Medical Research Archives

ESMED
European Society of Medicine

OPEN ACCESS

Published: July 31, 2022

Citation: Christopher Boorse, 2022. The Meaning of Clinical Normality, Medical Research Archives, [online] 10(7). https://doi.org/10.18103/mra.v10i7.2866

DOI https://doi.org/10.18103/mra.v10i7.2866

ISSN: 2375-1924

RESEARCH ARTICLE

The Meaning of Clinical Normality

Christopher Boorse*[1]

[1] Department of Philosophy, University of Delaware, Newark, DE, 19716 USA.

* cboorse@udel.edu

ABSTRACT

Many writers have called the term 'normal' highly ambiguous both in and out of medicine, especially between descriptive and normative meanings. But careful analysis shows that its ambiguity is much less than usually supposed. In fact, all correct nontechnical uses of 'normal' mean "typical" in some way – either typical, at least typical, or at most typical – and therefore express no value judgments except by contextual implication. The distinctive, purely medical use, as the opposite of 'pathological', is just a specialization of the second meaning, to at-least-typical biological part-function. As statisticians have often warned, one must not confuse this uniquely medical use with a general use formerly applied to clinical tests, in the term "normal range." That term is misleading because the reference ranges of clinical variables entail nothing about pathology, for three reasons: besides resting on an arbitrary choice of a 95% central range, they are derived from apparently healthy populations, and the variables' connection to underlying biological function can be very indirect. So the term "clinical abnormality" is best restricted to a diagnosed or diagnosable pathological condition. If so, true clinical normality contrasts with theoretical normality in some interesting ways: it may or may not correlate with disease severity; it is individual-relative; and it is partly determined by value judgments.

INTRODUCTION

Philosophers of medicine show a broad consensus that the term 'normal' is highly ambiguous, especially between facts and values. I shall argue that this consensus is mostly wrong, and the *Oxford English Dictionary* mostly right. As to general usage of 'normal', there are two main questions. One is whether it always means "typical" or "usual," or sometimes has other senses like "ideal." The other, related issue is whether its meaning includes a value judgment. My answers will be that today 'normal' always means typical in some sense, and so never entails ideality or any other evaluation. In §II and §III, I present three reasons why the "normal range" of clinical variables has no essential relation to the uniquely medical sense of abnormality -- pathologicity -- which is merely a specialization of one of the general senses to biological function. In §IV, I look at a few interesting twists in the truly medical concept of clinical abnormality: a patient's clinically diagnosed or diagnosable pathological condition.

I. THE MEANING OF 'NORMAL'

Physician and philosopher of medicine Edmond Murphy[1] famously distinguished "seven meanings of the word 'normal' important in medicine" (125). *[1]* (Page numbers are in parentheses, endnote numbers in italic brackets.) For the sake of clarity, he also suggested a "preferable" replacement for each. His seven alleged senses of 'normal' were as follows:

	Paraphrase	Domains of use	Preferable term
1.	Having a Gaussian distribution	Statistics	Gaussian
2.	Most representative of its class	Descriptive science (e.g., biology)	Average, median, modal
3.	Commonly encountered in its class	Descriptive science	Habitual, ordinary
4.	Most suited to survival and reproduction	Genetics, operations research quality control, etc.	Optimal or "fittest"
5.	Carrying no penalty	Clinical medicine	Innocuous or harmless
6.	Commonly aspired to	Politics, sociology, etc.	Conventional
7.	Most perfect of its class	Metaphysics, esthetics, morals, etc.	Ideal

I shall begin by dealing with Murphy's scheme, since he is at the polysemous extreme of semantic analysis, before passing to other writers on normality, especially Ian Hacking, and ending this section with my own view. Murphy's attempt to catalogue the many things speakers might mean by 'normal' is admirable, but I have many criticisms of his list. First, only 1-5 appear to be medical usages at all. Second, at least "meanings" 4 and 7 are, I suggest, clear errors. It is a mistake to think that 'normal' ever means "ideal." The now highly descriptivist *Oxford English Dictionary*[2] is a work so amazingly tolerant as to claim that 'reticent' can mean "reluctant," 'refute' "rebut," and 'alibi' "excuse." Yet the *OED* still has no sense for 'normal' that is near "ideal." Instead, under "general" uses it lists two current meanings. The usual one, it says, is "constituting or conforming to a type or standard; regular, usual, typical; ordinary, conventional." The second, said "of a person," is "physically and mentally sound; free from any disorder; healthy." *[2]* Some writers who use the word 'normal' may be confusing it with other words, such as 'normative', which can perhaps mean 'ideal'; but 'normal', I shall argue, cannot. *[3]* Thus 'normal' cannot mean "optimal" in genetics, as Murphy claims. The normal allele at a certain locus may in fact be the fittest one available in the gene pool, but, if so, that is a result of the efficiency of natural selection, not of the meaning of the term.

I shall argue that the *OED* entry, though unnecessarily complex, is very nearly right in its basic structure. In the first place, it rightly separates off a range of "technical" uses in mathematics and the sciences. Mathematicians are familiar with a long list of specific senses, which, as with 'complete', are all fully precise but vary with the subject matter.

In geometry or linear algebra a line may be normal (= perpendicular) to a plane; in algebra a subgroup may be normal (= closed under conjugation by group members); in logic, a modal system may be normal (= containing system K), and a propositional formula may be in disjunctive normal form (= a disjunction of line-descriptions); and so on through many other domain-specific meanings completely irrelevant to our topic. The mathematical use of 'normal' most relevant to medicine is, as Murphy says, "Gaussian," which is the familiar family of bell-shaped continuous probability distributions fixed by two parameters, mean μ and standard deviation σ. The *OED* also notes special technical meanings in various sciences. In chemistry, a normal solution of a substance has one gram per liter; a normal salt has no acidic hydrogen; and so on. Physics, geology, and biology are also credited with their own technical meanings.

As for other uses of 'normal', my conclusion will be this: they always reduce to "typical" or "usual" in some sense, and so never entail any value judgment -- though in context they can have evaluation as a pragmatic implication. *[4]* Once again, the typicality claim puts me in some disagreement with Murphy. His sense 3 ("habitual") is indeed appropriate to discrete characters (e.g., number of fingers). And one might think his sense 2 ("average") better suits continuous ones (e.g., height), but again I disagree. If 5' 9.1" is the mean height of American men, no one can properly call a height of 5' 9.2", or even 5' 11", abnormal. On the contrary, a man of exactly mean height is just as unusual as a day of mean low temperature for its date. *[5]* So, unless 'abnormal' means something other than "not normal," Murphy's #2 is just one more improper use of the term. Rather, for continuously distributed traits, the best one can do is to speak of a "normal range" surrounding the mean. How this range is, or should be, determined is the subject of §II.

Curiously, Murphy wholly omits what is, to me, the one distinctively medical sense of 'normal', as the antonym of 'pathological.' On my analysis[3-6] *[6]*, the normal-pathological distinction is the foundation of scientific medical thought. I have argued that the essence of the pathological is statistically subnormal biological functional ability in some part of the organism, relative to species, sex, and age. Thus, normality in the distinctively medical sense – the *OED*'s second general one – is a kind of statistical normality, just statistical normality of *function*. *[7]* (Medical sources also often call statistical anomalies of structure pathological, but I have argued[3] (565-66) that to do so makes the concept unanalyzable, only obscuring the fundamental meaning of the pathological: dysfunction.) Now clinicians also make great use of the idea of a "normal range" of a clinical variable such as temperature, blood pressure, platelet count, serum potassium, and so on. But in a salutary recent development for which Murphy, Feinstein, and others eloquently spoke, this term is being replaced by "reference range," for good reasons to be explored in the next section.

Besides Murphy, we can learn from the historical and philosophical analysis of several other writers on normality -- especially Ian Hacking, in his magisterial book *The Taming of Chance*[7] (chh. 13, 19-21). As Hacking (162-3), Jiri Vacha[8] (924), and Christel Fricke[9] (698) note, the word 'normal' descends to us from the Latin *norma* (itself derived from Greek), which meant simply a tool: the carpenter's square. And although Fricke, like Vacha and Hacking, believes the English 'normal' is ambiguous between descriptive and evaluative meanings (692), she is admirably clear about the value status of perpendicularity. A right angle is mathematically special, since "[t]here is only one way in which two straight lines can cross such that the angles between them are all the same" (698). Nonetheless, the practical value of this to builders is contingent on both "laws of nature" (gravity) and builders' purposes.

> Who needs a tool such as the *norma*? Rectangularity is an important property for the work of carpenters and other builders .... As Vitruvius pointed out long ago, walls should be erected strictly vertically – otherwise they could not stand without any additional support. ... Attributing the property of rectangularity to a material object is a matter of description and classification. The corresponding norm owes its authority to the architects' and builders' commitment to erecting stable buildings. (699)

Thus the original meaning of *norma* was purely descriptive, though it had nothing to do with the modern adjective's idea of typicality. But despite the contingent value of the perpendicular, already in classical Latin[10] we find *norma* also used to mean

"a rule, pattern, precept". *[8]* Although this use is only metaphorical, it foreshadows the alleged descriptive/normative "ambiguity" of 'normal' that Hacking stresses and that Fricke says scholars trace to this humble carpenter's tool. *[9]*

Hacking, in his inimitably breathless historiographic style, is eloquent about this supposed ambiguity. The word 'normal', he says,

> uses a power as old as Aristotle to bridge the fact/value distinction, whispering in your ear that what is normal is also all right. ... Nothing is more commonplace than the distinction between fact and value. From the beginning of our language the word 'normal' has been dancing and prancing all over it. ... The word ... is like that baneful Californian shrub, poison oak, which assumes whatever form resembles the environment. Now it is a creeper, crawling close to the earth, now a pleasant round bush five metres high ...; now it is red, now it is green, now it is leafless but the sap is running and itching to attack. ... [F]or much of the century before Durkheim, and ever since, we have regularly used 'normal' to close the gap between 'is' and 'ought'. Wrongly so, perhaps, but that is what the concept of normality does for us. (160, 163) *[10]*

Hacking bases this view on a rich and detailed historical inquiry, in which the most important figures are five: Broussais, Quetelet, Comte, Durkheim, and Galton. According to his historical researches, the first influential modern use of 'normal' to mean "typical" is in physician F.-J-.V. Broussais's 1828 medical treatise *De l'Irritation et de la Folie*. The main thesis of Broussais's "physiological" medicine was that "a diseased state simply is an irritated tissue or organ, which is nothing other than 'a normal excitation that has been transformed by an excess'" (165). In the same year the word passed into English (162), and was quickly popularized in France by Balzac and Comte (166). Comte hailed Broussais's principle that "the phenomena of disease are of essentially the same kind as those of health, from which they [differ] only in intensity"[11] (160). In his sociology, Comte proposed to extend this idea – that "the pathological is not radically different from the normal, but only an extension of the variation proper to a 'normal organism'" (166) -- to the social organism, diagnosing whole societies as normal or pathological. "But," Hacking notes,

> when Comte moved normality to the political sphere, he effected another twist. The normal ceased to be the ordinary healthy state; it became the purified state to which we should strive, and to which our energies are tending,

which Comte called the "true normal" (168).

> Comte thus expressed and to some extent invented a fundamental tension in the idea of the normal – the normal as existing average, and the normal as figure of perfection .... This is an even richer source of hidden power than the fact/value ambiguity that had always been present in the idea of the normal. (168)

Hacking sees a recurrent basic tension between a conservative view of the normal, exemplified by Durkheim, as the status quo which is right, and a progressive ideal of normality, ever striving for new excellence, exemplified by Comte and Galton (168-9).

In between Broussais's and Comte's books came two works by highly influential royal astronomer Adolphe Quetelet, his 1835 *Treatise on Man* and his 1844 *Recherches Statistiques*. To Hacking, Quetelet's most important idea was to reify the Average Man – the *homme type* – of a race or population (105, 107). Applying the established Gaussian theory of astronomical measurement errors first to measuring a single human being, and then to bell-shaped human variation in general, he "pass[ed] from a real physical unknown, the height of one person, to a postulated reality, an objective property of a population at a time, its mean height or longevity or whatever" (109). For Elliott Sober,[12] however, the most important feature of Quetelet's idea was a corollary of the measurement-error analogy: that all human variation is pathological deviation from a natural state.

> For Quetelet, the law of errors is still a law about errors, only for him

> the mistakes are made by nature, not by observers. ... [V]ariation in a population ... is the result of interferences confounding the expression of a prototype. (367)

By contrast, Francis Galton, in his 1869 *Hereditary Genius*, "used the [Gaussian] law of errors, but no longer viewed it as a law *about* errors" (368). And it was Galton who later (1889)

> introduced the name "normal law" as a more appropriate label .... Bell curves are normal; they are found everywhere, Galton thought. ... [They] need not represent mistakes made by fallible observers or by sportive nature. Regardless of the underlying etiology, *they are real*; they enter into explanations because the variability they represent is lawful and causally efficacious. (369)

By this date, Hacking says, the word 'normal' "just meant typical." *[11]*

Despite both Hacking and Sober's deep historical knowledge *[12]*, there are aspects of both accounts with which I must quarrel. As to Hacking, to begin with, I do not quite see how his description of Broussais's two-part principle is coherent. If "[t]he pathological was defined as deviation from the normal," while "[t]he normal is the centre from which deviation departs" (164), then it sounds as if Quetelet was right: since the center is a point, not a region, there would be no "variation proper to a 'normal organism'" (166). But perhaps Broussais's thesis was merely that all pathology is quantitative, not qualitative, deviation of a certain size from the mean. Also, as I shall explain in a moment, the contrast is not so sharp between Durkheim's and Galton's "two visions" (178) of normality. Hacking writes:

> What is the opposite of the normal? The abnormal, certainly. But for Galton the normal was characterized by the Normal curve; the abnormal was what strongly deviated from its mean. For Durkheim, the abnormal was called the pathological. In the end the abnormal is sick. For Galton, the abnormal is exceptional .... For Galton the normal was not good but mediocre. Some extremes were not pathological but superb. (178)

This paragraph manufactures unnecessary drama. We can, like the OED, recognize two senses of 'normal' here, whose relation is easy to explain, as I am about to show.

Finally, I suggest that both Hacking and Sober overestimate Quetelet's influence on contemporary medical thought. In claiming that Quetelet's application of the law of errors to population variation "was to determine the entire future of statistics" (109), Hacking fails to make clear that scientific biomedicine has no place at all for Quetelet's thesis that all human variation is pathological. As I have explained elsewhere, Sober makes the same error on an even grander scale: he uses Aristotle and Quetelet to question[12] (377-8), and ultimately reject[13] (160-1), the whole normal/pathological distinction as unscientific. For related reasons, I also dispute Hacking's overall thesis (1, 161-2) that his nineteenth-century writers replaced the Enlightenment concept of human nature with the concept of a normal person. On the contrary, as I argue elsewhere, human psychological nature – the normal functional architecture of the human mind -- is precisely the concept we need, and have implicitly always used, for the classic roles of human nature in science, ethics, politics, and law.

In any case, I believe we can state a simple, general semantic analysis of 'normal' without fear of Proteus, or of a stealthy attack by the raging sap of California poison oak. In fact, given Hacking's statement (184) that by 1888 'normal' just meant "typical," it is unclear that he has any serious disagreement with my position.

First, like the *OED,* we separate off technical uses with specific precise definitions. Second, on the key question of whether the nontechnical meaning of 'normal' entails a value judgment, there are three possible positions: (i) it always does; (ii) it sometimes does, but not always, so it is ambiguous; and (iii) it never does. We can quickly reject (i). In some contexts normality is good, but in some it is bad. To return to normal life after a Caribbean cruise or European vacation can be depressing. *[13]* And the statement 'Temperatures here have been abnormally high this season' has the same meaning when uttered in Miami in August and International Falls in January *[14]*, but if value

judgments are conveyed, they are opposite. Hence neither value judgment can be part of the statement's meaning.

As for option (ii) – fact-value ambiguity – it is unnecessary. True, in conversation, "X is normal" often conveys a value judgment. But there is no reason to make this part of its semantics, rather than a pragmatic implicature based on a contextual limitation to a topic where there is a value presumption that what is typical is good, or at least acceptable. An analogy might be 'common' and 'rare' in coin or gem collecting. 'Common' (outside aristocratic Britain) does not, of course, mean "worthless." But since collectors agree in viewing common coins or rocks as worthless, they draw this inference automatically. Similarly, like rarity in coin collecting, we assume intelligence to be valuable. So to tell a parent that a child homozygous for the phenylketonuria gene can still be of normal intelligence on a phenylalanine-free diet is, in effect, to praise the result. But that is not because 'normal' means "good" in this context. It is because of our background value assumption that intelligence is good, so it is good to be of at least typical intelligence – which is what "normal" intelligence is taken to mean here. In other cases, 'normal' can also mean "at most typical." If a Floridian hopes for the return to normal temperatures after an August heat wave, he is hoping for at-most-typical heat -- not hoping against atypical coolness. An urban politician in a crime wave who pledges a return to normal crime levels is not vowing to block historic lows.

On the other hand, we do not assume that, say, thyroxine secretion is good or bad in itself. Rather, physicians know that both too much and too little thyroxine are bad for a person, causing the syndromes of hyper- and hypothyroidism. So doctor and patient are relieved by a report of normal – that is, typical -- thyroxine levels. Thus, as Fricke[9] notes (697, 701), normality can be associated with either a bivalent (bad/good) or a trivalent (worse/good/better) value classification, depending on whether the thing measured is not, or is, thought good or bad in itself. Still, valuation differences are not the only source of ambiguity in normality. It is also quite acceptable to say that Newton and Einstein were abnormally – atypically -- intelligent, even though their superior intelligence was not abnormal in the sense of pathology. So 'abnormal' can be ambiguous even on a single subject with a fixed valuation.

Finally, these medical examples remind us that the uniquely medical use of 'normal' – which Hacking assures us (160) is the original modern meaning -- is as the opposite of 'pathological'. Although I would not say that it was clear in the 1820's, by the end of the century physicians were generally coming to see that the essence of the pathological is not mere statistical abnormality, but statistical *sub*normality of *function*. No doubt the normal/pathological distinction is so basic to medical thought that we can follow the *OED* in making "nonpathological" a separate general sense of 'normal', which I will henceforth call "true medical normality." But it rests merely on a contextual limitation of subject matter: to biological function. *[15]* And even with "nonpathological" as a separate sense of 'normal', our semantic analysis is nicely general:

> (i) In current nontechnical usage, 'normal' always means either "typical," "at least typical," or "at most typical" *[16]* (the latter two being my improvement on the *OED*'s account). Thus the term is ambiguous, with a matching ambiguity in 'abnormal' (= not normal).
>
> (ii) In the most common context where 'normal' means "at least typical" -- the truly medical meaning where 'normal' is the opposite of 'pathological' -- normality is, specifically, at-least-typical *biological function*. *[17]*
>
> (iii) 'Normal' never entails a value judgment.
>
> (iv) It does, however, often convey a value judgment -- by pragmatic implicature, given background value assumptions. *[18]*

Because of its medical origin, 'normal' may be especially likely to be preferred to 'typical' or 'usual' when such background presumptions apply. But these value judgments are rarely universal. Almost anything, even life and health, can be bad for you in some circumstances, even though at-least-typical function is almost always good.

I hope that this fairly tidy framework explains and preserves all defensible current uses, while eliminating confused or indefensible ones and

dissolving all mysteries. In any case, the remaining sections of this paper presuppose nothing but my analysis of pathologicity.[3-6]

II. REFERENCE RANGES

For the innumerable continuous variables used in clinical diagnosis, there is a standard methodology for determining the "reference range" reported in handbooks, textbooks, and lab reports. *[19]* In contrast to the risk-based disorders I survey elsewhere, this procedure has nothing to do with risk, or even directly with disease. First, a sufficiently large sample of apparently healthy people is chosen -- either healthy in general, or free of a particular kind of disease – and variable V is measured in each by some uniform method. Second, one sees whether the resulting data are close to a normal (= Gaussian) distribution. If so, the reference range is chosen as $[\mu - 2\sigma, \mu + 2\sigma]$. That is, in the old terminology, the "normal range" for this variable includes everything within two standard deviations of the mean. This range constitutes a central ~95% of the sample, with two "abnormal" tails of ~2.3% each. *[20]* If the data do not look bell-shaped, as Feinstein[14] (247) says is usually the case, there are two alternatives. First, one can test whether they do so after some transformation. For example, some variables turn out to have a "log-normal" empirical distribution, which means that the curve becomes Gaussian when each value is replaced by its logarithm. Or one might find that cube roots or arcsines do the trick. In such cases, one can find the interval $\mu \pm 2\sigma$ in the transformed data and then reverse the transformation to get a reference range. If the data cannot easily be transformed into a normal distribution, then simple percentiles can still be used to come up with the desired central 95%. *[21]*

Of course, if one applies this procedure to a sample of apparently healthy human beings in general, the resulting data are often "multimodal," as determined by various statistical tests. One reason for this is that the mean of many variables varies with sex, age, race, and pregnancy, not to mention a host of factors within each individual such as posture, recent meals, exercise, and current drugs ("preanalytic variability," Jacobs *et al.*[15] (15)), as well as time of day or position in other, noncircadian biological cycles ("biologic variability"). So, for example, the reference range for hematocrit (proportion of red cells to blood volume) comes out to be 37-48% in females but 42-52% in males (Wilson[16] 306); the blood uric-acid range is 2.3-6.6 mg/dl for females and 3.6-8.5 for males (*ibid.*, 579). That such differences in a sample are real is confirmed by significance testing. Serum calcium for adults reveals a range of 8.6-10.0 mg/dl, but 7.6-10.4 for newborns[15] (131). Adults under 60 have an ESR (erythrocyte sedimentation rate) of 0-15 mm/hr for men and 0-20 for women, but for those over 60 the ranges are 0-25 and 0-30 respectively (Ravel[17] 5). Pregnant women's mean uric acid and alkaline phosphatase, as percentages of the nonpregnant mean, are 68% and 90% at 12 weeks, but, like many other quantities, rise sharply throughout pregnancy to 120% and 347% at term (Jacobs *et al.*[15] 16). Hard exercise produces temporary rises in lactate, phosphorus, creatinine, leukocyte count, and many other variables, as well as declines in albumin, iron, and sodium (*ibid.*, 17). Merely taking blood samples when a patient is lying down can reduce total protein, albumin, cholesterol, triglyceride, and alkaline phosphatase levels by 9% from the same patient's results in upright position (Ravel[17] 4). And countless physiological quantities vary regularly and significantly throughout a day and night, such as creatinine, which increases 30% from 7 am to 7 pm (Jacobs *et al.*[15] 18).

Since the procedure just described, which I will call "the 95% procedure," is standard practice in laboratory medicine, let us consider some conceptual points about it before looking at a refinement. First, both Murphy and Feinstein object to the assumption that biological data should have a bell-shaped distribution. Murphy[1] writes:

> Why the term "normal" ever was applied to [the Gaussian curve] is not clear; but there is no reason for thinking that it has anything whatsoever to do with the word normal in any other sense. There is no reason at all why an attribute of "normal" people should have this distribution; indeed, it is usually impossible for it to do so, since this distribution has no limits, and most variables in man (height, weight, blood sugar, etc.) cannot assume negative values. (124)

Murphy is of course right on the last point: no biological distribution can literally be Gaussian. Otherwise, his judgment seems a bit harsh. Among many properties that make it central to probability theory, the normal curve is the asymptotic limit of

various distributions, such as the binomial and the Poisson, which arise naturally in biological models. For example, if the magnitude of a given character were determined by the alleles at 25 loci, with a population distribution at each locus of 50%-50% for the alleles determining presence or absence of one unit, the expected distribution of the character, Binomial (25, 1/2), would closely resemble a normal curve. But, theory aside, if clinical variables usually turn out non-normal in reality, reality must win.

A second point, made by everyone from Murphy and Feinstein to writers of laboratory manuals, is that the 95% procedure virtually guarantees that everyone has some clinical "abnormality." If multiple tests of **n** independent variables are given, each with probability 5% of a positive result, the probability of at least one positive result is as follows:

n	**Probability of at least one + (%)**
1	5
2	10
5	23
10	40
15	54
20	64
100	99
∞	100

Already at 15 tests, anyone is more likely than not to have a test result outside the reference range, and at 20 tests the probability is nearly 2/3. And if we regard the 5% on each test as abnormal, it follows, Murphy notes, that "a normal person is anyone who has not been sufficiently investigated" (123). Since he regards this conclusion as "patently absurd," he takes the computation as a reason not to regard the reference range as a normal range. But I disagree with his judgment of the conclusion: far from absurdity, it is an important truth. To a pathologist, no one is entirely normal, if only because nearly everyone has some skin lesion or other (cut, bruise, mole, insect bite, etc.), and atherosclerosis begins in childhood. Moreover, the skin lesions are all visible, so nearly everyone has some clinical disease as well.

Rather, the main objections to calling a reference range "normal" are two features inherent in the methodology used to derive it. First, it was derived from an apparently healthy population, with sick individuals excluded. But the method guarantees that 5% of that population will become "abnormal" at the end, seemingly contradicting the original assumption. Feinstein[14] writes:

> Having been selected for their normality, the people might be expected to remain "normal" after the numerical analyses, but the statistical procedure is remorseless. No matter how medically "normal" the people may have been, 5% of them must emerge as "abnormal" after the statistical partitions. (246)

Of course, one can reasonably assume one's initial judgments of health to be imperfect, with some seemingly healthy persons harboring hidden disease. As Ravel[17] says,

> A small but definite group of clinically normal persons may have subclinical or undetected disease and may be inadvertently included in the supposedly normal group used to establish normal values. (4)

But there is certainly no general reason to assume that these hidden sick constitute 5% of the population rather than, say, 0.1%, 1%, 10%, or 25%. So the second objection seems conclusive: choosing a reference range to exclude a symmetrical 5% is wholly arbitrary. *[22]* Feinstein writes:

> The usual statistical partition of the zone called the "range of normal" depends on three arbitrary judgments about proportions, location, and symmetry. With the first judgment, we decide that 1 in every 20 values ([5%] of the total array) are sufficiently uncommon to be regarded as "abnormal," i.e., beyond the "normal" zone. In the second judgment, we decide that

> this 95% zone of normality will be located in the central portion of the ranked array of numbers, rather than at one of the extremes. In the third judgment, we decide to place the 95% in the exact center of the distribution, so that the remaining 5% of the values are divided symmetrically, with 2.5% at one end and 2.5% at the other.
>
> These judgments are completely arbitrary. As Murphy points out, the strategy "contrary to popular opinion ... is *not* a recommendation of statisticians, and ... has no support from statistical theory." *[23]*

Feinstein goes on to claim that 95% limits in laboratory medicine arose from their common use in hypothesis testing and estimation. Now even if we waive all objections to classical statistics, in significance testing the demand for "p-values" < .05, as opposed to, say, < .25 or < .001, is a primitive effort to weigh the relative significance of Type-I and Type-II errors. Even so, no such pseudo-decision-theoretic analysis carries over to the clinical use of reference ranges as bounds of "normality." Ravel is right to say that 5% is just

> a deliberate compromise. A wider normal range (e.g., ± 3 SD) would ensure that almost all normal persons would be included within normal range limits and thus would increase the specificity of abnormal results. However, this would place additional diseased persons with relatively small test abnormality into the expanded normal range and thereby decrease test sensitivity for detection of disease. (4) *[24]*

But there is no principled reason for choosing 5%, or any other number, as best *in general*, so the compromise, however deliberate, is baseless.

In sum, the Gaussian distribution determines no "normal range" of any clinical importance. The misconception that it does may stem from a simple semantic confusion. The bell curve was dubbed "normal" by Galton because he believed it typical of biological distributions. Thus, it was the whole distribution that was supposedly normal, not any particular segment of its values. And insofar as it is the whole distribution that is typical, its extreme values are, of course, just as typical as its central ones. In any case, the moral of this section is that there is no reason to believe that a clinical measurement outside a standard reference range must be pathological in itself, or even good evidence for any pathological condition. On the contrary, not only can values outside the range be fully consistent with complete health, but reference-range methodology is almost certain to place some percent of the population in this category. I will let Robert Galen[18] have the last word:

> It is impossible ... to evaluate a data set of reference values and select a suitable reference interval that will be meaningful for the practice of medicine. The reference interval, no matter how derived statistically, tells us nothing about disease. (861)

While this is all we need for our discussion, before leaving the subject we should mention a different method – the one for which Galen and others crusaded – for fixing diagnostic ranges. For any test, whatever proportion of the population it makes positive, one can calculate its *sensitivity* and *specificity* in detecting some disease if one has some other, "gold-standard" way to diagnose that disease. The sensitivity of a test for disease D is its conditional probability of detecting D when D is present. In probability terms, Sens = Pr(+|D). Its specificity is its conditional probability of rejecting D when D is absent, or Spec = Pr(−|~D). Usually, there is an empirical tradeoff between the two virtues: a more sensitive test has more true positives, but also tends to have more false ones, hence lower specificity. At any rate, given the prevalence of D in the population, since prevalence is the prior probability Pr(D) that the patient has D, we can use Bayes' Theorem with Sens and Spec to calculate the predictive value of the test. The predictive value of a positive is just Pr(D|+), and the predictive value of a negative is Pr(~D|−). All this information is crucial for clinicians. For example, studies show that all too few doctors realize how prevalence affects the predictive value of a positive result. Even for tests of high sensitivity and specificity, a patient with a positive result is very unlikely to have a rare disease. As Ravel notes, if D's prevalence is 0.1%, or 1 in 1000 persons, even a test with sensitivity and specificity 95% each -- very good -- misclassifies 49 of every 50 patients with + results. But a positive

result for a similarly good test for a disease with prevalence 1%, 5%, or 10% would have predictive values 16%, 50%, and 68% respectively (1-2).

It is, of course, possible to design the reference limit of a test so as to achieve desired levels of sensitivity or specificity in detecting a particular disease. In this context, a positive test result does give probability information about disease. Moreover, genuine decision-theoretic calculations can then be made about the effect of adopting one or another reference limit as a trigger for clinical actions, *i*.e., further diagnostic tests or therapy. For example, the effects of different PSA levels as criteria for ordering prostate biopsies have been carefully evaluated. We need only note that limits obtained in this way are best described as *clinical decision limits* (Jacobs *et al.*[15] 16) and are very different from the kind of standard reference limits reported on lab tests, which come from the 95% method.

III. THE LINK BETWEEN CLINICAL VARIABLE AND PHYSIOLOGICAL FUNCTION

Clinical tests yield data observable in the office, diagnostic unit, or laboratory. But their sole interest lies in what evidence such data offer about health and disease, *i*.e., physiological function or dysfunction. Usually, the latter are internal processes not directly observable, so the inference from clinical test to underlying normal or abnormal function is indirect, and in some cases tortuous. This section offers a small zoo of examples *[25]* meant to stress the distinction between "normality" of a clinical test (which we have just seen is a confusing term) and truly medical normality: healthy function.

A very few clinical tests measure physiological function directly. Bleeding time is simply how long it takes for a clot to form in a test wound. Endoscopy can show patency or blockage of a tube like the colon, pylorus, oviduct, or bile duct. So, by the same token, dysfunction or disease can be directly observable. Further familiar examples are a stopped heart, lack of respiration, or unconsciousness. Biopsies can directly show cancer cells. And some test results show a normal reaction to disease: e.g. the presence of antibodies in the blood to, say, hepatitis B virus. Other tests are, at least, nearly direct measures of function or dysfunction. Blood pressure, observed in the usual way, is a precondition for circulation, the function of the heartbeat. Imaging techniques, like ultrasound, X-rays or MRIs, offer visual evidence mediated by technology. By these means one can watch the heart pumping, the blood flowing, or the air sacs of the lung filling with air.

But most clinical tests, especially blood tests, are not like this at all. To begin with, a test may measure an item that has a function, but not for the purpose of evaluating how well it is performing it. Information about organ damage is conveyed by "liver enzymes" such as alanine transaminase, aspartate transaminase (AST), and alkaline phosphatase, or "cardiac markers" such as troponin, AST, creatine kinase, and lactic dehydrogenase. Such tests are not necessarily tests of function, either of the organ itself or of the enzyme. Their usual point is only that elevated blood levels of the enzyme can indicate leakage from disease in the organ. It is not that there is too much of the item, but that it is in the wrong place. Partly similar is prostate-specific antigen (PSA): it has a reproductive function, but one completely irrelevant to PSA tests, which aim to detect leakage from, e.g., infection, or overproduction by a tumor. *[26]* Another common situation is excess of a precursor. For example, the free erythrocyte protoporphyrin test measures a biochemical, protoporphyrin, which is used in conjunction with iron in the last step of heme synthesis. Elevated levels therefore indicate iron deficiency. Thus, while protoporphyrin has a function, its elevation neither is nor causes a dysfunction; it merely indicates a dysfunction in iron balance. *[27]*

On the other hand, a great many clinical tests measure items that have no function at all, being only waste products of functional ones. Metabolites of a hormone may be used to assess the production of the hormone itself, as with 17-hydroxycorticosteroids, which are metabolites of glucocorticoids and so can be used to measure adrenal cortical function. An unusually high level of a metabolite of a hormone can also signal its overproduction by a tumor, as when an excess of 5-hydroxyindoleacetic acid reveals a serotonin-producing carcinoid. These metabolites, though clinically useful, have no function. In other cases, the level of a waste product in blood or urine is used to evaluate the function of the organ that degrades or excretes it. Thus the blood level of creatinine, a waste product of creatine phosphate, is used to test kidney function. Since the liver normally converts ammonia into urea for excretion by the kidneys, the blood level of ammonia is an indication of liver function and the blood level of urea nitrogen an indication of kidney function. Again, neither of these substances itself has any function.

In sum, most clinical data are not direct, or even fairly direct, measures of physiological function. In many cases, the item measured has no function; in other cases, even though the item has a function, the test measures item but not function. This is a second reason, independent of those in §II, to be careful not to confuse a test result outside a reference range with a pathological condition. In uniquely medical terminology, there really is no such thing as clinical normality, except in the sense of the lack of clinically detectable pathology. Clinical testing per se defines no true medical normality, both because "abnormal" results are consistent with perfect health (§II) and because clinical tests are usually only indirect tests of physiological function (§III).

IV. SOME QUIRKS OF TRUE CLINICAL NORMALITY

Insofar as medicine is a practical discipline, the physician's ultimate interest is not in mere pathology, but in pathology that merits treatment. This is a version of the category that I called "therapeutic abnormality" in my "grades of health"[4] (365). But pathology can rarely be treated unless it can be detected. So treatable disease is usually a subset *[28]* of a wider category of clinical disease, which I termed "diagnostic abnormality" (*ibid.*): a pathological condition that is diagnosed or diagnosable. Steven Tresker[19] has recently explored this concept -- using actual, not potential, diagnosis as his criterion -- and offers an interesting taxonomy of the broadest class of clinical conditions.[20] In this section, I add to his discussion a few interesting features of clinical abnormality. First, to be clinically evident often means that a disease process is more severe, but sometimes that it is less severe, and sometimes it has no relation to severity at all. Second, clinical normality is individual-relative: the same condition is evident in one patient but not another, often for extraneous reasons. And third, for this reason and others, on at least one version of clinical normality, it is -- like therapeutic normality but unlike theoretical normality -- a value-laden notion.

A. Clinical Evidence as a Measure of Severity

1. Clinical disease as more severe. Very often, becoming clinically evident marks a grade in the progress of a disease process. This is the usual situation with both tumors and chronic disease, not to mention most infections. In cancer, pancreatic tumors have a very low survival rate precisely because they produce only vague and nonspecific symptoms until the point where they are incurable. To a lesser degree, the same is true of lung cancer, though here the ready availability of X-ray evidence offers greater hope of early diagnosis and cure. But the same pattern also fits many chronic diseases, including those I have elsewhere discussed.[21] Coronary atherosclerosis typically must progress for a long time before it produces symptoms like chest pain, and is often wholly asymptomatic until a fatal heart attack. Emphysema is not clinically evident until at least a third of lung parenchyma has been destroyed, and signs of type-1 diabetes only appear when 90% of pancreatic-islet β cells are gone.

2. Clinical disease as less severe. On the other hand, the opposite situation can occur. A number of diseases have what one might call "herald signs," which are visible at the earliest stage, but then disappear as the disease becomes more entrenched and therefore more severe. The chancre of primary syphilis is a familiar example. At that stage the infection is still localized and highly treatable, whereas the later stage of latent syphilis is occult but more serious. Another example is the bull's-eye rash (*erythema migrans*) of acute Lyme disease, which is much easier to treat than the chronic, often clinically occult version of disseminated infection.

3. Clinical disease as unrelated to severity. Finally, being clinically evident sometimes has no implications for severity of disease. That is usually the case in dermatology, where most pathology is easily visible since the skin is the organism's surface. An obvious lesion may be no more significant than a less obvious one, either between diseases or within a single category. A common wart and a melanoma may be equally visible, but the latter is far more serious than the former. Nor is a bigger, more obvious wart a graver disease than a smaller one, since common warts are harmless regardless of size. The same could be true even for visible vs. invisible lesions: pityriasis versicolor may be either visible to the naked eye under a summer tan or invisible in winter pallor, but it is of no real importance in either case. (For brief discussion and references on these and a few other kinds of dermatopathology, see Boorse[21].)

B. Clinical Normality as Individual-Relative

Often, the identical condition is diagnosable in one patient but not another. One reason, of course, is patients' differing attitudes toward diagnostic testing, which falls under my

value-ladenness point C below. But irrespective of values, many diseases are diagnosed incidentally, by a clinical test aimed at unrelated phenomena. Tumors discovered in this way are often called "incidentalomas." The diagnostic route to such findings can be infinitely varied. A search on "incidental diagnosis" yields, among many others, the following four stories. (i) A pregnant woman's pancreatic cancer was discovered during a growth scan of her fetus. In this case, her pregnancy probably saved her life.[22] (ii) An asymptomatic, healthy 16-year-old girl, being pre-evaluated for sports, got an echocardiogram that showed a congenital teratoma in her pericardium.[23] Not only tumors, of course, but many other diseases can be accidentally diagnosed. (iii) A middle-aged man being evaluated after falling off his bicycle was found by X-ray to have a giant hydronephrosis.[24] (iv) A young woman's tetralogy of Fallot, a congenital heart malformation, was revealed by her chest X-rays following a car accident.[25] And (v) in 1950, a little boy named Christy was hosting a sizeable tapeworm, which made its appearance after the ether used in his tonsillectomy killed it. *[29]* In that case it was treatment, not diagnosis, that revealed the pathological condition.

C. Clinical Normality as Value-Laden

There are two obvious concepts of clinical pathology: pathology that has been diagnosed, or that could be. Tresker[19] chooses the first alternative, using "clinical disease" to mean "diagnosed theoretical disease" (3). This definition itself, like my analysis of theoretical disease, is value-free, since whether a specific patient has received a diagnosis is just a historical fact. However, Tresker notes that whether or not a theoretical disease gets diagnosed is the causal result of value judgments.

> Diagnosis can reflect a patient's or clinician's values because some tests (say, a prostate-specific antigen test) will not be ordered for a patient for whom the benefit-cost ratio is not favorable. Similarly, diagnostic criteria for some diseases may not include a certain test if that test is too expensive, not sensitive enough, or for any number of other reasons. Thus values determine whether a person has a BST-based clinical disease because they inform whether a diagnostic test should be performed, as well as informing or determining the criteria constituting the diagnosis itself. (4)

On the other hand, if one uses "clinical disease" to mean "diagnosable theoretical disease," the same point recurs on the conceptual level. At first sight, one might think there is a value-free notion of diagnosability: being technically detectable by the entire range of available diagnostic procedures. But this notion is irrelevant to medical practice. No doubt nearly all disease is detectable by first killing the patient and then doing an exhaustive autopsy. But this procedure, of course, is pointless and flagrantly violates medical ethics. Less obviously, so does indiscriminate use of diagnostic procedures on a living patient. Naturally, a doctor cannot greet each new patient by ordering every conceivable biopsy, cutting holes in each of his parts. But the same is true even of tests as benign as blood samples, X-rays, and so on. Like treatment, all diagnostic tests are actions on patients; all actions on patients are subject to value judgments; and nearly all diagnostic tests have costs as well as possible benefits, whether in money, risk, pain, or simply time. In fact, as we have learned from the now-extensive literature on overdiagnosis *[30]*, even true diagnoses can be worse than ignorance. For all these reasons, all diagnostic testing requires justification, making diagnosability a value-laden concept. So either causally, on Tresker's actual-diagnosis concept, or conceptually, on the diagnosability alternative, values partly determine what is a clinical disease.

V. CONCLUSION

We began by analyzing the meanings of 'normal', past and present. Legitimate uses of the term, I argued, now always mean "typical" in some sense: either typical, or at least typical, or at most typical. A special medical sense of abnormality is "pathological," which I have long held means a less-than-typical level of part-functional ability. This usage is not invariable even in medicine – after all, as Rautmann noted, we speak of the "normal course" of a disease (Vacha[26] 343). But to confuse purely statistical with distinctively medical senses of normality is a constant danger, and nowhere more so than in clinical testing. We saw that values outside the conventional 95% central reference range formerly called "normal" in no way entail pathology, for three separate reasons: (i) the whole distribution is derived from a seemingly healthy population, (ii) the $\mu \pm 2\sigma$ limits are wholly arbitrary, and (iii) clinical variables often have only a very indirect relation to underlying biological

function. Finally, true clinical normality – absence of a diagnosed or diagnosable disease – has several interesting features. That a disease is clinically evident can mark greater severity, but it may also mean lesser severity or nothing at all. Also, the same condition may be clinically evident in one patient but not another, for extraneous reasons; and therefore, unlike theoretical normality, clinical normality is either causally or conceptually tied to value judgments.

ENDNOTES

[1] 2nd ed. 145. The table originated in Murphy[27], 33. Vacha[8] (927-29) offers a similar but slightly different list of seven supposed senses to show the "semantic chaos" afflicting the term. I believe my discussion of Murphy's list, and of function, will suffice for Vacha as well.

[2] This dictionary also lists a third general sense – "having the function of prescribing a course of action or way of living; prescriptive" – but calls it obsolete. And it claims that 'normal' can mean "heterosexual", but that is just an obvious contextual limitation of one of the first two general meanings.

[3] In accord is Corina Strößner[28]: "The view that normality is essentially normative is a misunderstanding ... caused by the appearance of the words. ... [T]here are many normality statements which are clearly not normative" (795).

[4] On my side are both Strößner, just quoted in note 3, and Lara Huber[29]: "It is commonly acknowledged ... that scientific accounts of the 'normal' often are misunderstood as bearing normative conclusions simultaneously" (41).

[5] The *OED* reports that past meteorologists have used 'normal' as a synonym for 'mean', but calls this usage "rare." Unfortunately, it is very common among American TV weatherpeople, though in my view clearly wrong.

[6] In this paper, I will assume my own account of health and disease. For a general survey of philosophical analyses of health and disease, see Boorse.[30]

[7] Similarly, one might defend Murphy's sense 5 ("innocuous or harmless") -- which, as Jeff Jordan pointed out to me, is very common among clinicians -- as meaning normal *in risk*, that is, carrying no unusual penalties.

[8] Entry *norma*. The adjective *normalis*, however, still had only the literal meaning "made according to the square." A similar pattern is seen with *norma*'s geometric relatives *regula* (ruler) and *rectus* (straight), except that these words seem to have progressed further to the dead-metaphor stage.

[9] Fricke[9] (699) cites this judgment to Rolf[31], 24.

[10] Less floridly, Vacha[8] concurs: "Normality is the elusive Proteus whose countenance changes according to biological branch and philosophical conviction of authors" (923).

[11] 184. Hacking adds "and carried all the Comtian baggage with it." But I do not see how a word meaning "typical" can entail anything Comtian, or indeed anything that 'typical' does not itself entail.

[12] Another rich historical discussion is Vacha's analysis[26] of early-20th-century German writers on normality. Vacha mentions four advocates of a purely statistical concept (Rautmann, Bauer, Borchardt, Günther), who disagree somewhat over both the details of normal limits and their relevance to health. Also present were advocates of ideal norms (Hildebrandt) and of individual, person-relative ones (Grote). I omit these writers here for two reasons. First, Vacha judges (352-3) that they had almost no influence on current ideas of normality in world biomedicine, which stem instead from Galton and from English clinical biochemists. In the second place, my discussions of Hacking's writers, and of reference ranges in §II, make it clear enough what I would say about the pre-WW2 German debates.

I should also mention that I have not consulted Vacha[32] or Rolf[31], since they are not available in English translation.

[13] I thank Mark Greene for this example.

[14] Demonstratives like 'here' and 'this season' change their referents with speaker, time, and place, but presumably not their sense; and it is the sense of 'normal' that we are trying to analyze.

[15] Since, in relevant biological usage, functions always favor the organism's survival or reproduction, there is no such thing as excessive

function. As philosophers stress, 'function' does not mean just "effect" or "output": thyroid function is not just thyroxine secretion, but secretion of the right amount of it to regulate metabolism. For a thorough survey of philosophical analyses of biological function, see Garson.[33]

[16] An alternate view would be that 'normal' always means just "typical," but that when people say "normal," often what they really mean is "at least normal" or "at most normal." This view breaks the link between the general and uniquely medical senses, but otherwise seems not to affect my conclusions in this essay.

[17] In this usage 'normal' covers not just the middle range, but, contrary to Fricke[9] (700), both middle and upper ranges.

[18] Alternatively, if my thesis (iii) is wrong, I would argue that 'normal' sometimes means "good [or acceptable] *because* typical." It would then be what I call a "conclusory predicate," one that summarizes an argument.

[19] The ideas in this section are neither original nor, as far as I know, controversial. Except as otherwise cited, almost everything in my discussion comes from standard medical reference books, such as Jacobs *et al.*[15], Henry[34], and Ravel[17]. Summary handbooks on clinical tests include Pagana and Pagana[35], Wilson[16], and Sacher and McPherson[36]. Besides Murphy's *Logic of Medicine*[1], an invaluable collection of essays on statistical problems in medicine is Feinstein[14].

[20] If a full 5% abnormal range is desired, the interval must be $\mu \pm 1.96\sigma$.

[21] Feinstein's view[14] (248-9) is that the percentile method is preferable to the transformation method for non-normal distributions, since the required transformations are often biologically meaningless and there is no theoretical reason in the first place to expect Gaussian distributions or, except for ease of computation, to prefer them. For normal data, the percentile method automatically gives the same reference range.

[22] In particular, the choice of $\mu \pm 2\sigma$ as the boundaries of a region of the normal curve is mathematically arbitrary; these points have no mathematical significance that I know of. At $x = \mu$, of course, the function has a maximum with its first derivative being 0, while $x = \mu \pm \sigma$ are inflection points (changes of concavity) with the second derivative being 0. But, e.g., the points where the third derivative is 0 are μ and $\mu \pm \sigma\sqrt{3}$, not $\mu \pm 2\sigma$.

[23] Feinstein[14] 245-6, quoting Murphy[37]. Murphy had earlier called the idea of a "normal range" a "particularly distasteful error in which the statistician has somehow allowed himself to be embroiled."

[24] Note how Ravel shifts from one sense of 'normal' to another within the quote. (The second 'normal' means "healthy"; all the others mean "within the reference range.") His discussion is a good example of how confusing the older terminology was.

[25] Information on these examples and countless others can be found in any handbook of clinical tests, such as the ones listed above in note 19.

[26] In many other cases, of course, a test indicates both a dysfunctional level of some functional entity and some other dysfunction that is its cause. Thus, various tumors lead to measurable overproduction of hormones like ADH or calcitonin. This overproduction may have significant clinical effects, but the main point of the test is to find the cancer, not to evaluate hormone function. Another interesting example is haptoglobin, a protein produced in the liver which has the function of returning free hemoglobin to the liver for recycling. A low haptoglobin level might indicate liver dysfunction; or it might be the result of any condition which destroys red blood cells quickly. In the latter case, while the haptoglobin level is now too low for normal functional readiness, that is merely because it has been exhausted by the high demand for it from another disease process, e.g., hemolytic anemia.

[27] In my terminology[5] (387 n 14), it is neither pathologic nor pathogenic, but at best pathodictic.

[28] It is a proper subset, since – despite so many philosophers' neglect of this simple point -- not every known disease state should be treated.

[29] Author's personal memory. The surgeon heartlessly threw my dead worm away, although it obviously belonged to me.

[30] See Welch *et al.*[38].

REFERENCES

1 Murphy E A. *The Logic of Medicine*. Baltimore: Johns Hopkins; 1976. Second ed. 1997.

2 *Oxford English Dictionary, Website 2000*, entry 'normal'. Accessed Dec. 23, 2021.

3 Boorse C. Health as a theoretical concept. *Philosophy of Science*. 1977;44:542-73.

4 **----------**. A rebuttal on health. In Humber JM, Almeder RF, *What is Disease?* Totowa, NJ. Humana Press; 1997:1-134.

5 **----------**. Concepts of health. In D VanDeVeer, T Regan, ed. *Health Care Ethics: An Introduction*. Philadelphia. Temple University Press; 1987:359-93.

6 **----------**. A second rebuttal on health. *Journal of Medicine and Philosophy* 2014;39:683-724.

7 Hacking I. *The Taming of Chance*. NY: Cambridge University Press; 1990.

8 Vacha J. Biology and the problem of normality. *Scientia* 1978;113:823-46.

9 Fricke C. Questioning the importance of being normal – an inquiry into the normative constraints of normality. *Journal of Value Inquiry* 2015;49:691-713.

10 Lewis CL, Short C. *A Latin Dictionary*. London: Oxford University Press; 1879.

11 Comte A. *Système de Politique Positive*. Paris; 1851.

12 Sober E. Evolution, population thinking, and essentialism. *Philosophy of Science* 1980;47:350-83.

13 **----------**. *The Nature of Selection*. Cambridge, MA: MIT Press; 1984.

14 Feinstein AR. *Clinical Biostatistics*. St. Louis: C.V.Mosby; 1977.

15 Jacobs DS, DeMott WR, Oxley, DK, ed. *Jacobs & DeMott Laboratory Test Handbook*. 5th ed. Hudson, OH: Lexi-Comp, Inc.; 2001.

16 Wilson DD. *Manual of Laboratory & Diagnostic Tests*. NY: McGraw-Hill; 2008.

17 Ravel R. *Clinical Laboratory Medicine: Clinical Application of Laboratory Data*. 5th ed. Chicago: Year Book Medical Publishers; 1989.

18 Galen, RS. Predictive value and efficiency of laboratory testing. *Pediatric Clin North Am* 1980;27(4):861-9.

19 Tresker S. Theoretical and clinical disease and the biostatistical theory. *Studies in History and Philosophy of Biomedical Sciences* 2020;82:1-11.

20 **----------**. A typology of clinical conditions. *Studies in History and Philosophy of Biological and Biomedical Sciences* 2020;83:e1-e9.

21 Boorse C. Wakefield's harm-based critique of the biostatistical theory. *Journal of Medicine and Philosophy* forthcoming.

22 Quaresima P *et al*. Incidental diagnosis of a pancreatic adenocarcinoma in a woman affected by gestational diabetes mellitus: case report and literature review. *Am J Obstet Gynecol MFM* 2021 Nov;3(6):100471

23 Garcia AM *et al*. Late incidental diagnosis of intrapericardial teratoma – A case report. *Revista Portuguesa de Cardiologia* 2017;36(7-8):571e1-571.e4.

24 Faeez MS, Abdul Rashid AM. 'Accidental' diagnosis of a silent giant hydronephrosis. *Journal of Urological Surgery* 2018;5(3):206-08.

25 Verdoia M *et al*. Incidental diagnosis after a car accident. *JACC Case Reports* (2020); 2(15):2289-94.

26 Vacha J. German constitutional doctrine in the 1920s and 1930s and pitfalls of the contemporary conception of normality in biology and medicine. *Journal of Medicine and Philosophy* 1985;10:339-67.

27 Murphy EA. A scientific viewpoint on normalcy. *Perspectives in Biology and Medicine* 1966;9:333-48.

28 Strößner C. (2015). Normality and majority: towards a statistical understanding of normality statements. *Erkenntnis* 2015;80:793-809.

29 Huber L. Norming normality: on scientific fictions and canonical visualisations. *Medicine Studies* 2011;3:41-52.

30 Boorse C. Concepts of health and disease. In: Gifford F, ed., *Philosophy of Medicine*, vol. 16 of the *Handbook of Philosophy of Science*. Amsterdam: Elsevier; 2011:13-64.

31 Rolf T. *Normalität: Ein philosophischer Grundbegriff des 20. Jahrhunderts*. Munich: Wilhelm Fink Verlag; 1999.

32 Vacha J. *The Problem of Normality in Biology and Medicine*. 2nd ed. (In Czech). Prague: Avicenum; 1983.

33 Garson J. *A Critical Overview of Biological Functions*. Springer; 2016.

34 Henry JB. *Clinical Diagnosis and Management by Laboratory Methods*. Philadelphia: W.B.Saunders; 2001.

35 Pagana KD, Pagana TJ. *Mosby's Manual of Diagnostic and Laboratory Tests*. St. Louis, MO: Mosby; 1998.

36 Sacher RA, McPherson RA. *Widmann's Clinical Interpretation of Laboratory Tests*. 11th ed. Philadelphia: F.A. Davis; 2000.

37 Murphy EA. The normal, and the perils of the sylleptic argument. *Perspectives in Biology and Medicine* 1972;15:566-82.

38 Welch HG, Schwartz L, Woloshin S. *Overdiagnosed: Making People Sick in the Pursuit of Health*. NY: Random House; 2012.

.