Introduction

This issue of the Journal is dedicated to Alvan R. Feinstein, the father of clinical epidemiology [1]. We like to join those who honor his memory and acknowledge his invaluable contribution to methodology in clinical research. The debate is still open and calls for reassessment of the field.

As to psychometrics, the basis of measurement in psychiatry, Fava et al. [2] argue in this issue that it is an inadequate method that persists in the field despite its many flaws. Their observations and musings have profound implications, because many clinical trials have faltered on the shores of the insensitive Hamilton Depression Rating Scale (HAMD). With insufficient sensitivity and an inability to detect small changes, the HAMD has caused investigators to fail to reject the null hypothesis (to accept that no difference exists), with the result that potentially effective antidepressants are kept out of the US market.

An alternative, clinimetrics, is proposed as the basis to assess clinical phenomena, diagnostics, prognostics, and therapeutics. Psychometrics is the managed care of methods; all variables have the same weight, just as all physicians have the same value, regardless of experience, expertise, or judgement. Clinimetrics arises from clinical observation, from clinical methods, and, most important, from clinicians, with different variables having different weights. The acceptance of psychometrics may be a simple lack of clinical research training, an inferiority complex such that clinicians feel the need to adapt the methods of their better-trained psychometrician colleagues.

Per Bech [3] discusses the item response theory of Rasch, a theory that uses a sensible method to assess symptoms based on their prevalence in those with a disorder (clinical coherence) and the importance of those symptoms for clinicians to define severity (weighting of symptoms). Consistent with the item response theory, several versions of a foreshortened HAMD have been found more sensitive to change than the full HAMD.

Carlo Faravelli [4] questions the mathematical assumptions behind rating scales: that individual item scores can be summed to produce a total that has a linear relationship with the clinical phenomenon it was designed to measure. He notes that the same rating scale score can reflect a broad array of severity, i.e., many different individual item scores can add up to the same sum. To calculate the number of combinations and permutations of individual items that can sum to a HAMD of 30 is to realize the complexity of this total score. Furthermore, he argues convincingly that not only is a 50% reduction in baseline scores insufficient as a criterion of response, but even remission, the new standard for outcomes, falls short...
of clinical sensibility – that well-being should also be taken into account. Carlo Faravelli makes a plea for researchers to integrate clinical observation so that answers to questions can be interpreted rather than just taken at face value. He calls for valid measures, but stops there: how to get valid measures remains a mystery.

Paul Emmelkamp [5] argues that, in contrast to Fava and colleagues, psychometrics still has something to offer the field and that the assumptions that underlie clinimetrics need to be questioned and tested. He warns clinicians to be wary of the flawed evidence-based medicine movement, lest they be overly conservative and naive in their uncritical acceptance of research findings, echoing a concern that Feinstein [6] and Feinstein and Horwitz [7] had about meta-analysis. Other papers in this issue are concerned with topics which had been covered by Feinstein’s work.

Bouvard et al. [8] addressed the management of obsessive-compulsive disorder, integrating research evidence with the clinical experience; Goodwin et al. [9] evaluated the risk of depression in the community; Porcelli et al. [10] found that assessment based on clinimetric principles could entail prognostic value in patients with functional gastrointestinal disorders; Prins et al. [11] highlighted the role of social support in the persistence of complaints in chronic fatigue syndrome; Slade et al. [12] addressed the relationship between patient needs and quality of life, and Mataix-Cols et al. [13] dealt with the sensitivity of a commonly used rating scale.

**The Architect of Clinimetrics**

Alvan R. Feinstein had a simple mission: to raise the intellectual level of clinical research and maintain its importance during the ascendency of molecular biology. Not that he thought that molecular biology was unimportant, but rather that clinical research and clinical practice were legitimate and complex fields, worthy of being called a basic science of clinical medicine [14]. He challenged medicine as a gadfly, with a biting and sarcastic wit, writing neologisms that changed the way physicians thought and training a generation of clinical researchers through the Robert Wood Johnson Clinical Scholars Program at Yale University. His mission is particularly relevant to psychiatry today, with the explosion of knowledge in neuroscience, psychopharmacology, and genetics. We can have the most elegant ‘hard science’, but if clinical research uses the wrong measures to assess efficacy, if subjects included in clinical trials fail to represent the patients who are encountered in clinical practice, if diagnostic tests are not thought through so that clinicians can apply them properly, and if cause and effect are confused in epidemiological studies, then the results of ‘hard science’ become merely academic, without improving the lives of our patients [15].

Remember the dexamethasone suppression test (DST)? Remember how enthusiastically it was embraced by the field, with literally millions of tests done to diagnose depressed patients? The hard science was elegant: the diurnal variation of cortisol was abnormal in depressed patients, and something was wrong with the hypothalamic-pituitary-adrenal axis. Give dexamethasone by mouth the night before, and then measure the blood levels of cortisol the next day in the afternoon. Normal controls suppress their cortisol levels, while depressed patients maintain high levels of cortisol, i.e., they fail to suppress. Thus was born a diagnostic test that finally showed that a biological abnormality was the basis for depression. Sounded great, but the field started to realize that something was ‘rotten in Denmark’. The DST did not perform well as a diagnostic test. Not only were sensitivity and specificity insufficient for the test to be useful clinically, but physicians had no idea how to actually use the test. Ransohoff and Feinstein [16, 17] have written about the importance of the clinical spectrum of disease and how diagnostic tests needed to take into account different levels of clinical pathology. The DST was no exception and could not withstand his careful scrutiny [18]. The DST has since faded as a clinically useful diagnostic test for depression – but not to worry, many more are waiting to take its place.

Our field is much concerned with comorbid psychiatric conditions, but few are aware that Feinstein coined the term [19]. We, however, unknowingly misuse it. The original concept of comorbidity refers to having another disorder unrelated to the one of interest, e.g., congestive heart failure and cirrhosis. In contrast, psychiatric comorbidity according to DSM standards usually consists of closely related categorical disorders that may be part of the development of the disorder of interest, e.g., social phobia that precedes major depression. So are we looking at true comorbid conditions in psychiatry or instead examining the interrelatedness of psychiatric symptoms and total psychopathological load?

Feinstein, above all, strove for clarity in thought, in writing, in research, and in clinical work. He railed against the thoughtlessness rampant in clinical research and admonished the field for always seeking the holy grail of a statistically significant p value without predetermin-
ing the size of a difference between groups that would be considered clinically important [20]. He railed against ‘abstruseness’ of the statistics used to describe differences between two groups [21]. He argued to maintain the link between clinical phenomena and pathophysiology [22]. He defended the importance of clinical investigators who studied clinical phenomena [23]. He did not suffer fools gladly. As part of the Robert Wood Johnson Clinical Scholar Program, Feinstein presided over the lectures he gleefully and formally referred to as ‘Academic Toilet Training’ during which he taught presentation skills, grant writing, data management, and manuscript preparation. One of his group of students repaid him in kind and gave him a wooden toilet seat with a brass plaque that he proudly hung over the door of his office.

He was a prolific writer and was the single author for Clinical Judgment [24], Clinical Epidemiology: The Architecture of Clinical Research [25], Clinimetrics [26], and an introduction to multivariate statistics [27]. We owe a great debt to Feinstein and should strive to be thoughtful while we plan, conduct, and consume the results of clinical research.

Feinstein would have appreciated the papers in this issue. The authors take the field to task for using a flawed method to develop scales that are then used to assess efficacy, a polemic and debate that is pure Feinsteinian.

References

5. Emmelkamp PM: The additional value of cliniometrics needs to be established rather than assumed. Psychother Psychosom 2004;73:142–144.