

Clinical validation of diagnostic procedures (Mainly Imaging)

Michael L. Goris

Stanford University School of Medicine, USA

The first level of evaluation can be called *diagnostic efficacy*. The appropriate research question to ask in this phase, while the technology is just beginning to diffuse into clinical practice or when a substantive advance in the technology occurs, is "How well does the new technology detect specific disease conditions?" The measures of effectiveness are sensitivity, specificity, positive and negative predictive value (given a certain prevalence of disease in the study population, the likelihood that a positive or negative test result means that the patient does or does not have the disease), and receiver-operating characteristic curve analysis. The test has diagnostic efficacy if the test classifies the patient in the correct category. The correct category is called the diagnosis. Not all diagnostic techniques aim to properly classify the patients, but rather to predict either the outcome, or the best therapy to obtain the desired outcome (measurements of plasma cholesterol do not provide a diagnosis but a prognosis, staging is not diagnostic, but predictive). In view of this, validation can be based on the following approaches:

1. **Outcome analysis:** Outcome studies measure effectiveness rather than efficacy. The analysis of those data is complicated by the concatenation of increased detection (an increase in incidence) and the fact that early detection may not always predict progression. For breast cancer and mammography it did take a long time to show the beneficial outcome.
2. **Predictive power:** A taxonomic exact diagnosis may not be predictive. Consider that in some diseases the median survival time is n years: fifty percent die earlier, 50% later. The prognosis is not necessarily well defined by the diagnosis.
3. **Predicting the taxonomy:** The most relevant aspects of this approach are 1) that at some point there has to be a defining test 2) that a ground truth is assumed to exist and be known. The major problem is verification bias.
4. **Discriminating power:** The classic approach is to take 2 groups, equivalent in many aspects, classified either genetically or prospectively by evolution, and see if in the absence of clear symptoms they can be differentiated.
5. **Internalized:** Mainly in Image processing or acquisition.
6. **Regions of Interest:** Information extraction. Fourier analysis. Parametric imaging. Reconstruction. Multi-slice CT
7. **Equivalence:** The equivalence design is not altogether valueless since the new procedure may globally decrease the cost (expenses in material and personnel, pain and danger). Essentially however the cost due to an improvement cannot be demonstrated to increase. What can be demonstrated is an improvement in reproducibility of the interpretation if the method is automated or quantitative.

Biography

Michael L. Goris received his M.D. from the Katholieke Universiteit Leuven (Belgium) in 1967 and his Ph.D. in Medical Physics from UC Berkeley in 1972. He was board certified in Nuclear Medicine in 1972. In 1973 he became assistant professor in Radiology (Nuclear Medicine) at Stanford University School of Medicine. He became a full professor in 1984. He has published more than 125 papers in reputed journals. He is serving as Chief Editor of the OJMI. He is the author of 4 books on Image Processing, Diagnostic Characteristics, Nuclear Cardiology and the Mathematical basis of Nuclear Medicine procedures.

mlgoris@stanford.edu