Donald F Klein's final comment

Martin M. Katz: Clinical Trials with Antidepressants: How Changing The Model Can Uncover New, More Effective Molecules

Collated by Olaf Fjetland

Comments on salient aspects of reviews by Bech, Lader and Brown of the new book by Martin Katz are presented. Critical statements and rejoinders by Katz and Klein are interspersed.

The usefulness and novelty of the book's suggestions on how to revivify novel drug discovery is a central concern. That the novel drug discovery process has dried up furnishes the declared motive for Katz's new book.

Bech, an eminent statistician, well versed in psychopharmacology, has argued, like Katz, that the Hamilton Depression Scale is multi-dimensional. His contribution to this review of Katz's book is on the whole quite supportive. This is surprising since, in Per Bech's book, "Clinical Psychometrics," that I reviewed for INHN, he points out that Factor Analysis depends upon the rule of thumb selection of the number of factors that then is rotated (by various methods) to differing definitions of simple structure. Bech holds that these procedures do not flow from a logical basis that allows firm deductions or sampling inferences. This defect is affirmed by the lack of factor replicability across various samples. Strikingly, Bech argues that the ubiquitous factor analysis does not provide appropriate measures of change or a foundation for diagnosis. This critically challenges Katz's work, as well as the NIMH-sponsored RDoC manifesto for dimensional primacy via multivariate analysis. Bech supports scale analysis by using more modern approaches, including IRT, Rasch and Guttman approach. He does not claim that this subscale approach leads to novel drug discovery. Bech avoids a confrontation with Katz, who reciprocates. Bech does not reject the value of Hamilton Subscales. His team developed a six-item subscale believed to improve depression diagnostic specificity, as well as sensitivity to change (Timmerby et. al., 2017).

There is a superficial similarity to Katz's assertions about the utility of a component/dimensional approach for psychopharmacologic studies. Surprisingly, Bech does not

address the usefulness of Katz' components. Further, claims that componential analysis is required for identification of new, more effective drugs goes unremarked. Assertions such as "Antidepressants are not 'diagnosis-specific,' but are in their modes of action "component-specific'" seem ill-founded since Katz studied only depressed patients. Hotelling's principal component analysis technique allowed components to be "discovered," such as depressed mood, psychic anxiety psychomotor retardation, psychomotor agitation, hostility, somatization, interpersonal sensitivity, sleep or cognitive impairment. These components can then be parts of specific dimensions, namely (1) anxiety-agitation-somatization-sleep (2) depressed mood-retardation; and (3) hostility-interpersonal sensitivity. It is common in a Hotelling analysis that on the major first factor all loading variables are positive, while the second factor is bipolar. That is, some loading variables have positive and some have negative loadings. The British tradition uses only the contrast evident in the second factor. "In contrast, an American approach rapidly emerged in which factor analysis was used to identify as many factors as possible." Bech argues that these factors, even if "rotated to simplicity," cannot be represented by a total since they contain items relevant to both severity and group discrimination. This impairs their use both as change and diagnostic measures. Therefore, Factor scores derived from patients' status scores are not particularly sensitive to change as they bury relevant change sensitive items by many unaffected loading items. This problem has been described in a widely unnoticed paper (Klein and Fink 1963). An immediate problem with Katz's components is psychopathology coverage. For instance, where do hallucinations or delusions or mania or dementia fit? Bech does not address whether these dimensions differ from those produced by ordinary factor analyses or have some qualities that make them particularly useful for drug discovery.

Lader's extensive review reflects his expertise in epidemiology, pharmacology and nosology (Wilson and Lader 2015). He agrees with Bech and Katz that the FDA required clinical trial has limited legal purposes with regard to marketing and that artificial outcomes may serve those purposes. It is not stated that it is not the FDA that limits involved clinical trials, rather industry's profit-maximizing decision to restrict the extent of clinical trials to the economic minimum that passes FDA standards. Lader notes that Katz suggests decreasing waste by expanding the limited FDA requirements for an efficacy trial into complex measurements, including componential analyses, that will lay the groundwork for drug discovery and broadening the range of therapeutic

indications. Lader points out that expensive data gathering during this pre-marketing, Phase Three trials, runs the risk of extensive expenditure on an agent that proves a failure (as is currently frequent). “Caution is needed not to substitute one source of waste with another.” However, the pragmatic outcome measures mentioned are not suitable for outpatient practices. A realistic outcome determination depends upon the current medication profile, patient health status, functional abilities and symptomatic state, as well as social functioning, work and family engagement, over several time periods. It is not required by FDA’s sparse standards and is rarely done.

- Lader holds that the measures used in FDA approved trials do not reflect the superior clinical judgments that compare treatments by “a probabilistic analysis of the chance of obtaining a useful therapeutic response.” It's hard to see on what the clinician bases such estimates. Few methodologically sound studies compare various treatments and use a valid control group. Therefore, the clinician does not have enough sound information to support objective choices. The standard waiting list control is unsound. A diagnosis is made, but treatment is delayed. This generates entirely different emotions and expectancies than placebo treatment. It may increase anxiety. This artifact leads to an exaggerated difference between waitlist and active medication. Waitlists may result in covert protocol non-compliance by self-treatment Concurrent placebo-treated controls are necessary to establish efficacy. Requiring such a control group also needs a medication treatment arm to preserve blindness for the placebo-treated group. This increases trial complexity but justifies efficacy interpretations and public health relevance. The usual simple two-group design is frequently vulnerable to a covert allegiance effect. It does not prove efficacy or allow judgments of respective value with other treatments. The clinician does not have the ability or unbiased information needed for a probabilistic analysis, so Lader’s suggestion is reduced to a best guess. That patient is not sympathetic to both clinical and research prescriptions, which leads to covert non-compliance. This, as well as dropouts, destroys randomization. This grave problem produces misleading estimates of both efficacy and safety. It is rarely corrected.
- Lader believes that the emphasis on waste in premarketing studies is due to a confusion between the FDA’s narrowly defined regulatory choices, which justify economically

rigorous Pharma supported clinical trials, compared to science support where the unknown truths of the therapeutic situation justify wide exploration. Nobody notes that the FDA is a Federal Regulatory Agency debarred from generating knowledge unless closely tied to medication evaluation. Knowledge generation is NIH and NSF's turf.

- The new models, proposed by Katz, have not convinced Lader that these methods are vehicles for finding new antidepressants. He is, however, prepared, to await further developments in clinical trials research, which is an exasperating truism. Katz argues that the only direct way to relevant research is for investigators to apply his proposed "well researched" alternative methods in their studies. The cited studies do not address the validation of componential measures. Rather they supposedly support Katz's stand for decreasing the time span for a clinical evaluation from the usual six weeks to approximately two weeks.
- Brown, a clinical psychiatrist with extensive clinical trials experience and a focus on the importance of placebo, and in agreement with Katz, notes that remarkable biological advances have not produced an understanding of how brain processes can eventuate in depression. "The pharmaceutical industry comes up with 'new' antidepressants all the time, and they are launched with great fanfare. But these 'new' antidepressants are invariably me-too variants of older drugs."
- Brown agrees with Katz's suggestion that if a researcher has in hand a compound with novel psychotropic properties, our current system for evaluating psychotropic drugs makes it unlikely that its novel clinical effects would be detected, particularly if they were unexpected.

This claim is entirely out of keeping with the recognition of anticonvulsants as mood stabilizers, as well as the recent furor over the psychedelic ketamine's quick action. Clinical scientists have eyes, interviews, and often understanding. They can see beneficial changes in their patients before scale evaluations and modify these instruments appropriately. A model change is the expanded Hamilton Depression Scale from 17 to 21 items, allowing the distinctive features of atypical depression to be assessed. Scale composition is not a limiting factor on discovery. No psychiatric drug has been discovered by scale analysis.

Scales are used for validation purposes, as well defined concrete referents for patient evaluation. That is not the discovery process. Katz's central prediction is that scale refinement and extensions of clinical assessment by video recording will lead to discoveries that clinical observation misses. The logical analysis and positive pilot findings that justify investments in expensive programs are not presented. The face validity of Katz' program depends on confusing possible increases in scale reliability with a unique discovery process.

Katz adduces a small controlled study from his group that contrasts a medicated group vs. placebo. Strangely, the medicated group reported for independent analysis combines the separate randomizations of desipramine and paroxetine. No justification is given for this senseless procedure. Also, to bolster his support for a different model of depression by using a sample who were "soundly diagnosed" as depressed. It means the investigators stringently applied criteria for Major Depressive Disorder from some accepted source. Such criteria are usually based on variables that portray a diluted version of melancholia. Therefore, variance in depression's measurement becomes constricted, which is considered useful. But Factor analyses are effectively based on correlations or similar indices of coherence. The constricted variance of the depression variables also constricts correlations with depression towards zero. Analyses within depression may indicate various item groupings that are not relevant to depression diagnosis. Rather, they refer to depression modifiers. This view is supported by the labels of the proposed, three dimensions. Katz did not address this issue. Positive within drug analyses were criticized for an obscure presentation that could be swiftly rectified. However, obscurity persisted. Relying on reported inferential statistics appeared to support Katz's views on early onset of drug effect, the predictability of both major and absent benefit and radical shortening of clinical trials. Our presentation of difficulties with Katz' analyses provides candid examples of why solely relying on inferential statistics affords an inadequate basis for thoughtful. conclusions. The requested 2X2 data layout, as presented in "Martin M. Katz's response to Donald F. Klein's reply to Carlos Morra's comment" (INHN.Controversies.10.15.2015) presented in a parallel project (Martin M. Katz: Onset of antidepressant action) were insufficiently identified, as Leslie Morey agreed (Controversies 12.12.2015). The ambiguity is the uncertainty about which table row should be considered as early improvement. Assuming early improvement refers to row 2, this table roughly agrees with Katz's statement that "70% of patients showing early improvement would go on to respond at 6 or 8 weeks."

Hamilton Rating Scale

		Late	
		<50%	>50%
early	<20%	15	2
	>20%	8	25

(Note, 33 are predicted to do well, but only 27 (82%) did. Based on Katz's within drug analysis the drug is overvalued.)

One might be interested in the possibility that a very low pre-score would indicate a likely treatment shift. However, even better such a score should allow a drug free period of clinical watchful waiting.

The hopefully predictive correlation (0.6) between pre- and post measures, accounting for 36% of the variance, is generally considered too low for predictive use. Further problems remain. The "active drug" sample, N = 50, combines the Paroxetine study (N=24) with the DMI study (N=26). No justification is given. The combination of Paroxetine, picked as a serotonergic agent, and DMI as a noradrenergic agent requires a prior justification. Apparently, an increase in sample size was considered necessary.

Katz provided placebo data to Morey who shared it. This allows progress from a predictive study, derived entirely from within drug data, to an estimate derived from contrasting drug vs placebo.

	Drug	Placebo
Recover	27	6
Not Rec	23	13

Chi-square = 2.77 p=0.09 2Tailed

This analysis, focused on invalidating the null, does not have sufficient strength to be a useful predictor. The correlation, 0.6 found here, has 95% confidence limits of 0.39 ,0.72. So, the correlation's upper limit remains insufficient for predictive utility, even if one stacks the dice by an untrue assumption of sample bivariate normality. Katz's argument is questioned by the insignificant contrast between drug and placebo outcomes. Even strong findings, if derived from

a small data set, would call for large sample replication before allowing interpretation as sound predictions about the useful length of definitive clinical trials. That this insignificant, 6-week, drug vs. placebo contrast justifies the utility of a much shorter clinical trial is preposterous. Katz's claim that larger studies have already agreed with his conclusions needs more than an article reference. The exact analyses allowing parallel conclusions must be pointed out. I have failed to find them.

It is also illogical for large supposedly definitive trials to be followed by a small trial that could add nothing new. Katz replies that the large studies used total Hamilton scores whereas his small study was investigating componential scores. It follows that claims that componential analysis was backed up by large trials are incorrect.

There was no indication of the multiplicity of analyses picked over to show supportive analyses. It is well known that analyses based on within drug analyses are often meaningless. Adequate placebo controls and proper analyses are required for the correct understanding of real effects. The late partial release of detailed placebo data allowed the comparison of the medicated group to placebo. It was non-significant. This casts doubt on all of Katz's analyses, but this was denied by an assertion of trust in their own analyses. However, requested data allowing independent analyses were not made available.

To sum up, the reviews did not address major issues invalidating Katz's conclusions. Several illogical beliefs were not exposed. The illogic of a supposing tightening up such descriptions would somehow produce novel drugs was reviewed. It was almost unnecessary to review the data analyses since the logical framework was so impaired by the history of discovery. The persuasiveness of these propositions relies on a historical and logical confusion that increasing reliability somehow suffices for increases in discovery. The claimed relation to novel psychiatric drug discovery is not evidenced, but appeals to wishful thinking. However, Katz's data analysis, used to support his conclusions, proved to be, at least, questionable. Both book and reviews fail.

References:

Katz MM. Response to Klein's reply to Morra's comment. inhn.org. Controversies. October 15, 2015.

Klein DF, Fink M. Multiple item factors as change measures in psychopharmacology. Psychopharmacologia 1963; 4: 43-52.

Morey LC. Comment on the interaction between Katz and Klein. Controversies December 17, 2015.

Timmerby N, Andersen JH, Søndergaard S, Østergaard SD, Bech PA. Systematic Review of the Clinimetric Properties of the 6-Item Version of the Hamilton Depression Rating Scale (HAM-D6). Psychother Psychosom. 2017; 86:141- 149.

Wilson E. and Lader M. A review of the management of antidepressant discontinuation symptoms. Therapeutic Advances in Psychopharmacology 2015; 5:357-368.

January 11, 2018