Validation of Biologic Markers for Use in Research on Chronic Fatigue Syndrome

Paul A. Schulte

From the National Institute for Occupational Safety and Health, Centers for Disease Control, Cincinnati, Ohio

Unresolved aspects of chronic fatigue syndrome can be addressed by research involving biologic markers. These may be any molecular, biochemical, physiological, or other biologic parameter obtainable from biologic specimens. The use of biologic markers in research requires their validation as dependent or independent variables. Additionally, other characteristics of markers such as reliability of assays, background level, confounding factors, interpretations, and legal and ethical implications should be considered before the use of markers in research. A checklist is provided for evaluating a biologic marker before its inclusion in research.

It has been said that chronic fatigue syndrome (CFS) is a disease in search of laboratory confirmation, that it is necessary to identify biologic differences between those who have the syndrome and those who do not, and that the techniques of molecular biology should be applied to the evaluation of potential etiologic agents. All of these objectives require measurements on biologic specimens and, hence, will yield biologic markers. In this paper some of the methodologic issues involved in the use of biologic markers in epidemiologic research will be discussed, with the hope that such a review will be of value to those who design research on CFS. This presentation will highlight some of the important features reviewed in two previously published articles [1, 2].

The concept of biologic markers is not new. Measurements of biologic specimens, e.g., antibody titers, have been the hallmark of research on infectious disease and have been integral to research on most other diseases. What characterizes the current generation of biologic markers is the possibility of detecting smaller amounts of xenobiotic compounds in biologic specimens, of more subtle changes at the molecular or gene level, and of earlier changes in the natural history of a disease [1]. Biologic markers are also useful in distinguishing individuals who might develop clinical disease from those who might not [3, 4]. Biologic markers may be defined as any measurable molecular, biochemical, physiological, cytologic, morphologic, or other biologic parameter obtainable from biologic specimens. Hence, a biologic marker is a measurement of a biologic phenomenon and, as such, comprises two components referred to as "signal" and "noise" or "systematic effect" and "random effect" or "true score" and "error" [5].

Accuracy of Biologic Markers

In the measurement of a biologic marker, measurement errors need to be identified and controlled. Control refers to the limitation of measurement error. Failure to control these errors can lead to serious consequences, including the presentation of misinformation to study participants about their individual results and a reduction in the probability of detecting true associations between a biologic marker and the outcome. Reliability is the degree of stability of a measurement repeated under identical conditions [6]. In addition, failure to control for unreliability may lead to the following untoward consequences described by Fleiss [5]: need for increased sample size, systematic bias of correlations in the direction of underestimating them, and biased selection of cases in case-control studies.

To control unreliability in measurement, Fleiss recommends that it be standard practice to conduct a pilot reliability study before embarking on a major research undertaking in which measures known to be unreliable are used. Reliability may be improved by replicating measurement procedures on each study subject until a prespecified level of reliability is reached. The calculations of the reliability coefficient, i.e., the reliability pertaining to the number of measurements, are described in the Fleiss analysis [5]. A marker, in addition to being reliable, must be valid, i.e., it must represent the phenomenon being measured. The broader context of the validity question is whether a marker has a true and reproducible relationship with a precursor or outcome event. To determine validity, it is necessary to know that there is an association between a marker and an exposure or outcome event; the location, shape, and slope of the exposure or dose-response relationship for that association; the threshold of "no effect" level; and the positive predictive value (PPV) of the marker for the exposure or disease [7].

Often biologic markers will be considered for large-scale epidemiologic research without strict attention being paid to these requirements for validation. Such a problem may have occurred because researchers have not critically evaluated the initial research on the marker and the milieu in which this research was carried out [1, 8].

Early research on a marker often is derived from limited laboratory tests and clinical series of cases in which investigators have not been blind to conditions and in which selec-

Please address requests for reprints to Dr. Paul A. Schulte, National Institute for Occupational Safety and Health, Centers for Disease Control, Cincinnati, Ohio 45226.

Reviews of Infectious Diseases 1991;13(Suppl 1):S87–9

tion biases are likely, a frequent occurrence in research on tumor markers [8]. Observation bias also is likely if researchers look more carefully for markers among people with the disease than among those without the disease.

The ultimate criterion for a valid marker is whether it has a strong PPV. The PPV is the proportion of those with the marker who have the outcome of interest. The PPV of a marker for predicting an event has been described by Griffith et al. [9], Arezzo et al. [10], and Khoury et al. [11]: PPV = $[1/([1 + (1-a)(1-p)]/bp)$, where a = specificity, b = sensitivity, and p = disease frequency. As adapted from Khoury et al. [11], sensitivity (b) of the marker with respect to the disease is the conditional probability of the marker being carried given the presence of the disease, and specificity (a) is the conditional probability of the marker not being carried given the absence of disease.

With low disease frequency the PPV is highly responsive to changes in the specificity of the marker and less responsive to changes in the other parameters; e.g., in population studies of CFS, in which prevalence is low, the specificity of a marker will weigh heavily in determining the PPV. In studies of subjects among whom the prevalence of a biologic marker is high (i.e., 0.5), such as in a clinical series of patients already diagnosed with CFS, sensitivity becomes the more important operating characteristic.

The interrelationships among the parameters used for calculating the PPV are described by the equation $m = bp + (1-a)(1-p)$. Hence, Griffith et al. [9] suggest that it is desirable for each estimate of marker sensitivity, specificity, and frequency and disease frequency to be based on independent samples from the population being studied. The use of biologic markers in general and for research on CFS in particular requires collaboration across disciplines. This is evident in the range of expertise required to evaluate a marker from its introductory research in the laboratory to its utilization in the field. Different disciplinary groups must understand the need for collaboration and learn language that will permit effective communication so that the strengths and limitations of each discipline are known to collaborators. This is especially true when epidemiologic studies using biologic markers are designed. Although a marker may appear to be a validated candidate for use in the field, attention to study design, statistical power, choice of dependent and independent variables, the distributional form of the data, type of analysis, and clinical significance of findings are still necessary. The epidemiologist and the biostatistician are integral to addressing these questions.

It is tempting to believe that markers identified with technologies or laboratory tests are inherently valid and that researchers may use them in subsequent investigations without much pretesting or forethought. To avoid this temptation, it may be useful for researchers to consider the following features when assessing a candidate marker for study of CFS or any other condition.

Checklist for Evaluating a Biologic Marker

(1) What evidence is there that the marker is valid? (a) Was the initial research adequate? (b) What is the reliability of the marker? (c) What are the marker's sensitivity, specificity, and predictive value in the population to be evaluated?

(2) What type of marker will be used? Is it a surrogate, a predictor, or a correlate of the event of concern? (a) What is the natural history of the marker? (b) What influences intra- and interindividual variability? (c) What are the normal background and reference levels?

(3) What conditions are necessary for use of the marker? (a) Does specimen collection require invasive procedures? (b) What are the optimal specimen collection time, amount of specimen, and handling requirements? (c) How stable is the marker in storage?

(4) How will a marker be influenced by (or influence) study design and analyses? (a) Will the marker be used as a dependent or independent variable? (b) How will quantitative and categorical data be combined in multivariate analyses? (c) Will marker collection result in time-series data? (d) How will multiple markers be handled? How can they be combined into a useful index?

(5) What is the impact on the subject and on society of the use of a marker in a study? (a) How will the limitations and results be interpreted by the subjects? (b) How will they be communicated both in the informed consent document and in the dissemination of the findings? (c) What ethical and legal implications might result from the use of the markers?

If investigators can review this checklist before initiating research, it is likely that many of the problems that might arise with the use of biologic markers will be anticipated and prevented or ameliorated.

To evaluate how well a marker is validated, it would be useful to have a scale on which to rank it. One such qualitative rating scale for the validity of biologic markers can be adapted from the work of Busch et al., who reviewed tests for mutagens [12]. Eight levels, increasing in validity, have been delineated as follows: (1) the marker is "totally experimental" with complete uncertainty about health or exposure significance of results; (2) the marker is experimental, but theoretical reasons exist that suggest the marker will correlate with exposure or disease; (3) the marker may be found to correlate with exposure or disease, but significance of the data is still uncertain; (4) the marker probably correlates well with exposure or disease but conclusive data are not available; (5) the marker has been extensively studied and has been validated as a useful tool for monitoring exposure or diseases but gives an unexpected, positive response in 10% of people screened; (6) the marker has been extensively studied and has been validated as a useful tool for monitoring exposure or disease but gives an unexpected, negative response in 10% of people screened who have a history of chronic abnormal exposure; (7) the marker has been extensively studied and

has been validated as a useful tool for monitoring exposure or disease with no or rare false-positive and false-negative results; and (8) the marker has been validated and is completely predictive of exposure or disease.

Biologic markers may be quite useful in defining CFS, in determining its etiology and natural history, and in subsequently indicating the effectiveness of preventative or therapeutic actions. However, if such markers are to be used, they need to be valid indicators or their use can lead to inaccurate conclusions.

Biologic markers should not be considered a panacea in the quest for the identification and control of CFS or any other medical malady. Rather, they are merely additional tools that need to be added to the collection of tools and techniques already in use. It would be a mistake to shift emphasis toward the identification of biologic markers only and fail to utilize the descriptive, categorical, and phenomenologic information that historically has yielded successful leads in the study of many diseases.

References

1. Schulte PA. Methodologic issues in the use of biologic markers in epidemiologic research. Am J Epidemiol **1987**;126:1006–16
2. Schulte PA. A conceptual framework for the validation and use of biologic markers. Environ Res **1989**;48:129–44
3. Evans AS. Subclinical epidemiology: the first Harry A. Feldman Memorial Lecture. Am J Epidemiol **1987**;125:545–55
4. Omenn GS, Gelboin HV, eds. Banbury report 16. Genetic variability in responses to chemical exposure. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, **1984**
5. Fleiss JL. Statistical factors in early detection of health effects. In: Underhill DW, Radford EP, eds. New and sensitive indicators of health impacts of environmental agents. Pittsburgh: Center for Environmental Epidemiology, University of Pittsburgh, **1986**:9–16
6. Last JM, ed. A dictionary of epidemiology. New York: Oxford University Press, **1983**:91
7. Gann P. Use and misuse of existing data bases in environmental epidemiology: the case of air pollution. In: Kopfler FC, Craun GF, eds. Environmental epidemiology. Chelsea, Mich: Lewis Publishers, **1986**: 109–22
8. Tate H. Assessing tumour markers. Stat Med **1983**;2:217–22
9. Griffith J, Duncan RC, Hulka BS. Biochemical and biological markers: implications for epidemiologic studies. Arch Environ Health **1989**; 44:375–81
10. Arezzo JC, Schaumburg HH, Petersen CA. Rapid screening for peripheral neuropathy: a field study with Optacon. Neurology **1983**;33:626–9
11. Khoury MJ, Newill CA, Chase GA. Epidemiologic evaluation of screening for risk factors: application to genetic screening. Am J Public Health **1985**;75:1204–8.
12. Busch DB, Easterling D, Leventhal H, Bryan GT. Recontacting subjects in mutagen-exposure-monitoring studies. II. Results of a questionnaire study of mutagenesis researchers with review of the pertinent literature. Environ Mutagen **1988**;11:333–58