15]) which is most often used in clinical studies to assess the instantaneous risk at any time of a given study. So, it accounts for the reality that some subjects may drop out of the study before the event of interest happens, or that the study may end before all of the subjects experience the event (time-to-event analysis).

Note that these representations of probability are the standard forms found in medical literature. In the interests of accuracy and relevancy to the intended clinical audience, we follow the medical conventions.

Empirically determined probabilities across populations can come with a range of associated qualifiers. A probability determined from a clinical study lies within a confidence interval, and the study design/analysis may have been adjusted, or not, for certain factors (for example, age, sex, and so on). In order to be able to properly represent risk factors, these must be included – especially where the goal is to produce personalised risk calculations.

3.3 Observables

An event, a condition, a disorder or any other factor becomes a risk source when certain conditions are met. These conditions are associated with one or more observable, which is either environmental or a physical or mental property of the patient.

So, for example, medical evidence suggests that obesity is a risk factor for diabetes [16]. However, if we want to be able to report a certain probability measure, we have to define what obesity means for the sample population used to calculate the statistics. Common observables to quantify obesity include waist circumference, waist to hip ratio, waist to height ratio, body fat percentage and body mass index (BMI, i.e. body mass divided by height squared). In the particular systematic review and meta-analysis mentioned here, people with waist circumference between 79.3 and 107.5 have a risk ratio of 1.65 to develop diabetes (as compared to people with lower values for waist circumference).

Therefore, in order to describe properly a risk association we have to state a specific observable that provides a measure/description of the risk source and the specific condition or value of this observable. For the same risk factor, a number of different risk associations can be measured in the literature, each association corresponding to a different observable or a different observable condition or even different combinations of observables corresponding to different concurrent risk sources.

In another example, medical evidence suggests that obesity is a risk factor for coronary arterial disease [17]. In this particular systematic review and meta-analysis, men with a BMI between 25 and 30 kg/m^2 were found to have a risk ratio of 1.29 to develop coronary arterial disease (as compared to normal male of a BMI 18.5 to 25 kg/m^2). A different risk association (for the same risk) is found for men of a BMI greater than 30 kg/m^2, who present an elevated risk ratio of 1.72. The same evidence source shows that women with a BMI between 25 and 30 kg/m^2 have a risk ratio of 1.80 to develop coronary arterial disease and when their BMI is above 30 kg/m^2 the risk ratio is elevated to 3.10. Thus in this example, four different associations are described between the risk source of obesity (and age) and the outcome of coronary arterial disease.

3.4 Medical evidence provenance & quality

Risk factors are derived from population statistical studies. The cornerstone of evidence based medicine is that such studies are the only source of knowledge regarding clinical risk, and thus, for two reasons, it is important that evidence be represented explicitly in our model. The first is provenance: no one could (or should) trust data purporting to represent clinical knowledge without the ability to trace it back to its source. The second is the question of quality. Sources of evidence can range from small in vitro studies or case reports to large randomized clinical trials, to meta-analyses of systematic reviews. All these population studies carry a different level of evidence. In the past, various evidence ranking schemes have been used, to appraise quality of evidence, based on study design and methodology utilized; one such commonly used scheme is the grading system proposed by The Oxford Centre for Evidence-Based Medicine [18].

4. A MODEL OF RISK

4.1 Risk factor model

The term *risk factor* is used in the medical literature seemingly interchangeably to indicate a risk source and/or the particular association. For clarity in the proposed model, we use the term *risk element* to indicate the risk source and/or the outcome (as described above), and the term *risk association* to refer to a specific triple of "source-association-target" when coupled to a particular observable condition, its probability, and evidence source. Based on this description, primary concepts and their relationships are identified in the paragraphs below and shown schematically in Figure 1.

Risk Element: A risk association defines the (often causal) association of an agent (source risk element) to a health outcome (target risk element). This outcome is in most cases negative, and most often the (causal) agent is in itself a negative health outcome. In this sense, risk agents and their outcomes can be seen as instances of the same entity, called here 'risk element'. Risk elements include all the disorders/diseases, as well as any other risk causing agent as discussed in Section 3.1, e.g. demographic, genetic, behavioural, environmental, or even interventions (e.g. pharmaceutical substances, contrast agents).

Risk Association: The association of one risk element as the risk source with another risk element which is the negative outcome under certain conditions is a 'risk association'. Note that a source risk element can be associated to a target risk element with more than one risk association. This association is a rather complex one and is characterized by a number of other concepts:

- Association Type: The association can be of a certain 'association type'; most often, it is of type 'causes', but it can also be 'complicates', otherwise 'affects' or in the general case (and when there is no knowledge of a specific effect), 'is an issue in'. There are also cases where an agent can have a positive effect, that is "reduces" the risk of a negative outcome. Generally, a number of other semantic relationships as described in UMLS could be encountered here.
- Risk Ratio: The association is always accompanied by the likelihood of the negative outcome to occur. This likelihood is expressed as a 'risk ratio', that is the ratio of the probability of the negative outcome when the person is exposed to the risk agent over the probability of the

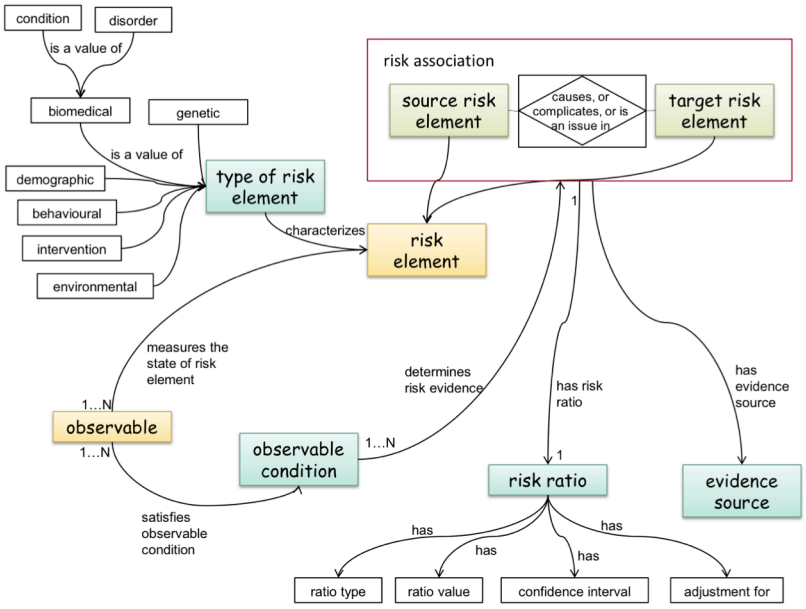

Figure 1 Basic concepts and their relationships

negative outcome when the person is not exposed to the risk agent. The risk ratio has a certain type (e.g. relative risk ratio, hazard ratio, etc.); a value (a positive real number, which when below 1 indicates a reduced risk compare to the control population and when greater than 1 indicates an elevated risk); a confidence interval; and a list of population characteristics for which the study was adjusted.

- Observables Condition: For the association to occur, certain circumstances should exist. These prerequisite circumstances relate directly to the existence of the risk agent (source risk target) and/or its severity, and/or any other specific conditions. These are reported via certain 'observables', that is, variables that can be measured or otherwise ascertained (e.g. biomarkers, biometric variables, biological signals and other non-biological factors e.g. environmental). The circumstances thus are ascertained via an explicit logical expression that involves observables; this logical expression is termed 'observables condition'.
- Evidence Source: Risk associations in medicine are determined from clinical studies as reported in evidence based medical literature. Thus each association is directly related to an 'evidence source' which is a specific scientific publication.

4.2 Standardized concept descriptions

To ensure that the model can be seamlessly integrated into existing medical information systems, we adopt the commonly used standards and controlled vocabularies in the description of the concepts presented above. For example, risk elements of type *biomedical* include an ICD-10 [19] classifier, of type demographic, a SNOMED-CT [20] classifier. Other controlled vocabularies used for risk elements of type environmental or intervention include SNOMED-CT, RxNorm [21], and EnvO [22]. Measurements and units follow the QUDT [23] and UO [24] ontologies. Evidence sources are described using their DOI and/or their PubMed identifier, while evidence level follows the OCEBM system [18]. In general, where available UMLS [25] codes are also used.

5. IMPLEMENTATION & EVALUATION

This model was used to capture scientific information on medical risk factors in the area of cardiorenal disease. Chronic cardiorenal disease is the condition characterized by simultaneous kidney and heart disease while the primarily failing organ may be either the heart or the kidney. Very often the dysfunction occurs when the failing organ precipitates the failure of the other. The cardio-renal patient (or the person at risk of this condition) presents an interesting case example for exploring risk factors, as (a) is a complex comorbid condition which involves and is affected by a number of related health disorders as well as lifestyle related factors; (b) chronic cardiorenal disease has an increasing incidence and a number of serious (and of increasing incidence) comorbidities, including diabetes and hypertension, and may lead to serious chronic conditions such as nephrogenic anemia, renal osteodystrophy, peripheral neuropathy, malnutrition, and various systemic diseases (e.g. rheumatoid arthritis, lupus erythematosus); and (c) prevention is of major importance. Good appreciation of risks therefore plays an important role for the various stages of cardiorenal disease evolution, from normal health condition, to chronic disease, to end-stage renal deficiency and/or heart failure.

To test and put the model into use, a group of 8 medical doctors (members of the CARRE project team) reviewed current medical literature to identify major risk factors related to cardiorenal syndrome. At this time, 98 different risk factors were identified and described formally using the proposed model. The descriptions resulted in 268 respective associations. There were 45 involved risk elements, corresponding to a total of 47 different observables. The evidence sources used were 62 scientific publications. The review methodology and the available descriptions in text (tabular) format are provided in CARRE Deliverable 2.2. available from the project site [26].

This process of testing and using the proposed model resulted in the following qualitative findings. The medical experts found the model straightforward to use to describe risk factors. The terminology used was found to be familiar and thus easy to understand and apply to describe risk factors found in the literature and also to read descriptions already produced by

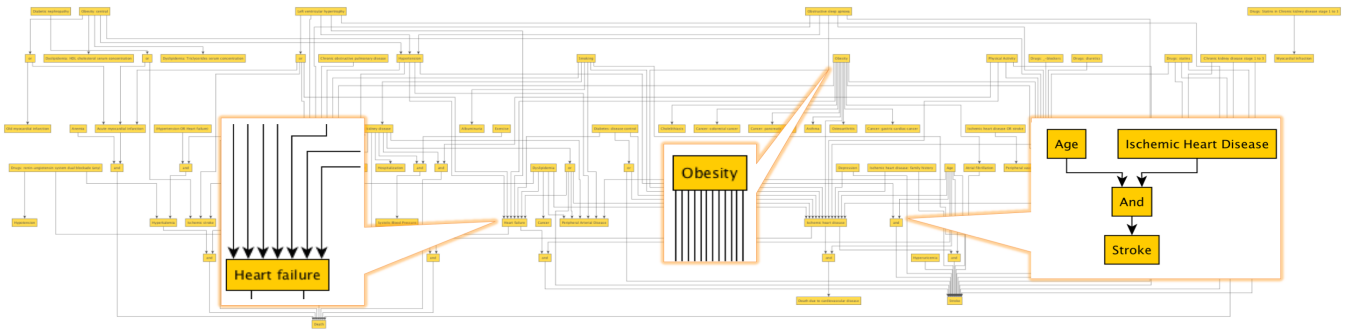

Figure 2 A visual overview of currently encoded risk factors, with some examples highlighted.

colleagues. The only difficulty identified related to expressing accurately and rigorously the observables' condition that has to be satisfied in order for a risk association to hold. Initially, medical experts were asked to produce this condition in the conventional way this is written in the literature, using natural language – which was a straightforward task. Subsequently, they were asked to reformat this condition using a logical operator expression (so that this expression can be easily translated to computer readable format). This task proved to be more cumbersome and required 1-2 hours training and testing before the medical experts could independently produce correct expressions.

In order to enable the open and seamless use and reuse of these described medical risk factors, we have developed an on-line web based system for their description. Also, the resulting risk factor descriptions are available as Linked Data [27], in the Resource Description Framework (RDF) format [28], via an open access RDF repository.

The web-based system was developed by customizing the Drupal content management system **Error! Reference source not found.** to reflect the structure of the model presented here, so that observables, evidence sources, risk elements and associations can be entered via web forms, and automatically translated to RDF. The system maintains referential integrity, so that if, for example, "diabetes" is entered as a risk element entity, then a risk association representing an observed link between diabetes and hypertension will refer to the existing diabetes risk element entity. Users are supported in the reuse of data already entered into the system by the user interface, which allows existing relevant entities to be selected via drop-down lists wherever possible. Customisation is straightforward: Drupal allows the definition of custom content types (e.g., risk factor) annotated with RDF terms, so that when data for a particular risk factor is entered, it can be made automatically available as RDF.

Figure 2 illustrates a projection of the various risk factors, as captured by the medical experts in the context of our project. The diagram contains 93 distinct risk factor associations. These associations are constructed combining 45 different risk elements, used either as source or target (or both). Highlighted in Figure 2 is the example of age *and* ischemic heart disease increasing a patient's risk of a stroke. It can also be seen how many risk elements increase the risk of heart failure, and how many elements are at an increased risk in obese patients.

6. DISCUSSION

The model presented in this paper enables the clinical experts to encode the risk associations between biological, demographic, lifestyle and environmental elements and clinical outcomes in accordance with evidence from the clinical literature. While the motivation and initial thinking regarding the model was focused on factors which increase the probability of negative consequences, the end result is equally as capable of modelling factors which decrease those probabilities, or which affect the probability of positive consequences. In other words, it is just as straightforward to represent, for example, an intervention with the potential to *lower* a patient's body mass index.

The encoded data is useful for reasoning about personalised medical risk, both actual and hypothetical. In particular, reasoning related to actual or potential comorbidities is supported by the model. By implementing the model using standard semantic technologies, it is possible to link both model and data to other clinical models (such as OCRe and OBX trial and data descriptions) and to external sources of data (e.g., environmental risk factors could be linked to open sources of environmental data).

Nothing in the model is specific to the motivating domain of cardiorenal conditions, and extension to risk factors relating to other domains of medicine is not anticipated to pose any problems. Extending to more 'distant' domains where evidence-based risk calculation is relevant (e.g., climate science) ought also to be practical. The model already accommodates different representations of probability, and so could be adapted to those representations suitable to the new domain's conventions, and the concept of "observable" is already generic. It would be necessary to extend the notion of evidence, and in particular, evidence quality, which is currently dependent on medical definitions.

The benefit of modelling risks explicitly in this way is that it gives a very easy to follow overview of the field of medicine under consideration, showing at a glance both which risks are increased by multiple factors, which factors lead to multiple risks, as well as which associations have received more (or less) research attention. Compared to existing risk prediction models, this approach has a significant advantage in being able to be expanded and updated easily as clinical knowledge increases and changes.

It should be noted that the model is intended to capture medical evidence as presented in current medical literature. The predictive accuracy of the model is thus directly linked to and depended on this knowledge. Figure 2 projects only the risk association relationship, excluding the other details and relationships encoded in the model. The full range of knowledge captured in the model supports a sophisticated analysis of the kind which is difficult and time-consuming to undertake manually. We intend to pursue such analysis in future work.

The difficulty with clinical scientific knowledge is the necessity to keep it accurate and up to date, and we are currently exploring the best process for curating a knowledge base of risk factors. However, the encoding task proved to be quite straightforward for our medical experts.

7. ACKNOWLEDGMENTS

This work was supported by the FP7-ICT project CARRE (Grant No. 611140), funded in part by the European Commission. We express our gratitude to all project team members for fruitful discussions.

8. REFERENCES

[1] Kong YM, Dahlke C, Xiang Q, Qian Y, Karp D, Scheuermann RH, Toward an ontology-based framework for clinical research databases, J. Biomed. Inform., 44(1):48-58, 2011.

[2] CDISC, http://cdisc.org, Accessed on: 24/07/2015.

[3] Sim, I., Tu, S. W., Carini, S., Lehmann, H. P., Pollock, B. H., Peleg, M., & Wittkowski, K. M. (2014). The Ontology of Clinical Research (OCRe): an informatics foundation for the science of clinical research. *Journal of biomedical informatics*, *52*, 78-91.

[4] Bhattacharya, Sanchita, et al. "ImmPort: disseminating data to the public for the future of immunology." *Immunologic research* 58.2-3 (2014): 234-239.

[5] Sim, I., Carini, S., Tu, S., Wynden, R., Pollock, B. H., Mollah, S. A., & Bakken, S. (2010). The human studies database project: federating human studies design data using the ontology of clinical research. *AMIA Summits on Translational Science Proceedings*, *2010*, 51.

[6] Sheridan, S., Pignone, M., & Mulrow, C. (2003). Framingham‐based tools to calculate the global risk of coronary heart disease. *Journal of general internal medicine*, *18*(12), 1039-1052.

[7] Boon, N., Boyle, R., Bradbury, K., Buckley, J., Connolly, S., Craig, S., ... & Wood, D. (2014). Joint British Societies' consensus recommendations for the prevention of cardiovascular disease (JBS3). *Heart*, *100*(Suppl 2), ii1-ii67.

[8] Woodward, M., Brindle, P., & Tunstall-Pedoe, H. (2007). Adding social deprivation and family history to cardiovascular risk assessment: the ASSIGN score from the Scottish Heart Health Extended Cohort (SHHEC). *Heart*, *93*(2), 172-176.

[9] Mrazek, P. B., & Haggerty, R. J. (Eds.) , 1994. Reducing risks for mental disorders: Frontiers for preventive intervention research: Summary. National Academies Press

[10] Kaplan S, The words of risk analysis, Risk Analysis, 17(4):407-417, 1997

[11] Offord DR, Kraemer HC, Risk factors and Prevention, EBMH vol 3, p. 71, 2000

[12] Kraemer, H.C., Kazdin, A.E., Offord, D. R., Kessler, R. C., Jensen, P. S., & Kupfer, D. J., Coming to terms with the terms of risk. Archives of General Psychiatry, 54(4), 337, 1997.

[13] National Library of Medicine (2009) Chapter 5 - Semantic Networks. UMLS Reference Manual. Bethesda, MD: U.S. National Library of Medicine, National Institutes of Health

[14] Crowson CS, Therneau TM, Matteson EL, Gabriel SE, Primer: demystifying risk - understanding and communicating medical risks. Nature Clinical Practice Rheumatology, 3(3, March 2007), 2007

[15] Bull K., Spiegelhalter DJ, Tutorial in Biostatistics, - Survival Analysis in Observational Studies, Statistics in Medicine, 16:1041-1074, 1997

[16] Kodama S, Horikawa C, Fujihara K, Heianza Y, Hirasawa R, Yachi Y, Sugawara A, Tanaka S, Shimano H, Iida KT, Saito K, Sone H., Comparisons of the strength of associations with future type 2 diabetes risk among anthropometric obesity indicators, including waist-to-height ratio: a meta-analysis. Am J Epidemiol. 2012 Dec 1;176(11):959-69.

[17] Guh DP, Zhang W, Bansback N., Amarsi Z., Birmingham CL, Anis AH, The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis, BMC Public Health, 2009 Mar 25;9:88

[18] Oxford Centre for Evidence-based Medicine Levels of Evidence (2011) Produced by J. Howick, I. Chalmers, P. Glasziou, T. Greenhalgh, C. Heneghan, A. Liberati, I. Moschetti, B. Phillips, H. Thornton, O. Goddard and M. Hodgkinson

[19] ICD-10: International Classification of Diseases v10, WHO, http://www.who.int/classifications/icd/en/

[20] SNOMED-CT: Systemized Nomenclature of Medicine – Clinical Terms, International Health Terminology Standards Development Organization, http://www.ihtsdo.org/snomed-ct/

[21] RxNorm: Normalized Names for Clinical Drugs, U.S. National Library of Medicine http://www.nlm.nih.gov/research/umls/rxnorm/

[22] EnvO: Environmental Ontology, http://environmentontology.org/

[23] QUDT: Quantity, Unit, Dimension and Type Ontologies, http://qudt.org/

[24] UO: The Ontology of Units of Measurement, OBO Foundry Initiative, https://code.google.com/p/unit-ontology/

[25] UMLS: The Unified Medical Language System, US National Library of Medicine, http://www.nlm.nih.gov/research/umls/

[26] CARRE http://carre-project.eu/

[27] Berners-Lee, T., Bizer, C., & Heath, T. (2009). Linked data-the story so far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1-22.

[28] Resource Description Framework http://www.w3.org/RDF/

[29] Drupal content management system https://www.drupal.org/