Formalization of indicators of diagnostic performance in a realist ontology

Adrien Barton[1,2,*] Régis Duvauferrier[2,3] and Anita Burgun[4]

[1]The Institute of Scientific and Industrial Research, Osaka University, Japan
[2]INSERM UMR 1099, LSTI, Rennes, France
[3]CHU de Martinique, Université Antilles-Guyane, France
[4]INSERM UMR 1138 team 22, Centre de Recherche des Cordeliers, Paris, France

ABSTRACT

We present a formalization of indicators of diagnostic performance (sensitivity, specificity, positive predictive value and negative predictive value) in the context of a realist ontology. We dissociate the indicators of diagnostic performance from their estimations and argue that the former should be represented in a first place in biomedical ontologies. Our formalization does not require to introduce any possible, non-actual entities - like the result a person would get *if* a medical test would be performed on her - and is therefore acceptable in an ontology built in a realist spirit. We formalize an indicator of diagnostic performance as a data item that is about a disposition borne by a group; the diagnostic value of this indicator is given by the objective probability value assigned to this disposition.

1   INTRODUCTION

1.1   Definition of indicators of diagnostic performance

Biomedical ontologies aim at providing the most exhaustive and rigorous representation of reality as described by biomedical sciences. A large part of medical reasoning concerns diagnosis and is essentially probabilistic. It would be an asset for biomedical ontologies to be able to support such a probabilistic reasoning.

Ledley & Lusted (1959)'s seminal article on Bayesian reasoning in medicine defines different kind of probabilistic entities. Consider for example the simple case of an instance of test of type A aiming at detecting if a patient in a group g has an instance of disease of type M[1]. The performance of test A in diagnosing M can be quantified by the positive predictive value of this test, hereafter abbreviated PPV, and generally defined as the proportion of people who have the disease among those who would be tested positive by A in g (that is, the proportion of true positives among positives); and by the negative predictive value, hereafter abbreviated NPV, and generally defined as the proportion of people who do not have the disease among those who would be tested negative by A in g (that is, the proportion of true negatives among negatives). Those two values provide the probability, once the result of test A is observed, that the patient has the disease M.

However, such positive and negative predictive values are typically not available in the scientific literature. Instead, they are generally computed from other probabilistic values, namely: the prevalence value of M in g, generally defined as the proportion of people who have the disease M in g, and hereafter abbreviated Prev(g,M); the sensitivity value of the test A for M in g, generally defined as the proportion of people who would get a positive result by A among those who have the disease M in g (that is, the proportion of true positives among diseased), hereafter abbreviated Se(g,A,M); and the specificity value of A for M, generally defined as the proportion of people who would get a negative result by A among those who do not have the disease M in g (that is, the proportion of true negatives among non-diseased), hereafter abbreviated Sp(g,A,M). As a matter of fact, these values are related through the following Bayesian equations:

$$\text{PPV}(g, A, M) = \frac{\text{Prev}(g, M)\,\text{Se}(g, A, M)}{\text{Prev}(g, M)\,\text{Se}(g, A, M) + (1 - \text{Prev}(g, M))\,(1 - \text{Sp}(g, A, M))}$$

$$\text{NPV}(g, A, M) = \frac{(1 - \text{Prev}(g, M))\,\text{Sp}(g, A, M)}{\text{Prev}(g, M)\,(1 - \text{Se}(g, A, M)) + (1 - \text{Prev}(g, M))\,\text{Sp}(g, A, M)}$$

In the wake of Ledley & Lusted (1959), the sensitivity and specificity values have often been considered as depending only on the pathophysiological characteristics of the disease, and thus as independent of the group of people under consideration. However, sensitivity and specificity values do in fact depend upon the group under consideration: this is the "spectrum effect" (Brenner & Gefeller, 1997; for a detailed explanation, see Barton, Duvauferrier & Burgun, 2015). Spectrum effect can be manifested, for example, as a dependence of sensitivity and specificity on the degree of severity of the disease in the group under consideration (Park, Yokota, Gill, El Rassi, & McFarland, 2005).

In the remainder of the articles, sensitivity, specificity, PPV and NPV will be called "indicators of diagnostic performance" and abbreviated "IDPs".

1.2   The challenge of representing indicators of diagnostic performance in an ontology

To the extent that they aim at representing biomedical knowledge and enabling medical reasoning, biomedical ontologies should provide a formalization of IDPs as well as

---

[1] These will be abbreviated in the following as "a test A" and "the patient has M".

the prevalence. This article will propose such a formalization in the context of the OBO Foundry (Smith et al., 2007), one of the most massive sets of interoperable ontologies in the biomedical domain, built on the upper ontology BFO.

The question of how probabilistic notions can be represented in ontologies has been tackled from different perspectives in the past. For example, da Costa et al. (2008) have proposed the new PR-OWL format that extends the classical OWL format; we take here a different approach, which does not aim at changing the OWL format. Soldatova, Rzhetsky, De Grave, & King (2013) have described a model in which probabilities can be assigned to research statements. We have proposed an alternative approach (Barton, Burgun, & Duvauferrier, 2012) in which we show how probabilities can be assigned to dispositions, upon which we are going to build here.

Sensitivity and specificity have been recently introduced in the Ontology of Biological and Clinical Statistics (OBCS; Zheng et al., 2014) as subclasses of *Data item* – a classification that we will endorse here, and extend to PPV and NPV. A data item, as defined by the Information Artifact Ontology (IAO), is intended to be a truthful statement about something. In order to formalize IDPs, one should thus clarify what entities in the real world they are about.

Sensitivity value[2], as we said, is generally defined as the proportion of people who would get a positive result by *A* among those who have the disease *M*. But note here the conditional structure: what is referred to is the proportion of true positives among diseased *if* A was performed on them. In practical situations, however, the sensitivity value will be estimated by performing the test on a sample of the population only – not the entire population g. This will lead to two difficulties. First, it will be necessary to differentiate clearly the IDPs' values from their estimations, and to determine which of those should be represented in a first place in an ontology – part 2 will be devoted to this issue. Second, possible-but-non-actual situations cannot be straightforwardly defined in a realist ontology like BFO; this problem will be explained and solved in part 3, by considering that an IDP is a data item about a disposition borne by an instance of group of individuals, whose probability value will be identified to the diagnostic value of the IDP. This will provide a formal characterization of IDPs.

2 THE INDICATORS AND THEIR ESTIMATIONS

2.1 Two limits for the estimations of indicators of diagnostic performance

Numerical estimations of IDPs face two limits (Barton et al., 2015). First, frequencies will be measured on a sample

judged to be representative of the population as a whole, and these values are then extrapolated to the frequencies in the entire population. Second, whether a given person has *M* or not cannot generally be known for sure, through reasonable means: sometimes, the only way to be certain is to perform an autopsy on the deceased patient. Consequently, a "gold standard" must be chosen, namely the best reasonable available diagnostic test[3]. If a patient gets a positive result to this gold standard test, one will conclude that he has the disease; if he gets a negative result, one will conclude that he does not have it.

For example, Park et al. (2005) estimate the sensitivity of the Neer test for diagnosing the impingement syndrome; their estimation is made on a sample of 552 patients considered as representative of the general population, using as gold standard surgical observation. The proportion of patients tested positive by the Neer test among *those who are tested positive by surgical operation* in the *sample* is considered as representative of the sensitivity value - which can be interpreted as the proportion of people who would be tested positive by the Neer test among *those who have an impingement syndrome* in the *whole population*. Similar estimation strategies hold for prevalence, specificity, PPV and NPV.

Note that the estimations of the values of the prevalence, sensitivity, specificity, PPV and NPV depend on both the sample and the gold standard; however, the real values of the prevalence, sensitivity, specificity, PPV and NPV, as defined above, depend neither on the sample, nor on the gold standard.

2.2 What should be represented in an ontology?

This being clarified, one can ask which entities should be preferably represented in an ontology: the IDPs' values, or their estimations?

For sure, we have no direct access to such IDPs' values; but this does not imply that they should not be represented in an ontology. To clarify why, consider an analogy: the measure of the ambient temperature by reading the height of a mercury column in a thermometer. Suppose that at a given time, this height is aligned with the sign "20 °C" written on the thermometer. In such a case, an ontology curator would be in a first place interested in formalizing the fact that the ambient temperature is 20°C, rather than in formalizing the fact that the mercury column in the thermometer is at the same height as the sign "20°C".

In a similar fashion, imagine that 65% of people are tested positive for a gold standard of *M* in a sample s of a population g. The ontology should then formalize in a first place the fact that 65% of the people in g have *M*, rather than the

---

[2] Note the distinction between a sensitivity and its value: a sensitivity is a data item, but its value is a number.

[3] Even if the gold standard consists in the naked-eye observation of a macroscopic disorder associated exclusively with this disease, this can still theoretically lead to a diagnostic error: any empirical evidence is defeasible.

fact that 65% of the people in s have a positive result to this gold standard. This estimation of this prevalence value may be false (it is indeed very likely to be false, strictly speaking), but future estimations will lead to its being corrected to bring it closer to the real value. As a matter of fact, realist ontologies are built according to a fallibilist methodology (Smith & Ceusters, 2010): they represent the state of the world according to our best knowledge at the present instant, and can be corrected as our knowledge of the world is refined.

That being said, it is possible to represent in an ontology the measurement process of a temperature involving the height of a mercury column in a thermometer. Similarly, one could represent the different estimation processes of the IDPs, and the results to which they led. Such processes are biomedical investigations, and should therefore be formalized in an ontology like OBI (Ontology for Biomedical Investigations, Brinkman et al., 2010), a prominent OBO Foundry candidate dedicated to these issues. This would be relevant in order to formalize in an ontology different estimations given by various samples and gold standards. However, medical practitioners are first and foremost interested in the IDPs' values themselves, rather than in their estimations, and thus we will deal here with the formalization of the former.

This clarification being made, we can now consider the second difficulty mentioned at the end of part 1, namely the formalization of possible-but-non-actual situations in BFO.

3 A FORMALIZATION OF INDICATORS OF DIAGNOSTIC PERFORMANCE IN APPLIED ONTOLOGIES

Sensitivity value has been interpreted as the proportion of people who would get a positive result to *A* among *M*'s bearers in *g*. This definition thus involves the condition of performing the test *A* on the members of g. As we said, such a condition is never realized, because the test is performed (at best) on a sample of the population, not on the whole population g: the performance of test *A* on *g*'s members is a *possible* (leaving aside practical difficulties), *non-actual* condition. Interpreting specificity, PPV, and NPV along the former lines would also imply such possible, non-actual conditions.

However, BFO is built according to the realist methodology, which implies that all the instances it recognizes should be *actual* entities (cf. Smith & Ceusters, 2010). Thus, one cannot represent directly such a possible-but-not-actual condition in an ontology based on BFO. In order to solve this difficulty, we will introduce a strategy named "randomization", enabling to formalize the probabilities of interest as assigned to an actual entity, namely a disposition. This strategy will enable to represent IDPs in a realist fashion, compliant with BFO's spirit.

3.1 From proportions to objective probabilities: the randomization strategy

We will explain first how the proportion of a subgroup in a group can be formalized as a probability value assigned to a disposition; this will help explaining later how the proportion of a subgroup in a group undergoing a possible, non-actual condition can be formalized along similar lines.

Dispositions are entities that can exist without being manifested; an example of disposition is the fragility of a glass, which can exist even when the glass does not break. We will use Röhl & Jansen's (2011) model of disposition in BFO, which associates to every instance of disposition one or several instances of realizations, and one or several instances of triggers (a trigger is the specific process that can lead to a realization occurring). In this model, the fragility of a glass is a disposition of the glass to break (the breaking process is the realization) when it undergoes some kind of stress (the process of undergoing such a stress is the trigger); this disposition inheres in the glass. Starting with the definition of these entities and their relations at the instance level, Röhl & Jansen proceed to formalize them at the universal level. We have shown in a former article (Barton, Burgun & Duvauferrier, 2012) how to adapt this model to probabilistic dispositions. Thus, an instance of balanced coin is the bearer of a disposition instance to fall on heads (the realization process) when it is tossed (the trigger process), to which an objective probability 1/2 can be assigned.

We will now apply this model to the situation at hand. Consider the prevalence Prev(g,*M*), which was defined above as the proportion of bearers of *M* in the actual population g. We can define the disposition $d_{g,M}$, borne by the group g, that a person randomly drawn in g has *M*. More specifically, let's write T_g the process "randomly drawing a person in g", and $R_{g,M}$ the process "drawing by T_g someone who has *M*": the triggers of $d_{g,M}$ are instances of T_g and its realizations are instances of $R_{g,M}$. Following the lines of Barton et al. (2012), one can thus define the probability assigned to the disposition[4] $d_{g,M}$, which is the probability of drawing randomly someone who has *M* in g. This probability is equal to the proportion of individuals who have *M* in g, that is, to Prev(g,*M*): as a matter of fact, if there are *e.g.* 10% diseased people in g, then the probability of drawing randomly a diseased person in g is 10%. Thus, the prevalence value can be identified to the objective probability assigned to the disposition $d_{g,M}$. We name this strategy the "randomization" of the proportion of *M*'s bearers among g.

---

[4] In Barton et al. (2012), a probability was assigned to a triplet (*d*, *T*, *R*) rather than to a disposition *d*, because we had to take into account disposition that may have several classes of triggers or realizations (that is, multi-trigger and multi-track dispositions, cf. Röhl & Jansen, 2011). However, in the present situation, $d_{g,M}$ is simple-trigger and simple-track: all its triggers are instances of T_g, and all its realizations are instances of $R_{g,M}$. Therefore, the probability value assigned to ($d_{g,M}$, T_g , $R_{g,M}$) can be, for practical matters, assigned directly to $d_{g,M}$.

The randomization strategy may not be necessary to formalize a prevalence, which characterizes a proportion in an actual group, and thus could be formalized as such in an ontology based on BFO. But this strategy can also be applied to proportions of people in groups subject to a *possible, non-actual* condition – and thus, be relevant to formalize sensitivity and other IDPs. As a matter of fact, the sensitivity value Se(g,*A*,*M*) was defined as the proportion of people who would get a positive result to *A* among *M*'s bearers in g. This value can be "randomized" as follows. We can define $d_{Se,g,A,M}$ as the disposition to draw someone randomly who is tested positive by *A*, among the individuals of g who have *M*. More specifically, let's define the process $T_{Se,g,A,M}$ as the "performance of test *A* on the individuals in g, and random draw of an individual among those who have the disease *M*"[5]; and the process $R_{Se,g,A,M}$ as the "drawing by $T_{Se,g,A,M}$ of someone who got a positive result to *A*". The triggers of $d_{Se,g,A,M}$ are instances of $T_{Se,g,A,M}$, and its realizations are instances of $R_{Se,g,A,M}$. One can then define the sensitivity value Se(g,*A*,*M*) as the objective probability assigned to this disposition $d_{Se,g,A,M}$: indeed, if there are *e.g.* 15% of the diseased people in g who would get a positive result by *A*, then the probability of randomly drawing someone who would get a positive test result by *A* among diseased people in g is equal to 15%.

Specificity value can be defined along similar lines, as probabilities assigned to actual dispositions borne by the group g noted $d_{Sp,g,A,M}$ (and similarly for the PPV and NPV). Although $d_{Se,g,A,M}$ and $d_{Sp,g,A,M}$ are both dispositions inhering in g, they have different triggers and different realizations; the process $T_{Sp,g,A,M}$ is the "performance of test *A* on the individuals in g, and random draw of an individual among those who *do not* have the disease *M*" and the process $R_{Sp,g,A,M}$ is the "drawing by $T_{Sp,g,A,M}$ of someone who got a *negative* result to *A*".

3.2 A formal model of indicators of diagnostic performance in ontologies

Let us now consider how to formalize these probability values in ontologies. First, a group g will be considered as any collection of humans (for more on collections, see Jansen & Schultz, 2010). $d_{Se,g,A,M}$ is a disposition individual inhering in the group g; and a probability value can be assigned to this disposition using a datatype property **has_probability_value**. Sensitivity of *A* for *M* in g will be denoted Se$_{g,A,M}$, and following OBCS, it will be defined as a data item. Thanks to our analysis above, we can now answer our original question, and state what this sensitivity is about: Se$_{g,A,M}$ **is_about** $d_{Se,g,A,M}$. We can also introduce a relation **has_diagnostic_value** that relates a sensitivity to its value.

In our framework, the (diagnostic) value of a sensitivity Se$_{g,A,M}$ is the probability value assigned to the disposition $d_{Se,g,A,M}$; this can be formalized by writing that if s is a sensitivity, then:

s **has_diagnostic_value** p ⇔ ∃ d ∧ d **is_a** *Disposition* ∧ s **is_about** d ∧ d **has_probability_value** p

As $d_{Se,g,A,M}$ is an individual, it cannot be related directly to the universals *A* and *M*. However, it can be related indirectly to them, by the following formalization. First, $d_{Se,g,A,M}$ can be seen as an instance of a disposition universal symbolized as $D_{Se,A,M}$, which has as trigger the processus universal $T_{Se,A,M}$: "performance of test *A* on the members of a group, and random draw of a person among those who have the disease *M*"; and as realization the process universal $R_{Se,A,M}$ defined as "drawing by $T_{Se,A,M}$ of someone who got a positive result to *A*". We can then introduce two new relations *sensitivity_disposition_of_test* and *sensitivity_disposition_for_disease* (abreviated as *se_of_test* and *se_for_disease*) such that $D_{Se,A,M}$ *se_of_test A* and $D_{Se,A,M}$ *se_for_disease M*. These two relations are introduced for pragmatic reasons of facility of use: on a foundational level, $D_{Se,A,M}$ and *M* (resp. *A*) could be related through a complex array of relations and entities that involve the relation *has_trigger* between $D_{Se,A,M}$ and $T_{Se,A,M}$, as well as a sequence of relations between $T_{Se,A,M}$ and *M* (resp. *A*). Such an analysis would raise theoretical issues though, as instances of $D_{Se,A,M}$ can exist even if no instance of *M* or *A* do exist. We would therefore face here issues similar to the ones addressed by Röhl & Jansen (2011) and Schulz et al. (2014).

Finally, we introduce a class *Sensitivity* that can be characterized as a subclass of *Data item*, which is related to a disposition through the above-mentioned relations:

s **instance_of** *Sensitivity* ⇒ s **instance_of** *Data item* ∧ ∃ d **instance_of** *Disposition* ∧ ∃ a **instance_of** *Test* ∧ ∃ m **instance_of** *Disease* ∧ s **is_about** d ∧ d **se_of_test** a ∧ d **se_for_disease** m

We can also introduce $Se_{A,M}$, the class of sensitivities of test *A* for disease *M* (in whatever group), which can be formalized as a subclass of *Sensitivity* related to *M* and *A* through the following relations:

s **instance_of** $Se_{A,M}$ ⇒ s **instance_of** *Sensitivity* ∧ ∃ d **instance_of** *Disposition* ∧ ∃ a **instance_of** *A* ∧ ∃ m **instance_of** *M* ∧ s **is_about** d ∧ d **se_of_test** a ∧ d **se_for_disease** m

Figure 1 summarizes this formalization of sensitivity (with universals in boxes, instances in diamonds, and the numerical value assigned by datatype properties in a circle). Specificity, PPV and NPV can be formalized along similar lines, as data items about dispositions related to tests and diseases through relations that could be labeled *sp_of_test*, *sp_of_disease*, *ppv_of_test,* etc.

---

[5] In general, we cannot determine in practice with certainty which individuals of g have *M*, and which do not; but the practical impossibility to realize this trigger does not preclude to define this entity.

4 CONCLUSION

We have thus provided a practically tractable formalization of IDPs in a realist ontology, which clearly dissociates IDPs from their estimations (which are relative to a sample and a gold standard). It also solves the difficulty of considering possible, non-actual conditions in a realist ontology based on BFO.

Note that IDPs raise also other theoretical issues. For example, one may want to aggregate two sensitivity values Se(g,A,M) and Se(g',A,M) assigned to two different groups g and g' in order to reach a finer assessment of the sensitivity in a larger group; how to do this is a question for the meta-analyst though, not the ontologist, who is first and foremost concerned with representational issues.

This model could then be extended in three directions. A first step would consist in formalizing the estimations of the IDPs, and their relations to a given sample and gold standard. Second, the relations *se_of_test* and *se_for_disease* could be reduced to basic relations and entities already accepted in the OBO Foundry. Third, it could be used by ontology-based diagnostic systems that would compute positive predictive values or negative predictive values from the prevalence, sensitivity and specificity values; more generally, it could be articulated with medical Bayesian networks.

As it takes into account the dependence of IDPs upon the group of people considered, it has the potential to contribute to the development of precision medicine (Mirnezami, Nicholson & Darzi, 2012), an emerging approach that takes into consideration patients characteristics and dispositions, including individual variability in genes, to offer more personalized preventive, diagnostic and therapeutic strategies.
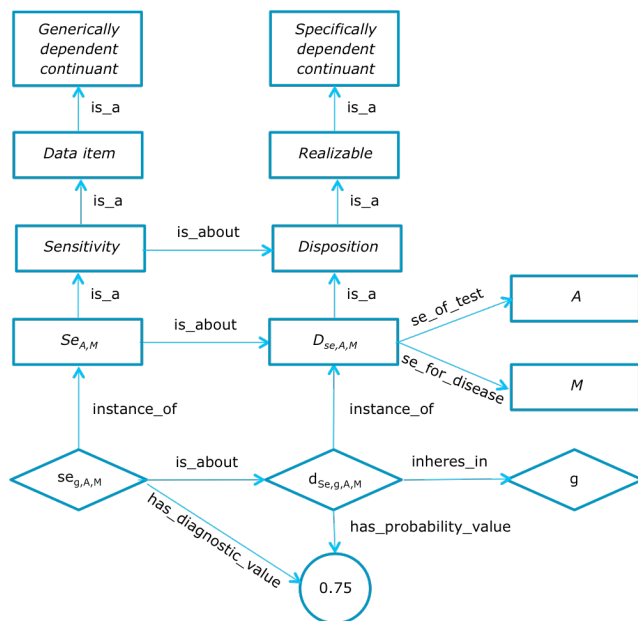
Figure 1 Sensitivity of a test *A* for a disease *M* in a group g with probability value 0.75

REFERENCES

Barton, A., Burgun, A., and Duvauferrier, R. (2012) Probability assignments to dispositions in ontologies. *Proc. 7th Int. Conf. Form. Ontol. Inf. Syst. FOIS2012* (M. Donnelly & G. Guizzardi, eds.), 3–14, Amsterdam: IOS Press.

Barton, A., Duvauferrier, R. and Burgun, A. (2015) *Une analyse philosophique des indicateurs de performance des tests diagnostiques médicaux*. Submitted manuscript.

Brenner, H. and Gefeller, O. (1997) Variation of sensitivity, specificity, likelihood ratios and predictive values with disease prevalence. *Stat. Med.*, **16** (9), 981–991.

Brinkman, R. R., Courtot, M., Derom, D., Fostel, J. M., He, Y., Lord, P., Malone, J., et al. (2010) Modeling biomedical experimental processes with OBI. *J. Biomed. Semant.*, **1 Suppl 1**, S7.

Costa, P. C. G. da, Laskey, K. B. and Laskey, K. J. (2008) PR-OWL: A Bayesian ontology language for the semantic web. In: *Uncertainty Reasoning for the Semantic Web I*, 88–107, Springer.

Jansen, L., and Schulz, S. (2011) Grains, components and mixtures in biomedical ontologies. *J. Biomed. Semant.*, **2 Suppl 4**, S2.

Ledley, R. S. and Lusted, L. B. (1959) Reasoning foundations of medical diagnosis. *Science*, **130**(3366), 9–21.

Mirnezani, M.R.C.S., Nicholson, J. and Darzi, A. (2012) Preparing for precision medicine. *N. Engl. J. Med.*, **366**(6), 489-491.

Park, H. B., Yokota, A., Gill, H. S., El Rassi, G., McFarland, E. G. (2005) Diagnostic accuracy of clinical tests for the different degrees of subacromial impingement syndrome. *J. Bone Joint Surg. Am.*, **87**(7), 1446–1455.

Röhl, J., Jansen, L. (2011) Representing dispositions. *J. Biomed. Semant.*, **2** (Suppl 4), S4.

Schulz, S., Martínez-Costa, C., Karlsson, D., Cornet, R., Brochhausen, M., and Rector, A. (2014) An Ontological Analysis of Reference in Health Record Statements. *Proc. 8th Int. Conf. Form. Ontol. Inf. Syst. FOIS2014* (P. Garbacz & O. Kutz, eds.), 289-302, Amsterdam: IOS Press.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., et al. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.*, **25**(11), 1251–1255.

Smith, B. and Ceusters, W. (2010) Ontological realism: A methodology for coordinated evolution of scientific ontologies. *Appl. Ontol.*, **5**(3), 139–188.

Soldatova, L. N., Rzhetsky, A., Grave, K. De and King, R. D. (2013) Representation of probabilistic scientific knowledge. *J. Biomed. Semant.*, **4**(Suppl 1), S7.

Zheng, J., Harris, M. R., Masci, A. M., Lin, Y., Hero, A., Smith, B., and He, Y. (2014). OBCS: The Ontology of Biological and Clinical Statistics in ICBO2014. *Houston, TX, USA*