

Update 42

Who was Alvan R. Feinstein?

Alvan R Feinstein (1925-2001) was the founding father of clinical epidemiology, and, as a result, much of his work finds its voice today through the Evidence-based Health Care Movement.

Feinstein was Professor of Medicine and Epidemiology at Yale University School of Medicine from 1969, Editor of the *Journal of Chronic Diseases* (1982-1988) and founder and Editor of the *Journal of Clinical Epidemiology* (1988-2001). He was author of about 800 scientific publications. Most of these were full papers in peer reviewed journals, of which Feinstein was first author in most cases. Feinstein published six authoritative books. *Clinical Judgment* (1967), *Clinical epidemiology: the architecture of clinical research* (1985), and *Clinimetrics* (1987) were breakthroughs in clinical science, and still a breeding ground for innovation. (1,2)

Much of Feinstein's work is extremely relevant to the chiropractic profession because the methodologies he helped develop, apply and critique, are fundamental to further research aimed toward exploring the validity and reliability of tests for subluxation detection and analysis.

Just before he died unexpectedly, on the 25th of October 2001 at the age of 75, Feinstein was completing a review paper (3) for the *Journal of Epidemiology and Community Health*. Now published the content of that review paper deserves careful reading by chiropractic researchers, academics and evidence-based practitioners.

The following are a number of quotes from Feinstein's final paper -

At the end of the second world war, the prime scientific goal in clinical medicine was to make a correct diagnosis. Its importance was demonstrated by the attention it received in medical education.

With diagnosis continuing to be the main focus today, major attention has been given to the use and accuracy of "marker tests" that can be less expensive or less invasive indicators of the definitive diagnosis. Although symptoms and physical signs are sometimes appraised as markers, most of the tests are done as technological procedures, entailing examination of such entities as blood, urine, smears, graphic tracings, and ultrasonograms.

As diagnostic medical technology proliferated after the second world war, mechanisms were needed for its evaluation.

A method and nomenclature for appraising diagnostic accuracy had been introduced in 1947. They calculated *sensitivity* as the proportion of diseased "cases" that were correctly identified by a positive test, and *specificity* as the corresponding proportion of non-diseased "controls" with an accurate negative test. What practising clinicians needed, however, were truly diagnostic indices that would be applicable to patients for whom a diagnosis was not yet established. Accordingly, an intricate mathematical procedure, was established to convert sensitivity and specificity into pragmatic *diagnostic* indices.

Further research, however, eventually showed that the entire calculational procedure was based on a wrong premise. The assumption that the values of sensitivity and specificity were constant for each disease and for each non-diseased control group, regardless of the spectrum of patients who were tested. This assumption was incorrect.

Regardless of the limited focus and erroneous premises of diagnostic research, and regardless of its academic popularity and publicational frequency, many of the studies have not been well done. (4,5) Tests that involved subjective decisions (by radiologists, pathologists, or other interpreters) have not always been checked for reproducibility. The objectivity needed to avoid "review bias" has seldom been achieved by arranging for "blind readings" when a second procedure is performed after the results of the first procedure are known.

The cases and controls in the academic studies have not always been chosen to be suitable representatives of the corresponding clinical conditions and to cover an appropriately wide spectrum of those conditions. Many studies have had a built in bias, sometimes called "work up" or "verification bias", produced because the definitive procedure is often ordered mainly for persons with a positive marker result, but not for those with a negative result.

Perhaps the most glaring flaw of the entire appraisal process, however, has been the persistent focus on accuracy of diagnosis. This focus was justified 60 years ago, but is no longer appropriate

Furthermore, some of the most important roles of technological tests today are in non- diagnostic clinical decisions. The results of the tests are often used in estimating prognosis, in choosing treatment, in appraising post-therapeutic status, and in changing treatment. Yet none of these activities is included in the procedures developed for appraising diagnostic efficacy. If the total clinical contributions of technological tests are to be suitably evaluated now and in the future, this methodological gap will have to be eliminated, with new appraisals developed for the currently unmet challenges.

An important step in the desired progress will require fundamental alterations in nomenclature and in methodology. In nomenclature, the results of the technological procedures should no longer be called "diagnostic tests", because the results are used for much more than diagnosis alone. A name such as technological tests would be more appropriate for the broad array of clinical decisions in which the data can be used.

The main points to be made in this essay are:

1. the current methods of evaluating marker tests are unsatisfactory;
2. suitable appraisal of the "definitive" test procedures has been generally overlooked; and
3. the most important contributions of the technological procedures today are for prognostic and therapeutic decisions (rather than for diagnosis alone), but these decisions are seldom specifically evaluated.

If the appropriate changes do not occur researchers will continue their misguided efforts; technological procedures will continue to be appraised inadequately and often misleadingly; and the medical world will continue spending huge sums of money for research that is often unsatisfactory, and for tests that are often ineffectively evaluated and applied.

References:

1. Knottnerus JA, Alvan R Feinstein. *J Epidemiol Community Health* 2002;56:322.
2. Spitzer WO. The teacher's teacher: a personal tribute to Alvan R Feinstein. *J Epidemiol Community Health* 2002;56:328-329.
3. Feinstein AR. Misguided efforts and future challenges for research on "diagnostic tests" *J Epidemiol Community Health* 2002;56:330-332.
<http://jech.bmjournals.com/cgi/content/full/56/5/330>
4. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. *JAMA* 1995;274:645-51.
5. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J Med* 1978;299:926-30.