Understanding the Role of Human Judgment in Applicant Selection for the Study of Medicine:

Achieving Share Values Using Interviews, Admission Committees, and Empirical Models

Clarence D. Kreiter, PhD

Visiting Professor

International Research Center for Medical Education –

University of Tokyo

Why Do We Use Human Judgment?

- We feel that important decisions about people should NOT be made mechanically.
- Value generally held in medicine
  - (e.g. profession judgment believed superior to decisions based on evidence-based practice).
- Value also strongly held in medical school admissions

Why Do We Use Human Judgment? (continued)

- Provides ability to consider the person as an individual rather than a number
- Provides power to decision makers
  - ("I don't want to lose my job to a computer!")
- Can serve to encourage financial donations

Why Do We Use Human Judgment? (continued)

- Public relations – recruitment
- Promote Diversity– (e.g. AAMC holistic review initiative - USA)
- To select applicant using qualitative / subjective methods
- Guide mission of medical education

How Do We Use Human Judgment in Admission?

- Interviews
- Admission Committees
- Defining Goals (mission statement)
- Selecting weights for combining information

Big Picture

- Some form of human judgment is inevitable
- Used in a wide variety of ways
- Current research addresses certain aspects of human judgment
- An evidence-based model is needed to optimally utilize human judgment

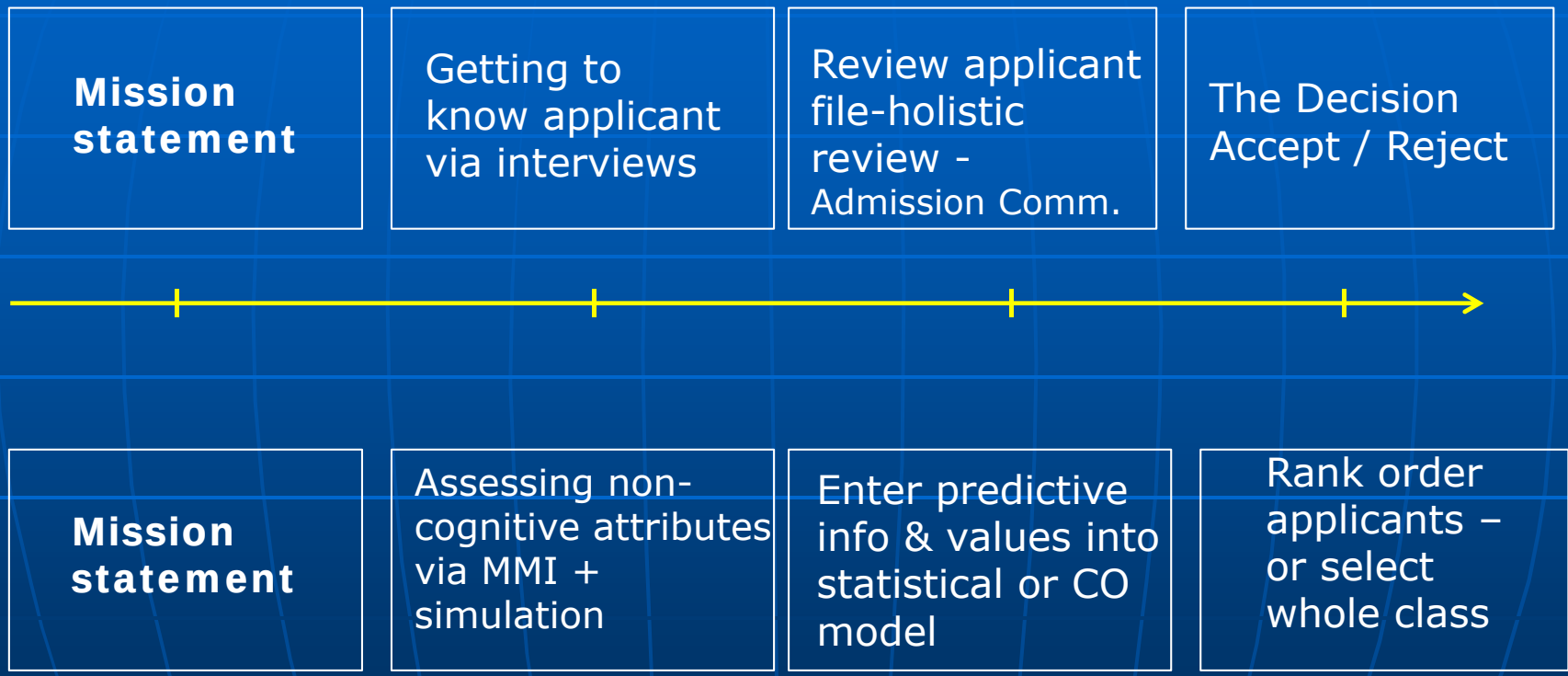
Research Questions

- What is the research evidence regarding human judgment in the selection process?
- How can human judgments be most effectively and efficiently incorporated into the admissions process?

Two Competing Models -

- Model 1* :
  - Going with the gut: (intuition / subjective judgment – values implicitly implemented)
    - *most widely used
- Model 2 :
  - An evidence-based model:(empirical / objective – values explicitly implemented)

Two models


Mission statement
Getting to know applicant via interviews
Review applicant file-holistic review - Admission Comm.
The Decision Accept / Reject
Mission statement
Assessing non-cognitive attributes via MMI + simulation
Enter predictive info & values into statistical or CO model
Rank order applicants – or select whole class

Mission Statement

- Selection Goals

  Identify Applicants:

  -who will succeed academically
  -who are psychologically fit
  -who are likely to engage in public service
  -who will make a contribution to Healthcare and Research
  -who will assume leadership roles in society
  -who are socially deserving
  -diversity
  -demographics

Two models

Mission statement	**Getting to know applicant via interviews**	Review applicant file-holistic review - Admission Comm.	The Decision Accept / Reject
Mission statement	**Assessing non-cognitive attributes via MMI + simulation**	Enter predictive info & values into statistical or CO model	Rank order applicants – or select whole class

The Interview

Questions Addressed Regarding the Medical School Preadmission Interview

- Does the interview work?
- What are the psychometric properties the interview?
  - Reliability
  - Validity
- Are there alternatives to interview?

Uses of the Interview

- Internationally most medical schools use the interview

Uses of the Interview (cont'd)

- Information gathering
  - Implies info beyond interview score conveyed to admissions committee-no documentation of this
- Decision making
  - Survey (self-report) studies suggests important role
  - Regression policy capturing suggests small role

Uses of the Interview (cont'd)

- Verification—no description of how to achieve this
- Recruitment—anecdotal—no evidence—further research needed
  - (Edwards, 1990)

Uses of the Interview (cont'd)

- Promote diversity
  - Used in conjunction with holistic review
  - Diverse faculty (interviewers) define diversity
- Identify Extreme Low Outlier
  - Conditional error (Stansfield & Kreiter, 2006)

Why the Interview?

- Perceived over-reliance on quantitative and cognitive information
- Easiest method of collecting non cognitive information
- Tradition

Why the Interview? (cont'd)

- Designers of admissions policy maintain it is necessary to meet the candidate.

- Too much numerical data.

  However, at most medical schools, interviews are translated into numbers. So still quantitative.

  -- Numerical summaries allowing an evaluation of validity and reliability.

  -- More numerical data.

Does the Interview Work?

- Is it validity?
- Modern validity theory requires a formal statement regarding the intended interpretation of the scores.
- For admissions testing the old fashion criterion model (i.e. predictive validity) is still the preferred approach. (Guion, 1998)

Does the Interview Work? (cont'd)

- Since the interview is translated into numerical scores, the validity question can be answered using statistical research.
- Relationship between reliability and validity.

$$\text{Validity} \leq \sqrt{r_{yy}}$$

Does the Interview Work? (cont'd)

$$\text{Validity} \leq \sqrt{r_{yy}}$$

- The maximum correlation of a measure with a perfectly reliable validity criterion cannot exceed the square root of the reliability.

Summary of Reported Reliability Since 1990

Study	Method	Coefficient	Effects	Blinded	Study Characteristics
Harasym et al. (10)	G Study/ Reinterview	9% p variance. G=.51 with 6 interviews	Rater & Occasions	Yes	Actors were used to portray applicant. Each interviewed 6 times
Collins et al. (11)	Interview/ Reinterview	Obtained r=.67	Occasions	Yes	Two different raters who reach consensus on each of two occasions using same questions
Shaw et al. (12)	Correlation between raters	r=.47 Blinded/.49 Not Blinded	Rater	Yes/No	Correlation between raters not impacted by academic information
Carrothers et al. (13)	Internal consistency alpha	Alpha=.66-.95	Items	Unknown	Interviewer form to measure emotional intelligence
Tutton et al. (14)	Internal consistency alpha	Alpha=.80	Items	Yes	Examines independent prediction provided by interview
Kulatunga-Moruzi et al. (15)	Correlation between raters	r=.66	Raters	Yes	Study assesses cognitive and non-cognitive predictors
Patrick et al. (16)	Inter-rater agreement	% agreement within one pt.=87-98%	Raters	Yes	Examines the structured interview
VanSusteren (17)	Inter-rater agreement	Kappa .13-.79	Raters	Yes/No	Academic file did not effect ratings. Kappa with two raters, but multiple methods of calculations
Eva et al. (6)	G Study	G=.65 Ten stations	Rater, Task, & Occasion	Yes	An OSCE style task and interview stations

Summary of Reported Reliability Since 1990 (cont'd)

- Lots of estimates of interview reliability

- The most informative reliability estimate would revolve around the question of how similarly would applicant's score if we repeated the entire interview process with new raters and questions on a different occasion.

Generalizability Study Results for the Multivariate ($r^{\bullet}:p^{\bullet}$) x $q^{\bullet}$ Design (Kreiter, Yin, Solow, Brennan 2004)

Facet	Occasion = 1		Occasion = 2		Occasion 1, 2	
	VC	SE	VC	SE	CovC	r^a
p	0.215 (26.5%)	0.057	0.124 (17.1%)	0.043	0.134	0.818
r:p	0.163 (20.1%)	0.034	0.148 (20.4%)	0.035		
q	0.000 [b] (0.0%)	0.002	0.008 (1.1%)	0.007		
pq	0.155 (19.1%)	0.028	0.091 (12.5%)	0.027		
rq:p	0.279 (34.3%)	0.024	0.355 (48.9%)	0.030		

Implications for Validity

- Reliability
  - .27
- Maximum Predictive Validity
  - .51

Predictive Validity

- Kulatunga-Moruzi and Norman (2002)
  - Criterion LMCC Part II (OSCE)
    - Communication Skills
    - Problem Exploration
  - Results
    - Low correlation with Comm. Skills (r = .24)
    - No unique contribution (incremental)

Predictive Validity (cont'd)

- Meta-Analysis by Goho and Blackman (2006)
  - 19 studies of academic prediction
    - Mean r was 0.06
  - 10 studies of clinical prediction
    - Mean r was 0.17

Improving Validity

- Structured vs. Unstructured
- Structured almost universally recommended, but….
- Structure misunderstood

Improving Validity (cont'd)

- Structure Reconsidered
  - Reliability (Kreiter et al. 2006)
    - Structure proved less reliable in G study
    - Spontaneous responses more likely with unstructured

What happens when we rely on interview data?

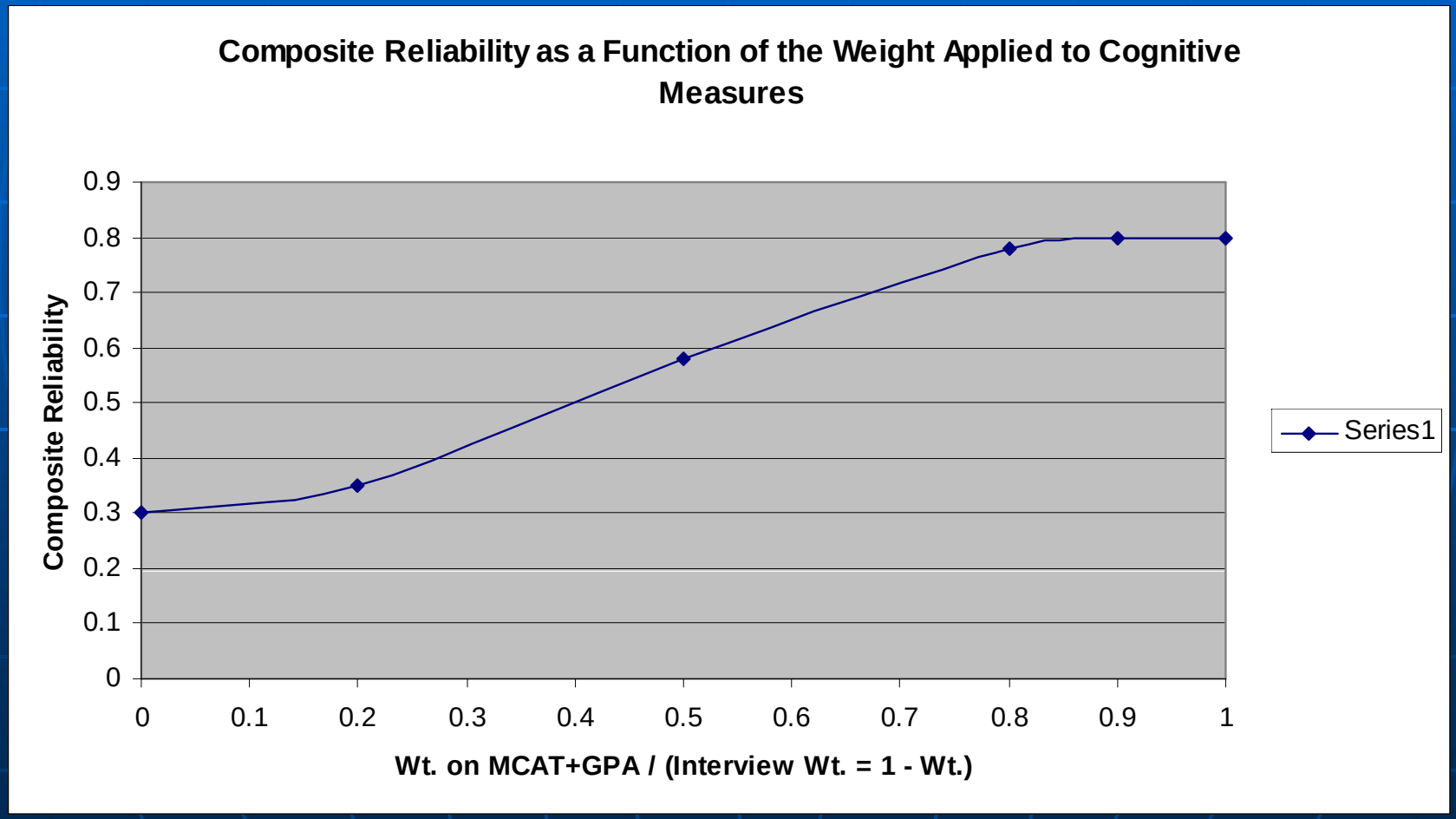
- Standardized Interview Scores
  - Reliability = .3
  - Mean = 50 and SD = 10
- Sum of Standardized GPA and MCAT (Cognitive)
  - Reliability = .80
  - Mean = 50 and SD = 10
- Correlation between Cog. And Interview
  - r = .15

Combining Cognitive and Interview Measures

$$r_c = 1 - \frac{\sum_v w_v^2(1-r_v)}{\sum_v w_v^2 + \sum_v \sum_{v'} w_v w_{v'} r_{vv'}}$$

Where:
r_c = the reliability of the composite score,
w_v = the weight for component v,
r_v = the sample estimate reliability for component v, and
$r_{vv'}$ = is the correlation between components v and v'.

Composite (Decision) Reliability Given Various Weights on Cognitive Measures



Interview Score as Supplement to Cognitive Measures

- Not useful for those above a set cut score (Albanese et al.) (Kreiter et al. 2006)
- Assign a very small weight in selection
- Best Advice – **Don't use traditional interview data as part of decision process!**

Alternatives to Interview

- MMI or simulation (Eva et al.) (Ziv)
- Multiple interviews (Kreiter & Axelson, 2009)
- Personality tests

Screening Candidates Using MMI and Simulation-Based Assessment

- Simulation of common medical encounters –no medical knowledge required
- Applying OSCE measurement method to Interview

MMI and Simulation Screening
(cont'd)

- Types of MMI and simulations
  - Actor portraying aggressive patient
  - Counseling a simulated friend who applicant has been told is a bus driver who often drinks on the job
  - Candidates work together to solve a common problem (i.e. solve a hospital budgeting problem)

MMI and Simulation Screening
(cont'd)

- Four Constructs Assessed
  - Communication Skills
  - Stress Management
  - Initiative/responsibility
  - Self-Awareness

MMI and Simulation Screening
(cont'd)

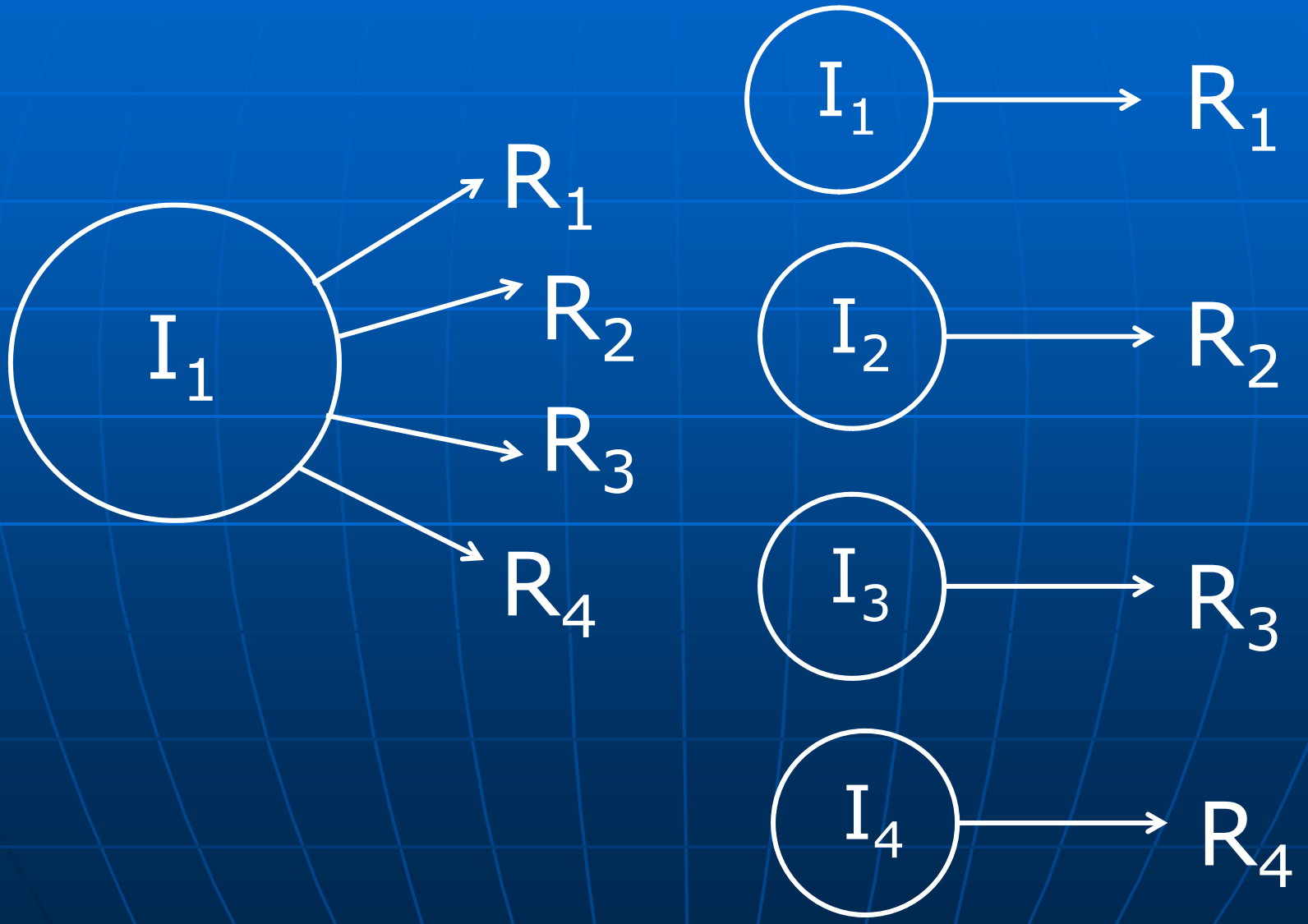
- Results
  - Reliability in .70 range for an across station calculation (8 stations)

- Validity studies are underway
  - Initial evidence is positive

Multiple Interviews

- One rater per interview much more effective than panel interview
- Multiple interviews cost the same as panel interview
- Much better reliability (Axelson & Kreiter)
  - G Studies

Panel Interview vs. Multiple Interviews



Personality Test

- Currently no positive evidence for using
- High Stakes personality test illogical and invalid

Two models

Mission statement	Getting to know applicant via interviews	**Review applicant file- holistic review Admission Comm.**	The Decision Accept / Reject
Mission statement	Assessing non-cognitive attributes via MMI + simulation	**Enter predictive info & values into statistical or CO model**	Rank order applicants – or select whole class

Third Step in Model

Admission Committee Holistic Review

Versus

Statistical Formula-Based Prediction or Constrained Optimization

How to decide who studies medicine?

- Evidence regarding prediction
- Values held by medical college
- How to balance values
- We know optimal selection important

How to Decide (cont'd)

- Many potential solutions can be regarded as 'optimal' depending on mission statement
- However we cannot justify 'sub-optimal'

Current Trends

- AAMC – holistic review
- U.S. Supreme Court – holistic review

Making the decision

- 50 years of research:

- 1. Meehl, P.E. (1954) *Clinical vs. statistical prediction: A theoretical analysis and a review of the evidence.* Minneapolis: University of Minnesota Press.

- 2. Schofield, W., Garrard, J. Longitudinal study of medical students selected for admission to medical school by actuarial and committee methods. *British Journal of Medical Education*, 1975;9:86-90.

- 3. Dawes, R.M., Faust, D., Meehl, P.E. Clinical versus actuarial judgment. *Science*, 1989;243:1668-74.

How Decision is Made

- Holistic vs. Actuarial (McGaghie & Kreiter, 2006) (logic of holistic admission)
  - Going with gut widely popular
  - Lots of solid evidence that it doesn't work

A second Look

- Why so popular?
- Are the social scientists missing something regarding holistic review?
- Could it still be valid in some way?

A second look (cont'd)

- Past research does not utilize all possible outcome variables
  - Social responsibility
  - Ethical characteristics
  - Leadership roles assumed

A second look (cont'd)

- Many aspects of file cannot be quantified.
  - Interpreting letters of recommendation
  - Family circumstances
  - Life history

A second look (cont'd)

- Could past research have missed something?
- Is possible that the benefits derived from committee decisions are not revealed in outcomes that are commonly or easily measured?

Nature of the Research

- For optimizing a defined measureable outcome, there is a consistent body of research demonstrating that statistical methods yield superior decisions to those generated by holistic judgments of raters. However, it is possible that the benefits of holistic committee decisions could impact other, unmeasured, outcome variables. If such benefits exist, they would necessarily appear as systematic variance in raters' scores beyond the portion captured by statistical approaches. (Kreiter & Axelson, 2011)

Why should we care?

- Validity of final decision is what is of paramount importance
- Combining information inappropriately can decrease predictive validity – compromise reliable and predictive measures

Why should we care? (cont'd)

- High Cost in Time and ¥
  - 74,000 in US interviewed and Admission Committee reviewed in US alone
  - Average 525 per school – each with 3 reviewers
  - Highest paid faculty do reviewing

Why should we care? (cont'd)

- Very Important Decision - Equivalent to deciding who will be tomorrow's physicians
  - Attrition < 3%

Final Decision is the most important and consequential

Validity Research Design Regarding Admission Committee Contribution

- (practical question) - Can we find any evidence that committee reviews serve a useful function?
- (scientific measurement question) – Can we generate validity evidence?

Validity Research Design Regarding Admission Committee Contribution

- We know from well established research that a committee inferior in selecting to maximize measureable outcomes.
- How about unmeasured outcomes?
- Of course if outcomes impossible measure, it is also impossible to evaluate the success of decision process

Validity Research Design Regarding Admission Committee Contribution

- A Descriptive Equation:

$$\text{Comm. Decision} = [X + (\beta_1 * \text{AdmTest}) + (\beta_2 * \text{HSGPA}) + (\beta_3 * \text{Interview})]$$

- A Difference Score

DiffSco = (*statistical rankings*) – (*committee member ranking*)

Calculation of Difference Scores

Statistical Descriptive Ranking		*Random Sample for One Committee Member*		*Committee Member Ranking-Raw*		*Difference from Statistical Rank*		*Difference Score Data*
450								
449		400 (6)		300 (5)		6-5		+1
.		300 (5)		400 (6)		5-6		-1
.		250 (4)		250 (4)		4-4		0
.		50 (3)		15 (1)		3-1		+2
.		20 (2)		50 (3)		2-3		-1
2		15 (1)		20 (2)		1-2		-1
1								

Validity Research Design Regarding Admission Committee Contribution

- Outcomes and interpretations
  - Rater agreement on difference scores equal zero – certain evidence that committee adds random error
  - Rater agreement on difference scores great than zero - implies but does not prove committee adds useful info.
  - Regardless nature of judgments very difficult know

Validity Research

- Such a study is crucial to demonstrating the validity of current techniques
- Most schools already have the data or some variant of this data.

Alternatives to Admission Committee

- Linear weighting model
  - Usually based on regression
  - Based on statistical prediction
  - Usually optimal in some sense (human judgment)
- Constrained Optimization
  - Optimal as defined by human judgments

Linear vs. Constrained Optimization

- Linear
  - Licensure Score =

$[c + (\beta_1 * \text{Test}) + (\beta_2 * \text{GPA}) + (\beta_3 * \text{other})]$

Versus

- Nonlinear

(Kreiter & Solow, 2002)
(Kreiter, 2002)
(Kreiter, Stansfield, James, Solow, 2003)

Constrained Optimization

- Linear and nonlinear programming
- Primum Solver®
- Easiest to think of as all possible combinations of applicants

Constrained Optimization

- Many opportunities to insert human judgment
- Shapes class characteristics rather than making each applicant satisfy all criteria
- Scott Page – Diverse Groups

Two models

Mission statement	Getting to know applicant via interviews	Review applicant file-holistic review Admission Comm.	**The Decision Accept / Reject**
Mission statement	Assessing non-cognitive attributes via MMI + simulation	Enter predictive info & values into statistical or CO model	**Rank order applicants – or select whole class**

Conclusions

- Considerable evidence that Model 2 outcomes are superior
- Many valid ways that human judgment can enter into selection process
- Many differences in how human judgment used in two models
- Also many invalid methods of incorporating human judgments

Discussion…….

- What are the biggest challenges / shortcoming in current admission procedures?
- Professor Nishigori

Question

- What is the biggest shortcoming in the methods used to select applicants for the study medicine in Japan?

References

1. Johnson, E.K. & Edwards, J.C. Current practices in admission interviews at US medical schools. Acad. Med. 1991; 66:408-412.

2. Nayer, M. Admission criteria for entrance to physiotherapy schools: How to choose among many applicants. Physiotherapy Canada 44: 41-46, 1992.

3. Kreiter, C.D., Yin, P., Solow, C. & Brennan, R.L. Investigating the reliability of the medical school admissions interview. Advances in Health Science Education, 2004; 9:147-159.

4. VanSusteren, T. J., Suter, E., Romrell, L.J., Lanier, L. & Hatch, R.L. Do Interviews Really Play an Important role in the Medical School Selection Decision? Teaching and Learning in Medicine, 11(2), 66-74, 1999.

5. Puryear, H.G. & Lewis, L.A. Description of the interview process in selecting students for admission to U.S. medical schools. Journal of Medical Education, 1981; 56: 881-885.

6. Kulatunga-Moruzi, C. & Norman, G.R. Validity of Admission Measures in Predicting Performance Outcomes: The Contribution of Cognitive and Non-cognitive Dimensions. Teaching and Learning in Medicine, 14(1), 34-42, 2002.

7. Salvatori, P. Reliability and validity of admissions tools used to select students for the health professions. Advances in Health Science Education, 6: 159-175, 2001.

References (cont'd)

8. Norman, G.R. The fiscal and psychic cost of admissions. Advances in Health Science Education, 6: 89-91, 2001.

9. Edwards, J.C., Johnson, E.K. & Molidor, J.B. The interview in the admission process. Academic Medicine, 65, 167-177, 1990.

10. Conway, J.M., Jako, R.A. & Goodman, D.F. A meta-analysis of interrater and internal consistency reliability of selection interviews. Journal of Applied Psychology, 80(5), 565-579, 1995.

11. Huffcutt, A.I. & Arthur, W. Hunter and Hunter (1984) revisited: interview validity for entry-level jobs. Journal of Applied Psychology, 79(5), 565-579, 1994.

12. Eva, K.W., Rosenfeld, J., Reiter, H.I., Norman, G.R. An admissions OSCE: the multiple mini-interview. Medical Education, 2004; 38(3):314-26.

13. Eva, K.W., Reiter, H.I., Rosenfeld, J., Norman, G.R. The ability of the multiple mini-interview to predict preclerkship performance in medical school. Academic Medicine, 2004; 79(10 Suppl):S40-2.

14. Salvatori, P. Reliability and validity of admissions tools used to select students for the health professions, Advances in Health Science Education, 6: 159-175, 2001.

References (cont'd)

- 15.Kreiter, C.D., Solow, C., Brennan, R.L., Ping, Y., Ferguson, K., Huebner, K. Examining the influence of using same versus different questions on the reliability of the medical school preadmission interview. Teaching and Learning in Medicine, 2006, 18(1), 4-8.
- 16.Guion, R (1998) *Assessment, Measurement, and Prediction for Personnel Decisions. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.*
- 17. Stansfield, R. B., Kreiter, C.D. Conditional Reliability of Admissions Interview Ratings: Extreme Ratings are the Most Informative, Medical Education, 41:31-37, 2007.
- 18. Kreiter, C.D. A Commentary on the Use of Cut-Scores to Increase the Emphasis of Non-Cognitive Variables in Medical School Admissions, Advances in Health Sciences Education, August 2006.
- 19. Kreiter, C.D., Stansfield, B., James, P.A. and Solow, C. A Model for Diversity in Admissions: a Review of Issues and Methods and an Experimental Approach. Teaching and Learning in Medicine, 15(2):116-22, Spring, 2003.
- 20. McGaghie, W.C., Kreiter, C.D. Holistic Vs. Actuarial Student Selection. Teaching and Learning in Medicine, 17(1):89-91, Winter 2005.

References (cont'd)

- 21. Albanese, M.A., Farrell, P. & Dottl, S.L. (2005) Statistical Criteria for Setting Thresholds in Medical School Admissions. Advances in Health Sciences Education 10: 89-103.
- 22. Albanese, M.A., Farrell, P. & Dottl, S.L. (2005) A Comparison of Statistical Criteria for Setting optimally Discriminating MCAT and GPA Thresholds in Medical School Admissions. Teaching and Learning in Medicine 17(2): 149-158.
- 23. Kreiter, C. D., Kreiter, Y. A Validity Generalization Perspective on the Ability of Undergraduate GPA and the Medical College Admission Test to Predict Important Outcomes, Teaching and Learning in Medicine, 19(2):95-100, 2007.
- 24. Goho, J., Blackman, A. The effectiveness of academic admission interviews: an exploratory meta-analysis. Medical Teacher, 28, 4, 335-340, 2007.
- 24. Ziv A.R., Moshinsky, A., Gafni, N., et al. MOR: a simulation-based assessment centre for evaluating the personal and interpersonal qualities of medical school candidates. Medical Education, 42(10):991-8, 2008.
- 25. Axelson, R.D., Kreiter, C.D. Rater and occasion impacts on the reliability of pre-admission assessments. Medical Education, 43(12):1198-1202, 2009.
- 26. Kreiter, C.D., Axelson, R.D. A proposal for evaluating the validity of holistic-based admissions processes – In press
- 27. Axelson, R.D., Kreiter, C.D., Ferguson,K.J., Solow, C.M., Huebner, K. Medical school preadmission interviews: Are structured interviews more reliable than unstructured interviews? Teaching and Learning in Medicine, *22(4):241-5, 2010.*

References (cont'd)

- 28. Kreiter, Stansfield, James, Solow (2003) A model for diversity in admissions: A review of issues and methods and an experimental approach. Teaching and Learning in Medicine 15(2): 116-122.
- 29. Kreiter, C.D. (2002) The use of constrianed optimization to facilitate admission decisions. Academic Medicine 77(2): 148-151.
- 30. Kreiter, C. D., Solow C. (2002) A statistical technique for the development of an alternate list when using constrained optimization to make admission decisions. Teaching and Learning in Medicine, 14(1):29-33.
- 31. The Difference – *How the power of diversity creates better groups, firms, schools and societies* – by Scott Page, Princeton University Press – © 2007