

Toward an Ontological Treatment of the Initiation, Realization, Recognition and Representation of Disease

Some Terminological Proposals

prepared in connection with the 2008 'Signs, Symptoms and Findings' and 'Infectious Disease Ontology' workshops

Richard H. Scheuermann and Barry Smith

Effective knowledge representation requires the use of standardized nomenclatures to ensure both shared understanding between people and interoperability between computer systems. Unfortunately, many existing biomedical vocabulary standards do not consistently distinguish between disease, clinical and pre-clinical manifestations of disease, and diagnosis. They therefore do not adequately support translational research applications. Here we outline a logical framework for describing the initiation and progression of disease and the associated evaluation procedures and information artifacts. We defend a view of disease as involving in every case some physical basis within the organism that bears a disposition toward malfunctioning and thus towards the initiation and execution of pathological processes. We provide a preliminary set of definitions, which are given in natural language form but are designed to be translatable into a formal language such as OWL-DL.

Introduction

The goal of this study is to develop a terminological framework that encompasses representations of diseases and disease processes, the phenotypic manifestations caused by the underlying disorders, and the ways these manifestations are recognized and interpreted in the clinic. Inspection reveals that the terms treated here have thus far not been adequately defined in standard terminology resources. The National Cancer Institute Thesaurus, for example, defines 'Clinical Course of Disease' as a subtype of 'Finding', which it defines as:

Objective evidence of disease perceptible to the examining physician (sign) and subjective evidence of disease perceived by the patient (symptom) (<http://nciterns.nci.nih.gov/NCIBrowser>).

This implies, however, that a disease course does not exist except as one or other form of evidence, and thus as perceived in some way. This example illustrates a common conflation between processes on the side of the patient on the one hand, and the observation of such processes in the context of medical care on the other. This conflation becomes

problematic when we try to connect translational data and information with the underlying biological processes being investigated.

Results

Defined and Undefined Terms

While it is generally good practice to provide precise definitions for the terms assembled in a terminology standard, in any logically coherent approach to the definition of basic terms of the sort that is attempted here, some terms must remain undefined in order to avoid circularity or infinite regress. The undefined terms used in what follows are of three sorts: either (i) they are non-technical terms derived from ordinary English; or (ii) they are technical terms derived from basic science (for example, 'organism', 'molecule'); or (iii) they are primitive terms specific to the domain of interest. Terms in group (iii) – specifically: 'physical basis', 'clinically abnormal', 'homeostasis', 'disposition', and 'realization' – require special attention. And, while, ex hypothesi, we cannot provide definitions for such terms, we can provide informal elucidations.

We begin by describing the notion of the 'physical basis' of health and disease. This physical basis, as we understand it, includes any collection of the physical components that make up the body, whether healthy or diseased, at any level of granularity from molecule to whole organism (e.g. a single nucleotide, a liver cell, an arm, a liver cell that has been infected by a virus, a polyp, a clot, a displaced disk). By 'bodily feature' we mean any member of the defined class comprising not only such bodily components but also associated bodily qualities and processes in an organism, both physiological and pathological.

We use 'clinically abnormal' to characterize those bodily features of or in an organism that are (1) causally linked to an elevated risk of pain or other feelings of illness, to dysfunction, or to enhanced morbidity, and which (unlike pregnancy or menopause) are not such as to belong to the life plan for an organism of the relevant type. A clinician will judge a bodily feature to be clinically abnormal only

where the elevated risk exceeds a certain threshold level of clinical significance. The determination of this threshold will reflect the training and experience of the clinician in question and thus may vary from clinical context to clinical context. It is worth noting that this treatment of 'normality' is distinct from those statistical treatments which do not take account of overlap of parameter values used to distinguish between normal and abnormal or of distribution extremes.

We use 'homeostasis' to designate the state in which the bodily processes of the organism are regulated 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 caused by external influences. During homeostasis it is as if the organism continually assesses its current state to determine if its bodily qualities fall within this range. If a quality is determined to have departed from the homeostatic range, the organism initiates processes designed to return the qualities to a value within this range. In some cases, homeostasis can be lost and then re-gained at a level that is judged to be clinically abnormal, for example in the case of adaptation to major injury. In other cases the organism will pass a point where it falls irreversibly outside the realm of homeostasis.

We use 'disposition' to designate that feature of a physical entity in virtue of which it participates in processes of a certain type whenever it is in circumstances in which certain preconditions are satisfied. We use 'realization' to refer to the corresponding processes. Examples are: the disposition of a glass vase to break when dropped; the disposition of an epithelial cell in the G2 phase of the cell cycle to become diploid following mitosis.

Definitions of Terms Referring to Entities on the Side of the Organism

In what follows, we pursue a physical view of disease as resting in every case on some (perhaps as yet unknown) physical disorder in the organism. When, for example, there is a persistent elevated level of glucose in the blood of an organism, this is because (1) some physical structure or substance in the organism is disordered (loss of beta cells in pancreatic islets) as a result of which (2) there exists a disposition (diabetes) for the organism to act in a certain abnormal way. The disposition in question is realized by the initiation and execution of specific pathological processes (diabetic nephropathy) whose manifestations can be recognized as signs of the disorder (proteinuria).

Homeostatic Range =def. – The range of values for a set of bodily feature types whose maintenance is continuously sought by an organism in the state of homeostasis (e.g. 65 – 110 mg glucose/dL serum).

Abnormal Homeostasis =def. – Homeostasis of a type that is clinically abnormal for an organism of a given type and age in a given environment (e.g. maintenance of high blood pressure).

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

Disorder =def. – A physical structure or portion of bodily substance that is clinically abnormal (e.g. a tumor, an infected cell, a prion molecule in the brain, a mutation in genomic DNA, endotoxin in blood).

It is disorders so defined which serve as the physical basis of disease. Disorders arise because a bodily structure is malformed, or because a bodily substance such as blood is affected by the presence of a pathogen or a toxin, in ways that lead to impairment of normal functioning.

Pathological Process =def. – A biological process in an organism that is a manifestation of a disorder (e.g. transient inflammation in response to bacterial infection).

Two types of pathological processes can be distinguished, according to whether (1) the underlying disorder changes the way a normal physiological function is realized, or (2) causes a pathological process that has no normal physiological counterpart (as for example in the case of acute inflammation).

Disease =def. – A disposition (i) to undergo pathological processes that (ii) exists in an organism because of one or more disorders in that organism (e.g. epilepsy as a disease that disposes to the occurrence of seizures (pathological process) due to an underlying abnormality in the neuronal circuitry of the brain (physical basis)).

Predisposition to Disease of Type X =def. – A disorder in an organism that causes an increased risk of acquiring the disease X.

Note that such a predisposition is a disposition on the part of the organism to acquire a further disposition. Diseases themselves may be predispositions to further diseases.

Etiological Process =def. – A process in an organism that leads to a disorder (e.g. toxic chemical exposure resulting in a mutation in the genomic DNA of a cell).

The above definitions imply that an etiological process works by creating the physical basis of some disposition to pathological processes. It is this disposition which is the disease. Some diseases are such that a patient can suffer from what is qualitatively the same disease on two distinct occasions – for example two bouts of influenza in successive years. We then say that the patient has two distinct disease instances of the same disease type. These successive bouts are differentiated by their etiology in the sense that their respective physical bases are caused by distinct processes.

Disease Course =def. – The sum of all realizations of a given disease instance.

Transient Disease Course =def. – A disease course (e.g. a bout of flu) that terminates in a return to normal homeostasis.

Chronic Disease Course =def. – A disease course that (a) does not terminate in a return to normal homeostasis and (b) would, in the absence of intervention, fall within an abnormal homeostatic range (e.g. intermittent seizures in an epilepsy patient).

Progressive Disease Course =def. – A disease course that (a) does not terminate in a return to homeostasis and (b) would, in the absence of intervention, involve an increasing deviation from homeostasis (e.g. malignant cancer).

Definitions of Terms Referring to Genetic Disorders

Genetic Disorder =def. – A disorder whose etiology involves an abnormality in the nucleotide sequence of a patient's genomic DNA.

Epigenetic Disorder =def. – A disorder whose etiology involves (1) a modification to the patient's genomic DNA which leads to alterations in the normal expression pattern of the genome, but is (2) not a change to the nucleotide sequence.

Constitutional Genetic Disorder =def. – A genetic disorder inherited during conception that is borne by all cells in the organism (e.g. mutation in the hexosaminidase A gene leading to Tay-Sachs disease).

Acquired Genetic Disorder =def. – A genetic disorder acquired by a single cell in an organism that leads to a population of cells within the organism bearing the disorder (e.g. point mutation acquired in the H-ras gene in colorectal cancer cells).

Constitutional Genetic Disease =def. – A disease whose physical basis is a constitutional genetic disorder.

Examples: *chronic*: color blindness, polydactyly; *progressive*: Down syndrome, Tay-Sachs disease.

Acquired Genetic Disease =def. – A disease whose physical basis is an acquired genetic disorder.

Examples: *chronic*: benign colonic neoplasia (here the physical basis is an APC mutation); *progressive*: malignant colon cancer (here the physical basis is a combination of APC, ras and p53 mutations).

Genetic Predisposition to Disease of Type X =def. – A predisposition to disease of type X whose physical basis is a constitutional abnormality in a patient's genome.

Examples: *chronic*: neurofibromatosis; *progressive*: Li-Fraumeni Syndrome.

Definitions of Terms Referring to Infections

Infection =def. – A disorder of a type which involves the presence of a pathogenic organism within the host organism leading to pathogen persistence and/or pathogen duplication.

Infectious Disease =def. – A disease caused by an infection.

Examples: *transient*: seasonal flu; *chronic*: genital herpes; *progressive*: Ebola-virus-mediated hemorrhagic fever.

Secondary Infection =def. – A disorder consisting in (1) the presence of a pathogenic organism within a host organism that (2) leads to infection in virtue of (3) a predisposition to disease that exists in virtue of (4) a prior infection with a different pathogenic organism (e.g. cryptosporidiosis in AIDS patients)

Definitions of Terms Relating to Clinical Evaluations

In many cases, organisms harbor disorders for many years before they manifest themselves in observable changes. These changes are frequently first recognized by patients (symptoms) and subsequently observed by clinicians (signs). Although the terms 'sign' and 'symptom' are frequently used in this way to distinguish the source of information, since both represent observable manifestations of disease, the distinction may be of limited utility. We believe that a more rigorous treatment of the distinction would be through the use of distinct evidence codes to reflect the source of a given body of clinical information (direct clinician observation, patient report, report of some other observer, laboratory test, and so on). However, because the distinction is made routinely by clinicians in the conduct of patient care, we include definitions for 'sign' and 'symptom' which are conformant with the other definitions here provided.

Sign =def. – A bodily feature of a patient that is observed in a physical examination and is hypothesized by the clinician to be of clinical significance.

Symptom =def. – A quality of a patient that is observed and can be observed only by the patient and is of the type that is hypothesized by the patient as a realization of a disease.

Clinical History =def. – A series of statements representing health-relevant features of a patient and of a patient's family.

Clinical History Taking =def. – An interview in which a clinician elicits a clinical history from a patient.

Physical Examination =def. – A sequence of acts of measuring bodily features of an organism performed by a clinician.

Measurements may occur with and without elicitation.

Laboratory Test =def. – A measurement assay that has as input a specimen derived from a patient, and as output a result that represents a quality of the specimen.

Laboratory Finding =def. – A representation of a quality of a specimen that is the output of a laboratory test and that supports an inference to an assertion about some quality of the patient.

Image Finding =def. – A representation of an image that supports an inference to an assertion about some quality of a patient.

Clinical Finding =def. – A representation of a bodily feature of a patient that is either an image finding or the output of a clinical history taking or of a physical examination.

Preclinical Finding =def. – A representation of a bodily feature of a patient that is (1) recorded by a clinician because the feature is hypothesized to be of clinical significance and (2) refers to features obtaining in the patient prior to their becoming detectable in a clinical history taking or physical examination.

Manifestation of a Disease =def. – A bodily feature of a patient that is (a) a deviation from clinical normality that exists in virtue of the realization of a disease and (b) is observable (including observable through elicitation of response and through the use of special instruments).

Preclinical Manifestation of a Disease =def. – A realization of a disease that exists prior to the

manifestation of signs and symptoms associated with that disease.

Clinical Manifestation of a Disease =def. – A realization of a disease that exists subsequent to the time when the disease becomes manifest in signs and symptoms.

Alternative definitions not using 'sign' or 'symptom':

Preclinical Manifestation of a Disease =def. – A realization of a disease that exists prior to its becoming detectable in a clinical history taking or physical examination.

Clinical Manifestation of a Disease =def. – A realization of a disease that is detectable in a clinical history taking or physical examination.

Clinical Phenotype =def. – A constellation of those types of bodily qualities that are associated with a disease at each stage of its development.

Note that, according to this definition, a clinical phenotype can exist without being observed. Indeed, as technology advances, our ability to *detect* the underlying components of the clinical phenotype expands.

Clinical Picture =def. – A representation of a clinical phenotype as instantiated in a given patient that is inferred from the constellation of laboratory, image and clinical findings available to the clinician about a given a patient at any given stage.

Diagnosis =def. – The conclusion of an interpretive process that has as input a clinical picture of a given patient and as output an assertion to the effect that the patient has a disease of such and such a type.

Discussion

The following figure summarizes the view of disease and diagnosis presented in the foregoing. As a result of an etiological process, a physical change occurs in the healthy individual giving rise to a disorder, which initiates the development of a series of manifestations that are initially undetectable without the use of special instruments (pre-clinical) and then become detectable as symptoms and signs. These clinical manifestations of disease constitute in their totality the clinical phenotype for the given disease as instantiated in this specific patient. They can be observed through physical examination and laboratory testing of specimens derived from the patient, the results of which can be recorded in the medical record as the clinical picture. The clinical picture is interpreted by the physician in arriving at a diagnosis, which serves in turn as the foundation for the development of a patient management plan.

Some advantages of this account of disease as dispositions rooted in physically disordered structures in the organism which are realized in pathological processes are that, in contrast to definitions of disease in terms of signs and symptoms or in terms of disease processes, it helps us to do justice (1) to the existence of pre-clinical manifestations of disease, (2) to the different combinations of disease and predispositions to disease in the form of elevated risk, and (3) to the fact that the disease course and the clinical picture may vary widely between patients even where the patients in question have the same disease.

Thus, this framework should support more sophisticated approaches to the treatment of patient data for clinical decision support and for inferential data mining on the part of the translational biomedical informatics research community.

Related Readings

1. Taylor CF, Field D, Sansone S-A, et al. Promoting coherent minimum reporting requirements for biological and biomedical investigations: The MIBBI Project, *Nature Biotechnology*, 2008;26:889-896.
2. Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ; OBI Consortium, Leontis N, Rocca-

Serra P, Ruttenberg A, Sansone SA, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology* 2007; 25 (11): 1251-1255. PMID: 17989687

3. Smith B, Ceusters W, Klagges B, Köhler J, Kumar A, Lomax J, Mungall C, Neuhaus F, Rector AL, Rosse C. Relations in biomedical ontologies. *Genome Biol*, 2005;6(5):R46. PMID: 15892874

4. Smith B., Kumar A., Ceusters W. and Rosse C., On Carcinomas and other pathological entities, *Comparative and Functional Genomics* 6 (2005); 379–387.

5. Schulz S, Johansson I. Continua in biological systems. *The Monist*, 2007; 90(4): 499-522

6. Tu, S.W., Tennakoon, L., OConnor, M., Shakar R., and Das, A.K. (accepted) Using an integrated ontology and information model for querying and reasoning about phenotypes: the case of autism. *Proceedings of the 2008 AMIA Annual Symposium*.

7. Williams, N. The Factory Model of Disease. *The Monist*, 2007; 90(4): 555-584.

