Response to reviewers
on
Biomarkers in the Ontology for General Medical Science
Werner Ceusters & Barry Smith

Congratulations: Your submission Biomarkers in the Ontology for General Medical Science (Id=67) was accepted as Full paper (Strong accepted) by the SPC.

[…]

Comments to authors[1]:

Reviewer 1:

"Biomarker" has become an important term in medical research. Therefore, a precise definition is desirable to what biomarkers are, taking into account that the term is used inconsistently in medicine, in other fields of science as well as a buzzword [https://twitter.com/medidata/status/537294693722107905].

==Pros==

* Topic is certainly of relevance and of interest to the conference

* Paper is reasonably well written regarding the language, as far as we can tell

→ We agree

==Issues==

A number of question arose:

1.) We see some issues in the definitions provided, apart from the fact that it is not fully clear whether or to what extent these should be seen as definitions in natural language or in an underlying formal language (which one)?

→ They are semi-formal natural language definitions, semi-formal in the sense that each meaningful phrase within a definition either denotes an entity which is defined or elucidated in other descriptions or expresses the existence of a relation that obtains between corresponding entities in the definition. We added 'semi-formal natural language' to the title of Table 1 to make this clear, as well in the results section.

Def. 1a (and 1c)

MATERIAL BIOMARKER =def.– BODILY COMPONENT instances of which are capable of being assessed objectively to determine what kind of PROCESSES either they or some QUALITIES that depend on them result from.

--> Why is there an "instances of which"? Shouldn't it be bodily components themselves that are capable of ...? The question applies analogously to Def. 1c of Process Biomarkers. Interestingly, Def. 1b directly refers to Bodily Quality inhering in a Bodily Component c - which thus can be assumed to be an individual.

[1] Reviewers' comments are in black font, authors' responses in blue font

→ This is merely a matter of phrasing. We replaced it with the following definition which qua form satisfies the preference of this reviewer, but describes the same portion of reality.

MATERIAL BIOMARKER =def.– BODILY COMPONENT capable of being assessed objectively to determine either (a) what kind of PROCESSES it results from, or (b) what kind of PROCESSES resulted in QUALITIES that depend on it.

Def. 1b:

QUALITY BIOMARKER =def.– BODILY QUALITY inhering in a BODILY COMPONENT c and in virtue of which c is an instance of MATERIAL BIOMARKER

-> Does that mean that any quality biomarker entails/requires the existence of a material biomarker? If so, we wonder whether this may lead to issues of unintendedly classifying entities as biomarkers.

→ Any quality depends on some independent continuant. Any bodily quality depends on some bodily component. If a bodily quality is a quality biomarker, then it is that quality which makes the bodily component being a material biomarker. That is explained in section 3.1

Sect. 3.1 suggests this coupling of quality biomarkers and material biomarkers. The question is whether the blood sample is adequately seen as a biomarker itself due to being used in measuring the quality biomarker, e.g. blood glucose level. E.g., the blood sample in itself may not serve "functional characteristics" that the glucose level does.

→ True. We agree that the original phrasing 'Thus for entities such as blood glucose and A1C levels – both qualities of portions of blood – we first define the relevant material biomarkers (definition 1a, Table 1) that indeed have the required capability of being observed (for example samples of blood taken from a patient)' may be interpreted as a claim on our side that blood samples are material biomarkers. That is not intended. It is the portion of glucose in the portion of blood in the patient's body which is the material biomarker. The blood sample itself is not as the part of the portion of blood that became a blood sample is not a bodily component anymore at the time it became a sample. But the sample allows for the observation to be made. We therefore rephrased the sentence above as such: 'Thus for entities such as blood glucose and A1C levels – both qualities of portions of blood – we first define the relevant material biomarkers (definition 1a, Table 1) that indeed have the required capability of being observed (for example ***through analysis of*** samples of blood taken from a patient).'

-> Why is this different from the case of process biomarkers in the definitions, where 1c does not mention any participating bodily component? Or is it not meant to be different, given Def. 3c, which appears analogous to 1b wrt "in virtue of which c is an instance of ..."?

→ Bodily processes depend by definition on some bodily component. See table 2. There is thus no need for explicitly mentioning this.

3a. MATERIAL BIOMARKER instances of which are capable of being assessed objectively to determine whether a DISEASE OF TYPE X inheres in the ORGANISM of which these instances are part.

--> what if there are entities of the same kind in a body, some of which are related to the disease, some not (e.g. lymphnodes)? Does this lead to problems?

→ No. But it would of course be wrong in such case to classify these lymph nodes – these specific instances in that patient – which are not related to an instance of DISEASE OF TYPE X as being material

biomarkers for that type of disease. Finding out for each lymph node what is the case, is an epistemic issue, not an ontological one.

"... not data obtained through such processes ..."

--> How could such an entity (data) be named?

→ This question is not clear. But 'data' according to BFO, are generically dependent continuants of type information content entity. They are not on the side of the patient.

In the OGMS definitions:

ABNORMAL HOMEOSTASIS =def. – HOMEOSTASIS that is clinically abnormal for an ORGANISM of a given type and age in a given environment.

NORMAL HOMEOSTASIS =def. – Homeostasis of a type that is not clinically abnormal.

--> Isn't this a circular definition? We'd say yes, if "clinically abnormal" encapsulates/reformulates "not clinically normal", but then "normal" is defined by "not clinically abnormal". Or can clinical abnormality be determined prior to normality?

→ No, there is clearly no circularity. See paper referred to as [2].

We note that the last paragraph of sect. 3.1. discusses clinical abnormality, but perhaps the link with the definition could be strengthened.

→ Indeed, and the reference to definition 1c is even given to make the link. No action taken.

Regarding the issue of reading OR as exclusive or - what is the language of the definitions that is assumed here? Is an OWL implementation behind, where the disjointness is expressed? Alternatively, if the definitions are stated in natural language, why not just write "either A, B, or C"?

→ The 'or' we refer to as being assumed by us to mean exclusive or (XOR) is – as explicitly stated in the paper – the 'or' used in the quote taken from the IOM's report. We cannot rewrite their report.

2.) The ontological nature of biomarkers remains somewhat unclear to us. On the one hand, it seems that the author suggest the same status for, e.g., 'material biomarker' as an universal …

→ we don't suggest that, it is explicitly stated in the paper: last paragraph of section 3.1. Furthermore, in Table 1 we added for each representational unit defined in the column marked 'T' an explicit annotation regarding its ontological status.

… as we'd expect for 'material entity', e.g. by this fragment:

"Second, we assumed that in requiring that biomarkers be 'objectively measured and evaluated' the IOM had in mind not that an entity becomes a biomarker after and because it has been measured and evaluated"

→ this sentence from the IOM's report has nothing to do with the ontological status of material biomarker.

On the other hand, biomarker still seem to have a contextual nature (with which we would agree), e.g. considering the sentence:

"This reflects the observation that the color of some bodily component might be a quality biomarker for some disease, even though the size or temperature of that component would not be."

→ No, context does not come into play here. Where does this reviewer see context?

For instance, this sentence leads us to the question whether it could be the case that a quality is a quality biomarker for some disease, but not for another?

→ Trivially! Glucose levels are not a quality biomarker for schizophrenia.

Would that affect the classification according to the proposal?

→ No.

Similarly, we do not see how contextual aspects/parameters are accounted for by the approach, e.g., as suggested to play a role in the sentence "Thus the glucose level in a person's blood 30 minutes after drinking a sugary soda would not qualify as biomarker.", while certain glucose levels probably count as biomarkers.

→ The point is that a blood sample taken from a person that drank a soda is not '... *capable of being assessed* ***objectively***...' as required by the definition, and therefor is not a diagnostic biomarker for a disease of some type.

This may further link to the requirement that the biomarker comes into existence during processes which are the realization of a disease of that type. What about a raising level of blood glucose, for instance, that at some point is traced back to a disease. Does that quality - as a disease biomarker - come into existence only then? If not, what is the difference between the blood glucose of the human and the biomarker?

→ We need to distinguish between two possible ways for some entity to become a biomarker. The questions of this reviewer made it clear that one such way was not well captured in the definition. An entity that becomes a disease biomarker can only acquire that status during a part of the disease course which realizes that disease. However, the entity might already have existed before that time, an example being the patient's blood glucose level: it becomes an instance of disease biomarker once exceeding a certain threshold, thus once becoming *clinically abnormal* as described in OGMS. Another example might be a bent spine: the spine existed prior to the patient's Bechterev disease, but it becomes a disease biomarker once sufficiently bent. Other entities, such as the tonic-clonic convulsions of an epileptic patient, come into existence during the disease course, and instanciate the disease biomarker universal from the very beginning of their existence. The phrase 'starts to exist' in the original definition was thus too strong as it suggested only this last case to be described. We therefore changed the definition

> **DISEASE BIOMARKER FOR DISEASE OF TYPE X** =def.– **BIOMARKER** which starts to exist during a ***part*** of the DISEASE COURSE which ***realizes*** the DISEASE OF TYPE X

into:

> **DISEASE BIOMARKER FOR DISEASE OF TYPE X** =def.– **BIOMARKER** instances of which either start to exist or become clinically abnormal during a ***part*** of the DISEASE COURSE which ***realizes*** the DISEASE OF TYPE X.

Note that in this case the extra phrase 'instances of which' cannot be avoided as otherwise the 2nd case would be read as '…biomarker which becomes clinically abnormal…' thus leaving open the interpretation that all such entities were already biomarker before becoming clinically abnormal.

However, the first question of this reviewer seems to suggest that he or she believes that an epistemic criterion is used in the definition of disease biomarker. That is not the case! An entity which qualifies as 'disease biomarker' acquires that status through a process that realizes the disease and ***at the time*** that process is unfolding. Of course, at that time, that fact might be unknown, and the presence of that disease only determined later. At that time, when the disease is diagnosed, the entity in question which ***was*** already a disease biomarker, becomes ***known*** as a disease biomarker.

3.) (more) structural issues

3a) The last lines of section 3.1 ("Biomarkers (in general, definition 1) ...") do not seem closely connected to the discussion about processes before. Perhaps they should go into a separate paragraph.

→ We did so

3b) The paper has some imbalances and lacks certain aspects, from our point of view. Imbalance concerns the length of the discussion compared to the sections "Results" and "Conclusion". ot quite justification in the discussion - just what it is like

Moreover, we'd suggest more justifications of the approach and comments on past or future evaluation of the results. E.g., it's not clear from the text whether or to what extent the proposed definitions have been applied to further examples, beyond those stated in the text. Regarding justifications, only little is provided even in the discussion, which rather describes the views of the authors (alone).

→ We 'd love to, but there is unfortunately a 5 page restriction. We will cover this in a subsequent journal paper.

The conclusions should be strengthened:

"We believe that our approach has a number of advantages, not least that it can be generalized easily to apply to a range of different sorts of biomarkers, including not only disease and diagnostic biomarker of a range of different types, but also image biomarkers, environmental biomarkers, toxicity biomarkers, and many more."

Which other advantages are there? Why is it easily generalizable?

→ Because the hard work was to get the first definitions accurate.

Finally, we suggest that at least some related work of _other_ authors is mentioned (not counting [1] here), to better understand the overall context that the paper lives in, e.g., regarding biomarker databases, other existing ontologies (or parts thereof, cf. e.g. the treatment in Chebi [[1]]).

→ We agree, but that is again not possible within the page restrictions. We did an analysis of existing work and will publish this in the journal paper.

[[1]] http://rgd.mcw.edu/rgdweb/ontology/view.html?acc_id=CHEBI:59163

→ And we will explain in that journal paper why the CHEBI definition – as so many others we found – is terribly inadequate. We believe it is not intellectually correct to just make a statement like that without thorough discussion. For this, there is no space.

Minor:

- BFO is not introduced as an abbreviation (not cited)

→ we provided the full name

- language: "that are indeed have"

→ corrected

- language: use of "which" vs "that" everywhere adequate?

→ has been addressed

Conclusion: Loosely quoting this submission, one could say: "The number of open issues in the paper is regrettable; but it does not mean that the paper itself is of no value."

→ Nice, although we disagree that so many issues are open, as witnessed by our comments.

Reviewer 2:

The authors propose definitions for biomarkers of various types, based on the OGMS ontology.

The paper is well-written and clear. Moreover the authors describes in details the assumptions of their approach.

The generality of the approach proposed is a very good result.

→ Ok, thanks.

Reviewer 3:

The paper is relevant to the topics of MIE15 conference. It presents a well-defined ontology for biomarkers. It reads very fluent, introduces the necessity for a clear understanding of the term "biomarker" and discusses the definitions of the proposed ontology in concise and well-structured way.

Given the theoretical nature of ontologies, the paper could benefit if you would mention or describe a real world example of how the improved discrimination and understanding of "biomarkers" could increase, the quality in the design of (bio)medical studies, in a pharmacological context or maybe in clinical routine care situations.

→ This is an important topic that we prefer to discuss elsewhere, given page restrictions.

Still, the contents of the paper are original and thus important to the field of study and the Medical Informatics community as well.

In my opinion - if presented at the conference - a related talk should highlight the practical implications of your findings to catch the audience. It definitely is a valuable contribution to MIE15!

Overall: I don't see any hard reasons to change the current version of the paper. Yet, a related talk might benefit by the aforementioned comments.

→ We will do so.