

RESEARCH

Open Access



# A universal diagnosis syntax

Carl-Fredrik Bassøe<sup>1\*</sup>

## Abstract

**Background** Diagnoses are crucial assets of clinical work and provide the foundation for treatment and follow up. They should be informative and customized to the patient's problem. Common prefixes, morphemes, and suffixes may aid the implementation of expressions that generate diagnoses.

**Results** Apt choices of symbols plays a major role in science. In this study, the variables e, o, and p are assigned to names of an etiological agent, a disorder, and a pathogenetic mechanism, respectively. The suffix -itis designates infections, allergies, inflammation, and/or immune reactions. Diagnoses (d) are generated by the formula  $d := e \& o \& p$  where '&' means concatenation and ':=' means assignment. Thus, with  $e := \text{'Staphylococcus aureus'}$ ,  $o := \text{'endocard'}$ , and  $p := \text{'itis'}$ ,  $d := e \& o \& p$  generates the diagnosis  $d = \text{'Staphylococcus aureus endocarditis'}$ . Diagnoses formed this way comply with common clinical diagnoses. Certain extensions generate complete, systematic medical diagnoses that are applicable to all medical specialties. For example, common medical prefixes, morphemes, and suffixes give rise to  $o = \text{'hypothyroidism'}$ ,  $o = \text{'tachycardia'}$ , and  $o = \text{'hypophagocytosis'}$ . The formula scales well with the developments in clinical medicine, systems biology, molecular biology, and microbiology. The diagnosis generating formula  $d := e \& o \& p$  requires meticulous analysis of the components of diagnoses plus the introduction of appropriate variables and terms. Terms partition on established clinical categories and adhere to established clinical nomenclature. The syntax generates universal medical diagnoses.

**Conclusions** The present study concerns a universal diagnosis syntax (UDS) that generates diagnoses using the formula  $d := e \& o \& p$  with several extensions described in the study. The formula is easy to learn and covers diagnoses in all medical specialties. The present work succeeded in creating diagnoses from the formula. The fundamental insight is that no matter how complicated a diagnosis is it can be generated by a systematic process, which adds terms one by one. UDS may have implications for medical education and classifications. The formula lays a foundation for structured clinical decision-making. Formulas are hallmarks of hard science. So,  $d := e \& o \& p$  anticipates a scientific medical revolution.

**Keywords** Medical syntax, Clinical diagnosis, Clinical language

## Background

It is hard to state the rules that govern diagnoses. Some diagnoses are ambiguous. Other seemingly incoherent strings of words are understandable utterances. In addition, there is a wide variety of diagnoses. The aim of this

work are diagnoses that are useful and allow perfectly understandable communication.

Current medical diagnoses are a blend of names of diseases, disorders, syndromes, and clinical findings. The names of diseases, disorders and syndromes consist of proper names and rigid descriptions of clinical entities. Proper names such as *schizophrenia* and *mononucleosis* are invented and *Conn syndrome* and *Hirschsprung's disease* are surnames. Such diagnoses cannot be constructed from morphemes that refer to the etiology, disorder or pathogenetic mechanisms. For this reason, proper names

\*Correspondence:

Carl-Fredrik Bassøe  
bassoe.carl@gmail.com

<sup>1</sup> EkviMed AS, Solfjellveien 7, 1389 Heggedal, Norway



do not count as diagnoses in the present sense of the word.

Systematic diagnoses such as *bacterial sinusitis*, *E. coli cystitis* and *myocardial infarction* are considered to be rigid (definite) descriptions. [1] All these diagnoses may refer to an etiology, one or more disorders and/or pathogenetic mechanisms. In this work, only rigid descriptions count as medical diagnoses. We also require a rather stringent but informal syntax that underlies informal systematic diagnoses such as *Acute Neisseria meningitidis meningitis*, *bacterial tonsillitis*, *Streptococcal tonsillitis*, *chemical alveolitis*, *hereditary spherocytosis* and *idiopathic pulmonary fibrosis* that are obtained from standard medical textbooks and classifications.

The same name is encountered in several diagnoses, but the names may refer to different objects and events. For example,  $\beta$ -cells are found both in the pancreas and pituitary gland. Also, the morpheme *itis* may point to inflammation, infections, allergies, and autoimmune reactions. This study aims to resolve ambiguities that arise from different meanings of terms in different contexts.

The history of medicine tells a long tale of misnomers, for example the obsolete diagnosis *pachymeningitis hemorrhagica interna* [2], p. 32]. The diagnoses *hysteria* and *neurasthenia* were common about a century ago but are rarely used today. Medicine has the capacity of clearing away misnomers, but some still remain. Furthermore, diagnostic terms demand a clear morphology. For example, the term *itiscys* is an incomprehensible misnomer, whereas *cystitis* is a well-formed expression that physicians immediately understand. The diagnoses generated in this study disallows misnomers.

Clinical findings (symptoms, signs, and supplementary investigations) are part of arguments used in clinical decision-making (CDM) to select and create diagnosis. For example, *dysphagia* and *dyspepsia* refer to unclear collections of clinical findings. The present work separates arguments from diagnostic conclusions (diagnoses) and does not allow collections of clinical findings as diagnoses. By doing this we can rid diagnoses of misnomers caused by mixing etiology, disorder and pathogenetic mechanisms with clinical findings.

At least 47 distinct ways for expressing *myocardial infarction* appear in clinical notes [3]. Because of the variety of such diagnoses, they deserve the label natural medical language. This article concerns the development of a universal diagnosis syntax (UDS) that standardizes diagnoses and assigns only one composite form to each diagnosis.

Logic, mathematics, and chemistry have profited from symbolic notations ([4], p. 13). Morphology and syntax fall within the domains of linguistics, logic, the philosophy of language and informatics ([1, 5] p, 32–3, 6,

7). An appropriate syntax for a formal language requires a vocabulary and a set of rules ([6] p.6–7). Transformational-generative grammars have proved useful for language production ([7, 8] p.21). Generative grammar evolved into formal language theory (FLT), which has a wide range of applications [9]. But no such symbol system is available to clinical medicine. This study introduces symbols and a formula that generates systematic medical diagnoses.

Widely used medical classifications such as the 10<sup>th</sup> International Classification of Diseases (ICD-10) [10] and the 2<sup>nd</sup> International Classification of Primary Care (ICPC-2) [11] essentially list proper names and descriptions of diagnoses. They are used in Electronic Patient Records (EPR) to select diagnoses and their associated codes. Any such classification lack important diagnoses. The lists last for some years. Extending and revising them is difficult and time-consuming, and backwards compatibility remains a serious problem.

Static classifications seem to be unsuited for the changing world of CDM. Combinatorial classifications such as the Systematic Nomenclature in Medicine (SNOMED) may overcome these problems [12]. They account for novel diseases and syndromes simply by adding new elements to existing lists and allow new elements to combine with existing elements. However, SNOMED is secluded from the public and the classification lacks an underlying clinical model. Since SNOMED's syntax has no semantics the classification cannot be validated.

Medical terminologies have advanced significantly in the last years ([13] p.124–35) but how to reduce the variety of disease definitions remains an important unsolved problem ([14] p.1, 16). Lack of an agreed infrastructure for terminology is identified as one of the major barriers to information interchange and integrate EPR with medical knowledge bases.

The Unified Medical Language System (UMLS) is an advanced effort towards the integration of biomedical terminologies [15, 16]. The Semantic Network, a component of the UMLS, is a structured description of core biomedical knowledge consisting of well-defined semantic types and relationships between them [17]. However, UMLS does not provide sufficient logic-based structures [18]. Concept-oriented and logic-based approaches are beneficial for creating categorical terminological structures.

The Generalized Architecture for Language Encyclopaedias and Nomenclatures in Medicine (GALEN) and UMLS are large thesauri [18]. One aim of these projects is to bridge the gap between different terminology systems using a conceptual model and mapping facilities to natural language expressions and coding schemas [17, 19, 20].

The GALEN project aims to bridge the gap between different terminology systems through a terminology server, which contains a conceptual model and mapping facilities to natural language expressions and coding schemas [19, 20]. Several projects have been launched to converge clinical terminologies towards a grand unified system [21, 22]. The complexity and high number of medical and health terminologies lead more recent projects to limit their attack on man-machine and machine-machine interoperability to limited domains [23–27]. Despite ongoing work toward shared data formats and linked identifiers, significant problems persist in semantic data integration across heterogeneous biomedical data sources [28]. It remains difficult to establish shared identity and shared meaning.

A formal language is characterized by its vocabulary and syntax ([1], p.26). The vocabulary consists of three basic expressions: logical terms, logical variables, and auxiliary signs such as brackets. There is also a set of rules which show how expressions can be combined to make new expressions. The meaning of expressions is determined by Frege's principle of compositionality: the meaning of a composite expression is wholly determined by the meanings of its component parts and the syntactic rule by which it was formed [1, 4]. There are opponents to this view ([29] p. 219–37) but for diagnoses we adhere to Frege's principle.

Systematic diagnoses can be decomposed into their component parts [30]. For example, *bacterial conjunctivitis* can be parsed into the etiology *bacteria*, the body part *conjunctiva* and the pathogenesis *itis*. Finally, a flexible combinatorial classification, which was based on the same components was implemented in the early days of EPR in Norway [31] and worked according to purpose in another EPR ([32, 33] p.212–4).

We tried a dynamic approach that lets physicians write ordinary textbook diagnoses into diagnosis fields in an EPR [34]. The diagnoses were some years later associated with an ICPC-2 code. Also, physicians could change the diagnosis name associated with a code when the name was inappropriate. That this option was used shows the advantage of being able to change diagnoses.

Various clinical specialties use the same diagnoses and operate within the domains etiology, disorders, and pathogenesis. This indicates that all specialties may use a common formula for generating diagnoses.

This study aims to base a universal diagnosis syntax (UDS) on a formula and clinically meaningful terms. The assignment of strings to variables is purely syntactic and the strings do not embody meaning by themselves [9]. The variables of the formula are instantiated with names of concepts that are derived from an empirical clinical

model [35]. The relationship between the present syntax and its semantics is investigated separately.

Health personnel that are not acquainted with formal languages or informatics need not worry. All the rules and terms are explained and aligned with standard school medicine. Use and mention of terms are distinguished typographically. Use: Lungs are in the chest. Mention: *Lungs* consists of five letters. Terms mentioned in formula are embraced by single brackets. Thus, in a formula *lung* is converted to 'lung'. Also, *Lungs* is a string of characters in [a-z, A-Z]. Rules for generating diagnoses show how to construct them and how to test them and show whether they are valid or not.

## Methods

This study aims at a syntax for systematic medical diagnoses. The definition of the syntax derives from the structure of definitions in first order logic [1]. The definition consists of a basic vocabulary and a formula for generating diagnoses. The formula is later extended to diagnoses in all medical specialties.

The design of the vocabulary and formula is iterative and evolved over many years. The vocabulary is derived from medical textbooks, reviews, and numerous searches on PubMed.

The setting is university studies (M.D., Master of Information Technology, and Master of Philosophy), practical work (surgery, primary care, occupational medicine, pathology, oncology), awarded specialties (general internal medicine and hematology), research experience (experimental medicine (PhD), medical informatics (PhD), and clinical research), and teaching as professor (internal medicine and health informatics).

The study involves no participants other than the author, and no drugs, other materials, processes, or statistical analysis.

## Results

### Universal diagnosis syntax (UDS)

Systematic diagnoses are character strings that name elements within the clinical domains etiology, disorders, and pathogenesis. UDS constitutes terms, variables, and rules that together generate medical diagnoses. That the character string 'abc' is assigned to the variable *v* is symbolized by  $v := 'abc'$  where  $:=$  is the assignment symbol. The equality sign  $=$  and the inequality sign  $\neq$  are used to check whether two-character strings are identical or not, or whether two variables refer to the same character string or not.

Concatenation joins character strings end-to-end. The symbol for concatenation is  $\&$ . For example, the strings 'abc' and 'abc' are concatenated by 'abc'  $\&$  'abc' into 'abcabc' and 'abc'  $\&$  'abc' = 'abc abc' since the left

hand ‘abc’ ends with a space. Since ‘abc’=‘abc’ whereas ‘abc’≠‘acb’ concatenation is not commutative. If  $v_1 := 'abc'$ ,  $v_2 := 'abc'$ , and  $v_3 := 'acb'$  then  $v_1 = v_2$ ,  $v_2 \neq v_3$ , and  $v_1 \neq v_3$ .

**Definition 1**

- i) The basic vocabulary of UDS consists of terms, connectives, and variables.
- ii) UDS connectives are :=, &, = and ≠. & takes precedence over :=.
- iii) Terms are strings of characters such as *bacteria* and *kidney*, which refer to the objects bacteria and kidney, respectively.
- iv) Terms obtain from the underlying sets of the etiology (E), disorder (D) and pathogenetic mechanisms (P), which derive from clinical model.
- v) The variables e, o, and p are assigned to terms from E, O and P, respectively. Thus, in  $e := 'bacteria'$  the string ‘bacteria’ from E is assigned to the etiology variable e. Thus, E, O, and P are different types.
- vi) The variables restrict the types of terms that can be assigned to them. The left hand side of the connective := holds one and only one variable and no terms and no other connectives. The right hand side of := contains one term or a concatenation of an arbitrary number of variables and terms. Alternative terms that can be assigned to one variable are discriminated by the sign |.
- vii) Diagnoses are generated by

$$d := e\&o\&p \tag{1}$$

- viii) Formula 1) defines the order of the variables uniquely from left to right. The variables are non-commutative, which means that  $d := e\&o\&p$  is a valid formula, but  $d := e\&p\&o$ ,  $d := o\&e\&p$ ,  $d := o\&p\&e$ ,  $d := p\&e\&o$ , and  $d := p\&o\&e$  are invalid. In other variables the first letter is uppercase letter followed by lower case.
- ix) Only strings generated by i) to vii) in a finite number of steps are diagnoses.

In order to increase readability ‘&’ associated with variables can be discarded whereby  $e\&o\&p$  is abbreviated to eop.

The variables e, o and p partition diagnostic terms into the three categories etiology, disorder and pathogenesis, respectively. In contrast to first order logic, the categorical distribution of descriptive terms such as *streptococcal*, *tonsil* and *fibrosis* are important to UDS diagnoses ([1], p.7). This empirical lexical prerequisite has no exceptions.

The following pseudocode generates diagnosis in UDS that accord with formula 1:

```

e := 'E.coli'
o := 'cyst'
p := 'itis'
d := eop
-----
d = 'E.coli cystitis'
    
```

Clearly, the same diagnosis is generated independently of the sequence of instantiation of e, o, and p. Therefore, formula 1) lends the same structure to all UDS diagnoses. All such diagnoses are well-formed formulas. All the variables in formula 1) are important for treatment and follow up. Definitions i-ix are assumed to hold for all clinical specialties.

Additional rules discriminate common types of diagnoses:

1. **Diseases** are complete and systematic diagnoses generated by the formula  $d := eop$  and requires that  $e \neq ''$ ,  $o \neq ''$ , and  $p \neq ''$ . For example, infections refer to the triple of microorganism, disorder, and immune reaction. Therefore, infections are diseases. By definition, hereditary disorders are diseases.
2. **Disorders** are incomplete systematic diagnoses generated by  $d := o$ .
3. **Syndromes** are incomplete systematic diagnoses generated by  $d := op$ . In UDS syndromes are not names of genetic disorders consisting of collections of clinical findings.
4. Diagnoses with unknown anatomical localization are generated by  $d := ep$  or  $d := p$ .
5. Etiological agents alone do not give rise to diagnoses. Hence,  $d := e$  is not a valid diagnosis. Etiological agents that are not translocated into a primary affected body part only qualify as risk factors.

Formula 1) is implemented from an algorithm and the knowledgebase contains complete lists of terms.

**Body parts—scope**

Diagnoses often describe the localization of a disorder or pathogenetic mechanism such as *bilateral crural edema* and *finger eczema*. Many dermatological disorders such as psoriasis are characterized by their regional distribution. The anatomical scope of such disorders is limited.

The scope of disorders is described by terms that name body parts. Scope partitions into segments *Seg* and regions *Reg*. The site variable *Site* determines absolute and relative locations. The ‘right kidney’ is an absolute reference. In contrast, in ‘tumor to the left of the right kidney’ the tumor is located relative to the right kidney. Recursive use of *Site* is allowed. Thus,  $Site := 'above'$  can be followed by  $Site := Site \& 'left\ kidney'$ . The vertical bar | discriminates alternatives. Typical examples are:

Seg := 'C1' | 'C2' | 'C3' | 'C4' | ...

Reg := 'frontal' | 'occipital' | ... | 'abdominal' | 'crural' | ...

Site := 'right' | 'left' | 'bilateral' | 'in' | 'left of' | 'to the right of' | 'to the left of' | 'to the right of' | 'lateral to' | 'medial to' | 'proximal to' | ...

The scope is considered by modifying formula 1) into

$$d := \text{Seg}\&\text{Reg}\&\text{Site}\&\text{eop} \tag{2}$$

In case of *crural edema* Reg := 'crural' and p := 'edema'. The etiology has not been investigated and is empty, i.e., the initialization leaves Seg=Site=e=o="". Therefore, formula 2) leads to d:= 'crural edema', which is obviously an incomplete diagnosis. Note that *crural swelling* is a sign, but *crural edema* requires a clinical inference to count as a pathogenetic mechanism.

The variable o incorporates morphemes of names of body parts held by the variable N. The variable N<sub>X</sub>, holds the morpheme of an organ system, organ, organ part, tissue, or a cell. The index X is instantiated by P, T and S, which discriminate the variables that hold names of parenchyma N<sub>p</sub>, tubes N<sub>T</sub> and slits|cavities N<sub>S</sub>, respectively. Thus,

N<sub>p</sub> := 'hepat' | 'neph' | 'dermat' | 'bone' | 'tendon' | 'myocard' | 'ligament' | 'intervertebral disc' | 'meniscus' | 'cornea' | 'lens' | 'corpus vitreum' ...

N<sub>T</sub> := 'Eustachian tube' | 'coledoch' | 'salping' | 'arterio' | 'aortic' ...

N<sub>S</sub> := 'cholecyst' | 'pyelo' | ...

The tissue N<sub>p</sub>, cell line N<sub>L</sub>, cell variable N<sub>C</sub> and their terms are:

N<sub>I</sub> := 'squamous cell' | 'lymphoid' | 'connective tissue' | 'adipose tissue' | 'osteo' | 'fibro' | 'lipo' | ...

N<sub>L</sub> := 'erythropoiesis' | 'spermatogenesis' | ...

N<sub>C</sub> := 'β-cell' | 'α-cell' | 'B-lymphocyte' | 'parietal cell' | 'adipocyte' | 'fibroblast' | 'mast cell' | 'neuron' | 'purkinje cell' | 'striated muscle cell' | 'basal cell' ...

The subcellular level cover subtler disorders. In the variable N<sub>CY</sub> the index C points to cells and Y to morphemes of organelles and other subcellular structures. The variables and their corresponding terms are:

N<sub>YO</sub> := 'dendrite' | 'perikaryon' | 'axon' | 'synapse' | 'cytoskeleton' |

'microtubule' | 'mitochondria' | 'Golgi apparatus' | 'nuclear double membrane space' | 'cell nucleus' | 'vacuole' | 'lysosome' | 'endosome' | 'phagosome' | 'sarcoplasmic reticulum' | ...

The variables can be concatenated into N := N<sub>p</sub>N<sub>L</sub>N<sub>I</sub>N<sub>C</sub>N<sub>O</sub>, which allows all possible combinations of morphological disorders in subcellular structures of cells in a tissue residing in an organ.

Some diagnoses generated with the variables N<sub>p</sub>, N<sub>T</sub> and N<sub>S</sub> are concatenated with suffixes like 'ism', and 'itis'.

For example, 'gonad' & 'ism' becomes 'gonadism', 'thyroid' & 'ism' is 'thyroidism' and 'thyroid' & 'itis' contracts to 'thyroiditis'. The formula secures that suffixes correspond with the actual etiology, disorder, and pathogenesis.

### Direction of change—Dir

Morphological and functional directions of change are denoted by well-known clinical prefixes. The direction may differ in the etiology, disorder, and pathogenesis. Also, morphology and function may require different directions of change even in one and the same diagnosis. The direction of change is assigned to the variable Dir<sub>X</sub><sup>Y</sup>. Dir<sub>E</sub>, Dir<sub>O</sub><sup>Y</sup>, Dir<sub>T</sub><sup>Y</sup>, Dir<sub>S</sub><sup>Y</sup> and Dir<sub>P</sub> are used with Dir<sub>P</sub><sup>Y</sup> the etiology E, and parenchyma O, tube T, or slit S disorder, and pathogenesis P, respectively. For the etiology and pathogenesis the superscript Y is ignored as in Dir<sub>E</sub> and Dir<sub>P</sub>.

Dir<sub>X</sub><sup>M</sup> and Dir<sub>X</sub><sup>F</sup> discriminate morphology M from function F. Some other X and Y values are set aside for specific parenchyma functions and pathogenetic mechanisms. For example, *hypoglycemia due to hyperinsulinemia* belongs to the pathogenesis and requires X := 'P'. In this context, Y=B and Y=H refer to metabolites and regulators, respectively. Therefore, Dir<sub>O</sub><sup>B</sup> := 'hypo' and Dir<sub>O</sub><sup>H</sup> := 'hyper' depend on context. Accordingly, the scope of Dir<sub>X</sub><sup>Y</sup> is limited to the term it associates with. The direction of change is assigned to the variable Dir<sub>X</sub><sup>Y</sup> as follows:

Dir<sub>X</sub><sup>Y</sup> := 'a' | 'an' | 'hypo' | 'normo' | 'hyper' | 'meta' | 'dys' | 'neo' | 'extra' | 'brady' | 'tachy' | 'localized' | 'generalized' | 'paroxystic' | 'seizure' | 'epi' | 'allo' | 'anti' | 'iso' | 'deficiency' | 'para' | 'elevated' | 'diminished' | ...

These prefixes pertain to the disorder variable o in formula 1) and 2) and describe the direction altered morphology and/or function. Crude disorders of function are generated by O<sub>p</sub><sup>F</sup> := Dir<sub>O</sub><sup>F</sup>&N<sub>p</sub>. The variable F<sub>T</sub> is a tube function. Diagnoses of tube functions are generated by O<sub>T</sub><sup>F</sup> := N<sub>T</sub>&Dir<sub>T</sub><sup>F</sup>&F<sub>T</sub>. Thus, if N<sub>T</sub> := 'arterial', Dir<sub>T</sub><sup>F</sup> := 'hypo' and F<sub>T</sub> := 'tension' then the diagnosis is *arterial hypotension*. If a term starts with a vocal then 'a' is automatically replaced with 'an'. The terms associated with Dir<sub>X</sub><sup>Y</sup> are always selected from the list above, which is implemented in the knowledge base.

### The disorder variable o

The diagnosis *Staphylococcus aureus myocarditis* does not tell whether the infestation causes heart failure, arrhythmia or has no adverse effect on heart function. In fact, the diagnosis is incomplete with regard to heart morphology and function. Complete diagnoses require variables and terms that inform on the morphology and function of the affected body part.

The variable  $O_X^Y$  additionally discriminates parts of the general organ.  $X=P, X=T, X=S, X:=I, X:=L, X:=C$  and  $X:=O$  refer to parenchyma, tubes, slits|cavities, tissues, cell lines, cells, and cell organelles, respectively. Thus, diagnoses of morphology and function in particular body parts are given by

$$O_X^Y := N_X S_V K L D i r_X^Y Y_X^Z \tag{3}$$

where  $Y=M$  and  $Y=F$  means morphology and function, respectively.  $S_V$  represents the suffixes  $S_E, S_O$  or  $S_P$  that can be attached to  $N_X$ , for example  $S_P :=$  'lithiasis' attaches to  $N_O =$  'nephro' in *nephrolithiasis*.

The variables  $K$  and  $L$  only concern names of metabolites and regulators. Otherwise,  $K :=$  ",  $L :=$  " and 3) simplifies to  $O_X^Y := N_X \& D i r_X^Y \& Y_X^Z$ . The index  $X$  pertains to the body parts as described above. The superscript  $Z$  handles specific functions and holds a single character (see below). In morphological diagnoses  $Z :=$  " and formula 3) simplifies to  $O_X^M := N_X \& D i r_X^M \& M_X$  and  $O_X^F := N_X \& K \& L \& D i r_X^F \& F_X^Z$ , respectively. For example, disorders of parenchyma morphology are generated by  $O_P^M := N_P \& D i r_X^M \& M_P$ .

The lungs, liver and kidneys are composite organs that consist of several parenchyma, tube systems, and slits/cavities. Disorders in composite organs may concern more than one part of the organ. This applies to many liver diseases accompanied by hepatocyte swelling that give rise to both intrahepatic cholestasis and loss of liver functions. All these changes should be explicitly described. Very complex disorders can be generated by

$$\begin{aligned} O^M &:= O_P^M O_T^M O_S^M O_I^M O_L^M O_C^M O_O^M \\ O^F &:= O_P^F O_T^F O_S^F O_I^F O_L^F O_C^F O_O^F \end{aligned} \tag{4}$$

In most clinical situations only one variable is instantiated, and the others remain empty. For example, common disorders such as *myocardial hypertrophy* and *arterial hypertension* are generated by  $O^M := O_P^M$  (with  $N_P :=$  'myocard',  $S_O :=$  'ial',  $D i r_O^M :=$  'hyper',  $M_P :=$  'trophy') & ' and ' &  $O^F := O_T^F$  (with  $N_T :=$  'arter',  $S_T :=$  'ial',  $D i r_T^F :=$  'hyper',  $F_T :=$  'tension') respectively, and keeping the other variables in formula 4) empty.

$D i r_X^Y$  can be considered as a function in itself. Parentheses are included in UDS to limit the scope of  $D i r_X^Y$ . Therefore,  $D i r_X^Y()$  may range over more than one morphology and function. For example,  $D i r_O^M(M_P F_P) = D i r_O^M \& M_P \& D i r_O^F \& F_P$ . Thus, if  $D i r_O^M =$  'hypo' then  $D i r_O^M(M_P F_P) =$  'hypo' &  $M_P$  & 'hypo' &  $F_P$ .

The direction of morphological and functional disorders of one body part do not always correlate. For example, goiter and hypothyroidism are characteristic of iodine deficiency. Another example is hepatocellular carcinoma with hepatic hypofunction. Hence, accurate

diagnoses may need descriptions of both morphology and function. The variable  $o$  in formula 1) and 2) holds the variable  $O^Y$  that pertains to morphology if  $Y=M$  and function if  $Y=F$ . This opens for the possibility to name causal relations. The diagnoses of compound disorders are given by

$$o := O^M \text{ \& 'causing' \& } O^F \tag{5}$$

A typical example is  $o =$  'hepatic carcinoma causing hepatic hypofunction'.

### Parenchyma and connective tissue

*Morphology* Morphological parenchyma diagnoses stored in  $O_P^M$  are generated by  $X := P, Z :=$  " and formula 3). Since  $Y=M$  the following variables hold morphological terms:

$N_P := M_1 := M_C :=$  'tumor' | 'adenoma' | 'oma' | 'pathy' | 'trophy' | 'plasia' | 'sarcoma' | 'carcinoma' | 'adenocarcinoma' | 'papillary carcinoma' | 'fissure' | 'fracture' | 'genesis' | 'degeneration' | 'porosis'....

In the unspecific diagnoses all  $D i r_O^Y :=$  ". Thus, if  $N_P :=$  'thyroid' and  $M_P :=$  'tumor' then formula 3) gives  $O_P^M =$  'thyroid tumor'. Alternatively, if  $N_P :=$  'thyroid' and  $M_P :=$  'adenoma' then  $O_P^M =$  'thyroid adenoma'. More serious tumor disorders are given by  $D i r_O^M :=$  'neo' and  $M_P :=$  'plasia', where for example  $N_P :=$  'thyroid' leads to  $O_P^M =$  'thyroid neoplasia'. Unknown morphology, i.e.,  $O_P^M :=$  " is also allowed.

Certain benign morphological tissue diagnoses are generated by 3) with  $X := I$  and  $M_1 :=$  'oma'. Then  $N_I :=$  'lip' gives  $O_I^M =$  'lipoma' and if  $N_I :=$  'fibr' we get  $O_I^M =$  'fibroma'. These simple concatenations cover a wide variety of benign tumors. In contrast, if  $N_I :=$  'lymph' we get  $O_I^M =$  'lymphoma'. The latter covers tumors with varying degrees of malignancy, and it is necessary to specify the degree of the malignancy (see below).  $M_1 :=$  'sarcoma' allows the generation of 'liposarcoma', 'fibrosarcoma' and 'rhabdomyosarcoma'. The concatenation with region or site generates the anatomic localization of other *omas* and *sarcomas* in accordance with the anatomic scope of the disorder.

The tissue and cell type are generated by simple rules. If  $N_I :=$  'squamous' and  $D i r_O^M :=$  'neo' then  $M_1 =$  'carcinoma'. In this case and with  $N_P :=$  'pulmonary' formula 3) gives  $O_P^M =$  'pulmonary carcinoma'. Likewise, if instead  $N_P :=$  'gastric' we get  $O_P^M =$  'gastric carcinoma'.

The brain is often considered to be something very special and complex. However, morphological brain disorders behave as other organs and tissues. If  $N_C :=$  'astrocyt' and  $M_P :=$  'oma' then formula 3) gives

$O_p^M$  = ‘astrocytoma.’ ‘anaplastic neuroblastoma.’ is generated by the same principle.

Alzheimer’s disorder is generated by  $Reg :=$  ‘frontal’,  $N_p :=$  ‘cortical’,  $Dir_O^M :=$  ‘’ and  $M_p :=$  ‘degeneration’ that gives  $o := O_p^M$  = ‘cortical degeneration’ and  $d =$  ‘frontal cortical degeneration’ because  $Seg :=$ ,  $Site :=$ ,  $e :=$ , and  $p :=$ . Clearly, Alzheimer’s is not (yet) a disease since the etiology is unknown ( $e :=$ ).

Alzheimer’s disorder has a characteristic morphology, but frontal brain atrophy diagnosed by CT- and MRI-scans may have other causes. The uncertainty is accounted for by selecting the more unspecific terms  $Reg :=$  ‘frontal’,  $Dir_O^M :=$  ‘a’ and  $M_p :=$  ‘trophy’, which generates  $O_p^M$  = ‘cortical atrophy’ and  $d =$  ‘frontal cortical atrophy’. The meanings of *atrophy* and *degeneration* differ.

Morphological cell line disorders are stated and recognized by formula 3) with  $X=L$ . Thus, when  $N_L :=$  ‘myelo’,  $Dir_p :=$  ‘dys’ and  $M_L :=$  ‘plasia’ the disorder is  $O_L^M$  = ‘myelodysplasia’.

In ICD-10 ‘Myelodysplasia’ is used for developmental disorders of the central nervous system only and does not occur as a hematological disorder [10]. This is quite unfortunate because hematologic myelodysplasia is a fairly common disorder. In this study the difference between the hematologic and neurologic myelodysplasias is clarified by using the different variables  $O_L^M/N_L^M$  and  $O_p^M/N_p^M$ , respectively. Similar ambiguities are resolved by using the same principle.

Cells exhibit morphological disorders  $M_C$  that are characterized by formula 3) with  $X=C$  and

$M_C =$  ‘surface membrane leakage’ | ‘lysis’ | ‘necrosis’ | ‘proliferation’ | ‘apoptosis’ ...

If  $N_C :=$  ‘hemo’,  $Dir_C^M :=$  ‘’ and  $M_C :=$  ‘lysis’ then  $O_C^M =$  ‘hemolysis’. Slightly more complicated is  $N_C :=$  ‘Hepatocellular’,  $Dir_C^M :=$  ‘’ and  $M_p :=$  ‘surface membrane leakage’ which gives  $O_C^M =$  ‘Hepatocellular surface membrane leakage’.

Today, cell proliferation and apoptosis are measured on many tissue samples. Disorders of these processes are accommodated for by for example  $N_C :=$  ‘B-lymphocyte’,  $Dir_C^M :=$  ‘hypo’ and  $M_p :=$  ‘apoptosis’, which gives  $O_C^M =$  ‘B-lymphocyte hypoapoptosis’, which is observed in certain lymphomas.  $\beta$ -cell hyperplasia is a rare disorder, but it is encountered in clinical practice. The diagnosis is generated by  $N_C :=$  ‘ $\beta$ -cell’,  $Dir_C^M :=$  ‘hyper’ and  $M_p :=$  ‘plasia’. However,  $\beta$ -cells are encountered in various body parts such as the pancreas and pituitary gland. The organ term  $N_p :=$  ‘pancreatic’ and  $N_L := N_L :=$  ‘’ in  $O_C^M := N_p N_L N_L N_C Dir_C^M M_p$  eliminates the ambiguity.

The treatment of leukemia depends on the affected cell line and arrested maturation stage. Highly specific

drugs are used in promyelocyte leukemia (AML-M3). The systematic name of the disorder is ‘promyelocyte neoplasia’. In order to comply with clinical praxis, we reformulate using the rule: IF  $N_C =$  ‘promyelocyte’ and  $Dir_C^M =$  ‘neo’ THEN  $Dir_C^M :=$  ‘’ and  $M_p :=$  ‘leukemia’, which generates ‘promyelocyte leukemia’. The rule works for other leukemias and lymphomas too (not shown).

**Function** Names of disorders of function are specified by the variable  $O_X^Y$  with  $Y := F$  and  $X := P, X := T, X := S, X := I, X := L$  or  $X := C$  and formula 3). Accordingly,  $Dir_O^F :=$  ‘hypo’,  $F_p :=$  ‘function’ and ‘ $N_p :=$  ‘gonad’ generates the rather unspecific diagnosis  $O_p^F =$  ‘gonad hypofunction’. But such diagnoses are quite useful. *Hepatocyte hypofunction* means that the secretion of albumin and coagulation factors, and the conjugation of bilirubin and secretion of bile acids are all reduced. Thus, the diagnosis sums up a lot of information in two words. Similar diagnoses for overall function such as *pulmonary hypofunction*, *myocardial hypofunction*, and *hypogonadism* follow the same rule. Also, unspecific polymorphonuclear neutrophilic granulocyte (PMN) hypofunction is diagnosed as  $O_C^F =$  ‘PMN hypofunction’ suffices in some clinical situations.

But many important disorders require higher resolution of the clinical problem at hand. Typical instances are isolated hormone deficiencies and isolated hyperbilirubinemia without other signs of liver failure. Also, the emergency discrimination between heart failure due to myocardial contraction hypofunction and heart rhythm disturbance is crucial for the selection of treatment and prognosis. Therefore, function disorders partition into genomics  $F_p^G$ , mechanical  $F_p^M$ , electric  $F_p^E$ , metabolic  $F_p^B$ , humoral regulation  $F_p^H$ , optic  $F_p^O$ , thermal  $F_p^R$ , immune  $F_p^I$ , and mental  $F_p^N$ .

**Genomic diagnoses** Names involved in genome diagnoses are varied. We have to cover chromosome abnormalities, gene families [36], point mutations [37], and translocations [38]. The morphemes are:

$N_G :=$  ‘Chr 1’ | ... | ‘Chr 46’ | ‘Philadelphia chromosome’ | ‘X0’ | ‘XXY’ | ‘G506A’ | ‘*SLC37 family*’ | ‘t(8;21)(q22;q22.1)’ ...

The molecular changes are defined for the actual parenchyma  $p$  by.

$F_p^G =$  ‘mutation’ | ‘aneuploidy’ | ‘translocation’ | ...

**Mechanical functions** Terms describing mechanical functions are:

$F_p^M$  := 'function ina' | 'contraction' | 'relaxation' | 'sliding' | 'rolling' | 'systole' | 'diastole' | 'stability' | 'tonia' | 'spasticity' | 'rigidity' | 'shivering' | 'kinesia' | ... ..

We limit the discussion to hypertonia. *Hypertonia* refers to motor disorders of the nervous system and divides into spasticity, dystonia, and rigidity [39]. *Hypertonia* is derived from  $Dir_O^F$  := 'hyper',  $F_p^M$  := 'tonia' and  $Dir_O^F F_p^M$ . *Dystonia* is expressed by  $Dir_O^F$  := 'dys' and  $F_p^M$  := 'tonia'. The other alternatives are simply  $Dir_O^F$  := " combined with  $F_p^M$  := 'spasticity' or  $F_p^M$  := 'rigidity'.  $N_p$  in  $o := N_p Dir_O^F F_p^M$  specifies the anatomical location of the disorder.

**Electric functions** The surface membrane of all cells has electric functions. For practical reasons, we limit the cell types to striated and smooth muscle, and nerve cells. Disorders of frequency are described by the prefixes and the words fibrillation and flutter.

$F_p^E$  := 'function ina' | 'frequency' | 'amplitude' | 'axis' | 'arrhythmia' | 'fibrillation' | 'flutter' | 'amplitude' | 'synchrony' | 'spike' | 'epilepsy' .....

For example, let  $N_p$  := 'atrial'. Then  $Dir_O^F$  := " and  $F_p^E$  := 'fibrillation' generates  $O_p^E$  = 'atrial fibrillation'. If  $Dir_O^F$  := 'extra' and  $F_p^E$  := 'systole' then  $O_p^E$  = 'atrial extrasystole'. Also Let  $N_p$  := 'sinus',  $Dir_O^F$  := 'tachy' and  $F_p^E$  := 'arrhythmia' then  $O_p^E$  = 'sinus tachyarrhythmia'.

For ventricular arrhythmias  $N_p$  := 'ventricular'. Then  $F_p^E$  := 'flutter' generates  $O_p^E$  = 'ventricular flutter'. If  $Dir_O^F$  := 'extra' and  $F_p^E$  := 'systole' then  $O_p^E$  = 'ventricular extrasystole'. Also, let  $N_p$  := 'AV-node',  $Dir_O^F$  := 'brady' and  $F_p^E$  := 'arrhythmia' then  $O_p^E$  = 'AV-node bradyarrhythmia'.

Next take some cerebral arrhythmias.  $F_p^E$  := 'epilepsy' can generate four types of epilepsy by selecting either  $Reg$  := 'frontal',  $Reg$  := 'temporal',  $Reg$  := 'parietal' or  $Reg$  := 'occipital'. The diagnosis of various specific epileptic types can be differentiated by adding  $F_p^E$  terms that are selected from EEG measurements (not shown) see [42].

**Metabolism and regulators** The variable K and L in formula 3) is introduced for terms naming metabolites and regulators. The two are often related such as glucose and insulin. Also, two regulators TSH and thyroxin (T4) are closely related. The variables  $N_X$ ,  $F_p^X$  and  $F_C^X$  allow  $X=B$  or  $X=H$ . Therefore, K and L occur as  $K := Dir_B^F N_B$ ,  $K := "$ ,  $L := Dir_H^F N_H$ , and  $L := "$ . The terms are:

$N_B$  := 'glyc' | 'glycogen' | 'triglyceride' | 'DNA' | 'purine' | 'cytosine' | 'albumin' | 'myosin' | 'actin' | 'pigment' | 'melanin' | 'rhodopsin' .....

$N_H$  := 'insulin' | 'thyroxin' | 'FSH' | 'NGF' | 'TGFβ' | ...  
 $F_p^B$  and  $F_C^B$  determine the scope of  $O_p^F$  and  $O_C^F$ . Overall metabolic parenchyma function is described by  $F_p^B$ .

$F_p^B$  := 'function ina' | 'secretion' | 'absorption' | 'intracellular metabolism' | 'receptor activity' | 'membrane channel activity' | 'concentration' | ...

$F_p^H$  := 'function ina' | 'secretion' | 'receptor activity' | 'concentration' | ...

In this list *glyc* is the term for *glucose* used in metabolic diagnoses such as *hyperglycemia*. Also, formula 3) generates diagnoses as varied as *adipocyte triglyceride hyposecretion* and *hematopoietic stem cell DNA synthesis*.

Since the β-cells' major product is insulin, type I diabetes is the disorder given by  $N_C$  := 'β-cell',  $N_H$  := 'insulin',  $Dir_H^F$  := 'hypo' and  $F_C^B$  := 'secretion' diagnosed as  $O_C^H$  = 'β-cell insulin hyposecretion' using  $O_C^E$  :=  $N_C N_H Dir_H^F F_C^B$ . Note that the diagnosis *diabetes type I* lacks terms for etiology and pathogenesis and is incomplete.

**Optic** The terms are:

$F_p^O$  := 'function ina' | 'refraction' | 'absorption' | 'transmission' | 'metropia' ...

The diagnosis *hypermetropia* is generated by assignments  $Dir_O^F$  := 'hyper',  $F_p^O$  := 'metropia'. The same goes for all refraction disorders. Color vision diagnoses are accounted for by genome and metabolic function.

**Thermal functions** The terms are:

$F_p^R$  := 'function ina' | 'thermia' | ...

The most common diagnoses are *hypothermia* and *hyperthermia*.

**Immune functions** The terms are:

$F_p^I$  := 'function ina' | 'immuno' | 'tolerance' | ...

Compliance with immunological, hematological, and certain other clinical uses of diagnoses requires the introduction of a simple rule. The first character in 'immuno' is a space. In this context the space means that if  $Dir_O^F$  := 'hypo' then *hypo* is switched with *deficiency* to obtain *immunodeficiency*.

**Cell functions** A more precise diagnosis is necessary in specialties such as hematology, infectious diseases, endocrinology, and immunology. Drilling down into the depth of medical complexity requires exact names for disorders of cell function. Hence, particular functions are assigned to the variable  $F_C^Z$  where  $Z=M$ ,  $Z=B$ ,  $Z=H$  or  $Z=E$ .



$F_C^M$  := ‘phagocytosis’ | ‘killing’ | ‘chemotaxis’ | ‘adhesion’ | ‘rolling’ | ‘sliding’ | ...  
 $F_C^B$  :=  $F_P^B$   
 $F_C^H$  :=  $F_P^H$   
 $F_C^E$  := ‘resting membrane potential’ | ‘polarization’ | ‘repolarization’ | ‘conduction’ | ‘resistance’ | ...

With regard to disorders of mechanical functions of monocytes, macrophages, dendritic cells and PMN, for example, we have  $O_C^F$  = ‘PMN hypophagocytosis’  $O_C^F$  = ‘PMN hypochemotaxis’ and  $O_C^E$  = ‘monocyte hypoadhesion’.

Within the scope of  $F_C^E$ ,  $Dir_O^F$  = ” means the healthy electrophysiological process.  $Dir_O^F$  := ‘hypo’ and  $Dir_O^F$  := ‘hyper’ allow for  $O_C^F$  to be *hypopolarization* and *hyperpolarization*, respectively.

**Organelle functions** Many clinical disorders are caused by disorders of intracellular functions, for example inborn metabolic storage diseases. They can be catalogued using the functions  $F_O^Y$  where Y accounts for the type of function as used above with parenchyma and cells. For example,  $F_O^M$  := ‘rolling’ | ‘sliding’ suffices to describe the rolling of vesicles along axon fibers and the sliding of myosin along actin in contracting striated muscle cells. The number of known combinations is huge. Therefore, only one  $F_O^B$  alternative is summarized.

The cell name variable  $N_C$  determines the scope of the functions  $F_O^Y$ . A large group of disorders are characterized by  $N_O$  := ‘mitochondria’,  $Dir_O^F$  := ‘hypo’ and  $F_O^E$  := ‘resting membrane potential’, which means  $O_O^E$  = ‘mitochondria hyporestingmembranepotential’. The diagnosis may perhaps seem cumbersome, but it is precise.

Appropriate diagnoses may demand the inclusion of organelles. Organelle disorders are generated from the organelle name  $N_O$  and structural  $M_O$ , and functional mechanical  $F_O^M$ , metabolic  $F_O^B$  and/or electric  $F_O^E$  disorders. If the function of an organelle type is disturbed in one cell line then that association is made, e.g., lysosomal enzyme  $N_S$  deficiency ( $Dir_O^F$  := ‘a’,  $F_O^B$  := ‘concentration’) in cell  $N_C$  of parenchyma  $N_P$ . In general, disordered organelle function involves a substance  $N_S$  and occurs within the scope of a cell type and a parenchyma  $N_P$ . The formula  $o := N_P N_C N_O N_S Dir_O^F F_O$  deriving from 3) takes the scope of substance disorders into account.

**Mental functions** The following are diagnostic terms pertaining to conscious mental functions covering cognition and emotions:

$F_P^N$  := ‘thinking’ | ‘cognition’ | ‘intelligence’ | ‘mnesia’ | ‘perseveration’ | ‘psychosis’ | ‘anxiety’ | ‘depression’ |

‘perseverance’ | ‘mood’ | ‘syntax’ | ‘calculia’ | ‘algebra’ | ‘praxia’ | ‘short term memory’ | ‘remembering’ | ‘tranquility’ | ‘anxiety’ | ‘sadness’ | ‘melancholy’ | ‘osmia’ | ‘geusia’ | ‘acusis’ | ‘tinnitus’ | ‘vertigo’ | ‘compulsion’ | ‘aggression’ | ...

Amnesia and hypermnesia are generated by  $F_P^N$  := ‘mnesia’ and  $O_P^F$  :=  $Dir_O^F F_P^N$ . Alzheimer’s disorder can be given the diagnosis ‘frontal brain hypomnesia’ with  $N_P$  := ‘frontal brain’,  $Dir_O^F$  := ‘hypo’,  $F_P^N$  := ‘mnesia’ and  $O_P^F$  :=  $N_P Dir_O^F F_P^N$ . The systematic equivalent diagnosis to Korsakoff’s psychosis is ‘bilateral hippocampal afuction’ generated by Site := ‘bilateral’,  $N_P$  := ‘hippocampal’,  $Dir_O^F$  := ‘a’,  $F_P$  := ‘function’. The term *praxia* may seem a little odd but is needed to generate *apraxia*.

Sensory receptors transmit to afferent nerves and the mind perceives and records the stimuli. Vertigo, for example, is defined as consciousness of disordered orientation of the body in space. Several diagnoses, e.g., amaurosis, anosmia, ageusia, parageusia, deafness, and hyperacusis are relevant to afferent neural pathways and their place in 3) is immediate.

**Connective tissue**

Connective tissue is part of the general organ. The morphology of fibroblasts, fibrocytes, and mast cells is treated like other cells. *Lung fibrosis* is equivalent to *lung fibrocyte hyperplasia and collagen hypersecretion*. The systematic equivalent to *liver cirrhosis and liver fibrosis* is *liver fibrocyte hyperplasia and collagen hypersecretion*. In a disease classification *fibrosis* is simpler than *fibrocyte hyperplasia and collagen hypersecretion* and the former can replace the latter in a classification. However, the complex term is easier to acknowledge in automated CDM.

**Tube systems**

**Morphology** Terms describing morphological disorders of tube systems are:

$M_T$  := ‘stenosis’ | ‘occlusion’ | ‘dilatation’ | ‘diverticulum’ | ‘aneurysm’ | ‘ectasia’ | ‘perforation’ | ‘erosion’ | ‘ulcer’ | ‘fistula’ | ‘rotation’ ....

Formula 3) with  $N_T$  and  $M_T$  give rise to well-known diagnoses such as coronary stenosis, carotid occlusion, aorta aneurysm, bronchiectasia, and gastric erosion and -perforation. With  $N_P$  := ‘hepatic’,  $N_T$  := ‘canaliculi’,  $Dir_T^M$  := ‘a’,  $M_D$  := ‘genesis’ we get the rare syndrome  $O_T^M$  = ‘hepatic canaliculi agenesis’.

**Function** Function disorders of tube systems are assigned to the variable

$F_T :=$  'tension' | 'resistance' | 'flow rate' | 'containment' | 'leakage' | 'hemorrhage' | 'epistaxis' | 'petechia' ...

The direction of change  $Dir_T^F :=$  'hypo' or 'hyper' is naturally concatenated with  $F_T :=$  'tension', which through  $O_T^F := Dir_T^F F_T$  gives 'hypotension' or 'hypertension'. This assignment covers a vast number of functions in various tube systems. If  $N_T :=$  'arterial',  $Dir_T^F :=$  'hypo' and  $F_T :=$  'tension' formula 3) immediately gives 'arterial hypotension'. Naming conventions are independent of the particular organ: hypertension is hypertension whether the affected tube is in the cardiovascular system, genitourinary tract, the bile tract or the gastrointestinal tract. These similarities across organs profoundly simplify the generation of diagnoses.

**Slits and cavities**

*Morphology* Morphological terms that refer to concepts in a model for morphological disorders of cavities and slits are.

$M_S =$  'hydro' | 'pneumo' | 'hemato' | 'pyo' ...

$M_S$  generates an exception in formula 3) because  $O_S^M := N_S Dir_S^F M_S$  and  $Dir_S^F :=$  " gives rise to thoraxhydro, for example, which is a misnomer. Therefore, if  $X = S$  then  $O_X^M := N_X Dir_S^F M_X$  is automatically converted to  $O_S^M := M_S N_S$ . The modified formula generates diagnoses such as *hydrothorax*, *pneumothorax*, *hematopericard* and *pyonephrosis*.

*Function* Some of the terms are:

$F_S :=$  'compliance' | 'pressure' | 'restrictive' | ...

The peritoneal and pleural slits allow smooth movement of its organs, bowels, abdominal wall, and diaphragm against each other. The terms allow diagnoses such as restrictive pericarditis.

**The etiology variable e**

The classes beneath the root node of the etiology hierarchy are shown in Table 1. They provide the roughest etiology terms that can be assigned to e. The leaf nodes are individual etiological agents like particular mutations and infectious agents. Terms located between the first-level node and the leaf nodes have an intermediary specificity such as *gram-negative bacteria*. The etiology influences the choice of suffix  $S_E$  that modifies the names  $N_X$  of body parts, tissues, and cells in 3).

Formula 2) allows the diagnoses *viral myocarditis* and *Coxsackie virus B1 myocarditis* depending on the diagnostic accuracy. The suffix 'itis' signifies the presence of a microorganism or an allergen in the primary affected

**Table 1** Terms and the variable e used in the etiology segment of diagnoses.  $S_E$  is the suffix variable

e	$S_E$
'Prothrombin G20210A mutation' ...	'pathy'; 'osis'
'Trauma'   'Car accident'   'fall'   'knife stab'   'bite' ...	'vulnus'   'fracture' ...
Substance name	'intake'
'Infectious'   micro-organism name	'itis'
'Allergic'   allergen name	'itis'
'Thermal'   'frost'   'burn' ...	'itis'
'Radiation'	'osis'
'Social'   social problem name	'pathy'
'Essential'   'idiopathic'	"   'osis'

body part. For example, with formula 2) and  $e :=$  '*Staphylococcus aureus*',  $N_p :=$  'myocard' and  $S_E :=$  'itis' we get  $d :=$  '*Staphylococcus aureus* myocarditis', which means presence of the bacteria in the myocardium, in accordance with the clinical model. The principle obviously extends to hereditary-, allergic- and radiation etiology.

Nutrients and vitamins slightly complicate matters. The intake of a substance may be too high (elevated) or too low (diminished) such that

$Dir_E =$  'Elevated' | 'Diminished'

In the context of nutrition  $e := Dir_E^F \& N_B \& 'intake'$ . Thus, if  $N_B :=$  'glucose' and  $Dir_E :=$  'Elevated' then  $e =$  'Elevated glucose intake'. The formula easily discriminates the unspecific *Elevated sugar intake* from the specific *Elevated glucose intake* and applies equally well to vitamins and other substances. The affected organ, tissue and cells are modified to  $N_p \& S_E$ ,  $N_t \& S_E$  and  $N_c \& S_E$ , respectively, where  $S_E :=$  'osis' (Table 1).

Complex sociopsychosomatic disorders can also be expressed by formula 2). Let  $e :=$  'Social',  $N_p :=$  '*neur*',  $p :=$  " ,  $S_E :=$  'osis'. Then 3) generates the unspecific diagnosis *Social neurosis*. These etiologies accurately describe clinical facts and are amenable to automated generation of the etiological part of diagnoses.

**The pathogenesis**

The pathogenesis links a primary affected body part (source) with some secondary affected body part (targets). The link and the target together are abbreviated by p. The source of the link derives from parenchyma o. The pathogenesis may be one-way from source to target. Typical examples are emboli and metastases. Alternatively, there is a two-way link from source to target and back to the source. The source and target may be the same or differ. If they are the same then the pathogenesis only describes the link. Otherwise, the pathogenesis

**Table 2** Terms used to describe pathogenetic mechanisms. The variables are p and the suffix  $S_p$ . The sign ‘|’ denotes alternatives

Pathogenesis	p	$S_p$
<b>Connective tissue</b>		
Intercellular space	‘edema’ ‘dehydration’   ... ‘sclerosis’ ...	“  ‘osis’
<b>Hemostasis</b>		
	‘hemat’ ‘thrombosis’  ‘bleeding’ ‘petechia’ ‘hematoma’ ‘hemorrhage’...	‘oma’
<b>Circulatory system</b>		
	‘inflammation’  ‘ischemia’ ‘infarction’	“ ‘ial’ ‘al’
<b>Immune reactions</b>		
Type 1	‘IR1’	‘itis’
Type 2	‘IR2’	‘itis’
Type 3	‘IR3’	‘itis’
Type 4	‘IR4’	‘itis’
Type 5	‘IR5’ ‘granuloma’	‘itis’
<b>Intercellular metabolism</b>		
Metabolite name		‘osis’
Precipitates, stones		‘lithiasis’
<b>Humoral regulators</b>		
Regulator name		“ ‘osis’
<b>Nervous system</b>		
Afferent	‘psychosomatic’	‘osis’
Efferent	‘somatopsychic’	
Transmitter	‘hyper $N_H$ secretion’  ‘hypo $N_H$ secretion’  ‘hyper $N_H$ reabsorption’  ‘hypo $N_H$ reabsorption’ ...	

contains the term of the target too. These properties are determined by the actual context. Terms used with p are summarized in Table 2.

As shown in 3) and Table 2 the suffix  $S_p$  modifies names of organs, tissues, and cells in the same way as  $S_E$  (Table 1). In most cases  $S_E = S_p$ . However, it might occur that  $S_E \neq S_p$ . Then, as seen by comparing Table 1 with Table 2, the etiology conflicts with the pathogenesis. The conflict has to be resolved by returning back to CDM and possibly dividing the clinical problem at hand on two or more different diagnoses.

For the localization of p we only need the suffix variable  $L_H$ . Fortunately, several suffixes and terms are already in use.

$L_H :=$  ‘aemia’ | ‘uria’ | ‘oma’ | ‘myelia’ | ‘suggulation’ | ‘emesis’ ....

Note that the variable L in 3) is within the scope of disorder, while  $L_H$  is within the scope of the pathogenesis. The scope precludes the confusion of L with  $L_H$ .

**Extracellular space (ECS)**

In *lung fibrosis* the fibrosis is considered to be the source  $o := N_p N_I S_p$  where  $N_p :=$  ‘lung’,  $N_I :=$  ‘fibr’ and  $S_p :=$  ‘osis’. We know that lung fibrosis causes lung hypofunction. Thus, we have the same source and target organ and  $N_p =$  ‘lung’ in the generation of o and p. The partial diagnosis p := ‘causing’ & ‘lung hypofunction’ is generated by p := ‘causing’ &  $N_p \text{Dir}_O F_p$  where  $\text{Dir}_O :=$  ‘hypo’ and  $F_p :=$  ‘function’. These assignments complete the partial diagnosis  $op =$  ‘lung fibrosis causing lung hypofunction’ as required.

Sclerosis means that calcium is deposited on extracellular fibers or precipitated with extracellular cholesterol and other negatively charged lipids. The composite term *sclerosis* could have been generated by  $N_C :=$  ‘scler’ and  $S_p :=$  ‘osis’, but sclerosis is a composite material and not a cell. Therefore, *scler* cannot be assigned to  $N_C$ . The assignments p := ‘scler’ and  $S_p :=$  ‘osis’ would require the concatenation of a pathogenetic mechanism with a suffix ( $pS_p$ ) that violates the formula 3). Therefore, we choose p := ‘sclerosis’ (Table 2) and generate *arteriosclerosis* by  $o := N_T :=$  ‘arterio’ & op.

Precipitates are marked by the suffix lithiasis. As with sclerosis, their localization is given by op. *Urolithiasis* means a stone somewhere in the urinary tract. *Cholecystolithiasis*, *nephrolithiasis* and *phlebolithiasis* constrain the location to the gall bladder, the kidneys and veins, respectively, and are generated as shown above.

Edema and dehydration are common disorders of ECS. For *edema* we choose the appropriate location in 3) p := ‘edema’. With  $\text{Reg} :=$  ‘conjunctival’ we derive *conjunctival edema*. The cause of the edema, e.g., renal failure, is given by o.

**The vascular and lymphatic circulation**

Terms that can be assigned to the variables p and  $S_p$  are found in Table 2. For example, the unspecific diagnoses *myocardial infarction* and *cerebral ischemia* are generated from op where  $o = N_p$ . The terms are selected during CDM. Also, emboli occlude arteries, so we may have  $M_T :=$  ‘occlusion’, but this is excessive.

CDM may lead to the unspecific cerebral ischemia (transient ischemic attack (TIA)) or infarction, or the more specific occlusion of a particular cerebral artery such as *left medial cerebral artery embolus*. I prefer the latter, which is obtained by  $N_T :=$  ‘left medial cerebral artery’. In the present example,  $O_T^M := \text{Site}N_T M_T$ ,

i.e.,  $O_T^M$  = 'left medial cerebral artery occlusion.' Formula 5) requests both the morphology and the affected function(s), which in the present example are linguistic and motor functions. Here,  $O^F$  := 'aphasia and right hemiparesis' so formula 4) gives *left medial cerebral artery occlusion causing aphasia and right hemiparesis*.

**Metastasis** Metastasis is readily described by for example  $o$  := 'colon carcinoma' and  $p$  := 'with ' & RegSite $N_p O_p^M O_T^M O_S^M O_I^M$  ' metastasis'. When one site is chosen the others are set empty. For example, if  $N_p$  := 'pulmonary' then Reg = Site =  $O_T^M = O_S^M = O_I^M =$  ". This instantiates  $op$  as *colon carcinoma with pulmonary metastasis*.

**Embolism** Emboli originate as thrombi at the source site and become stuck at the distant target site. Here we have  $o$  :=  $N_T$  & ' thrombosis' and  $p$  := ' with embolus to '  $N_p$  ' and '  $N_p$  & ' infarction,' which among others translates into *deep vein thrombosis with embolus to lung and lung infarction*.

The variable  $p$  depends on the link, i.e., the vascular and lymphatic path that determine the target. If the source in  $o$  is  $N_S$  := 'left heart auricle' the path and the target is given by  $p$  = ' embolus to ' &  $N_T$  and the affected  $N_p$  is uniquely determined by  $N_T$ . Alternatively (and rarely), if the source is not the heart then it is a vein, or the venous plexus of the pelvis and the path is through the foramen ovale. In that case,  $p$  = ' embolus from ' &  $N_T$  ' through foramen ovale'. The term *heart* discriminates between the two alternatives.

**Bacteremia, viremia, fungemia and septicemia** The source is always the infected body part described by the variables  $e$  and  $o$  and  $L_H$  := 'emia'. In bacteremia, viremia and fungemia the microorganisms spread with the blood stream but do not harm secondary affected organs. Thus,  $p$  :=  $eL_H$ , for example *streptococcemia* and *nocardemia*.

Septicemia, sepsis syndrome, septic shock and systemic inflammatory response syndrome (SIRS) are unclear clinical notions. Precision is vastly increased by accurate description in the variable  $o$ , which gives the primary affected body site and its altered function(s). Then, the secondary affected body sites and their hypofunctions are clearly defined in the same way as *lung fibrosis* above. We are then able to describe the assumed cause of tachycardia and hypotension as required for septic shock. Note that the duration and prognosis are derived parameters, and not assumed in the sepsis diagnoses. The pathogenetic formulas eliminate ambiguities inherent in existing diagnoses describing the spread of microorganisms.

**Intercellular metabolism**

Some conventional diagnoses are constructed from the names of organs, cells, etiology, and pathogenesis, but most such diagnoses are incomplete. The diagnosis *hyperglycemia*, for example, leaves open alternative pathogenetic mechanisms like *hypoinsulinemia*, *hypercortisolemia* and *hypersomatotropinemia*, and a series of other pathogenetic mechanisms. Again, hypoinsulinemia can be caused by disorders like insulin hyposecretion and insulin resistance. In addition, the diagnosis *hyperglycemia* ignores etiological causes such as autoimmune reactions against  $\beta$ -cells, alimentary hyperglycemia, chemical etiology (alcohol-induced pancreatitis) and hereditary glucose channel mutations. Accordingly, even seemingly structured conventional diagnoses often lack the explanatory power that is necessary for adequate therapy and follow up.

Let the variable  $N_B$  hold the names of substances and metabolites. The concentration of a substance  $N_B$  is measured at localization  $L_H$  and described by  $p$  :=  $N_B L_H$ , which generates for example *glucosuria*. *Hyperglycaemia* is derived from  $p$  :=  $Dir_p N_B L_H$  and the appropriate instantiation of the variables. The formula readily generalizes to other metabolites (fatty acids, enzymes, electrolytes, etc.).

There is a close relation between metabolic disorders  $O_P^A$  and  $O_C^A$ , and metabolic pathogenesis  $p$ . The name of the secreted and absorbed substances  $N_B$  is the same as that measured in peripheral blood. We saw above that  $O_C^F$  :=  $N_C N_H Dir_H^F F_C^H$  can describe type I diabetes and made the assignment  $o$  :=  $O_C^F$ . From above we have  $p$  :=  $Dir_p N_B L_H$  that is modified to  $p$  := ' causing ' &  $Dir_p N_B L_H$  in the present context. In general,  $op$  describes the causal relations between cell disorders and observed blood concentrations, for example *hepatocyte albumin hypoproduction with albumin hyposecretion causing hypoalbuminemia* and *hepatocyte bilirubin hypoconjugation causing hyperbilirubinemia*.

Formula 3) relates the etiology to organ failure and the pathogenesis and incorporates etiological and pathogenetic explanations. It also supplies a bridge between diseased systems and the names that describe the disease. Accordingly, diagnoses constructed by formula 3) and its extensions are likely to be complete.

**Immune reactions (IR)**

Immune reactions are denoted by concatenating organ names with the suffix  $S_p$  = 'itis'. Thus, thyroiditis is an IR of the thyroid. Its cause depends on the etiology (microbial, allergic, or hereditary (autoimmune)) obtained from the variable  $e$ . A variety of diagnoses are derived from formula 3) simply by  $d$  :=  $e \& o \& p$  where  $S_E = S_p$ ,  $e$

is an appropriate etiologic agent,  $o := N_X S_P$  and  $p$  drives from Table 2. The variable  $p$  makes the IR type explicit such as with  $p = \text{'IR4'}$ . Formula 3) can be extended with morphology and functions as described above.

$S_E = S_P$  with  $e = \text{'}$  or  $p = \text{'}$  or both leaves ambiguous diagnoses. For example, the diagnosis *hepatitis* means different things in different contexts that have different therapeutic and prognostic consequences. For example, *Viral hepatitis B* and *Viral hepatitis C* requires different treatment and follow up than *autoimmune hepatitis*. Likewise, *conjunctivitis IR1* is very different from *conjunctivitis IR3*. The former probably has an allergic etiology whereas pyogenic bacteria usually cause the latter. Therefore, clinicians should always aim for complete diagnoses with  $e \neq \text{'}$ ,  $o \neq \text{'}$  and  $p \neq \text{'}$ . In fact, the diagnosis is deficient whenever the suffix 'itis' is not followed by the type of immunological reaction: either colitis IR3 (ulcerating colitis) or colitis type IR4 and IR5 (Krohn's disease). Soon, it might be possible to pass names  $N_H$  of disorders of interleukins, cytokines, etc. into IR diagnosis, which is easily realized with formula 3).

### Hemostasis

Bleeding from the stomach and kidneys manifest themselves as *hematemesis* and *hematuria*, respectively. In this context we choose  $p := \text{'hemat'}$  and generate the *hematemesis* and *hematuria* recursively by  $p := pL_H$ . *Hematemesis* is generated by  $L_H := \text{'emesis'}$  and *hematuria* by  $L_H := \text{'uria'}$ . Reg and Site are used to localize suggestions and hematomas.

In thrombosis the source and target are in the same organ. Two common examples are  $o := N_T$  'stenosis'  $p = \text{'causing 'N}_p$  ' ischemia' and  $o := N_T$  'thrombosis'  $p = \text{'causing 'N}_p$  ' infarction' that are instantiated to *coronary stenosis causing myocardial ischemia* and *coronary thrombosis causing myocardial infarction*.

### Neural pathways

The nervous system links together social events, the brain, and the mind to other organ systems. These links allow sociopsychosomatic and somatopsychosocial diseases and disorders, respectively.

**Psychosomatic** Complex psychosomatic and somatopsychic diseases are expressed by formula 3) with  $e := \text{'Social'}$ ,  $N_p := \text{'neur'}$ ,  $p := \text{'}$ ,  $S_p := \text{'osis'}$ , which by  $d := eoS_p p$  becomes *social neurosis*. The etiology  $e = \text{'social'}$  implies that the cause of neurosis is due to some social problem, and not the consequence of a mental disease with another etiology such as heredity or drugs. The assignment formalizes the notion of sociopsychosomatic

disease and fits the biopsychosocial clinical model (CM) [35].

The general formula for psychosomatic disorders is  $op$  where  $o := O_p^F$  is the function of some part of the brain or mind and  $p := \text{Eff}O_X^F$ . The variable Eff types the disorder as mediated by a named efferent part of the nervous system and  $O_X^F$  is the secondary affected organ. For example,  $\text{Eff} := \text{'N. X'}$  describes the involvement of the vagus nerve and  $O_p^F := \text{'bradycardia'}$ . The diagnosis could be simplified to  $p = \text{'}$  causing ' $O_X^F$  (see above) and adding a path suggests a therapeutic option.

Naming the social problem specifies the etiology (Table 1). *Hyperventilation syndrome* is a paradigmatic example of a psychosomatic disorder with unknown etiology but does not tell whether it is a primary or secondary disorder. Table 1 identifies possible etiologies. Also, the pseudoalgorithm generates possible differential diagnoses  $\text{Seg}O_p^M = \text{'pontine tumor'}$  and  $\text{Segeo}S_E = \text{'pontine viral encephalitis'}$ .

Anxiety and panic disorders are other debilitating disorders with varied etiology and pathogenesis. Several brain regions are involved. For example, amygdala may be the primary affected body part and  $p$  may involve tachycardia, hyperventilation, etc. as described above.

**Somatopsychic** The constant assigned to the variable  $o$  designates the disorder in the primary affected body part. In somatopsychic disorder information is passed along some afferent nerve (Aff) and is made conscious in short-term memory  $O_p^F$ . Accordingly, we have  $p := \text{'causing 'Aff}O_p^F$ . Then  $op$  may describe situations such as *gastric adenocarcinoma causing epigastric pain* but recall that epigastric pain is a symptom and not part of UDS.

### Etiology and anatomical localization

Epstein-Barr virus infects B-lymphocytes and spreads rapidly to many organs and tissues [40]. Therefore, infectious mononucleosis involves many anatomical sites. The systematic diagnosis is  $d = \text{'Epstein Barr virus B-cell hyperplasia'}$ . The same principle applies to other widely distributed cell types.

### Heterogenous etiology

Immune deficiencies are the ultimate cause of many infections. A prominent example is AIDS. Other relapsing infections are due to granulocyte disorders. A relatively rare disorder is mutation A392G neutrophil hypofunction (Diagnosis 1) involved in *Staphylococcus*

*aureus meningitis IR3* (diagnosis 2) and *Staphylococcus aureus carditis IR3* (diagnosis 3). The general syntax for these diagnoses is: Diagnosis 1 causes diagnosis 2 and diagnosis 3. Each of the diagnoses is generated by formula 3).

**Examples of use of UDS**

Three clinical examples of combinations of e, o, and p are compared with diagnosis used in medical practice (Table 3). Gluten allergy mainly affects the small intestine and that is focused here. X-ray exposure induces bone marrow hypoplasia. X-rays also affect rapidly proliferating cells including intestinal epithelium and other tissues (not shown). The etiology of distorted body image will in many cases remain unknown and the etiology of these cases remains idiopathic. As described above one diagnoses (diagnosis 1) may cause another diagnosis (diagnosis 2) which is illustrated by the complex disorder anorexia nervosa. Malnutrition reduces intestinal uptake of numerous nutritional elements as reflected in the pathogenesis.

The conversion of three typical disease names to UDS is shown in Table 4. The etiology of Wilson’s disease is *ATP7B gene* mutation that primarily reduces

the secretion of copper with concomitant hepatocyte damage and hypofunction. This disorder causes copper precipitates in the iris edge and lenticular nucleus degeneration and hypofunction. Through neural pathways the latter affects the function of many body parts.

The hereditary etiology of Huntington’s disease is well known. As shown in Table 4 the disorder manifests itself morphologically, and by motor, cognitive, emotional metabolic disturbances. Depletion of BDNF [41] and altered behaviour are examples of regulator and psychosomatic pathogenesis, respectively.

The etiology of multiple sclerosis often remains idiopathic, but in recent years Epstein-Barr virus is assumed to cause damage oligodendrocytes causing demyelination. In clinical situations the discrimination between idiopathic and Epstein-Barr virus etiology has to be made on an individual patient basis. The disorder is worsened by IR4.

**Extensions to UDS**

Formula 3) is sufficient for many purposes but complete diagnoses describe the time course (Dur), prognosis (Prog) and degree (Deg). This naturally accomplished by extending formula 3) to

**Table 3** Relationships between e, o, and p and commonly used clinical diagnoses

e	o	p	common clinical diagnosis
Gluten allergy	Small intestinal mucosal atrophy with malabsorption	IR2a	Gluten induced celiac disease
1.idiopathic 2.malnutrition	1.Distorted body image 2.Adipose tissue and striated muscle atrophy	1.psychosomatic 2.hypoglycemia, hypokalemia, ...	anorexia, bulimia nervosa
x-ray exposure	bone marrow hypoplasia	fibrosis	Radiation induced bone marrow failure

<sup>a</sup> Immune reaction type 2 characterized by gluten antibodies

**Table 4** Conversion of typical disease names to UDS

Typical disease name	e	o	p
Wilson’s disease	<i>ATP7B</i> gene mutation	1.hepatocyte lysis, hypoplasia and hyposecretion 2.iris edge 3.lenticular nucleus degeneration and hypofunction	1.IR4, cirrhosis <sup>a</sup> , metabolic pathogenesis (hypoceruloplasminemia, hypoalbuminemia, hyperbilirubinemia, ... 2.Copper precipitate 3.Neural pathways
Huntington’s disease	dominant huntingtin gene hyperconcatenation of cytosine-adenine-guanine (CAG) repeats	1.striatal degeneration and atrophy 2.hypocognition, 3.psychosis, compulsions, aggression,... 4.Huntingtin with > 36 glutamine residues	1.cortical brain-derived neurotrophic factor (BDNF) depletion 2a. dyskinesia 2b. late behavioral hypoactivity
multiple sclerosis	1.essential 2.Epstein-Barr virus	Central nervous demyelination and nerve cell hypofunction	IR4 <sup>a</sup>

<sup>a</sup> Immune reaction type 4 characterized by lymphocyte infiltration

$$d := \text{DurSegRegSiteProgeoSpDeg} \quad (6)$$

### Time course

Time course is specified by the following alternatives:

Dur := 'peracute ' | 'acute ' | 'subacute ' | 'chronic ' | 'chronic remittent ' | 'chronic progressive ' | 'paroxystic ' | ...that explicitly state the rapidity of onset and duration of diseases and disorders. The naming convention places time course to the far left in diagnoses as seen with example *acute appendicitis* and *chronic hepatitis* as in (6). Terms such as chronic, chronic remittent and chronic progressive are used with liver diseases.

### Prognosis

The prognostic terms *benign* and *malignant* can stem from pathology or survival. The former is closely related to morphological changes such as a malignant neuroblastoma, whereas the latter is determined by clinical observations. UDS avoids the pathological-anatomical use of malignant. The term neoplasia and anaplasia are preferred. Accordingly, the terms benign and malignant are reserved for survival. This use eliminates an ambiguity that plagues many textbooks and scientific articles. Prognosis is described by the variable

Prog := 'benign ' | 'semi-malignant ' | 'malignant ' ....

With Prog := 'malignant ', e = 'hereditary ', o := 'thyroid neoplasm', Suff := " and p := " the diagnosis reads d = 'malignant hereditary thyroid neoplasm'. Likewise, the formula can generate d = 'malignant hyperthermia'.

### Degree—severity

Some classifications allow for the concept of degree [42]. Degree is common to many clinical scales such as grading the level of functioning [43], the severity of heart failure [44] and liver function [45–47]. Highly differentiated, little differentiated and undifferentiated tumor that justifies the diagnosis are placed in the pathologists' narratives. At the level of diagnoses verbal expressions of the degree of differentiation are better replaced with a systematic degree in the range 0–5 as practiced in cytology (see below).

Terms such as *preleukemia*, *subclinical celiac disease*, *subclinical hypothyroidism*, and *preclinical*-, *subclinical*- and *latent diabetes mellitus* are frequently encountered in medical textbooks. The concepts *preclinical*, *subclinical*, and *latent* imply a disorder without clinical manifestations. They describe similar states, and all can be substituted by degree 0, which announces abnormal supplementary observations but no symptoms and signs. Manifest disease implies the presence of symptoms or

signs. Manifest disease is often graded in the four degrees {1, 2, 3, 4}. Deg is the variable denoting the degree:

Deg := 'degree 0' | 'degree 1' | 'degree 2' | 'degree 3' | 'degree 4'

Typical examples that illustrate the variety of generated diagnoses are *cardiac insufficiency degree 3*, *hypertension degree 4*, *cervical squamous cell dysplasia degree 4* and *astrocytoma degree 3*.

One might object that the concepts *preclinical*, *subclinical*, and *latent* have a subtler meaning. But this information is already covered by the terms assigned to the time course, etiology, disorder, and pathogenesis.

Time course, prognosis and degree are interrelated. A fifth number in ICD-10 can be used to describe the progression of the disease from continual to episodic with a progressive deficit, episodic with a stable deficit, episodic remittent, incomplete remission, other and uncertain prognosis. Not surprisingly, the purpose of such a mixture of degree, change in degree and time course has been questioned [48]. I have not encountered a medical classification that makes the connections explicit as is done in formula (6).

### Discussion

d:=e&o&p is the first formula that generates systematic diagnoses. It rests on standard international clinical concepts and language. The formula and its extensions generates diagnoses for all medical specialties, whereby it counts as a universal medical diagnosis syntax (UDS). UDS may have implications for medical education and classifications and may create a foundation for structured clinical decision-making. Formulae are the hallmark of the hard sciences. Therefore, d:=e&o&p moves clinical medicine into the domain of hard science.

### UDS is open ended

The boost of medical knowledge challenges medical classifications and poses serious problems for medical education and CDM. The information explosion often leads to the re-evaluation of the etiology, disorders of structure and function of body parts, and the pathogenesis of known diseases. UDS opens for reasoning about the structure and content of diagnoses and predicts new diseases and syndromes. The ICD10, ICPC-2, and DSM V contain historical and some outdated diagnoses. UDS does not offer preformed diagnoses; diagnoses depends on actual knowledge and actual patients.

Disease is said to mean so many different things that it cannot be applied to psychiatry [49]. However, formula (6) adapts well to psychiatric diagnoses. The complexity of (6) with extension explains why diagnosis seems to be such a heterogeneous concept.

The present work shows that complete systematic diagnoses can be constructed from the names of organs, cells,

etiology, organ disorder and pathogenesis according to formula 6). The diagnosis formulae covers social, psychological, and biological aspects of diagnoses. In spite of its wide applicability, it has few variables. The few variables make the formula easy to use, control, and implement. The extended formula may also facilitate communication and reinforce agreement among physicians.

#### Excluded from UDS

The development of generative grammar in the nineteen seventies and eighties distinguished logical form from grammatical form. The logical form provided the syntactic component without the aspect of meaning of sentences ([1], p.23). Montague disagreed and claimed that there is no important difference between natural languages and the artificial language of logicians ([1], p.23). In UDS formula 6 defines the syntax. The categories delimit the content of the variables. Nothing outside a valid form and appropriate terms are allowed. Syntactic errors and category mistakes are easy to spot. For example, the diagnosis *pathenceph* cannot be generated, and *Staphylococcus aureus* is not a tissue. In contrast, medical terminologies sometimes does not distinguish between assertions and pseudoassertions. UDS excludes misleading diagnoses, but to see this requires interpretation into the semantics of a clinical mode [35].

#### Clinical findings

Clinical findings consist of symptoms, signs, and supplementary investigations. They are used to ground diagnoses during medical decision-making. The ICPC-2 allows clinical findings to serve as diagnoses [11]. However, EPR can separate diagnoses and clinical findings. The latter can be extracted from clinical narratives and laboratory files and are used in automated medical decision-making [30, 33, 50, 51]. Clinical findings are suited for justification of diagnoses but not as diagnoses. None of the formulas in UDS have a symptom variable. 'algia' in fibromyalgia and myalgic encephalopathy (ME) refers to a symptom. Accordingly, fibromyalgia and ME cannot be derived in UDS.

#### Misnomers

A clear understanding of clinical terminology requires the elimination of misnomers. *Dysfunction* is not a misnomer. It is much used as a crude diagnosis such as *gastric dysfunction*, *dysphagia*, and *dyspepsia*. *Dysphagia* may refer to a collection of symptoms such as regurgitation and/or pain and/or .... However, such diagnoses disregard the specificity of clinical functions. The cure is to eliminate all terms where 'dys' is associated with clinical findings.

Obvious misnomers cannot be generated by 6). Itiscys is incomprehensible but cystitis is immediately understood.  $d:=eoSp$  allows 'cyst' & 'itis' while 'itis' & 'cyst' cannot be generated. Formula 6) constrains the possible term combinations to those derivable in school medicine.

In EPR the diagnoses are stored separately from the clinical findings. Therefore, formula 6) can be compared with the justification, which consists of the clinical criteria [30]. This procedure allows systematic check of the accuracy and precision of diagnoses. In addition, any arbitrary diagnosis will not pass the constraints imposed by  $d:=eoSp$ . For example, the diagnosis 'hemiplegia due to coronary artery occlusion' fails because the knowledgebase rests on a clinical model (see below), but 'hemiplegia due to right medial cerebral artery occlusion' will pass. The formula combined with the clinical model limits well-formed diagnoses to those that both fit clinical observations and empirical medical knowledge.

#### Analysis and parsing

In a previous article, we used diagnoses to set up a semantic analysis of clinical narratives [30]. We assumed the semantics to result from parsing. Analysis divides sentences into words and word groups according to their syntactic function. In UDS analysis divides diagnoses into terms and groups them according to their place in diagnoses. Parsing categorizes individual terms in diagnoses. The categories of formula 6) are uniquely determined by their place in the diagnoses. Thus, the present work shows that UDS may fuse analysis and parsing into one process.

#### Standards of acceptability

The present method requires independent assignment of strings to the variables e, o, and p. This might theoretically seem to be an impossible task in clinical practice. However, several studies suggest the contrary. First, the formula  $d:=eop$  works in hematology and infectious medicine [52]. Second, the clinical findings corresponding to the suffix 'itis', and terms pertaining to selected organs can be obtained by automatic parsing of clinical narratives from EPR [30]. Third, names of entities pertaining to e, o and p can be automatically parsed from diagnoses in free text form [30, 51]. Preliminary data also suggest that formula 3) is useful for the analysis of social problems [33]. Finally, a combinatorial classification based on  $d:=eop$  was successfully built into an EPR [31, 35]. Taken together, these results indicate that the independent assignment of strings to e, o and p is feasible, but further independent studies are required.



### Limitations and disadvantages of UDS

This study is limited to a syntax generating diagnoses. Comments to clinical and laboratory work are limited to examples (Tables 3 and 4 with comments). This study does not comment on the semantics of UDS, i.e., does not interpret variables, terms, and connectives into a model. Aspects of semantics and models have been described [35, 53]. Further work on models of health and disease is in progress. As a result, UDS may be interpreted into the model of disease. Diagnoses are goals in CDM that may be generated by UDS, but CDM also involves clinical findings and together they are too voluminous to be treated here.

UDS may be a cumbersome disease classification in EPR. Existing implementations of ICPC-2 and ICD10 are more effective and labor-saving. However, the number of entries in ICPC-2 and ICD10 are limited, many rare disorders are not covered, and it is laborious to maintain them. A combination of UDS with ICPC-2 and ICD10 may solve these problem, but that remains to be shown. Finally, UDS may be hidden in a decision support system that suggests diagnoses.

### Conclusions

This study generates systematic diagnoses using standard medical concepts, language, and the syntax used in the classifications ICPC-2 and ICD10. The formula  $d := eop$  with extensions give rise to universal medical diagnoses. The generation of diagnoses allows the diagnoses to be tailored to the medical problem at hand and relieves physicians from squeezing the patients' clinical findings into some classification. In addition, novel combinations of terms may predict novel diseases.

Ambiguous diagnosis, clinical findings and misnomers that may lead to confusion and disagreement are excluded from UDS. Leibniz's dream of a symbolism for all human thought implies that arguments could be resolved by calculation ([54] p.94). The results presented here suggest that  $d := eop$  may have such a clinical role.

### Acknowledgements

Thanks are due to Ragnhild Bassøe Gundersen for help with the manuscript.

### Authors' contributions

The author has done all the work presented in this article.

### Author's information

The author is retired specialist in hematology and general internal medicine. He is Ph.D. (medicine), Ph.D. (informatics) and Master of Philosophy. He has worked as consultant in hematology and internal medicine and as professor in hematology/internal medicine and health informatics.

### Funding

No funding except for the author's contribution.

### Availability of data and materials

NA.

### Declarations

#### Ethics approval and consent to participate

Not applicable

#### Consent for publication

NA

#### Competing interests

The authors declare no competing interests.

Received: 31 July 2022 Accepted: 15 June 2023

Published online: 31 July 2023

### References

- Gamut LTF. Logic, language and meaning, vol. 1. Introduction to logic. Chicago: University of Chicago Press; 1991.
- Jones FA, editor. Richard Asher talking sense. London: Pitman Medical; 1972.
- Curé OC, Maurer H, Shah NH, Le Pendu P. A formal concept analysis and semantic query expansion cooperation to refine health outcomes of interest. *BMC Med Inform Decis Mak*. 2015;15 Suppl 1:S8. [cited 2021 December 2]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4460622/>
- Frege KA. An introduction to the founder of modern analytic philosophy. Oxford: Blackwell; 2000.
- Robins RH. A short history of linguistics. London: Longman; 1997.
- Miller A. Philosophy of language. London: Routledge; 2004.
- Chomsky N. Language and the mind. San Diego: Harcourt Brace; 1972.
- Chomsky N. Knowledge and language. New York: Praeger; 1986.
- Fitch WT, Friederici AD. Artificial grammar learning meets formal language theory: an overview. *Philos Trans R Soc Lond B Biol Sci*. 2012;367(1598):1933–55. <https://doi.org/10.1098/rstb.2012.0103>. Available from: <http://rstb.royalsocietypublishing.org/content/367/1598/1933.long>
- ICD-10. 2015 [cited 2015 December 1]. Available from: <http://apps.who.int/classifications/icd10/browse/2016/en>
- ICPC-2. 2015 [cited 2015 December 1]. Available from: <http://www.who.int/classifications/icd/adaptations/icpc2/en/>
- SNOMED. 2018 [cited 2018 May 25]. Available from: <http://www.snomed.com>
- Cimino JJ, Zhu X. The practical impact of ontologies on biomedical informatics. *Methods Inf Med*. 2006;45(Suppl):1.
- Boorse CA rebuttal on health. In: Humber JM, Almeder RF (editors). What is disease? Totowa: Humana Press; 1997.
- Humphreys BL, Lindberg DA. The UMLS project: making the conceptual connection between users and the information they need. *Bull Med Libr Assoc*. 1993;32(4):281–91. Available from: <https://doi.org/10.1055/s-0038-1634945>.
- McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A, Srinivasan S. UMLS knowledge for biomedical language processing. *Bull Med Libr Assoc*. 1993; 81(2):184–94. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC225761/>
- Kashyap V. The UMLS Semantic Network and the Semantic Web. *AMIA Annu Symp Proc*. 2003;2003:351–5.
- Zhang L, Perl Y, Halper M, Geller J, Cimino JJ. An enriched unified medical language system semantic network with a multiple subsumption hierarchy. *J Am Med Inform Assoc*. 2004;11(3):195–206. [cited 2017 November 2]. Available from: <https://doi.org/10.1197/jamia.M1269>
- Carlsson M, Ahlfeldt H, Thurin A, Wigertz O. Terminology support for development of sharable knowledge modules. *Med Inform (Lond)*. 1996;21(3):207–14. Available from: <https://doi.org/10.3109/14639239609025358>.
- Rector AL, Rogers JE, Zanstra PE, Van Der Haring E. OpenGALEN: open source medical terminology and tools. *AMIA Annu Symp Proc*. 2003;2003:982.

21. Rogers JE, Price C, Rector AL, Solomon WD, Smejko N. Validating clinical terminology structures: integration and cross-validation of Read Thesaurus and GALEN. *Proc AMIA Symp.* 1998;845–9. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2232201/>
22. Spackman KA, Campbell KE. Compositional concept representation using SNOMED: towards further convergence of clinical terminologies. *Proc AMIA Symp.* 1998;740–4. [cited 2015 August 7]. Available from: <https://pubmed.ncbi.nlm.nih.gov/9929317/>
23. Boscá D, Maldonado JA, Moner D, Robles M. Automatic generation of computable implementation guides from clinical information models. *J Biomed Inform.* 2015;55:143–52. [cited 2017 November 13]. Available from: <http://www.sciencedirect.com/science/article/pii/S1532046415000696>
24. Ceusters W, Smith B. Biomarkers in the ontology for general medical science. *Stud Health Technol Inform.* 2015;210:155–9. [cited 2020 January 20]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/25991121>
25. Komenda M, Schwarz D, Švančara J, Vaitšis C, Zary N, Dušek L. Practical use of medical terminology in curriculum mapping. *Comput Biol Med.* 2015;63:74–82. [cited 2017 September 2]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/26037030>
26. Marc DT, Zhang R, Beattie J, Gatewood LC, Khairat SS. Indexing Publicly Available Health Data with Medical Subject Headings (MeSH): An Evaluation of Coverage. *Stud Health Technol Inform.* 2015;216:529–33. [cited 2019 June 22]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/26262107>
27. Seitinger A, Rappelsberger A, Leitich H, Binder M, Adlassnig KP. Executable medical guidelines with Arden Syntax-Applications in dermatology and obstetrics. *Artif Intell Med.* 2016;30:321–9. [cited 2018 June 2]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27686851>
28. Livingston KM, Bada M, Baumgartner WA Jr, Hunter LE. KaBOB: ontology-based semantic integration of biomedical databases. 2015;16:126. [cited 2020 December 12]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4448321/>
29. Popper KR. *The open society and its enemies.* London: Routledge; 2011.
30. Rasmussen J-E. Bassøe C-F Semantic analysis of medical records. *Meth Inform Meth.* 1993;32(1):66–72.
31. Bassøe C-F, Sørli WG. EPR records and forms in primary health care. *Tidsskr Nor Legeforen.* 1983;103:1270–4.
32. Bassøe C-F. A combinatorial diagnostic system for general practice. *Proceedings of the 11th Conference of the World Organisation of National Colleges, Academies and Academic Associations of General Practitioners/Family Physicians, London, 1986.*
33. Bassøe C-F. A combinatorial diagnostic system for general practice: Evaluation of the social impact of disease by a computerized medical record. In: Hansen R, Solheim BG, O'Moore RR, Roger FH (editors): *Lecture notes in medical informatics.* Springer-Verlag, Berlin, 1988.
34. Botsis T, Bassøe CF, Hartvigsen G. Sixteen years of ICPC use in Norwegian primary care: looking through the facts. *BMC Med Inform Decis Mak.* 2010;10:11. [cited 2012 July 7]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2848129/>
35. Bassøe C-F. *Combinatorial clinical decision-making.* PhD dissertation, Department of Information Science and Media, Faculty of Social Sciences, University of Bergen, Norway, 2007. ISBN 978–82–308–0457–5. [cited 2022 December 23]. Available from: <https://www.ekvimed.no/post/thesis-combinatorial-clinical-decision-making>
36. Cappello AR, Curcio R, Lappano R, Maggiolini M, Dolce V. The Physiopathological Role of the Exchangers Belonging to the SLC37 Family. *Front Chem.* 2018;6:122. [cited 2020 October 2]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5913288/>
37. Jadaon MM. *Epidemiology of Activated Protein C Resistance and Factor V Leiden Mutation in the Mediterranean Region.* *Mediterr J Hematol Infect Dis.* 2011;3:e2011037. [cited 2015 September 1]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3251907/>
38. Vakiti A, Mewawalla P. *Cancer, Leukemia, Myeloid, Acute (AML, Erythroid Leukemia, Myelodysplasia-Related Leukemia, BCR-ABL Chronic Leukemia).* *Allegheny Health Network Cancer Inst.* 2018. [cited 2018 June 29]. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK507875/#article-25443.s2>
39. Jethwa A, Mink J, Macarthur C, Knights S, Fehlings T, Fehlings D. Development of the Hypertonia Assessment Tool (HAT): a discriminative tool for hypertonia in children. *Dev Med Child Neurol.* 2010;52:e83–7. e83–7. [cited 2014 December 1]. Available from: <https://doi.org/10.1111/j.1469-8749.2009.03483.x>
40. Kurth J, Spieker T, Wustrow J, Strickler GJ, Hansmann LM, Rajewsky K, Küppers R. EBV-infected B cells in infectious mononucleosis: viral strategies for spreading in the B cell compartment and establishing latency. *Immunity.* 2000;13(4):485–95. [cited 2015 October 3]. Available from: [https://doi.org/10.1016/s1074-7613\(00\)00048-0](https://doi.org/10.1016/s1074-7613(00)00048-0)
41. Jurcau A. *Molecular Pathophysiological Mechanisms in Huntington's Disease.* *Biomedicines.* 2022;10(6):1432. [cited 2022 July 3]. Available from: <https://doi.org/10.3390/biomedicines10061432>
42. Edwards N, Honemann D, Burley D, Navarro M. Refinement of the Medicare diagnosis-related groups to incorporate a measure of severity. *Health Care Financ Rev.* 1994;16:45–64. [cited 2016 October 13]. Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4193498/>
43. Karnofsky Scale. 2018 [cited 2018 July 12]. Available from: <http://www.anapsid.org/cnd/diagnosis/karnofsky.html>
44. NYHA. *New York Heart Association (NYHA) Classification.* 2018 [cited 2018 July 12]. Available from: <https://manual.jointcommission.org/releases/TJC2016A/DataElem0439.html>
45. Wang YY, Zhong JH, Su ZY, Huang JF, Lu SD, Xiang BD, et al. Albumin-bilirubin versus Child-Pugh score as a predictor of outcome after liver resection for hepatocellular carcinoma. *Br J Surg.* 2016;103:725–34. [cited 2019 September 9]. Available from: <https://doi.org/10.1002/bjs.10095>
46. Hiraoka A, Kumada T, Kudo M, Hirooka M, Tsuji K, Itobayashi E, et al. Albumin-Bilirubin (ALBI) Grade as Part of the Evidence-Based Clinical Practice Guideline for HCC of the Japan Society of Hepatology: A Comparison with the Liver Damage and Child-Pugh Classifications. *Liver Cancer.* 2017;6:204–215. [cited 2019 June 3]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5473065/>
47. Johnson PJ, Berhane S, Kagebayashi C, Satomura S, Teng M, Reeves HL, et al. Assessment of liver function in patients with hepatocellular carcinoma: a new evidence-based approach—the ALBI grade. *J Clin Oncol.* 2015;33(6):550–8. [cited 2017 November 30]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4322258/>
48. Kringlen E. Diagnostikk som ideologi [Diagnosis as an ideology]. *Tidsskr Nor Legeforen.* 1995;115:630–2.
49. Scadding JG. The semantic problems of psychiatry. *Psychol Med.* 1990;20:243–8.
50. Bassøe C-F. The skinache syndrome. *J R Soc Med.* 1995;88(10):565–9 [cited 2012 January 13]. Available from: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1295357/>
51. Bassøe C-F. Automated diagnoses from clinical narratives: A medical system based on computerized medical records, natural language processing and neural network technology. *Neural Networks* 1995a;8:313–319. [cited 2014 September 29]. Available from: <http://www.sciencedirect.com/science/article/pii/089360809400076X>
52. Bassøe C-F. Neutrophil functions studied by flow cytometry. In: Yen A, editor. *Flow cytometry: Advanced Research and Clinical Applications.* Boca Raton: CRC Press; 1989. p. 95–148.
53. Bassøe CF. Representing health, disorder and their transitions by digraphs. *Stud Health Technol Inform.* 2008;136:133–8.
54. Derbyshire J. *Unknown quantity.* Washington: Joseph Henry Press; 2006.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.