Epidemiological Research: Terms and Concepts

O. S. Miettinen

Epidemiological Research: Terms and Concepts

Springer

Prof. Dr. O. S. Miettinen
McGill University
Cornell University
1020 Pine Avenue West
Montreal, QC H3A 1A2
Canada
olli.miettinen@mcgill.ca

ISBN 978-94-007-1170-9 e-ISBN 978-94-007-1171-6
DOI 10.1007/978-94-007-1171-6
Springer Dordrecht Heidelberg London New York

Library of Congress Control Number: 2011925997

Preface

In his *The Joys of Yiddish* (1970), Leo Rosten delights the reader with this vignette:

> A bright young *chachem* told his grandmother that he was going to be a Doctor of Philosophy. She smiled proudly: 'Wonderful. But what kind of disease is philosophy?'

The reader here might wonder what a chachem is, but not after reading Rosten: it's a great scholar, a clever and wise learned person. On the other hand, though, the reader here, like that grandmother, likely presumes to know what a *doctor* and a *disease* are; but (s)he might do well consulting a suitable source on this too – a contemporary medical dictionary won't do (cf. sect. I – 1. 1 here) – insofar as (s)he is enough of a chachem to really care about the meanings (s)he associates with words.

According to *The American Heritage Dictionary of the English Language*, a dictionary is, in one of the four meanings of the word, "A book listing words or other linguistic items in a particular category of subject with specialized information about them: *a medical dictionary*." The listing generally is alphabetical in its ordering.

This book is, mainly, a dictionary in that meaning, with *terms* of American English – and some of their initials-based abbreviations also – addressed in alphabetical order; but it also is designed – by the proposed selection from among the terms and the proposed ordering of these – to function as a textbook in an introductory course on epidemiological research (see Introduction below). A term is, in logic and also in that word's usage here, a word or a composite of words – such as 'epidemiological research' – that can be the subject or predicate of a proposition [1].

The "specialized information" that this book gives about the terms it covers is their meanings, that is, the *concepts* to which they refer. This information is, in part, merely descriptive of prevailing terms and the concepts to which they refer; but it commonly also is quasi-prescriptive, conveying my opinion of what the term or the concept, or both, ought to be. A concept is the essence of a thing (entity, quality/quantity, relation); it is true of every instance of the thing and unique to it [1].

A concept is specified by its *definition*, which – in an ideal definition at least – posits the concept's "proximate genus" together with its "specific difference" within this genus [1]. For epidemiological research the proximate genus obviously is

research; but the specific difference – true of every instance of this research and unique to this particular genre of research – scarcely is a matter of shared understanding among, even, teachers of this research. In this book, many of the definitions are supplemented by explications of their meanings, as I see them – extensive explications, even.

With that "medical dictionary" term as a paradigm, this book can be said to be, in part, an epidemiological-research dictionary. This research is mostly about the causal origin – *etiology*, *etiogenesis* – of illness, knowledge of which is seen to be critically important for the practice of epidemiology – meaning, of community medicine, community-level (rather than clinical) preventive medicine. For another part, this book is a dictionary of related, meta-epidemiological *clinical research*, for clinical medicine. Both of these address, mostly, *rates* of the occurrence of events and states of ill-health/illness, and therefore clinical researchers, too, now tend to develop their concepts and terms – and principles, even – largely by studying epidemiological research.

Etiologic research has undergone enormous growth in the most recent half-century; and along with this there has been major development also of the very concepts and terms of epidemiological research. Around 1960 it was, still, conceptually unsurprising for an authoritative "committee on design and analysis of studies" [2] to declare that "The ideal epidemiological study would be based on probability samples from a very large population . . . ," even though no-one conducting 'bench research' was dreaming of probability samples of very large populations of particular types of rodents; and it was both conceptually and linguistically unsurprising that a major authority on the theory of epidemiological research classified etiologic studies in the broadest categories of "descriptive" and "analytic" studies [3].

To help epidemiologists keep up with the conceptual and linguistic – English-language – developments concerning epidemiological research, the International Epidemiological Association has been sponsoring *A Dictionary of Epidemiology*, which now is in its fifth edition [4]; but I consider this program to have been less successful than it really could, and should, have been [5, 6]. The examples below should suffice to justify this judgment.

The I.E.A. dictionary defines *etiology* as: "Literally, the science of causes, causality; in common usage, cause." This is all that it says about this central concept in epidemiological research. But if etiology indeed were, "literally," the science of causes (in truth there is no such science, just as morphology, e.g., is not a science), then, for example, tautology would be, literally, the science of unnecessary repetition [7]; and if any given cause of an illness were, in common medical conception, an "etiology" of that illness, then the word 'etiology' would have no justification for inclusion in the lexicons of medicine (in addition to that less abstruse term 'cause'). But the medical concept of etiology actually is, as I let on above, that of causal origin – etiogenesis – of illness [8]; and in respect to any given cause, etiology thus is the role of this cause in the initiation and/or advancement of the genesis of the illness.

That dictionary describes *etiologic study* as being of three common types: "cross-sectional, cohort, and case-control" (*sic*). To the respective definitions of these are associated quite extensive annotations; but none of these is about the fundamental misconceptions that those types of study represent, nor about the linguistic awkwardness of that triad of terms; and there are no words about the necessary, singular essence of etiologic studies that is dictated by logic [7, 9].

Rate – a concept as central as any to etiologic research as well as to the practice of epidemiology (as community medicine) – that dictionary defines as "an expression of the frequency with which an event [*sic*] occurs in a defined population," as though there were no rates for the occurrence of states of illness; that is, it defines rate as though there were no rates of prevalence (of states) in addition to rates of incidence (of events). And indeed, concerning prevalence rate the dictionary declares: "It is a proportion, not a rate." Yet the dictionary presents without any critical annotation "adult literacy rate," "response rate," and "survival rate," for example, each of them a proportion. (Cf., e.g., 'unemployment rate' in general English.)

Such confusion in this context is compounded by faltering logic. For example, while the dictionary says that "All rates are ratios, calculated by dividing a numerator ... by a denominator ... ," the numerator and denominator are said to be "components of a rate" rather than inputs into its calculation; and yet another "component" is said to be "the specified time in which events occur," as though this were not already involved in the inputs into the rate's calculation (insofar as it actually is involved at all; in proportion-type rates it isn't).

Illness – a recurrent concept and term in the foregoing, already, as it too is very central – the I.E.A. dictionary addresses under "disease" only. What it there says does not reflect familiarity with these two concepts and terms in medicine, including awareness of the confusion about them that now prevails even in eminent, "authoritative" dictionaries of medicine [10, 11].

Illness, I propose, should be understood to be any ill-health, and the term thus to be the English-language counterpart of *Krankheit* in German and *maladie* in French. Disease (L. *morbus*) in the medical usage of the term is but one of the principal subtypes of illness, the others being defect (L. *vitium*) and injury (Gr. *trauma*) – as has been explained elsewhere [10, 11], without any objection from the I.E.A. or any other source.

Epidemiological research is, by the very nature of its objects, *statistical* research; and as such, it commonly is testing of etiologic hypotheses. The central statistic now derived from the data of such studies is the null P-value. The I.E.A. dictionary defines *P-value*, in the meaning of the null P-value, as: "The probability that a test statistic would be as extreme as observed or more extreme if the null hypothesis were true." This conception of the essence of the null P-value – with probability taken to be its proximate genus – is quite unfortunate, even though very common. For, it underpins the common, serious misconception that the null P-value is the probability that the 'null hypothesis' is true.

A preferable conception of the null P-value, ignored in the I.E.A. dictionary, is this [12]: a statistic so derived that on the 'null hypothesis' its distribution is uniform

in the 0-to-1 range – so that $Pr(P < \alpha) = \alpha$ – and that, in addition, on the hypothesis proper (when the 'null hypothesis' is not true) its distribution, still within the 0-to-1 range, is shifted to the left – so that $Pr(P < \alpha) > \alpha$. At issue is the very same statistic as is addressed by the I.E.A. dictionary definition, but now defined in terms of its intended distribution, analogously with the way a 'confidence interval' generally is conceptualized and defined. With this definition of the null P-value there is no propensity to think of it as the probability that the 'null hypothesis' is true.

When a dictionary of epidemiology is I.E.A.-sponsored, it shouldn't commonly be questionable or even definitely wrong in its contents, the central ones in particular; for this does not promote sound development of the theory – concepts, principles, and terminology – of epidemiological research [5, 6]. In fact, any dictionary of epidemiology that claims to be authoritative – as the I.E.A. dictionary has done, its current Editor describing it as representing "a high level of scientific and intellectual rigor" (Preface) – may even retard progress. It may stifle critical reflection on the terms and concepts it presents.

For optimal development of the terms and concepts specific to epidemiological *research*, needed is, first, an alternative dictionary, periodically updated, in which the content is presented in the spirit of *propositions*, for the community of epidemiological researchers not simply to believe (or contradict) but to "weigh and consider" – the readers thus heeding an important precept of Francis Bacon on how to read [13]. I compiled this dictionary in response to that need, still otherwise unfulfilled.

The next need is for pursuit and attainment of practical *consensus* about the thus-presented terms and concepts in the community of the researchers. For, as Isaiah Berlin put it [14],

> where the concepts are firm, clear and generally accepted, and the methods of reasoning and arriving at conclusions are agreed between men (at least the majority of those who have anything to do with these matters), there and only there is it possible to construct a science, formal or empirical.

This insight applies to epidemiological research as well, even though this research does not constitute a science unto itself but is, instead, imbedded in various medical sciences – neuroscience/neurology and cardiology, for example.

With possible terms and concepts of epidemiological research put forward as propositions for the researchers to weigh and consider, they are intended to constitute a starting point for the development of consensus about the terms and concepts – through *public discourse* about them. This discourse naturally is a matter of public presentations of criticisms about the propositions, for a start; and these criticisms need to be responded to in a manner more constructive than that of the current Editor of the I.E.A. dictionary [15].

For that public discourse to really take place, extensively and in a timely fashion, there should be formed and maintained a *forum* dedicated to this, somewhere in cyberspace. At issue would be, as in this dictionary, *general terms and concepts*, exclusive of those specific to particular areas of subject-matter of the research. (The

I.E.A. dictionary addresses some of the latter, thus only nibbling at the enormous number of terms and concepts of epidemiological research across the various disciplines of medicine.)

Epidemiological research is concretely purposive – rather than merely interest-driven – when its aim is to advance the *knowledge-base* – scientific – of the practice of epidemiology – of community medicine, that is. Such research is quintessentially 'applied' [16]. It addresses rates, etiologically and otherwise; for, rates of the occurrence of illness are the objects of the practice of epidemiology: the practitioner's concerns are to know about them in the cared-for population and, then, to control them (by the means of community-level preventive medicine).

The knowledge-base of (the practice of) *clinical* medicine is not about rates but about probabilities [16]; but research on probabilities is about (probability-implying) rates. Therefore, quintessentially 'applied' clinical research is *meta-epidemiological* in nature, and it thus has become a concern of teachers of epidemiological research [16, 17]. It is a concern in the I.E.A. dictionary, and so it also is here.

"In a man's life dreams always precede deeds. Perhaps this is because, as Goethe said, 'Our desires are presentiments of the faculties latent within us and signs of what we may be capable of doing . . . we crave for what we already secretly possess. Passionate anticipation thus changes that which is materially possible into dreamed reality' " [18].

In my life ever since medical-school graduation half-a-century ago, I've had the dream of reaching true understanding of the theory of the research that would best serve to advance the knowledge-base of medicine, of genuinely scientific medicine [16, 19]. Having devoted my entire career to this pursuit, I've been craving, all along, for access to genuinely scholarly dictionaries of medicine, both clinical and community medicine, and of directly practice-serving medical research, both clinical and community-medical research. But these dreams have not really come true, to what I'd consider a reasonable approximation.

In an effort to make the dreamed reality come closer in respect to epidemiological research and meta-epidemiological clinical research also, I now launch this dictionary as the initial step. But the reality actually will have come about only when all of the orientational issues addressed in the foregoing, and numerous others, have been brought to secure closure – tentatively at least – by public discourse among epidemiological and clinical researchers and – not to be forgotten – those whose careers are dedicated, in part at least, to the advancement and teaching of the theory of epidemiological and/or related clinical research.

In the meantime, students of epidemiological research need to critically weigh and consider the terms and concepts – or concepts and terms – of epidemiological research as these are being taught, even if the teaching be done by the most senior of epidemiologic academics (myself included [17]).

Introduction

The I.E.A. dictionary [4] addresses almost 2,000 terms and, of course, the meanings of these. Those beginning with A run from *Abatement* to *Axis*, the B-terms from *Background level*, *rate* to *Burden of disease*, the C-terms from *Calibration* to *Cyst count*, . . .

From presentations like this, so William James taught, "We carve out order by leaving the disorderly parts out." I leave out, for example, "*Age* The WHO recommends that age should be defined by completed units of time, counting the day of birth as zero"; "*Gender* 1. In grammar, . . . 2. The totality of culturally constructed . . . about males and females and sometimes their sexual orientation"; "*GDP* Gross domestic product"; "*Goal* A desired state to be achieved within a specified time"; "*Interval* The set containing all numbers between two given numbers"; "*Justice* 1. A morally defensible distribution of benefits and rewards in society. . . . 2. In law, the successful administration of the rule of law"; and "*Sex ratio* The ratio of one sex to the other."

As another means to enhance order, I organize the terms and concepts into three main parts of this book, only one of these focusing on epidemiological research proper. This expressly epidemiologic part is preceded by one with separate sections for *medicine*, *science* in general, and *statistics* as plain statistics; and this epidemiologic part is followed by a related one, specific to *meta-epidemiological clinical* research.

Moreover, in the two parts focusing on the research that is expressly at issue here, I invoke a further measure to counteract entropy – chaos – of concepts. I arrange to have three separate sections within each of them: one of them covers select introductory terms and concepts, another one is more-or-less specific to *objects* of study, and the third one is more-or-less specific to *methods* of the research.

The coverage here is not confined to what remains from the entries in the I.E.A. dictionary. But where the topic is addressed in that other dictionary as well, this is indicated by an asterisk attached to the term, for the reader to be able to know that what is presented here is a second opinion to that expressed in the I.E.A. dictionary, a second opinion that may not accord with that first opinion.

When a dictionary of epidemiological (and meta-epidemiological clinical) research is organized in this way, its use as a handbook (à la I.E.A. dictionary) can

be challenging due to uncertainty about where – in which one(s) of the sections – the term can be found. To obviate this, a single Index at the end of this book lists all of the terms in the alphabetical order, and for each of them it gives the section reference(s) for the definition (and whatever Notes may be associated with it).

Relevant, well-formed concepts and the appropriate terms to express them are the essence of the contents in an ***introductory course*** on epidemiological (and meta-epidemiological clinical) research [17]; and this leads to a particular concern in the ordering of the topics, one that is central to philosophy: Chief among Aristotle's enormous accomplishments has been seen to be the establishment of the basis for correct thinking in terms of the rules of logic (*analytika)* that also enable discourse (*logos*) to become most productive – given also the development of suitable language for use as an enabling instrument of thought with a logically ordered structure and for the expression of this. And Kant's central thesis – very productive – was that the human mind confers a structure for knowledge through the concepts ("categories") that it brings to the acquisition of experience and to learning from it. In line with these extraordinarily fruitful ideas, Robert Roberts and W. Jay Wood in their *Intellectual Virtues* (2007) make the point that, "The real goal of philosophy, perhaps unachievable but still ideal, is reduction, the derivation of all the concepts in a given field from a single, key concept."

In this book, obviously, the introductory key concepts are those of medicine, science, and statistics; and then comes the core concept, the essence of epidemiological research, which is extended to that of meta-epidemiological clinical research. In an introductory course nominally addressing epidemiological research, the teacher would do well placing this research in its context, defining all five of those key concepts; and (s)he needs to try to follow a *logical sequence* in the introduction of the concepts of epidemiological, and of meta-epidemiological clinical, research, endeavoring to 'deduce' the subordinate concepts, sequentially, from the core concept.

I offer a suggestion for the sequence of the concepts' entry into an introductory course on epidemiological research. This I do in what follows the Index: Hierarchy of Concepts.

Acknowledgements

Miguel Porta had a major role in inspiring me to take on the task of developing this first edition of a new dictionary of epidemiological research, as an alternative to the dictionary now being published, in successive editions, by the International Epidemiological Association, with Porta the current Editor [4]. I really am very grateful to him, however unwitting may have been his inspirational role [15].

Given the total agenesis of my keyboard-related skills, *Rebecca Fuhrer*, my Departmental Chairperson, arranged for the requisite help. This was provided by *Kierla Ireland* in the main, with great skill and dedication, and with good cheer to boot. These two persons' role in this book's development was critical, and I much appreciate it.

Some local colleagues read a near-final draft of this book, and I am grateful for their suggestions – those of *Igor Karp* and *James Hanley* in particular.

In due deference to a giant of "the proper study of mankind" (I. Berlin; [14]), I underscored above (in the Preface) the importance of pursuing convergence of the concepts of epidemiological research to generally agreed-upon ones, and of the use of apposite terms also. And to this end, so I said, there should be public discourse among epidemiological researchers and the theoreticians of their work. For an undelayed opener of that discourse, I asked four highly esteemed colleagues to produce commentaries on a near-final draft of this book. These colleagues were: *James Hanley* (development and teaching, especially, of the statistics of epidemiological and clinical research), *Albert Hofman* (practice and teaching of epidemiological and clinical research, and editing a journal of epidemiology), *Andrew Miles* (philosophy of public health and editing a journal on clinical medicine in public health), and *Dimitrios Trichopoulos* (practice and teaching of epidemiological research and book-editing on teaching of epidemiology). They were, of course, free to engage co-authors of their own choosings. Their commentaries presumably will appear in the fullness of time, locus yet to be determined.

Insofar as some of these colleagues will respond to this call for public discourse, they will act the way genuine scholars do act, on this level already. But more to the point, they presumably will do so upon having read the draft of this book in the spirit which Francis Bacon counseled all of us to adopt: "Read not to contradict, nor to believe, but to weigh and consider" [13]. They thus will make a notable contribution to the mission here, and for this I thank them in advance already – on behalf of epidemiological researchers at large.

Contents