

Biomarkers in the Ontology for General Medical Science

Werner CEUSTERS^{a,1} and Barry SMITH^b

^aDepartment of Biomedical Informatics, University at Buffalo, NY, USA.

^bDepartment of Philosophy, University at Buffalo, NY, USA.

Abstract. A great deal of recent work has been devoted to the topic of biomarkers as aids to diagnosis, prognosis and treatment evaluation. Basing our work on the Ontology for General Medical Science (OGMS) and on the specifications provided by the Institute of Medicine (IOM), we propose definitions for biomarkers of various types. These definitions provide a formal representation of what biomarkers are in a way that allows us to remove certain ambiguities and inconsistencies in the documentation provided by the IOM.

Keywords. Biological Markers, Biological Ontologies, Metaphysics

Introduction

The Institute of Medicine (IOM) defines a biomarker as: *‘a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to therapeutic intervention’* [1]. Biomarkers can be used for diagnostic purposes, to assess the severity, burden or extent of disease, to predict future onset or course of disease (prognostic), and to provide information about treatment effectiveness. Biomarkers are also used in the design of diagnostic classifications and in the formulation of diagnostic criteria.

For biomarkers to play such a prominent role in advancing information-driven biomedical research there must be a uniform understanding among researchers and designers of such classifications about what biomarkers precisely are and a coherent classification of the different types of biomarker. The problem is that such a common understanding does not yet exist. Consider, for example, the inconsistent use of the term ‘biomarker’ in IOM’s own report [1, p2], for instance in: *‘Cholesterol and blood sugar levels are biomarkers, as are blood pressure, enzyme levels, measurements of tumor size from MRI or CT, and the biochemical and genetic variations observed in age-related macular degeneration.’* Here the IOM conflates *characteristics on the side of the patient with measurements* of such characteristics. It also leaves unclear whether by *‘measurement’* it means: (a) the process of measuring an entity on the side of the patient, or (b) the data obtained through such a process. In the same report the need is expressed *‘to develop a transparent process for creating well-defined consensus standards and guidelines for **biomarker development, validation, qualification, and use** to reduce the uncertainty in the process of development and adoption’* [1, p18

¹ Corresponding Author: Werner Ceusters, Department of Biomedical Informatics, University at Buffalo, 923 Main street, Buffalo NY 14203, USA; E-mail: ceusters@buffalo.edu.

emphasis added]. But what does ‘*development*’ mean here? We can imagine what it is to ‘develop’ a measuring process – for example by refining protocols or instrument error tolerances. But what would it mean to ‘develop’ *enzyme levels* or *tumor sizes*?

That the terminology around biomarkers is inconsistent is a problem which the IOM recognizes in its own report [1, p22]. The fact that it then itself contributes further to such inconsistency is regrettable; but it does not mean that the report itself is of no value. Rather, it means that the term ‘biomarker’ needs to be defined in a way that will ensure unambiguous use in the future, and to this end the definition must be anchored in an ontology which carefully distinguishes the types of entities referred to in standards and guidelines for biomarker selection, validation, qualification, and use.

1. Methods

The Ontology for General Medical Science (OGMS) is designed to represent the entities involved in a typical clinical encounter, including both diseases and their causes and manifestations, as well as diagnostic acts and other entities pertaining to the ways diseases are recognized and interpreted in the clinic. OGMS was designed to avoid the sort of conflations often encountered in medical discourse between entities on the side of the patient (or studied organism) on the one hand – for instance diseases, tumors, changes in temperature – and the *evidence* for the existence of such entities on the other [2]. A basic axiom of OGMS is that every disease rests on some (perhaps as yet unknown) physical basis. When, for example, there is an elevated level of blood glucose in a specific patient, then this is because (1) some physical structure in the organism is disordered (e.g. abnormal pancreatic beta cells – the *disorder*) as a result of which (2) there exists a disposition (e.g. diabetes mellitus – the *disease*) for the organism or its parts to act in a certain abnormal way. This disposition is realized in pathological processes (e.g. of diminished insulin production) including processes that can be recognized as *symptoms* and *signs* of the disorder (e.g. polydipsia) and serve as targets of assays such as lab tests or imaging procedures. We used the OGMS definitions for a variety of clinically relevant types of entities in order to construct a definition for ‘*biomarker*’ which allows this term to be used in ontologies designed in accordance with the principles of the Open Biomedical Ontologies (OBO) Foundry [3].

2. Results

Table 1 contains semi-formal definitions which we believe represent in an ontologically adequate way the biomarkers (and certain related entities) described informally by the IOM. Table 2 contains the definitions from OGMS upon which Table 1 draws.

3. Discussion

When subjecting assertions of domain experts to *post hoc* ontological analysis it is not uncommon to encounter phrases that reveal ambiguities of expression, sometimes to the degree that raises the question whether a literal interpretation is even achievable in

Table 1: Proposed semi-formal natural language definitions for biomarkers of different sorts

	T	Definition
1	C	BIOMARKER =def.– MATERIAL BIOMARKER , QUALITY BIOMARKER or PROCESS BIOMARKER .
1a	U	MATERIAL BIOMARKER =def.– BODILY COMPONENT <i>instances of</i> which are <i>capable of</i> being assessed objectively to determine what kind of PROCESSES either they or some QUALITIES that <i>depend on</i> them <i>result from</i> .
1b	U	QUALITY BIOMARKER =def.– BODILY QUALITY <i>inhering in</i> a BODILY COMPONENT <i>c</i> and in virtue of which <i>c</i> is an <i>instance of</i> MATERIAL BIOMARKER
1c	U	PROCESS BIOMARKER =def.– BODILY PROCESS <i>p instances of</i> which are <i>capable of</i> being assessed objectively to determine whether <i>p</i> is a <i>realization of</i> NORMAL or of ABNORMAL HOMEOSTASIS.
2	C	DISEASE BIOMARKER FOR DISEASE OF TYPE X =def.– BIOMARKER instances of which either start to exist or become clinically abnormal during a <i>part</i> of the DISEASE COURSE which <i>realizes</i> the DISEASE OF TYPE X
3	C	DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X =def.– MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X , QUALITY DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X , or PROCESS DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X
3a	U	MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X =def.– MATERIAL BIOMARKER instances of which are <i>capable of</i> being assessed objectively to determine whether a DISEASE OF TYPE X <i>inheres in</i> the ORGANISM of which these instances <i>are part</i> .
3b	U	QUALITY DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X =def.– BODILY QUALITY which <i>inheres in</i> a BODILY COMPONENT <i>c</i> and in virtue of which <i>c</i> is an <i>instance of</i> MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X
3c	U	PROCESS DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X =def.– BODILY PROCESS in which a BODILY COMPONENT <i>c participates</i> and in virtue of which <i>c</i> is an <i>instance of</i> MATERIAL DIAGNOSTIC BIOMARKER FOR DISEASE OF TYPE X .
4	U	PROGNOSTIC BIOMARKER FOR DISEASE OF TYPE X =def.– DISEASE BIOMARKER FOR DISEASE OF TYPE X which starts to exist during a <i>part p</i> of a DISEASE COURSE <i>d</i> which <i>realizes</i> the DISEASE OF TYPE X and where <i>p</i> is such that it <i>regulates</i> the further unfolding of <i>d</i> .

Legend: T = defined class (C) or universal (U). **BOLD SMALL CAPS** = generic entities defined in this table. **SMALL CAPS** = generic entities defined in other BFO-based ontologies; **bold italics** = relations.

Table 2: Relevant OGMS definitions

DISORDER =def. – causally relatively isolated combination of BODILY COMPONENTS that is (a) clinically abnormal and (b) maximal, in the sense that it is not a <i>proper part of</i> some larger such combination.
PATHOLOGICAL PROCESS =def. – BODILY PROCESS that is a <i>manifestation of</i> a DISORDER.
DISEASE =def. – DISPOSITION (i) to undergo PATHOLOGICAL PROCESSES that (ii) <i>inheres in</i> an ORGANISM because of one or more DISORDERS <i>in</i> that ORGANISM.
DISEASE COURSE =def. – The totality of all PROCESSES through which a given DISEASE instance <i>is realized</i> .
MANIFESTATION OF A DISEASE =def. – BODILY FEATURE of an ORGANISM that is (a) a deviation from clinical normality that exists in virtue of the <i>realization of</i> a DISEASE and (b) is observable.
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.
BODILY COMPONENT =def. – MATERIAL ENTITY <i>within or on</i> the surface of the body of an ORGANISM.
BODILY PROCESS =def. – PROCESS unfolding <i>in or on</i> the body in which BODILY COMPONENTS <i>participate</i>
BODILY QUALITY =def. – QUALITY which <i>inheres in</i> a BODILY COMPONENT.
BODILY FEATURE =def. – BODILY COMPONENT, BODILY QUALITY or BODILY PROCESS.

Legend: **SMALL CAPS** = generic entities defined in this table or in BFO; **bold italics** = relations

principle. When we constructed the definitions displayed in Table 1, we found that we needed to make three assumptions with respect to such issues.

First, we assumed that the IOM intended biomarkers to be entities *on the side of the patient*, and not (for example) processes on the part of the clinician or data obtained through such processes including what can be seen or measured for example in radiographic images. 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, but rather that it

was a biomarker already prior to observation because of certain properties it has intrinsically. Thus the A1C level in a person's blood is a diagnostic biomarker for whether that person suffers from Type II Diabetes whether or not this A1C level has actually been measured. It is, in other words, the *observability* and *measurability* – rather than the having been subjected to observation and measurement – which are the necessary conditions for something to be a biomarker. In similar vein, we do not interpret the word '*indicator*' as used in the definition to constitute an epistemic condition, nor do we believe that the IOM considers biomarkers to be fiat entities which acquire their status (or are 'developed') as biomarkers by mere decree.

Finally, we assumed that the logical disjunction expressed by the 'or' in the list of processes for which the IOM definition asserts biomarkers to be an indicator has to be interpreted as an exclusive or (XOR). Thus we assume that the IOM would not accept as biomarker some entity *e* from which it cannot be determined whether *e* is the result of a normal or of a pathological process, and so forth. Thus the glucose level in a person's blood 30 minutes after drinking a sugary soda would not qualify as biomarker.

3.1. Three disjoint types of biomarkers

The view of biomedical reality offered by the OGMS framework would allow us – at first sight – to substitute the vague term '*characteristic*' in the IOM definition with the precisely defined OGMS term '*bodily feature*'. Entities that qualify as bodily features are instances of one or other of the following three disjoint types (1): bodily components which are either parts of the body (as in the case of pancreas cells, blood, insulin receptors) or in the interior or surface of the body (such as pathogens, toxins, body flora); (2) bodily qualities such as body temperature or blood sugar concentrations; and (3) bodily processes in which bodily components participate, whether normal (transmembrane flux, urinating), pathological (reduced insulin production), or induced through pharmacologic intervention.

A problem with this reading, however, arises in light of our assumption that biomarkers must in every case be observable: that is they must be capable of being identified through observation, for instance in a process of measurement. Some biomarkers are also 'assessable' – for instance we can assess them by determining through what kinds of processes they came into existence. They are capable, in this sense, of being assessed. No problem arises from this assumption for material biomarkers, since material entities in the Basic Formal Ontology (BFO) are unproblematically able to serve as the bearers of (observability, and many other) dispositions. Processes and qualities in BFO do not have dispositions in the BFO and OGMS sense of this term. To resolve this problem, however, it is necessary only to point out that processes and qualities are observable in this sense: that corresponding material entities serve as bearers for the corresponding dispositions. 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). We then define the quality biomarkers themselves (definition 1b) in terms of these material biomarkers. Not any quality of a material biomarker qualifies as a quality biomarker on this approach, but only those qualities in virtue of which the underlying bearer is a material biomarker. 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.

The solution for processes qualifying as biomarkers (definition 1c) is to define them by referring to their participants – material entities that are assessable as participating in processes of normal or abnormal homeostasis. Homeostasis itself is a disposition of the whole organism (or of some causally relatively isolated part of the organism) to regulate its bodily processes in such a way as (1) to maintain bodily qualities within a certain range or profile and (2) to respond successfully to departures from this range. For a bodily feature to be clinically abnormal, it must be the case that: (1) it is not part of the life plan for an organism of the relevant type, (2) it is causally linked to an elevated risk either of pain or other feelings of illness, or of death or dysfunction, and (3) it is such that the elevated risk exceeds a certain threshold level.

Biomarkers (in general, def. 1) constitute a defined class which is such that (1) its members are instances of precisely one of the three disjoint biomarker universals (1a, 1b, or 1c), and (2) it forms a subclass of the OGMS class ‘bodily feature’. Examples of what is then ‘*measurable*’ for instances of each of these three subtypes are, respectively: (1a) time of appearance and disappearance, length of time in existence; (1b) ratios, lengths, volumes, concentrations; (1c) pulse rate, respiration rate [4].

3.2. Purpose-specific biomarkers

The IOM recognizes the importance of biomarkers for certain specific purposes such as (1) diagnosis, i.e. to ‘*definitively establish the presence of disease*’, (2) classification, i.e. to ‘*classify patients by disease subset*’, and (3) prognosis, to ‘*predict the probable outcome of disease to determine the aggressiveness of treatment*’ [2, p24]. Unfortunately it does not define what it means to be a biomarker for each of the corresponding types. Definitions 2-4 from Table 1 have been created to that end.

Our definition for ‘disease biomarker’ lines up directly with the IOM’s view on biomarkers for disease classification: such biomarkers must be such as to allow determination of presence of a specific type of disease. This is obtained by using ‘biomarker’ as genus and adding as differentia the requirement that the biomarker comes into existence during processes which are the realization of a disease of that type.

4. Conclusion

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.

References

- [1] *Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease*, ed. C.M. Micheel and J.R. Ball. 2010, Washington, D.C.: The National Academies Press. 314.
- [2] Scheuermann, R.H., W. Ceusters, and B. Smith, *Toward an Ontological Treatment of Disease and Diagnosis*, in *Proceedings of the 2009 AMLA Summit on Translational Bioinformatics, San Francisco, California, March 15-17, 2009*. 2009, American Medical Informatics Association. p. 116-120.
- [3] Smith, B., et al., *The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration*. *Nature Biotechnology*, 2007. **25**: p. 1251-1255.
- [4] Smith, B., *Classifying processes: An essay in applied ontology*. *Ratio*, 2012, **25**: p. 463-488.