Theory of signs and statistical approach to big data in assessing the relevance of clinical biomarkers of inflammation and oxidative stress

Pietro Ghezzi[a,1], Kevin Davies[a], Aidan Delaney[b], and Luciano Floridi[c,d]

[a]Brighton & Sussex Medical School, Brighton BN1 9RY, United Kingdom; [b]School of Computing, Engineering and Mathematics, University of Brighton, Brighton BN2 4GJ, United Kingdom; [c]Oxford Internet Institute, University of Oxford, Oxford OX1 3JS, United Kingdom; and [d]Alan Turing Institute, London NW1 2DB, United Kingdom

Biomarkers are widely used not only as prognostic or diagnostic indicators, or as surrogate markers of disease in clinical trials, but also to formulate theories of pathogenesis. We identify two problems in the use of biomarkers in mechanistic studies. The first problem arises in the case of multifactorial diseases, where different combinations of multiple causes result in patient heterogeneity. The second problem arises when a pathogenic mediator is difficult to measure. This is the case of the oxidative stress (OS) theory of disease, where the causal components are reactive oxygen species (ROS) that have very short half-lives. In this case, it is usual to measure the traces left by the reaction of ROS with biological molecules, rather than the ROS themselves. Borrowing from the philosophical theories of signs, we look at the different facets of biomarkers and discuss their different value and meaning in multifactorial diseases and system medicine to inform their use in patient stratification in personalized medicine.

biomarkers | oxidative stress | inflammation | epistemology | disease

Molecular biomarkers can be defined as substances whose levels correlate with some pathological process. They are used as diagnostic or prognostic indicators and to monitor the progression of disease in clinical trials (surrogate biomarkers) but also to formulate theories of disease. We recently discussed the problems related to the use of molecular biomarkers of oxidative stress (OS) (1, 2) in the context of the clinical translation of the OS theory of disease (3). The present paper attempts to develop a theoretical framework for defining the different types of biomarkers using those of OS or inflammation as examples.

Inflammation and the Cytokine Theory of Disease

Inflammation is largely mediated by activation of a gene expression profile, largely mapping to the transcription factor NF-κB, including inflammatory cytokines, adhesion molecules, and enzymes involved in the synthesis of prostaglandins and nitric oxide. The most recent breakthrough in this field was the identification of inflammatory cytokines that led to the approval of specific inhibitors of IL-1, IL-6, IL-17, and TNF for the therapy of several inflammatory diseases, particularly rheumatoid arthritis (RA).

Following identification of TNF as an inflammatory mediator (4), the availability of specific assays for its measurement in biological fluids of patients with RA were instrumental in the translation of that hypothesis into the clinical arena (5) and the formulation of a "cytokine theory of disease" (6). Just 14 y after the original finding, anti-TNF antibodies were a top selling biological drug.

The OS Theory of Disease and the Use of Indirect Biomarkers

OS is due to elevated levels of reactive oxygen species (ROS) (7) Since the 1956 paper postulating a "free radical theory of aging" (8), OS has been implicated as a causal factor in many diseases (3), but unlike the cytokine theory of disease, this has not led to any significant therapeutic breakthroughs.

A major problem with the OS theory is that it is practically impossible to measure ROS in biological fluids because of their very short half-life, ranging from microseconds for superoxide and nanoseconds for hydroxyl radicals (9). This difficulty applies of course to the measurement of ROS in patients and their biological fluids or biopsies, while generation of ROS by cells cultured in vitro can be measured more accurately.

This is reminiscent of high-energy physics, where the short life of subatomic particles is such that these are detected only though the traces they leave during their decay. Likewise, OS biomarkers include oxidation products of biological molecules (2). Unlike cytokines, we cannot measure ROS, but we measure "signs" that stand for them instead, and in this paper, we try to look at biomarkers from the perspective of different philosophical theories of signs.

To do so, we will first recapitulate some key aspects of the main theories on signs that might be relevant to our problem and attempt to draw links and similarities with biomarkers of inflammation and OS and their use in formulating theories of disease.

Concepts Arising from Various Theories of Signs

In the theory of signs of Charles Pierce (https://plato.stanford.edu/entries/peirce/), a sign is "something which stands to somebody for something in some respect or capacity" (10). For Pierce, a sign is a component of a triad along with the interpretant and the object, or the "thing" the sign *stands for*. For instance, smoke is a sign for fire. If we (the interpretant) see smoke coming from behind a mountain, we assume that there is a wildfire somewhere (Fig. 1).

Significance

Biomarkers are widely used in medicine. This study focuses on the use of biomarkers in the formulation of mechanistic hypotheses and their clinical use. Starting from different philosophical theories of signs, the study highlights the importance of networks in their meaning and value. The study also views biomarkers as endowed or not with action (proxies or signs) and suggests a perspective in the translational use of biomarkers in patient stratification and mechanistic studies.

---

[1]To whom correspondence should be addressed. Email: pietro.ghezzi@gmail.com.
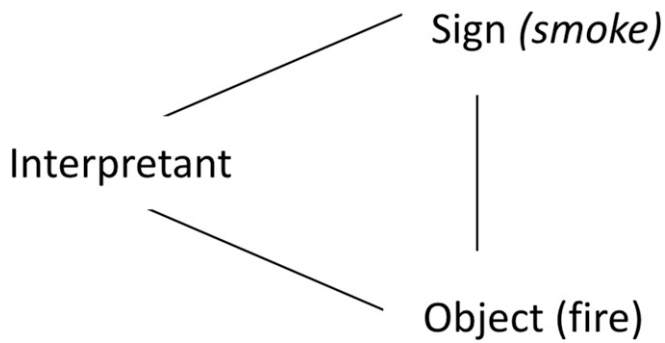
Fig. 1. The semiotic triad.

Pierce distinguishes different types of signs. While the smoke in Fig. 1 is called an *index*, a street sign depicting a fire is also a sign that resembles, but is not physically related to, the object, which is an *icon*. A sign can *stand for* an event or an object, and one can distinguish different types of signs based on the relationship he or she has with the object.

The linguist Ferdinand de Saussure saw the language as a system of signs, where signs are related to other signs within the system, which led him to develop the concept that signs have a *meaning* (signification) but also a *value*. For instance, the knight in the game of chess is a sign and has a meaning, the gentleman/soldier. However, it also has a value, but to appreciate that we need to see the knight in relation to the *network* of the other pieces in the game, its position on the chessboard, and the rules of the game. Although the difference between *value* and *meaning* in de Saussure is not completely clear, we suggest that, for the use we will make of this concept, we consider value as something related to quantity (justified in this by the fact that de Saussure also used the example of the face value of a coin). We will develop later this concept that signs have a value that depends on the interactions among them.

Floridi (11) recently analyzed the differences between signs and proxies. Floridi's concept of proxy is different from that of Wittgenstein, where "being proxy for" means "denoting," focusing instead on signs that exert an action and on the direction of this action. Let us take the example of when, in Italian and French law, it is the major who can perform a wedding, but a city councillor can *act on behalf of* the mayor. This has been defined as a vicarious relation, where P is acting for R (the mayor). However, the direction of the action is one way, from R to P; a wedding conducted by a councillor is perfectly valid, but if we shoot the councilor, the mayor will not die. If this was the case, then the councillor would be a proxy for the mayor.

This difference is outlined in Fig. 2.

What we want to retain here is that signs can be classified based on their functional relationship with the object and that a proxy relationship goes both ways and you can interact on the proxy to affect the object to which it refers.

In this analysis, Floridi distinguishes further between two types of signs with a vicarious relation: those that are also signs that refer to R, in the sense they have a semiotic relationship with R, and signs that can act on behalf of/replace R but do not refer to it. The latter are defined as *surrogates* (e.g., a PIN pad is a surrogate for a traditional key as it can perform the same function to open a door but has no resemblance to a physical key).

Another concept that we would like to develop is that of the degrees of separation between the sign and the object for which it stands, using again the example of smoke as a sign for fire. If there is a fire behind a mountain, the park ranger will see the smoke from an observation tower and call the fire department. Let us assume we are sitting in a café from a point from which we cannot see either the fire or the smoke. If we see several fire trucks with lights and sirens, we assume there is a fire, even if we cannot see smoke. Now, let us say we are at home and we do not live by the main road. We do not see the fire, we do not see the smoke, and we cannot see the fire trucks, but we hear several sirens going by; we can still assume that there is a fire somewhere.

We could say that, in this system, while smoke is directly related to fire (in the previous scheme), the passage of the fire trucks is a sign that is one degree of separation from the object of fire. Likewise, the sound of the sirens will be two degrees of separation from the fire.

$$\text{Fire} \to \text{smoke} - (\text{ranger on observation tower})$$
$$\to \text{fire trucks} \to \text{siren}.$$

Importantly, the more degrees of separation there are from the object, the more the sign is *ambiguous* (less specific). The sign that is closest to the object, smoke, normally really means a fire. Four or five speeding fire trucks (one degree of separation) are a good indicator that there is a fire, but they may be there for an accident. Hearing several sirens (two degrees of separation) is not a very specific sign of a fire, as the sirens may be police cars or ambulances called for an accident. Thus, the degrees of separation between the object and the sign are important in the specificity of the signification.

The Examples of Diabetes, Inflammation, and OS

We will try to analyze two examples of molecular biomarkers used to define disease mechanisms, bearing in mind the points retained from the discussion above, particularly the concept of proxy and that of degrees of separation.

Let us start with a simple example. At the origin of diabetes, there is a lack of production of, or response to, insulin. As a result, blood glucose increases, leading to a number of pathological consequences. The blood level of glucose is a diagnostic biomarker for diabetes. In 1976, Cerami and colleagues (12) discovered that exposure to high blood glucose resulted in the formation of glycated hemoglobin (HbA1c; Fig. 3). Both high levels of glucose and HbA1c are signs of diabetes. However, HbA1c is, for kinetic reasons, a better diagnostic biomarker (13), even if it is more distant (two degrees of separation) from the mechanism of disease.

Let's look at the same scheme from the perspective of the identification of the mechanism of disease, or its therapy. High glucose has pathological consequences, and lowering its levels is good for the patient. However, drugs specifically targeting HbA1c would not modify the disease. While it may be a good *surrogate biomarker* in developing antidiabetic drugs, HbA1c is not a *proxy* and acting on it will not act on the disease. In fact, some drugs, such as dapsone, reduce the levels of HbA1c, but this does not make diabetic patients better (14).

A more complicated picture can be drawn for inflammatory diseases, such as RA. To make our point, we have drawn an oversimplified scheme of the inflammatory network in RA, involving only a few mediators (Fig. 4). In this scheme, autoimmunity induces Th17 cells that, through the cytokine IL-17,
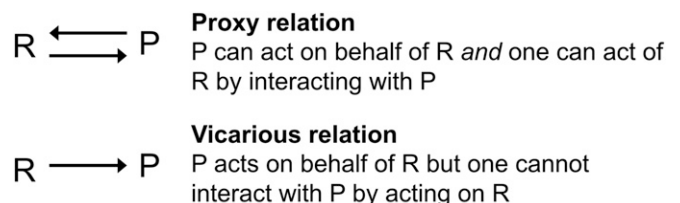

Proxy relation
P can act on behalf of R *and* one can act of R by interacting with P

Vicarious relation
P acts on behalf of R but one cannot interact with P by acting on R

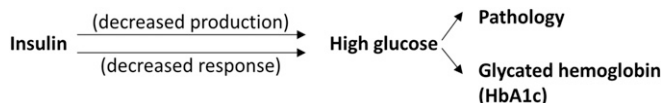**Fig. 2.** Direction of activity in proxies.

Fig. 3. Simplified scheme of the pathogenesis of diabetes.

activate macrophages to produce inflammatory cytokines (IL-1, IL-6, and TNF). The IL-1 produced by macrophages activated earlier is also a major inducer of IL-17. These will cause swelling or tissue damage, either directly or via prostaglandins such as PGE$_2$. IL-6 also acts on the liver to induce acute-phase proteins such as C-reactive protein (CRP) and fibrinogen, and to decrease the synthesis of albumin.

IL-17 is directly related to autoimmunity, one degree of separation, while IL-1 and TNF are two degrees of separation away, IL-6 both two and three, and PGE$_2$ and CRP four. Although the cytokines listed (IL-1, IL-6, IL-17, and TNF) are elevated in RA patients (15), only CRP is an established biomarker of disease (16) and is used both for diagnosis and as a surrogate marker in clinical trials.

In terms of disease mechanism, however, IL-17, IL-1, IL-6, and TNF are proxies. Acting on them, we act on the disease, and antibodies that neutralize these cytokines are approved therapies for RA. On the other hand, CRP is normally viewed as a protective molecule, and its absence, in animal models, is not protective against arthritis (17, 18); therefore, CRP can be considered a sign, possibly a surrogate, but not a proxy.

Let us now consider OS. According to its definition (7), OS occurs when the concentrations of ROS increase. One of the most reactive ROS is the hydroxyl radical (OH·), which reacts with lipids to initiate a chain reaction called lipid peroxidation and nucleic acids. Its short half-life forces us to use biomarkers that are the result of its interaction with those biological molecules, and we will consider two of them.

Fig. 5A shows how OH·, by initiating lipid peroxidation, results in the generation of malondialdehyde (MDA; probably the most used OS biomarker). Another biomarker used as an indirect indication of OH· formation is 8-OH-guanosine (8-OH-G; Fig. 5B). From the point of view of the distance, MDA is at >6 degrees of separation from OH (several chemical reactions are required), while 8-OH-G has only two degrees of separation. This distance raises the problem of the specificity of these biomarkers. In fact, MDA is produced not only by OH· but also during arachidonic acid metabolism and is used as a marker of cyclooxygenase (19).

Another question is whether OS biomarkers can be proxies. The main problem with the OS theory of disease is the difficulty of measuring the causative agents, ROS, unlike the cytokine theory of disease or the germ theory of disease, where cytokines and microbes can be measured in biological fluids. This means that none of the biomarkers of OS used in most of the studies are proxies.

Mathematical and Statistical Method for the Validation of Mechanistic Biomarkers and the Problem of Multicausality

When novel biomarkers are proposed, they are usually measured in a cohort of patients compared with healthy people or patients with unrelated disease. Then, a statistical comparison of the levels among two or more groups is done using the classical frequentist statistics, which calculates the probability that there is no difference between the two groups (the theoretical null hypothesis) (20). Another approach is to assess the probability that the biomarker levels correlate with the severity of a disease, a clinical score, or an outcome. The limitations of approaches based on the null hypothesis significance testing, with its dichotomization of evidence, have been discussed recently (21).

The size of the patient population investigated is also decided on the basis of conventional "power analysis," usually designed to meet the goals of the classical significance testing (22).

The problem of validating biomarkers that we can use to build theories of disease, be they proxies (such as TNF or IL-6) or just signs (such as CRP or 8-OH-G), is that while the usual statistical approach may be useful for surrogate biomarkers, we can apply it to mechanism biomarkers only for diseases that have a single cause, while many diseases are multifactorial, a result of multiple causes. This is typically represented by the so-called Rothman's causal pies (23). We could assume, for instance, that RA develops when a number of component causes add up to form a sufficient cause (the pie). However, the same disease could develop by a different combination of component causes to make a different sufficient cause.

OS has been implicated as a causal component in many diseases, including RA (3). Clearly, it must be a component cause because not all people with evidence of high OS develop RA. The question then arises, If OS is a slice in the causal pie for RA, is it a necessary component cause? That would mean that OS is always needed for the disease to develop, and therefore, all patients with RA will have OS (in red, Fig. 6A).

However, it is also possible that OS is a component cause only in one of the many possible sufficient causes, as in this case OS will be present in some patients but not in others. In this case, not all of the patients with RA will have OS; only those with one specific sufficient cause (Fig. 6B). These two possibilities are represented along with the third possibility (Fig. 6C), whereby OS is, alone, a sufficient cause of RA and is present in all patients.

If we measure biomarkers of OS in these patients, the classical statistical approach using the null hypothesis significance testing would only give a positive result in the case of mechanisms A and C. In the case depicted in Fig. 6B, if OS was implicated only in some patients, we would not find a significant difference in the level of biomarkers in a cohort of patients compared with healthy controls or a correlation with disease severity/outcome. Classical statistics has ways to analyze subpopulations of patients by stratification for covariates (e.g., gender, ethnicity, age etc.), but this would not help.

What is needed here is to first validate that higher levels of a biomarker signify that OS is present in a specific patient and then use that biomarker to identify those RA patients that have OS. It is difficult to name a clinically validated biomarker of OS
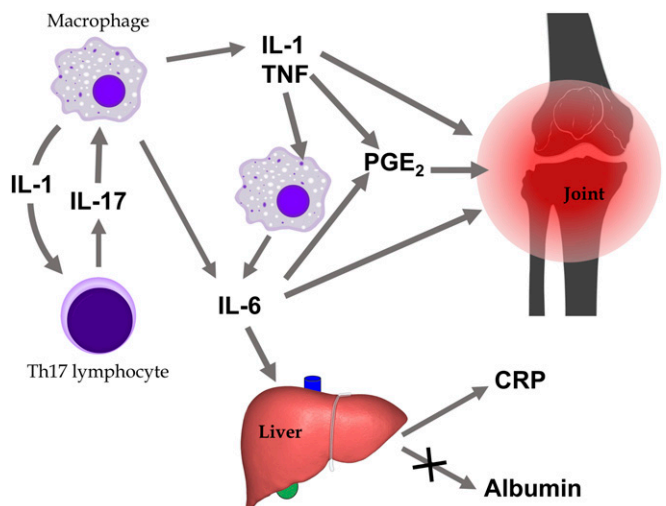

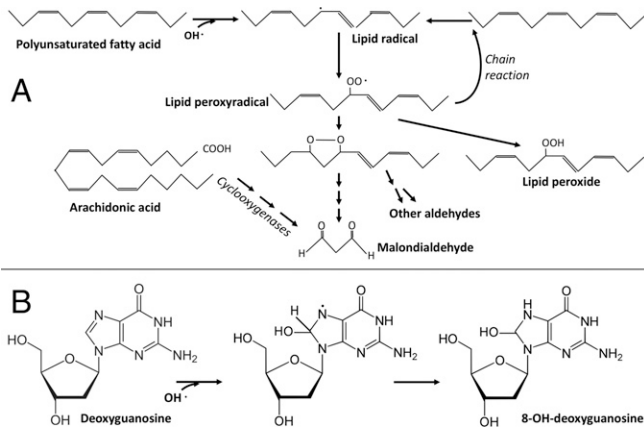**Fig. 4.** Simplified pathogenesis of RA (pictures: Wikimedia commons).

Fig. 5. Reactions leading to the formation of (*A*) MDA and (*B*) 8-OH-G.

Fig. 6. Different causal models for the OS theory of disease. Oxidative stress (in red) can be a component cause necessary but not sufficient (*A*) or not necessary (*B*). It can also be a sufficient cause of disease (*C*).

because there are no clinical conditions where OS is present in all patients (although radiation toxicity and ozone toxicity would get close) (24). Considering the additional problem of lack of specificity, it would be appropriate to use a different approach. We need to confirm that, in each patient, higher ROS levels are produced by measuring several OS biomarkers: We need to disambiguate the meaning of a sign. Exactly as it happens when we disambiguate the meaning of words, we need to look at the context. If we read a paper where MDA is measured in a study on diet and OS, we know its meaning is as a biomarker of OS, while if the context is that of a drug acting on platelet activation, we can extrapolate that MDA is intended as a biomarker of arachidonate metabolism.

In patients, to find the value of a biomarker and to validate it as an indicator of OS, one approach to disambiguate its meaning could be to look at another OS biomarker—another sign in the same system. Thus, in biomarkers, we need to see them in the context of other biomarkers in the same system also to disambiguate their meaning, not just to appreciate their value.

Along the line of "systems of signs," another possibility is that, before any statistical analysis, we visualize the data in a network, both to identify potential patient subpopulations and to disambiguate the meaning of the biomarkers measured.

We performed a proof-of-principle analysis of the levels of various OS biomarkers in different diseases (2). The data, obtained from published literature, consisted of a spreadsheet with different diseases and, for each of them, the OS biomarker that was reported elevated in that disease. These data were analyzed using hierarchical cluster analysis to visualize the similarities between different diseases in terms of pattern of OS biomarkers (2).

A second type of visualization is based on a network analysis, originally used to identify the "diseasome," by which different diseases are connected based on genetic risk factors, subsequently extended to protein–protein interactions (25–27).

Using published data (2), we have created a disease network based on the common OS biomarkers, as described in *Methods*. The result is shown in Fig. 7, which depicts the vicinity of diseases based on commonly elevated biomarkers.

This type of analysis, applied to biomarkers related to different disease mechanisms (inflammation, OS, autoimmunity, viral antibodies, etc.), could be used to identify diseases with common pathogenic mechanisms. It could also be applied to patients with a single disease to identify patients with different causal components to assign each patient to a hypothetical Rothman's causal pie. This approach might then be followed by correlation analysis and classical significance testing to assess whether differences in the levels of biomarkers among different subgroups are robust. Along this line, a Bayesian latent hierarchical model
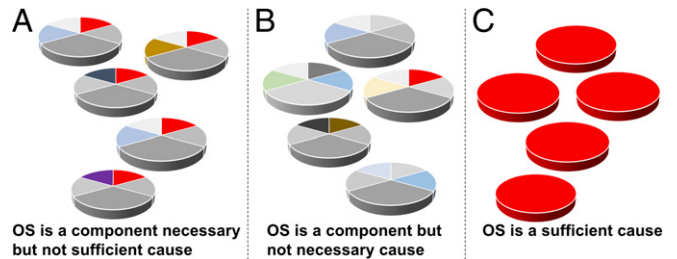
has been proposed to identify metapatterns of biomarkers in a different cohort to facilitate hypothesis generation (28).

Constraints in Biomarker Discovery

Several conventions limit the study of biomarkers that identify causal components of disease. Statistical analysis based on the null hypothesis significance testing is a prerequisite for publication in what has been defined a "lexicographic decision rule" (21). Funding agencies and ethical committees require a standard "research methodology" involving the use of power analysis to identify patients' group size, tailored to satisfy classical significance testing. For this reason, most of the studies on the diseasome published so far have been done analyzing published literature or databases rather than with laboratory or clinical studies.

Another constraint is the drive to develop diagnostic/prognostic biomarkers to meet real clinical needs (such as diagnosis of cancer) or drug development needs, where surrogate biomarkers are used for the approval on new drugs instead of clinical outcomes (29). These constraints favor the discovery and validation of biomarkers that often do not provide an indication on the mechanism of disease.

Let us go back once more to the OS theory of disease and the various possibilities outlined in Fig. 6. It is possible that OS is a causal component in RA, but it is also possible that it is a consequence of the disease—for instance, secondary to inflammation or joint damage. If OS is a causal component and the disease has multiple causes as in Fig. 6*B*, we will not be able to
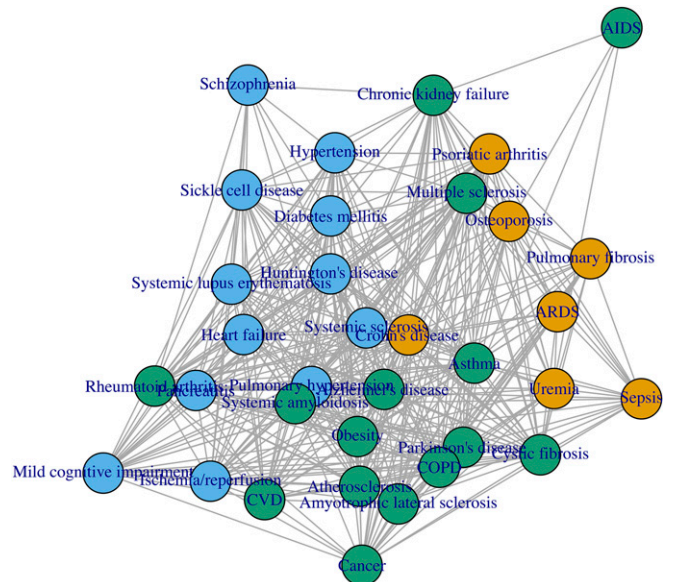

Fig. 7. Disease vicinity based on OS biomarkers, based on data from ref. 2.

find a statistically significant difference between a healthy group and a disease one. However, if OS was not a cause but a result of the disease—that is, a "biochemical symptom" or a causally irrelevant by-product—then it would be present in all patients and, possibly, correlate with disease severity. Paradoxically, in multicausal disease, the more distant the biomarker is from the mechanism of disease, the more likely it is that it will be validated (and published) using classical significance testing. For instance, in Fig. 4, CRP is a better biomarker than IL-17 or TNF. Conventional statistical analysis and study design will favor the identification of a causally irrelevant prognostic biomarker even when the study was aimed at investigating the mechanism of the disease.

Artificial Intelligence, Big Data, and the Problem of the Interpretant

For Pierce, the interpretant in Fig. 1 is a human mind, and all of the discussions in this paper implied a researcher/practitioner interpreting the meaning. If a biomarker is the carrier of a message, in medical semiotic, "an observation that does not convey a message is not a sign" (30). However, -omics and big data may generate messages not intelligible to a human. An example is the use of gene expression profiling to identify arthritis patients that would benefit from biologicals. These studies analyzed the expression of tens of thousands of genes in drug-responsive and nonresponsive patients and identified algorithms, derived from the expression of up to 200 genes and cluster analysis, that correlate with drug responsiveness (31, 32). Only a few of these genes will have a place in the theories about the mechanism of disease, because the function of many of them is not known, and the cluster will have no meaning for the human interpretant.

This will require an entirely different approach and raises the challenging question on whether the theories of signs and the concept of "meaning," and their application to biomarkers, need to be modified when interpreting big data. This is similar to the example cited by Evans on GPS navigation: "The physical world is translated into a database of instructions and distances, and interpreted by the application into a route to follow. … The application then presents the world back to the user in a mediated form … which is used to navigate the route" (33). Clearly an algorithm, or cluster of genes and transcripts, can represent a surrogate sign, as it can act on behalf of/replace the disease (and be used, for instance, as secondary outcomes in a clinical trial) without referring to it.

While the focus on biomarkers generated by high-throughput techniques has been on their diagnostic and prognostic use, their use in building mechanistic theories of disease will require a different theoretical and statistical approach. This is important in mechanism-based patient stratification and in the translation of the concept of personalized medicine.

Methods

The data were processed in GNU R and visualized as a graph. Thereafter, the graph was saved in a format that can be loaded into Gephi for exploratory analysis. The source data are represented as a matrix where each row in the matrix represents whether each of the 20 biomarkers indicates the presence of a disease. In general, either a biomarker is elevated in a disease, according to the small sample of the literature analyzed, or it does not. The binary nature of these observations lends themselves to computing the similarity of diseases using a Jaccard index (34). The Jaccard distance between diseases is normalized between 0 and 1. The more similarity between diseases, the closer their Jaccard distance is to 0. The computed matrix of Jaccard distances is then transformed into a matrix of edge weights and presented using a force-directed graph layout algorithm (35). In our graph layout, the diseases are visualized as graph vertices. The edges of the graph are weighted with the inverse of their Jaccard distance. Graph edges with a weight of 1 indicate a strong attraction between the vertices, whereas graph edges with a weight of 0 indicate no attraction between the vertices. The graph layout is computed using the DrL algorithm, and the results are saved to a Gephi file. The visualization code is available at https://github.com/AidanDelaney/DiseaseViz.

1. Ghezzi P, et al. (2017) Oxidative stress and inflammation induced by environmental and psychological stressors: A biomarker perspective. *Antioxid Redox Signal*, 10.1089/ars.2017.7147.
2. Frijhoff J, et al. (2015) Clinical relevance of biomarkers of oxidative stress. *Antioxid Redox Signal* 23:1144–1170.
3. Ghezzi P, Jaquet V, Marcucci F, Schmidt HHHW (2017) The oxidative stress theory of disease: Levels of evidence and epistemological aspects. *Br J Pharmacol* 174:1784–1796.
4. Beutler B, et al. (1985) Identity of tumour necrosis factor and the macrophage-secreted factor cachectin. *Nature* 316:552–554.
5. Feldmann M, Maini RN (2010) Anti-TNF therapy, from rationale to standard of care: What lessons has it taught us? *J Immunol* 185:791–794.
6. Tracey KJ (2007) Physiology and immunology of the cholinergic antiinflammatory pathway. *J Clin Invest* 117:289–296.
7. Sies H (2015) Oxidative stress: A concept in redox biology and medicine. *Redox Biol* 4:180–183.
8. Harman D (1956) Aging: A theory based on free radical and radiation chemistry. *J Gerontol* 11:298–300.
9. Kehrer JP (2000) The Haber-Weiss reaction and mechanisms of toxicity. *Toxicology* 149:43–50.
10. Peirce C (1955) Logic as semiotic: The theory of signs. *Philosophical Writings of Peirce*, ed Buchler J (Dover Publications, New York).
11. Floridi L (2015) A proxy culture. *Philos Technol* 28:487–490.
12. Cerami A (2012) The unexpected pathway to the creation of the HbA1c test and the discovery of AGE's. *J Intern Med* 271:219–226.
13. American Diabetes Association (2017) Standards of medical care in diabetes-2017 abridged for primary care providers. *Clin Diabetes* 35:5–26.
14. Polgreen PM, Putz D, Stapleton JT (2003) Inaccurate glycosylated hemoglobin A1C measurements in human immunodeficiency virus-positive patients with diabetes mellitus. *Clin Infect Dis* 37:e53–e56.
15. Holt I, Cooper RG, Denton J, Meager A, Hopkins SJ (1992) Cytokine inter-relationships and their association with disease activity in arthritis. *Br J Rheumatol* 31:725–733.
16. Aletaha D, et al. (2010) 2010 rheumatoid arthritis classification criteria: An American college of rheumatology/European league against rheumatism collaborative initiative. *Arthritis Rheum* 62:2569–2581.
17. Jones NR, et al. (2011) Collagen-induced arthritis is exacerbated in C-reactive protein-deficient mice. *Arthritis Rheum* 63:2641–2650.
18. Jiang S, Xia D, Samols D (2006) Expression of rabbit C-reactive protein in transgenic mice inhibits development of antigen-induced arthritis. *Scand J Rheumatol* 35:351–355.
19. Moncada S (1982) Prostacyclin and arterial wall biology. *Arteriosclerosis* 2:193–207.
20. Pencina MJ, D'Agostino RB, Vasan RS (2010) Statistical methods for assessment of added usefulness of new biomarkers. *Clin Chem Lab Med* 48:1703–1711.
21. McShane B, Gal D, Gelman A, Robert C, Tackett J (2017) Abandon statistical significance. arXiv:1709.07588.
22. Gelman A, Carlin J (2014) Beyond power calculations: Assessing type S (sign) and type M (magnitude) errors. *Perspect Psychol Sci* 9:641–651.
23. Rothman KJ (1976) Causes. *Am J Epidemiol* 104:587–592.
24. Gerschman R, Gilbert DL, Nye SW, Dwyer P, Fenn WO (1954) Oxygen poisoning and x-irradiation: A mechanism in common. *Science* 119:623–626.
25. Menche J, et al. (2015) Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347:1257601.
26. Barabási AL, Gulbahce N, Loscalzo J (2011) Network medicine: A network-based approach to human disease. *Nat Rev Genet* 12:56–68.
27. Goh KI, et al. (2007) The human disease network. *Proc Natl Acad Sci USA* 104:8685–8690.
28. Huo Z, Song C, Tseng G (2017) Bayesian latent hierarchical model for transcriptomic meta-analysis to detect biomarkers with clustered meta-patterns of differential expression signals. arXiv:1707.03301.
29. Wang B, Kesselheim AS (2015) Characteristics of efficacy evidence supporting approval of supplemental indications for prescription drugs in United States, 2005-14: Systematic review. *BMJ* 351:h4679.
30. King LS (2014) *Medical Thinking: A Historical Preface* (Princeton Univ Press, Princeton).
31. Stuhlmüller B, et al. (2016) Genomic stratification by expression of HLA-DRB4 alleles identifies differential innate and adaptive immune transcriptional patterns–A strategy to detect predictors of methotrexate response in early rheumatoid arthritis. *Clin Immunol* 171:50–61.
32. Toonen EJ, et al. (2012) Validation study of existing gene expression signatures for anti-TNF treatment in patients with rheumatoid arthritis. *PLoS One* 7:e33199.
33. Evans L (2011) Object-oriented philosophy: The nature of relations between humans and computational objects. *Proceedings of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, eds Kazakov D, Tsoulas G (University of York, York, UK). Available at http://www.aisb.org.uk/aisbpublications/93-aisb/publications/proceedings/91. Accessed February 6, 2018.
34. Jaccard P (1901) Étude comparative de la distribution florale dans une portion des Alpes et des Jura. *Bull Soc Vaud Sci Nat* 37:547–579.
35. Martin S, Brown W, Klavans R, Boyack K (2008) Distributed recursive (graph) layout. *SAND Reports* 2936:1–10.