III – 3. TERMS AND CONCEPTS OF OBJECTS OF STUDY

Diagnostic probability function – See 'Gnostic probability function' (Notes 2 and 5-7).

Diagnostic test's properties – The two aspects of a test's diagnostic informativeness, constituting the objects of research on the test [16]:

1. *Post-test informativeness* – Given the post-test diagnostic profile, the extent to which the test result influences the (correct) diagnostic probability.

Note 1: If this were to be an object of study, at issue would be the *likelihood ratio* for each of the test results, contrasting illness present to illness absent, specific for each of the pre-test profiles. Given the generally very great multiplicity of these LRs, study of them is prohibitively impractical; but it is unnecessary besides.

Note 2: The diagnostician need not know how to move from the pre-test probability to the post-test probability, given that the test result has been obtained. In the face of this updated diagnostic profile, (s)he merely needs to know what the (correct) post-test probability is. And to this end, needed is research on the *post-test DPF*, for the domain of the decision node about the test's use. The test result has its degree of post-test informativeness represented in this function (along with those of the pre-test indicators), but this role is no concern of the diagnostician, given the availability of the test's result.

2. *Pre-test informativeness* – In the 'decision node' about the test's use, the probability that the test result would provide for transition from an 'inconclusive' pre-test probability to a 'conclusive' post-test probability.

Note 3: DPFs are needed for situations in which diagnostic probability-setting is needed; and these are situations in which a decision about an action – diagnostic testing, treatment, or referral – is to be taken. The presenting complaint prompts a defined aggregate of fact-findings from history and physical examination, leading to the profile for *clinical* diagnosis. This diagnosis – commonly quite 'inconclusive' – is prone to raise the question about invoking a particular test (possibly a composite of component tests); and for this decision needed is knowledge about the probability that the post-test diagnosis would be 'conclusive' – high or low enough for the decision about treatment or referral.

Note 4: For a study on a test's pre-test informativeness, a set of possibly 'conclusive' ranges of post-test probability needs to be defined (for the practitioner to

O. S. Miettinen, *Epidemiological Research: Terms and Concepts*,
DOI 10.1007/978-94-007-1171-6_10, © Springer Science+Business Media B.V. 2011

be able to focus on a chosen one, or pair, of these). For any given one of these, the object of study is the probability of a given range of post-test probability as a function of the pre-test indicators. (The dependent variate is an indicator of the range, the post-test DPFs realization falling in it.)

DPF – Diagnostic probability function.

Endpoint – See section III – 4.

EPF – Etiognostic probability function.

Etiognostic probability function – See 'Gnostic probability function' (Notes 3 and 5-7).

*Explanatory trial** – See 'Explanatory versus pragmatic trial.'

Explanatory versus pragmatic* trial* – A duality in the purpose, and its corresponding intervention contrast, of a clinical trial.

1. In an *explanatory* trial the object of study is the effect of a particular *agent* of intervention (typically a medication, or a surgical alteration of structure such as installment of a coronary artery bypass graft). The causal contrast thus is, conceptually, between the presence and the absence of the agent, but operationally it is between two treatments: treatment with the agent and treatment without the agent – between *verum* and *placebo/sham* treatment. Treatment with the agent is a practical necessity (for introduction of the agent to its recipient), while identical treatment without the agent – placebo/sham treatment, that is – is necessary for prevention of confounding by extraneous (non-agent) aspects of the treatment. An added reason for this pseudo-treatment can be the need for blinding the study.

2. In a *pragmatic* trial the object of study is the effect of a particular *treatment*, without regard for what in the treatment produces the effect. The causal contrast thus is between this treatment (defined by an algorithm for the entire duration of follow-up) and an alternative for this – both of them candidates for being someone's treatment of choice. (The choice is for the recipient to make, whenever possible.)

Gnostic probability function – The generic object of quintessentially 'applied' clinical research [16].

Note 1: Corresponding to the three genera of gnosis – diagnosis, etiognosis, and prognosis – the respective GPFs are *diagnostic* probability function (DPF), *etiognostic* probability function (EPF), and *prognostic* probability function (PPF).

Note 2: A DPF addresses the probability – prevalence – of the illness at issue, in a particular domain of patient presentation (chief complaint, . . .), as a joint function of a set of diagnostic indicators (specifying the diagnostic profile, the union of the risk and manifestational subprofiles of this).

Note 3: An EPF addresses the probability of causal/etiogenetic role for an antecedent (that was present in lieu of its defined alternative) for an illness (that

is present), in a particular domain of the occurrence of the illness, as a joint function of a set of etiognostic indicators (specifying etiognostic profile, based on modifiers of the causal rate-ratio in the domain).

Note 4: A PPF may address, for a particular domain of prognostication, either the risk of a particular, still-absent illness developing; or it addresses the course of an already-existing illness – focusing on probability of sickness from the illness in the course of it, an outcome of the course, a complication of the illness, or an adverse effect of intervention on the course. With an outcome of the course of an illness a possible exception, the probability is a function of prognostic time; but it also is a function of the prognostic indicators at prognostic T_0 and the choice of intervention.

Note 5: A GPF is *scientific* if, and only if, it is rational in its (designed) form (for the domain at issue), and its content (of that form) derives from (gnostic) research.

Note 6: A scientific GPF can be, merely, the *result* of a gnostic study (original or derivative); but alternatively – and preferably – it expresses scientific *knowledge* (informed by gnostic research). GPFs of the latter type are the basis of (the practice of) truly scientific clinical medicine.

Note 7: A GPF is *quasi-scientific* if, and only if, it is scientific (rational) in form but its content represents experts' tacit knowledge without any input from gnostic research. (Cf. 'Gnostic expert paneling' in sect. III – 4.)

GPF – Gnostic probability function.

Intervention – See section I – 1. 2 and 'Outcomes research' (Note 2) in section III – 2.

Note: In the U. S. National Institutes of Health there has developed the pernicious habit of thinking about diagnostic testings as interventions [16]. (See 'Outcomes research' in sect. III – 2 and 'Screening' in sect. I – 1. 2.)

*Likelihood ratio** – See 'Diagnostic test's properties' (Note 1).

Link – Causal connection.

Note: The word, in this meaning, is commonly used by science reporters in the public media. It does not deserve to be used by researchers.

*Overparametrization/overfitting** – Concerning a regression model, the involvement of an excessive number of parameters in relation to the amount of information in the data.

Note 1: In logistic regression the here-relevant measure of the amount of information is $Np(1-p)$, where N is the number of datapoints and p is the proportion of these with $Y = 1$, and the measure of overparametrization is the number of parameters (Bs) in the model in proportion to $Np(1-p)$, this exceeding 0.05 or so.

Note 2: Given that the model (for a GPF) needs to be one with a rather large number of parameters, at issue here actually is *undersizing* of the study rather than overparametrization/overfitting of the model.

Note 3: The consequence of this undersizing is an unusual type of *bias*: relatively high values of the fitted function are biased upward, relatively low values being biased downward.

Note 4: Correction for the undersizing bias is termed *shrinkage*.

*Placebo** – See 'Explanatory versus pragmatic trial.'

Post-test informativeness/probability – See 'Diagnostic test's properties.'

PPF – Prognostic probability function.

*Pragmatic trial** – See 'Explanatory versus pragmatic trial.'

Pre-test informativeness/probability – See 'Diagnostic test's properties.'

Prognostic probability function – See 'Gnostic probability function' (Notes 4-7).

*Receiver operating characteristic curve** – "A graphic means for assessing the ability of a screening or diagnostic test to discriminate between persons with and without the target disorder. For an ordinal or continuous diagnostic test, the ROC curve depicts the plot of all pairs of sensitivity and 1-specificity (false-positive probability) over all possible or chosen cutoff values" [4].

Note 1: A screening test is a diagnostic test, the initial one in pursuing diagnosis (rule-in) about a latent case of a particular illness.

Note 2: A diagnostic test's "ability to discriminate" – informativeness – is not a property of the test in isolation. It is a *marginal* property, conditional on the pre-test diagnostic profile. (Cf. 'Diagnostic test's properties.')

Note 3: Suitable measures of that informativeness are result-specific *likelihood ratios* conditional on the pre-test profile – insofar as such a multitude of parameters is considered relevant and practical to address. (Cf. 'Diagnostic test's properties.')

Note 4: For both the ROC and LR outlooks, a suitable *alternative* is needed. This alternative is study of the pre-test and post-test diagnostic probability functions and, also, the distribution of the test result conditionally on the pre-test profile. (See 'Diagnostic test's properties.')

ROC curve – Receiver operating characteristic curve.