What is a risk? A formal representation of risk of stroke for people with atrial fibrillation

Adrien Barton[1,*] Ludger Jansen[2], Arnaud Rosier[3] and Jean-François Ethier[1,3,*]
[1] GRIIS, Université de Sherbrooke, Québec, Canada
[2] Ruhr University Bochum and University of Rostock, Germany
[3] INSERM UMR_S 1138 Eq 22, Paris Descartes University, Paris, France

ABSTRACT

We propose a framework for the representation of medical risks in the context of the OBO Foundry using the Web Ontology Language (OWL). The framework is developed for the use case of risk of stroke for people with atrial fibrillation, for which we distinguish three classes of dispositions: the atrial fibrillation disease; the risk of stroke for a human who has atrial fibrillation; and the risk of stroke over 12 months for a human who has atrial fibrillation. The latter is quantified by risk estimates, which are informational entities extracted from documents – such as journal articles – and to which epistemic probability values can be assigned. We discuss the reference-class problem (i.e., the possibility to have several risk estimates with different epistemic probabilities for the same individual, depending on the reference class the risk estimate is based on) and clarify the philosophical hypotheses on which this dispositional framework is based.

1 INTRODUCTION

Risks of adverse outcomes are ubiquitous in the medical domain and have a central importance. An older population with complex combinations of chronic diseases and many medications makes simple deterministic treatment decisions difficult. Instead, clinicians need to assess, manage, and balance risks much more explicitly than ever before. It would therefore be valuable if ontologies aiming at adequately representing medical knowledge could formalize such risks. This paper contributes to this aim by proposing a framework for the representation of risks, illustrated by the risk of stroke in people with atrial fibrillation. More specifically, we propose a representation of absolute risks, such as a 3.2% risk of stroke over 12 months for people with atrial fibrillation (Nielsen *et al.*, 2016).

This formalization is expressed in the Web Ontology Language (OWL), in the context of the OBO Foundry (Smith *et al.*, 2007). The OBO Foundry is one of the most comprehensive collections of interoperable ontologies in the biomedical domain, built on the upper ontology Basic Formal Ontology (BFO) 2.0 (Arp, Smith & Spear, 2015). A few ontologies have formalized the notion of medical risk; see Uciteli *et al.* (2016) for a recent account (though not in the context of the OBO Foundry), as well as a review of former accounts. However, there is currently no comprehensive account of the notion of risk in the OBO Foundry ontologies.

One principle of BFO is the strict separation between universals and their instances. We write names of instances as well as relations between instances in bold, and names of universals and defined classes in italics. When first introduced, names of universals will be prefixed by the name of the source ontology (e.g., "BFO:*Disposition*"), unless the context makes it obvious.

The OBO Foundry compliant Ontology of Biological and Clinical Statistics (OBCS; Zheng *et al.*, 2016) defines *Absolute risk* as a subclass of IAO:*Information content entity* ("*ICE*" for short). However, arguably, a person with atrial fibrillation has a risk to get a stroke independently of whether or not there exists some *ICE* estimating his risk to get a stroke. The risk itself has rather a dispositional character: an instance of risk of an adverse outcome of type *A* may be realized by an instance of *A*, but it may also never be realized; however, whether it is realized or not, the risk still exists. For this reason, this paper formalizes risks as dispositions that can be estimated by a specific kind of *ICE*, risk estimates, and the risk probability values as assigned to these risk estimates.

We begin by distinguishing two types of dispositions: the disease of atrial fibrillation on the one hand, and the risk of stroke of a human who has atrial fibrillation on the other (Sect. 2). We then show how a probability can be assigned to a risk of stroke in 12 months for a human with atrial fibrillation (Sect. 3). A discussion and conclusion follow.

2 DIFFERENTIATING RISK DISPOSITION AND DISEASE

The OGMS (Ontology for General Medical Science) considers a *Disease* as a BFO:*Disposition* (Scheuermann, Ceusters & Smith, 2009). Röhl & Jansen (2011) developed an axiomatisation of dispositions in the context of BFO. In this model, a disposition is a BFO:*Dependent continuant* that **inheres_in** his bearer (which is the **bearer_of** this property), a *Material entity*, and may be realized (**realized_in**) via a process. The realization process has the material entity as a participant (**has_participant**), and the disposition is triggered by (**has_trigger**) some event or process. Finally, according to

* To whom correspondence should be addressed: ethierjf@gmail.com; adrien.barton@gmail.com

BFO, a disposition **has_material_basis** some entity. For example, the fragility of a glass is formalized as a disposition inhering in the glass, that may be realized by a breaking process when some form of stress (the trigger) happens; moreover, the fragility exists because of some molecular structure of the glass, which is its material basis.

The OGMS model considers *Disease* as a disposition **realized_in** a *Disease course* that has as parts some *Pathological process*. The material basis of a disease is a *Disorder* in the organism. For example, the disease epilepsy is seen as a disposition to have a disease course composed by various epileptic crises (pathological processes), because of some disorder in the brain.

The OGMS model is applied to cardiovascular diseases by the Cardiovascular Disease Ontology (CVDO; Barton *et al.*, 2014). It formalizes the atrial fibrillation disease *Atrial fibrillation* as a disposition realized by a disease course that has as parts some processes of atrial fibrillation:

Atrial fibrillation subClassOf *Disease*

Atrial fibrillation process subClassOf *Pathological process*

Atrial fibrillation subClassOf (**realized_in** some *Disease course* and (**has_part** some *Atrial fibrillation process*))

This matches to an ambiguity in the natural language term "atrial fibrillation": it refers sometimes to a pathological process of atrial fibrillation (namely, irregular, uncoordinated contractions of the atria of the heart), and sometimes to a disease – a disposition exceeding a given threshold (Scheuermann *et al.*, 2009) to atrial fibrillation processes.

Consider a human **Jones**, who has an atrial fibrillation disease ("AF" for short) $\mathbf{af}_{\mathbf{Jones}}$ – an instance of the universal *Atrial fibrillation*. Jones is an instance of $Human_{AF}$, the class of humans with atrial fibrillation:

$Human_{AF}$ equivalentClass *Human* and (**bearer_of** some *Atrial fibrillation*)

Suppose that $\mathbf{af}_{\mathbf{Jones}}$ leads to the process instance $\mathbf{stroke}_{\mathbf{Jones}}$ (an instance of the universal *Stroke*) through the following scenario. There is some fibrosis in Jones' atrial myocardium (the disorder $\mathbf{atrium\_fibrosis}_{\mathbf{Jones}}$), which is the material basis of the disposition $\mathbf{af}_{\mathbf{Jones}}$:

$\mathbf{af}_{\mathbf{Jones}}$ **inheres_in Jones**

$\mathbf{af}_{\mathbf{Jones}}$ **has_material_basis** $\mathbf{atrium\_fibrosis}_{\mathbf{Jones}}$

$\mathbf{af}_{\mathbf{Jones}}$ is realized by a (long) disease course $\mathbf{af\_course}_{\mathbf{Jones}}$, that encompasses various pathological processes, including several episodes of atrial fibrillation ($\mathbf{pp}_{\mathbf{af,1}}$,...,$\mathbf{pp}_{\mathbf{af,n}}$):

$\mathbf{af}_{\mathbf{Jones}}$ **realized_in** $\mathbf{af\_course}_{\mathbf{Jones}}$

For every $i \in [1,n]$: $\mathbf{af\_course}_{\mathbf{Jones}}$ **has_part** $\mathbf{pp}_{\mathbf{af,i}}$

These pathological processes lead to the development of a blood clot in Jones' atrium, which is a new disorder. This blood clot is the bearer of a disposition to dislodge and migrate, which is at some point realized by the process of the blood clot migrating to the brain. The migrating clot has then a disposition to get stuck in a cerebral artery - by contrast to dissolve. When it gets stuck, the blood flow is blocked and $\mathbf{stroke}_{\mathbf{Jones}}$ happens.

On top of the disposition $\mathbf{af}_{\mathbf{Jones}}$, Jones is also the bearer of another disposition: the risk of stroke $\mathbf{risk}_{\mathbf{Jones},\mathit{Stroke}}$, which is realized by his stroke:

$\mathbf{risk}_{\mathbf{Jones},\mathit{Stroke}}$ **inheres_in Jones**

$\mathbf{risk}_{\mathbf{Jones},\mathit{Stroke}}$ **realized_in** $\mathbf{stroke}_{\mathbf{Jones}}$

$\mathbf{risk}_{\mathbf{Jones},\mathit{Stroke}}$ is an instance of $Risk_{AF,Stroke}$, the class of risks of stroke for humans with atrial fibrillation, which is itself a subclass of *Risk*:

Risk subClassOf *Disposition*

$Risk_{AF,Stroke}$ subClassOf *Risk*

$Risk_{AF,Stroke}$ subClassOf (**inheres_in** some $Human_{AF}$)

$Human_{AF}$ subClassOf (**bearer_of** some $Risk_{AF,Stroke}$)

$Risk_{AF,Stroke}$ subClassOf (**realized_in** only *Stroke*)

An instance of $Risk_{AF,Stroke}$ may be realized by one or several instance(s) of stroke, or may remain unrealized. To clarify the triggers of $Risk_{AF,Stroke}$, we need to use BFO:*History*. BFO defines the history of a material entity as the "process that is the sum of the totality of processes taking place in the spatiotemporal region occupied by a material entity or site, including processes on the surface of the entity or within the cavities to which it serves as host" (Arp, Smith and Spear, 2015). We define *History-part* as the class of temporal parts of the history of any material object:

History-part equivalentClass (**part_of** some *History*)

In our formalization, $Risk_{AF,Stroke}$ is triggered by any *History-part* of its bearer:

$Risk_{AF,Stroke}$ subClassOf (**has_trigger** some *History-part*)

This way, risks of stroke are dispositions that are always triggered. However, $Risk_{AF,Stroke}$ is not a sure-fire disposition (that is, a disposition that is always realized when triggered), but a tendency (that is, a disposition that is not always realized when triggered; Jansen 2007; Röhl & Jansen, 2011).

The material basis of $\mathbf{risk}_{\mathbf{Jones},\mathit{Stroke}}$ is a disorder that has as part Jones' fibrosis, but also other entities. As a matter of fact, Jones can have a stroke because of his atrial fibrillation, but also because of various random or progressive factors, such as the regular senescence of his blood vessels.

risk$_{\mathbf{Jones},Stroke}$ **has_material_basis** some (*Disorder* and **has_part atrium_fibrosis**$_{\mathbf{Jones}}$ and **has_part senescent_blood_vessels**$_{\mathbf{Jones}}$)

Thus, the disease **af**$_{\mathbf{Jones}}$ and the risk **risk**$_{\mathbf{Jones},Stroke}$ are not the same disposition: even if they both inhere in Jones, they have distinct material basis and distinct realizations – namely, **af_course**$_{\mathbf{Jones}}$ vs. **stroke**$_{\mathbf{Jones}}$. (The OGMS model leaves open the question whether pathological processes that are caused - or partially caused - by earlier pathological process of a disease course are also part of this disease course; thus, it is an open question whether **stroke**$_{\mathbf{Jones}}$ is a part of **af_course**$_{\mathbf{Jones}}$; see Barton *et al.* 2014 for a discussion. In any case, those two instances are distinct entities.)

3 PROBABILITY ASSIGNMENTS TO RISK DISPOSITIONS

We can now turn to the representation of a probability assignment to a risk of stroke. For this purpose, we first discuss the entity characterized by a probability assignment, then discuss the nature of probabilities at play, and finally formalize the probability assignment to a risk estimate.

3.1 What kind of entity do risk probabilities characterize?

Nielsen *et al.* (2016) studied a nationwide cohort of patients for which the overall ischemic stroke rate was 3.20 per 100 person-years. However, this value 3.2% does not relate only to a proportion in this cohort: we can infer from this information that an unspecified human from the same population with atrial fibrillation has a 3.2% probability to have a stroke over 12 months – even, of course, if he was not one of the patients in the cohort. Therefore, as we will now argue, the probability 3.2% also characterizes a certain property of $Human_{AF}$: their risk to have a stroke over 12 months.

Dispositions are natural targets for probability assignments (Barton et al. 2012). However, we cannot assign the probability 3.2% to the disposition $Risk_{AF,Stroke}$. As a matter of fact, 3.2% is the probability of a human with AF to have a stroke *over 12 months* – but $Risk_{AF,Stroke}$ does not have any ontological connection with 12-months-long processes.

To solve this issue, let's define $History\text{-}part_{12m}$ as the subclass of *History-part* with a 12 months-long duration. Let's now introduce $Risk_{AF,12m,Stroke}$ the class of risks of a human with AF to get a stroke over 12 months, that we also formalize as a *Risk* – and therefore, a disposition:

$Risk_{AF,12m,Stroke}$ subClassOf *Risk*

Like with $Risk_{AF,Stroke}$, there is an instance of $Risk_{AF,12m,Stroke}$ inhering in any person with AF; and those instances can only be realized by a stroke:

$Human_{AF}$ subClassOf (**bearer_of** some $Risk_{AF,12m,Stroke}$)
$Risk_{AF,12m,Stroke}$ subClassOf (**realized_in** only *Stroke*)

By contrast to $Risk_{AF,Stroke}$ which is triggered by all history-parts of its bearer, $Risk_{AF,12m,Stroke}$ is only triggered by all 12-months-long history-parts of its bearer:

$Risk_{AF,12m,Stroke}$ subClassOf (**has_trigger** some $History\text{-}part_{12m}$)

In order to determine how a probability can characterize $Risk_{AF,12m,Stroke}$, we need to clarify the ontological status of probabilities.

3.2 What are the probabilities?

Standardly, objective and epistemic interpretations of probabilities are distinguished (Hájek, 2012). Objective probabilities are meant to characterize the world independently of our knowledge of it, while epistemic interpretations consider probabilities to describe our knowledge of the world: epistemic probabilities can be defined as degrees of belief or degrees of confidence.

Consider for example a biased coin that has three times more chances to fall on heads than on tails. The objective probability of the coin falling on heads is ¾, and its objective probability to fall on tails is ¼. If Mr. Green knows about the coin's bias, he should assign epistemic probabilities with the same values: ¾ to heads and ¼ to tails; this is a consequence of a principle of rationality called the "principal principle" (Lewis, 1980). However, if Mr. White is not aware of this bias and thinks that the coin is balanced, he would assign epistemic probabilities ½ to heads and ½ to tails. Epistemic probabilities can be operationalized as rational, hypothetical betting coefficients – that is, coefficients indicating which odds should be considered as acceptable by the agent to bet on the occurrence of heads or tails (Maher, 1997).

Suppose that the 3.2% value would be an objective probability that could be assigned to the disposition $Risk_{AF,12m,Stroke}$. We would then have:

$Risk_{AF,12m,Stroke}$ subClassOf (**has_objective_probability** 0.032)

This would imply that for every **r instance_of** $Risk_{AF,12m,Stroke}$:

r has_objective_probability 0.032

Thus, all people with AF would have an objective probability 3.2% to have a stroke over 12 months. However, this cannot be the case: many people with AF have a lower or higher objective probability to have a stroke over 12 months, depending on various factors (see below the section 4.2 on $CHADS_2$ and CHADSVASC scores). Therefore, we would rather interpret 3.2% as an epistemic probability that characterizes the rational degree of confidence, given the evidence provided by Nielsen *et al.* (2016), that a person with AF would have a stroke over 12 months. We will now propose a formalization along those lines.

3.3 Epistemic probability assignment

We define the relation **object_of** as the inverse of IAO:**is_about** (Ceusters & Smith, 2010), which relates an *ICE* to what it is about. We introduce the class *Risk estimate* as a subclass of *ICE*. Let $\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$ be the following instance of *Risk estimate*: the estimate of the risk of a human with atrial fibrillation to have a stroke over 12 months, extracted from the article Nielsen *et al. (*2016)[1]:

$Risk_{AF,12m,Stroke}$ subClassOf (**object_of** $\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$)

To represent the assignment of the probability 3.2% to this risk estimate, we use two OBI relations (which are currently being formalized by the OBI development team), the object property **has_value_specification** (that relates an information content entity to an OBI:*Value specification*) and the datatype property **has_specified_value** (that relates a value specification to its numerical value). As a shortcut, let's introduce here the datatype property **has_value** defined as **has_value_specification** o **has_specified_value**. We can then write:

$\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$ **has_value** 0.032

Informally, if *R* is a risk and **re** is a risk estimate, (*R* subClassOf **object_of re**) and (**re has_value p**) mean together that according to the risk estimate **re**, it is rational to assign an epistemic probability p to the risk *R*.

The evidence for the estimate is documented in **Nielsen_et_al._(2016)**, an instance of IAO:*Journal article* (which is, in turn, a subclass of IAO:*Document*). To formally relate this journal article with $\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$, we introduce a relation **extracted_from**, whose domain is *Information content entity* and whose range is *Document*. If **r** is a risk estimate and **j** is a document, **r extracted_from j** implies that **j** participates in a IAO:*Planned process* whose specified output is **r**:

extracted_from subRelationOf **(is_specified_output_of** o **has_participant)**

Then, we can state:

$\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$ **extracted_from Nielsen_et_al._(2016)**

Altogether, the relations we have introduced here and in section 3.1 mean that a risk estimate is extracted from Nielsen *et al.* (2016), according to which it is rational to assign a 3.2% epistemic probability to the risk of stroke over 12 months for a person who has AF (in the absence of additional information).

[1] As in OWL the term "value" is used in a class restriction to introduce an individual after an object property, this could more specifically be written as

$Risk_{AF,12m,Stroke}$ subClassOf (**object_of** value $\textbf{risk\_estimate}_{AF,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$)

4 DISCUSSION

4.1 The reference class problem

As mentioned in section 3.2, all patients with atrial fibrillation do not have the same objective probability of stroke. The $CHADS_2$ score (congestive heart failure, hypertension, age ≥ 75 years, diabetes mellitus, stroke) is a tool that has been developed to predict the risk of stroke in patients with atrial fibrillation by stratifying patients into risk groups (Gage *et al.*, 2001). It was later expanded into the CHA_2DS_2-VASc score (Lip *et al.*, 2010; written "CHADSVASC" from now on), which includes three additional risk factors: vascular disease, age 65-74 years, and female sex.

Let $Human_{AF2}$ be the class of humans with atrial fibrillation and a CHADSVASC score of 2 ("AF2" for short). We can introduce a class of dispositions $Risk_{AF2,12m,Stroke}$, the risk of stroke over 12 months for people with AF2:

$Risk_{AF2,12m,Stroke}$ equivalentClass ($Risk_{AF,12m,Stroke}$ and **inheres_in** some $Human_{AF2}$)

Nielsen *et al.* (2016) state that the rate of stroke over 12 months among patients in the sample who had AF2 was 1.97%. Therefore, there is an instance $\textbf{risk\_estimate}_{AF2,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$ such that:

$Risk_{AF2,12m,Stroke}$ subClassOf **object_of** $\textbf{risk\_estimate}_{AF2,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$

$\textbf{risk\_estimate}_{AF2,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$ **has_value** 0.0197

Suppose that **Jones** has AF2. **Jones** is the bearer of the risk to get a stroke over 12 months $\textbf{risk}_{\textbf{Jones},12m,Stroke}$. Since **Jones instance_of** $Human_{AF2}$, his risk to get a stroke over 12 months is an instance of the class of risks to get a stroke over 12 months for people with AF2:

$\textbf{risk}_{\textbf{Jones},12m,Stroke}$ **instance_of** $Risk_{AF2,12m,Stroke}$

and therefore:

$\textbf{risk}_{\textbf{Jones},12m,Stroke}$ **object_of** $\textbf{risk\_estimate}_{AF2,12m,Stroke,\textbf{Nielsen\_et\_al.\_(2016)}}$

Moreover, since $Human_{AF2}$ subClassOf $Human_{AF}$, **Jones instance_of** $Human_{AF}$. Therefore, his risk to get a stroke over 12 months is an instance of the class of risks to get a stroke over 12 months for people with AF:

$\textbf{risk}_{\textbf{Jones},12m,Stroke}$ **instance_of** $Risk_{AF,12m,Stroke}$

and therefore:

$\mathbf{risk}_{\mathbf{Jones},12m,Stroke}$ **object_of**
$\mathbf{risk\_estimate}_{AF,12m,Stroke,\mathbf{Nielsen\_et\_al.\_(2016)}}$

Thus, $\mathbf{risk}_{\mathbf{Jones},12m,Stroke}$ is the object of two different estimates with two different probability values (0.032 and 0.0197), based on two different reference classes (AF or AF2): this is the reference class problem (Hájek, 2007). This is ontologically sound: if Dr. Khan only knows that Jones has AF, it is rational for him, based on Nielsen *et al.* (2016), to assign a probability 3.2% to the risk that Jones will have a stroke over 12 months; and if Dr. Patel knows in addition that Jones has a CHADSVASC score of 2, then it is rational for him, based on Nielsen *et al.* (2016), to assign a probability of 1.97% to this risk.

However, this raises practical difficulties. It might seem at first sight rational, for a computer system who has the information that Jones has a CHADSVASC score of 2, to always give precedence to the 1.97% risk estimation over the 3.2% estimation, as it is based on more specific factors. However, other criteria may matter. For example, if both values had been obtained from different studies, the 3.2% could be considered as a more reliable value for other reasons – such as a smaller 95% confidence interval.

Moreover, different articles relating about different cohorts or samples might give risk estimates with different values of the same risk class. They might give also risk estimates for risk classes that are not included into each others. Suppose that Jones has atrial fibrillation and is a smoker, and that we know two data from two different cohorts: the probability p_{AF} that someone with atrial fibrillation will have a stroke during ten years; and the probability p_{Smoker} that a smoker will have a stroke during ten years. There is no easy way to decide which epistemic probability is the best to estimate Jones' risk, or how they should be weighted in a common probability estimate. Note however that this is a classical issue for probabilistic reasoning, independent of the ontological representation chosen here.

4.2 Articulating objective and epistemic probabilities

We have seen earlier that we could not formalize in OWL 3.2% as an objective probability assigned to $Risk_{AF,12m,Stroke}$, as it would imply that every instance of this risk ($\mathbf{risk}_{\mathbf{Jones},12m,Stroke}$, $\mathbf{risk}_{\mathbf{Hubbard},12m,Stroke}$, etc.) would have the same objective probability – which is false.

An alternative reading would be to interpret the objective probability 3.2% in line of Barton, Burgun & Duvauferrier (2012) as assigned only to the universal $Risk_{AF,12m,Stroke}$, but not to its instances – a conception that is not straightforwardly implementable in OWL. Informally, this assignment would be elucidated as follows: in a hypothetical, representative sequence $\mathbf{seq_0}$ of instances of $Risk_{AF,12m,Stroke}$ inhering in hypothetical instances of $Human_{AF}$, the proportion of those risks who are realized – that is, the proportion of those humans who have a stroke over 12 months – tends towards 0.032 as the size of the sequence tends towards infinity.

This conception raises the issue of what it means to have a *representative* sequence of hypothetical instances of $Risk_{AF,12m,Stroke}$ inhering in hypothetical instances of $Human_{AF}$. Indeed, several factors can influence the risk of having a stroke – in particular those involved in the CHADSVASC score: hypertension, age ≥ 75 years, etc. It makes sense to speak of a representative sequence of instances only by reference to an actual population $\mathbf{pop_0}$: to be representative of $\mathbf{pop_0}$, the sequence $\mathbf{seq_0}$ should involve the same proportions as in $\mathbf{pop_0}$ of people with hypertension, of people older than 75 years, etc. But this implies that the probability 0.032 would characterize the actual population $\mathbf{pop_0}$ (that is, a collection of human particulars – cf. Jansen & Schulz, 2011) rather than a subclass of *Human*. This is a possible orientation, pursed by Barton, Ethier, Duvauferrier & Burgun (2017) to formalize indicators of diagnostic performance.

This article has used an alternative interpretation of probabilities as epistemic in nature. This interpretation makes the formalization simpler in the present context, as it enables to relate a probability estimate to a universal of human. Future work will need to discuss further the articulation between the objective and epistemic probability views, and compare the strengths of each.

4.3 Generalization of this formalization

Note that this formalization can be adapted to represent a risk during a process that is not characterized by its duration (such as 12 months), but by some other characteristics. Imagine for example that we want to represent the probability p that a human with AF would have a stroke during a hospitalization process; we would then introduce the risk $\mathbf{risk}_{\mathbf{Jones},Hospitalization,Stroke}$ that Jones would have a stroke during a hospitalization process, and assign the probability p to its risk estimate (the class of triggers of this risk would be the class of history-parts of Jones that temporally span any of his hospitalization process).

5 CONCLUSION

This article has shown how a specific risk and its various probability estimates could be formalized in the context of the OBO Foundry. We took the example of the risk of stroke for people with AF over 12 months $Risk_{AF,12m,Stroke}$, which was formalized as a disposition. The article introduced $\mathbf{risk\_estimate}_{AF,12m,Stroke,\mathbf{Nielsen\_et\_al.\_(2016)}}$, related (by the relation **extracted_from**) to the instance of *Journal article* **Nielsen_et_al._(2016)**. It was also related (by the relation **is_about**) to the risk $Risk_{AF,12m,Stroke}$, which was itself related to the following relevant classes: humans with atrial fibrillation $Human_{AF}$ (by the relation **inheres_in**); 12-months-long history-parts $History\text{-}part_{12m}$ (by the relation **has_trigger**); and *Stroke* (by the relation **realized_in**).

This representation of risk of stroke for patients with atrial fibrillation could also be used to stratify patients into risk groups by computing their CHADSVASC score, using e.g. SWRL rules (Rosier, 2015). This would also provide formal definitions of classes $Human_{AF1}$, $Human_{AF2}$, etc.

This paper has shown how a specific example of probability assignment to a risk – the risk of stroke over 12 months of a patient with atrial fibrillation – could be formalized. Future work will need to systematize, using OBO-Foundry relations, the relations involving the classes *Risk* or *Risk estimate*. Elaborating on the work of Barton & Jansen (2016), a relation of disposition-parthood could also be introduced to represent the connection between the risk of a human with AF to have a stroke ($Risk_{AF,Stroke}$) and the risk of a human with AF to have a stroke over 12 months ($Risk_{AF,12m,Stroke}$).

The formalization presented in this paper relies on two hypotheses. The first is that for every class of material objects *O*, and every classes of processes *T* and *R*, there exists a class of dispositions *D* that inheres in *O*, has *T* as maximally specified class of triggers and *R* as maximally specified class of realizations; and *O*, *T* and *R* together constitute the conditions of identity of this disposition. This hypothesis was used to define $Risk_{AF,12m,Stroke}$ (from the classes $Human_{AF}$, $History\text{-}part_{12m}$ and *Stroke*) as a class different from $Risk_{AF,12m}$, which has a different maximally specified class of triggers (*History-part*). The second hypothesis is that a realization **r** of a disposition **d** can happen during a trigger **t** – not necessarily just after the trigger ended. Thus, $\mathbf{risk}_{\mathbf{Jones},12m,Stroke}$ could be realized during a 12-months-long history part that acted as a trigger. Finally, this formalization raises the wider philosophical issue whether the risk of stroke $Risk_{Human,Stroke}$ could be classified as a kind of disease, given OGMS definition of disease.

ACKNOWLEDGEMENTS

We thank three anonymous reviewers for their feedback. AB acknowledges financial support by the "bourse de fellowship du département de médecine de l'université de Sherbrooke" and the "CIHR funded Quebec SPOR Support Unit".

REFERENCES

Arp, R., Smith, B., & Spear, A.D. (2015) *Building Ontologies with Basic Formal Ontology*. Mit Press.

Barton, A., Burgun, A., & Duvauferrier, R. (2012) Probability assignments to dispositions in ontologies. In Donnelly, M. & Guizzardi, G. (eds), *Proceedings of the 7th International Conference on Formal Ontology in Information Systems (FOIS2012)*. IOS Press, Amsterdam, pp. 3–14.

Barton, A., Ethier, J.-F., Duvauferrier, R., & Burgun, A. (2017) An ontological analysis of medical Bayesian indicators of performance. *J. Biomed. Semant.*, **8**, 1.

Barton, A. & Jansen, L. (2016) A modelling pattern for multi-track dispositions for life-science ontologies. In Loebe, F., Boeker, M., Herre, H., Jansen, L., & Schober, D. (eds), *ODLS 2016 - Ontologies and Data in Life Sciences*, *CEUR Workshop Proceedings*. p. H.1-2.

Barton, A., Rosier, A., Burgun, A., & Ethier, J.-F. (2014) The Cardiovascular Disease Ontology. In Garbacz, P. & Kutz, O. (eds), *Proceedings of the 8th International Conference on Formal Ontology in Information Systems (FOIS 2014)*. IOS Press, Amsterdam, pp. 409–414.

Gage, B.F., Waterman, A.D., Shannon, W., Boechler, M., Rich, M.W., & Radford, M.J. (2001) Validation of clinical classification schemes for predicting stroke: results from the National Registry of Atrial Fibrillation. *Jama*, **285**, 2864–2870.

Hájek, A. (2007) The reference class problem is your problem too. *Synthese*, **156**, 563–585.

Hájek, A. (2012) Interpretations of Probability. In Zalta, E.N. (ed), *The Stanford Encyclopedia of Philosophy*, Winter 2012. edn.

Jansen, L. (2007) Tendencies and other realizables in medical information sciences. *The Monist*, **90**, 534–554.

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

Lewis, D. (1980) A Subjectivist's Guide to Objective Chance. In *Studies in Inductive Logic and Probability*. University of California Press, pp. 83–132.

Lip, G.Y., Nieuwlaat, R., Pisters, R., Lane, D.A., & Crijns, H.J. (2010) Refining clinical risk stratification for predicting stroke and thromboembolism in atrial fibrillation using a novel risk factor-based approach: the euro heart survey on atrial fibrillation. *Chest*, **137**, 263–272.

Maher, P. (1997) Depragmatized Dutch book arguments. *Philos. Sci.*, **64**, 291–305.

Nielsen, P.B., Larsen, T.B., Skjøth, F., Overvad, T.F., & Lip, G.Y. (2016) Stroke and thromboembolic event rates in atrial fibrillation according to different guideline treatment thresholds: a nationwide cohort study. *Sci. Rep.*, **6**, 1–7.

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

Rosier, A. (2015) Raisonnement automatique basé ontologies appliqué à la hiérarchisation des alertes en télécardiologie. (doctoral dissertation, Rennes 1)

Scheuermann, R.H., Ceusters, W., & Smith, B. (2009) Toward an ontological treatment of disease and diagnosis. In *Proceedings of the 2009 AMIA Summit on Translational Bioinformatics*. San Francisco CA, pp. 116–120.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., … & Lewis, S. (2007) The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotech*, **25**, 1251–1255.

Smith, B. & Ceusters, W. (2015) Aboutness: Towards foundations for the information artifact ontology. In *Proceedings of the 6th International Conference on Biomedical Ontology*. Presented at the ICBO 2015, CEUR Workshop Proceedings, Lisbon, Portugal, pp. 1–5.

Uciteli, A., Neumann, J., Tahar, K., Saleh, K., Stucke, S., … Herre, H. (2016) Risk Identification Ontology (RIO): An ontology for specification and identification of perioperative risks. In Loebe, F., Boeker, M., Herre, H., Jansen, L., & Schober, D. (eds), *ODLS 2016 - Ontologies and Data in Life Sciences*, *CEUR Workshop Proceedings*. p. G.1–7.

Zheng, J., Harris, M.R., Masci, A.M., Lin, Y., Hero, A., Smith, B., & He, Y. (2016) The Ontology of Biological and Clinical Statistics (OBCS) for standardized and reproducible statistical analysis. *J. Biomed. Semant.*, **7**, 53.