

**Setting Priorities on Waiting Lists:
Point-Count Systems as Linear Models¹**

David C. Hadorn, MD and the Steering Committee of the Western Canada Waiting List Project (WCWL)*

Dr. Hadorn is Research Director of the Western Canada Waiting List Project; hadorn@dnai.com

*Members of the Steering Committee appear at the end of this article.

Reprint requests to: Dr. Tom Noseworthy, Department of Public Health Sciences, 13-103 Clinical Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2G3;
tom.noseworthy@ualberta.ca

Funded under Health Canada's Transition Fund.

Revised 14 August 2000.
Submitted for publication.

¹ Production of this document has been made possible by a financial contribution from the Health Transition Fund, Health Canada. The views expressed herein do not necessarily represent official policy of Health Canada.

Abstract

The Western Canada Waiting List Project (WCWL) is a federally funded initiative designed to develop tools and methods for standardising and managing waiting lists.¹ Like similar projects in New Zealand and the UK, the principal approach adopted by WCWL was the development of point-count measures for assessing patients' clinical urgency and relative priority, based primarily on (1) the severity of patients' conditions and (2) the extent of benefit expected from wait-listed services. Point-count measures like these function as linear models from a statistical perspective. This paper describes several theoretical and practical implications of this functional relationship.

Determining Priority among Patients on Waiting Lists

Fairness of access to wait-listed services is a prime concern of Canada's publicly funded health system. At present, waiting times vary tremendously from doctor to doctor and from hospital to hospital, and the length of patients' waits for services is not always commensurate with the severity of their conditions. In principle, patients with more urgent conditions should receive services ahead of those with less urgent conditions, and patients with approximately the same degree of urgency should wait about the same length of time regardless of where they happen to live. Moreover, it should be possible to assess whether and to what extent this ideal is being met.

In a companion paper,² we describe how clinical urgency is a function of the severity of patients' conditions and the extent to which treatments are expected to be beneficial. Relative priority, in turn, derives from a combination of urgency and non-clinical factors (e.g., ability to work, live independently, or care for dependents).

The time-honoured method for denoting relative urgency on waiting lists has been to assign patients to one of a few broad categories, such as "emergent," "urgent," and "routine." A "semi-urgent" category is sometimes added. However, the vagueness of such category descriptions does little to reduce the inherent subjectivity of global urgency judgements, nor is it possible using this approach to prioritise patients within each broad category.

An alternative and more powerful method for assessing and comparing patients' relative urgency on waiting lists has been used in New Zealand for several years as part of a national project to replace waiting lists with booking systems.³ This method operates by assigning points to patients based on the severity of clinical findings (e.g., degree of pain or limitation in motion) and on considerations of expected treatment benefit. A similar approach is being tested in the UK.^{4,5} The Western Canada Waiting List Project (www.wcwl.org) has developed such measures in five clinical areas: MRI scanning, hip and knee replacement, cataract surgery, general surgery procedures, and children's mental health. Figure 1 shows the hip and knee replacement criteria.



HIP AND KNEE REPLACEMENT PRIORITY CRITERIA Revised 26 June 2000 with weights

PLEASE PRINT CLEARLY

Provincial Health Care Number: _____

Patient Age: _____ Sex: [circle one] **M** **F**

[Tick one box] Left Hip Right Hip Left Knee Right Knee

[Tick one box] Primary Revision

Diagnosis: _____

Surgeon's Name: _____ Phone: _____

Date: _____

Patients must be on appropriate non-surgical treatment prior to evaluation (e.g. medications, walking aids, shoe inserts)

PLEASE CHECK THE BOX THAT MOST ACCURATELY DESCRIBES THE PATIENT'S CURRENT SITUATION

1. Pain on motion (e.g. walking, bending): *

- 0 None/mild
- 6 Moderate
- 13 Severe

2. Pain at rest (e.g. while sitting, lying down, or causing sleep disturbance): *

- 0 None
- 3 Mild
- 8 Moderate
- 11 Severe

* Take into account usual duration, intensity, and frequency of pain, including need for narcotic vs. non-narcotic medication.

3. Ability to walk without significant pain :

- 0 Over 5 blocks
- 0 1-5 blocks
- 4 <1 block
- 7 Household ambulator

4. Other functional limitations (e.g. putting on shoes, managing stairs, sitting to standing, sexual activity, bathing, cooking, recreation or hobbies):

- 0 No limitations
- 4 Mild limitations (able to do most activities with minor modifications or difficulty)
- 11 Moderate limitations (able to do most activities but with modification or assistance)
- 19 Severe limitations (unable to perform most activities)

5. **Abnormal findings on physical exam related to affected joint (e.g. deformity, instability, leg length difference, restriction of range of motion on examination):**
 0 None/mild
 5 Moderate
 10 Severe
6. **Potential for progression of disease documented by radiographic findings (e.g. recurrent dislocation, x-ray evidence of protrusion, significant bone loss, component wear, impending fracture):****
 0 None
 4 Mild
 11 Moderate
 20 Severe

** Predominantly applies to revisions, use in primary cases only in special circumstances (e.g. ligament instability, bone loss)

7. **Threat to patient role and independence in society (i.e. ability to work, give care to dependants, live independently (difficulty must be related to affected joint)):**
 0 Not threatened but more difficult
 10 Threatened but not immediately
 20 Immediately threatened or unable

8. **All things considered, how would you rate the urgency or relative priority of this patient?**
 (Draw a line across the scale.)

|-----|
 Not Urgent at all Extremely Urgent
(just short of an emergency)

9. **In your clinical judgement, what should be the maximum waiting time for this patient?**
 Number of weeks Number of months
 _____ **OR** _____

10. **In your practice how long would it take this patient to have the surgery done from the time you first see the patient?**
 Number of weeks Number of months
 _____ **OR** _____

Comments on the form or process used to complete form: _____

The use of such point count measures provides a relatively transparent and consistent method for assigning priority to patients on waiting lists. In addition, the quantitative output available using this method provides a method for adjusting waiting lists based on clinical case mix, which is an essential step toward assessing and comparing waiting lists.

In view of the increasing use of point count measures for assessing patients' priority on waiting lists, a closer look at the nature of these measures seems warranted. In particular, the relationship between priority criteria and statistical linear models has not been discussed in the literature up to this point, nor have the implications of this relationship been described.

Linear Models

Point-count systems are familiar to most people. Such examples as point-count immigration systems, hotel rating schemes, and whimsical creations like the financial "misery index" (the sum of the inflation rate, unemployment rate, and prime interest rate), all have a certain face validity. From a statistical perspective, such systems function as statistical *linear models*, which are additive scoring systems consisting of two or more factors divided into various levels (e.g., "degree of pain – none, mild, moderate, or severe"), each level being assigned a numerical weight. (The term "linear" is statistical jargon for "additive.") Weights are assigned in accordance with the relative importance of that factor/level in determining the overall outcome or parameter of interest.

Linear models are used throughout medicine. Two well-known examples are the APACHE scoring system⁶ and Mortality Prediction Model,⁷ which are used to predict mortality in intensive care units and other settings. These systems incorporate relevant clinical variables (e.g., blood pressure, serum electrolytes) into additive, point-count formulas. Higher scores generally reflect greater mortality risk. Mortality prediction models are also widely used to adjust for case-mix when comparing mortality rates (e.g., for coronary bypass graft surgery) across surgeons or hospitals.^{8,9}

Another well-known clinical point-count system is the Apgar Score, which is widely used in neonatal assessment. This model incorporates five clinical factors (viz., heart rate, respiratory effort, color, muscle tone, responsiveness to stimuli) consisting of three levels (e.g., poor, fair, good), which are weighted (0, 1, or 2 points) to permit a total score to be calculated for each patient. In this case, higher scores indicate better prognoses.

Researchers have developed linear, point-count models in many clinical (and non-clinical) situations^{4,5,10,11,12,13} and such models have proved more accurate than physician judgement in diagnosing acute abdominal pain,¹⁴ myocardial infarction,^{15,16} streptococcal tonsillitis,^{17,18} pneumonia,¹⁹ intracellular vs. extracellular causes of jaundice,²⁰ presence of ankle fracture,²¹ survival after diagnosis of Hodgkin's disease,²² or coronary artery disease.²³ Similar models are also used in a wide variety of non-clinical settings, including prediction of graduation from college, business bankruptcies, and divorce.^{24,25}

These models are, as a rule, more accurate than human predictors. This is not surprising, as it has been known for decades that human cognitive performance is limited with respect to

complex multi-variable judgement and prediction tasks.²⁶ The superior accuracy of even simple statistical prediction models is observed even when those models are derived directly from the judgement strategies of the same physicians whose performance is later compared to the model's.²⁷ This occurs because physicians often deviate from the models in non-systematic and inconsistent ways.^{28,29} Predictive accuracy almost always suffers as a result.

How Are Linear Models Developed?

Linear models are developed in one of two basic ways. The statistically preferred method uses a technique called regression analysis, in which information is collected on a large number of patients prior to an outcome of interest occurring (e.g., 30-day mortality; presence of appendicitis as confirmed at surgery). The statistically optimal weighting scheme is deduced from the observed relationship between the predictor variables and this outcome.

The second basic method for developing linear models is known as judgement analysis. Despite being less formal, models developed this way have predictive power nearly equivalent to those developed using regression analysis.^{25,26,30} Using this method, experts familiar with the relevant judgement tasks are asked to think out loud while formulating a diagnosis or prediction. By doing so, observers may readily identify the predictor variables used by these experts. For example, guidance counselors asked to predict a student's probable success in graduate school would likely invoke such factors as GPA and GRE. Similarly, a physician would likely emphasize a patient's age, sex, and rebound tenderness when determining the probability that a given instance of abdominal pain represents a ruptured appendix.

These informally developed models are called "improper" because the variable weights are not selected using standard regression techniques. However, as described by Dawes, improper models perform at near optimal levels because different linear composites of a constant set of variables tend to correlate highly. For example, the correlation between $X + 2Y$ and $2X + Y$ is 0.80 even when X and Y are uncorrelated.²⁵ Accordingly, "The whole trick is to know what variables to look at and then know how to add," as Dawes and Corrigan concluded in 1974.²⁷

As noted above, additive priority criteria are similar to mortality prediction models and to the Apgar score insofar as all of these models depict the severity of patients' conditions, as reflected in symptoms, physical findings, and test results. In addition, priority criteria incorporate, where appropriate, considerations of expected benefit, i.e., the extent to which treatments (are expected to) ameliorate the painful symptoms, reductions in functional abilities, and risk of premature mortality produced by these conditions.² As such, they can be considered "expected-benefit prediction models."

Validating Priority Criteria

How might priority criteria be tested for validity? At least three forms of validity must be distinguished in this context. First, the criteria can be considered valid to the extent they make sense to clinicians, insofar as they capture the important factors clinicians use to assess urgency and priority. Such face validity is probably the most important form of validity. Second, the criteria might be considered valid to the extent they reflect clinicians' quantitative judgements of

relative urgency. Efforts to assess this form of validity have taken place in New Zealand and Canada, as described below. Finally, priority scores could be considered valid to the extent they reflect differences in actual clinical outcomes between patients who do (or do not) receive the wait-listed services. Higher scores should be associated with greater degrees of net benefit. This is probably the “gold standard” against which priority scores should be assessed.

Lacking outcome information in most cases, however, the validation of priority criteria is generally restricted to an assessment of face validity and of concordance with clinical judgements of urgency. Such an approach was used in New Zealand,³ where priority scores were compared with global subjective judgments of urgency arrived at by experienced clinicians on a case-by-case basis in prospective series of patients. Global judgements were expressed either on a scale of 0 – 100 or in terms of what a “reasonable waiting time” would be for each patient (in days). The priority scores were derived using weights that had been established by consensus at panel meetings prior to this validation exercise. The correlation of global urgency ratings with priority scores was then calculated, and the results used to adjust the weights assigned to the various criteria.

A similar approach was taken in the WCWL, except that no preliminary scores were assigned to the various levels and criteria prior to the collection of prospective data. Raters simply placed tick marks next to the appropriate levels on each criterion (independent variables) on a prospective series of patients, and also indicated overall clinical urgency for each patient using a visual analogue scale, anchored at “not urgent at all” and “extremely urgent,” as the dependent variable. This approach reduced the bias associated with assigning an overall urgency rating after having first calculated a priority score (as in the New Zealand approach). Regression analysis of the responses permitted calculation of a set of weights that best reflected (or predicted) clinicians’ urgency scores.

Any such approach to validation is ultimately of uncertain value, however, given the subjectivity and low inter-observer reliability of overall urgency judgements. In particular, statistical measures of correlation between priority scores and clinical judgement must be viewed with caution. For example, low R-squared values (a common statistical measure of explanatory power) do not necessarily indicate that tools lack validity. Rather, low values may reflect, in large part, the fact the clinicians’ unaided judgements of urgency contain so much “noise” that they can be only incompletely modeled using statistical techniques. As such, statistical correlations of criteria scores with subjective judgements of urgency cannot provide the kind of validity available elsewhere, e.g., for mortality prediction models, which have a definitive endpoint available against which to assess accuracy.

Despite this limitation, it is useful to conduct analyses of the kind just described. By comparing priority scores against clinical judgements, a process of “bootstrapping” can begin, in which the analysis of global judgements improves the quality of those judgements, which in turn permits the clearer identification of factors that can be used to assist (and to assess) future judgements, and so on. This is a useful process that can potentially improve (e.g., make more consistent) clinical judgements of urgency and relative priority.

Empirical efforts that collect data of the kind just described are worthwhile from at least two other perspectives. First, clinicians gain experience with the tools and the process of filling out the forms, thus permitting an assessment of acceptability and face validity. Second, comparing criteria responses with overall judgements of urgency can reveal when criteria need to be added, deleted, or modified in order to make sense clinically. As noted earlier, clinicians are quite adept at selecting predictor variables—it is the case-by-case integration of these variables that is more difficult, which is where linear models excel.

One final point is worth making concerning the idea that priority criteria might somehow be validated against an indicator of “real urgency.” Like most linear models in medicine, priority criteria function as *clinical indexes*, which are often used in medicine to *define* clinical entities. For example, degrees of heart failure are defined in terms of Killip or New York Heart Association criteria and levels (e.g., Class I, II, III). Alvan Feinstein explains why the clinical indices themselves often constitute the definition for what the indexes are measuring:

Thus, there are no definitive standards for such indexes as the Apgar Score, the TNM [tumour, nodes, metastases] stages of cancer, the Glasgow Coma Scale, the New York Heart Association Functional Classification, the Katz Index of Activities of Daily Living, or any of the diverse clinical indexes that are cited in such global scales as **trace, mild, moderate, severe**, or **0, 1+, 2+, 3+, 4+**. In all of these examples, the rating given to the phenomenon is what we use because no definitive measurement exists.^{31 (p. 191)}

Regarding the validity of clinical indices, Feinstein notes that, “particularly for indexes that are intended to reflect the complex observations and dissected intuitions of clinical experience, face validity is often the most important attribute of the index.”^{30 (p. 156)} Priority criteria are indeed designed to reflect complex observations and dissected clinical intuitions of relative urgency and priority. Face validity is, therefore, of paramount importance in attempting to “validate” these criteria. From a statistical perspective, one linear model is pretty much as good as another (provided they share most criteria), as described above. This situation is not likely to change until meaningful information on actual patient outcomes becomes available.

Are Patients Who Score the Same Really the Same?

As discussed above, linear models produce total scores based on the sum of several sub-scores or criteria weights. Thus, the same total score can be arrived at by different routes. For example, Student A might have a higher SAT score than Student B, but a lower GPA, in which case total predictive scores (e.g., of graduation from college) might be the same (or very similar) for these two individuals. However, they would clearly be distinguishable.

In the case of priority criteria, most scores (except zero and 100) can be attained in a variety of ways. For example, Figure 2 describes findings from two dissimilar patients (using the WCWL hip and knee replacement criteria shown in Figure 1):

Figure 2. Two patients with identical scores despite being distinguishable

<i>Criterion</i>	<i>Description (Patient A)</i>	<i>Point s</i>	<i>Description (Patient B)</i>	<i>Point s</i>
Pain on motion	Severe	13	Moderate	6
Pain at rest	Moderate	8	Mild	3
Walking ability	< 1 block	4	Household ambulator	7
Functional limitations	Moderate	11	Severe	19
Abnormal findings (reduced range of motion)	Moderate	6	Severe	10
Potential for progression	Mild	4	Moderate	11
Ability to work, give care, or live independently	Immediately threatened	20	Threatened but not immediately	10
<i>Total score</i>		66		66

In this example, Patient A has a higher degree of pain both on motion and at rest relative to Patient B, as well as a more immediate threat to core role functions. This discrepancy is counterbalanced by a greater degree of restriction of walking and functional limitations in Patient B relative to Patient A, as well as more abnormal findings on examination and a great potential for progression. The result is an identical priority score.

Are these two patients really the same in terms of priority? How could we tell? As described earlier, absent a gold standard, the score itself becomes an indicator of severity, which must then be subjected to face validity testing. Of course, it could be argued that patients A and B are not really the same in terms of relative urgency, despite their being characterized by findings resulting in similar priority scores. This issue (of subjective similarity despite different ways of arriving at scores) has been discussed for many years. In 1947, LL Thurstone, one of the pioneers of subjective measurement theory and practice, considered the case of two introverts who scored the same on a scale of introversion but were nonetheless distinguishable in terms of how they manifested their introversion. Thurstone claimed that it would make no more sense to object to this situation than to object to the statement that two men have the same income on the grounds that their income derives from different sources.³²

Similarly, psychologist Paul Meehl described “certain misconceptions held by the more tender-minded clinicians which prevent clear thinking”:

For instance, there is still the misconception that mathematical descriptions of persons in terms of scores *require* that persons achieving identical scores should be identical or indistinguishable with respect to the traits so quantified. We sometimes hear this view expressed by such statements as “A human being is more than just a set of numbers.”...A cannon ball falling through the air is “more than” the equation $S = \frac{1}{2}gt^2$, but this has not prevented the development of a rather satisfactory science of mechanics.³³

Such considerations notwithstanding, it could happen that two patients with nearly identical scores are viewed as not being similar in priority for reasons that are clear and compelling reasons. This would imply a fault in the criteria or the in weights assigned to them. Or perhaps the weighting system is generally valid but breaks down in certain cases. For this reason, among others, it is important that the criteria be re-evaluated and revised periodically.

Summary

This paper has described certain technical considerations regarding point-count priority measures. Because these measures are in effect statistical linear models, they are well grounded in both theory and practice. Priority criteria in effect constitute “expected benefit prediction models.” Pilot testing is useful to ensure that the appropriate criteria are incorporated and to provide clinicians with relevant experience. Ultimate validation of the criteria must come through large-scale studies that determine the actual outcomes of patients. In the meantime, clinical sensibility is the principal basis for assessing the validity of priority criteria.

Word count: 3177

Members of the Steering Committee of the Western Canada Waiting List Project

Dr. Tom Noseworthy, Professor and Chair, Department of Public Health Sciences, University of Alberta, Edmonton, Alberta. Chair.

Dr. Morris L. Barer, Director, Centre for Health Services and Policy Research, and Professor, Department of Health Care and Epidemiology, University of British Columbia, Vancouver, British Columbia;

Dr. Charlyn Black, Co-Director, Manitoba Centre for Health Policy and Evaluation, and Associate Head and Associate Professor, Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba;

Ms. Lauren Donnelly, Acting Executive Director, Acute and Emergency Services Branch, Saskatchewan Health, Regina, Saskatchewan;

Dr. Isra Levy, Director, Health Programs, Canadian Medical Association, Ottawa, Ontario;

Mr. Steven Lewis, Partner, Access Consulting Ltd., Saskatoon, Saskatchewan;

Mr. John McGurran, Director, Western Canada Waiting List Project, and Research Associate, Department of Public Health Sciences, University of Alberta, Edmonton, Alberta;

Dr. Mark C. Taylor, Assistant Professor, Department of Surgery, University of Manitoba;

Mr. Darrell Thomson, Director, Economics and Policy Analysis, British Columbia Medical Association, Vancouver, British Columbia;

Mrs. Barbara Young, Regional Utilization Consultant, Clinical Evaluation Services, Calgary Regional Health Authority, Calgary, Alberta.

References

1. McDonald P, Shortt S, Sanmartin C, Barer M, Lewis S, Sheps S. Waiting Lists and Waiting Times for Health Care in Canada: More Management!! More Money??. Ottawa: Health Canada; July 1998.
2. Hadorn DC and the Steering Committee of the Western Canada Waiting List Project. Setting priorities for waiting lists: defining our terms. *CMAJ* 2000; 163:857-60.
3. Hadorn DC, Holmes AC. The New Zealand priority criteria project: Part I: Overview. *BMJ* 1997; 314:131-4.
4. Managing Orthopaedic/Ophthalmic Demand for Equity in Localities: The MODEL Project. Mid-Hampshire RHA, England. Draft 1999.
5. Frankel S, Eachus J, Pearson N, Greenwood R, Chan P, Peters TJ, et al. Population requirement for primary hip-replacement surgery: a cross-sectional study [see comments]. *Lancet* 1999; 353:1304-9.
6. Knaus WA, Draper EACH, Wagner DP, et al. APACHE II: A severity of disease classification. *Crit Care Med* 1985; 13:818-29.
7. Lemeshow S, Teres D, Pastides H, et al. A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985; 13:519-25.
8. Multicenter Postinfarction Research Group. Risk stratification and survival after myocardial infarction. *N Engl J Med* 1983; 303:331-6.
9. Merrilees MA, Scott PJ, Norris RM. Prognosis after myocardial infarction: results of 15 year follow up. *BMJ* 1984; 288:356-9.
10. Bonsel GJ, Klompmaaker IJ, Van't Veer F, et al. Use of prognostic models for assessment of value of liver transplantation in primary biliary cirrhosis. *Lancet* 1990; 335:493-510.
11. Eisenberg M, Hallstrom A, Bergner L. The ACLS score: predicting survival from out-of-hospital cardiac arrest. *JAMA* 1981; 246:50-2.
12. Lemeshow S, Teres D, Avrunin JS, et al. Predicting the outcome of intensive care unit patients. *J Am Stat Assoc* 1988; 83:348-56.
13. Wasson JH, Sox HC, Neff RK, et al. Clinical prediction rules: applications and methodological standards. *N Engl J Med* 1985; 313:793-9.
14. DeDombal FT, Horrocks JC, Walensley A. Computer-aided diagnosis and decision-making in the acute abdomen. *J Coll Phys Lond* 1975; 9:211-23.

15. Pozen MW, D'Agostino RB, Mitchell JB. The usefulness of a predictive instrument to reduce inappropriate admissions to the coronary care unit. *Ann Int Med* 1980; 92:238-42.
16. Goldman L, Weinberg M, Weisberg M, et al. A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain. *N Engl J Med* 1982; 49:1927-31.
17. Cebul RD, Poses RM. The comparative cost-effectiveness of statistical decision rules and experienced physicians in pharyngitis management. *JAMA* 1986; 256:3353-8.
18. Tompkins RK, Burnes DC, Cable WE. An analysis of the cost-effectiveness of pharyngitis management and acute rheumatic fever prevention. *Ann Int Med* 1977; 86:481-92.
19. Heckerling PS, Tape TG, Wigton RS. Clinical prediction rule for pulmonary infiltrates. *Ann Intern Med* 1990; 113:664-70.
20. Boom R, Chavez-Oest J, Gonzalez C, et al. Physicians' diagnoses compared with algorithmic differentiation of causes of jaundice. *Med Decis Making* 1988; 8:177-81.
21. Diehr P, Highley R, Dehkordi F. Prediction of fracture in patients with acute musculoskeletal ankle trauma. *Med Decis Making* 1988; 8:40-7.
22. Einhorn HJ. Expert measurement and mechanical combination. *Org Beh Hum Perf* 1972; 7:86-106.
23. Lee KL, Pryor DB, Harrel FE, et al. Predicting outcome in coronary disease: statistical models versus expert clinicians. *Am J Med* 1986; 80:553-60.
24. Dowie JJ, Elstein A. *Professional Judgement: A Reader in Clinical Decision Making*. Cambridge: Cambridge University Press; 1988.
25. Arkes HR, Hammond KR. *Judgement and Decision Making: An Interdisciplinary Reader*. Cambridge: Cambridge University Press; 1986.
26. Meehl PE. *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. Minneapolis: University of Minnesota Press; 1954.
27. Dawes RM, Corrigan B. Linear models in decision-making. *Psychol Bull* 1974; 81:95-106.
28. Sawyer J. Measurement and prediction, clinical and statistical. *Psych Bull* 1966; 66:178-200.
29. Dawes RM. The robust beauty of improper linear models. In Kahneman D, Slovic P, Tversky A. *Judgement Under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press; 1982. Originally in *American Psychologist* 1979; 34:571-82.

30. Dawes RM. You can't systematize human judgement; dyslexia. *New Directions for Methodology of Social and Behavioral Science* 1980 (4); 67-78. Reprinted in Dowie and Elstein, ref 21, p.150-62.
31. Feinstein AR. *Clinimetrics*. New Haven (CT): Yale University Press; 1987.
32. Thurstone LL. *Multiple Factor Analysis*. Chicago: University of Chicago Press; 1947.
33. Meehl PE. General remarks on quantification of clinical material. In Arkes HR, Hammond KR. *Judgement and Decision Making*. Cambridge: Cambridge University Press; 1986. P. 549-60.