



**Canadian Perspectives:  
Applied Behavioral Science in  
Support of Intelligence Analysis**

David R. Mandel

Thinking, Risk, & Intelligence Group

Adversarial Intent Section

DRDC Toronto



Defence Research and  
Development Canada

Recherche et développement  
pour la défense Canada

Canada



## Acknowledgements

- Sponsorship:
  - Capt(N) Barber, D. Int. Capability, CDI
- Partnership:
  - Alan Barnes, D. Middle East & Africa Div., International Assessment Staff, PCO
  - WO Ron Wulf, Canadian Forces School for Military Intelligence
  - DRDC Toronto: Lianne McLellan, Natalia Derbentseva, & Nada Pavlovic
- Presentation to Public workshop of the National Academy of Sciences Committee on Behavioral and Social Science Research to Improve Intelligence Analysis for National Security, May 15, 2009, Washington, D.C.
- Correspondence to: [David.Mandel@drdc-rddc.gc.ca](mailto:David.Mandel@drdc-rddc.gc.ca)



## Outline

- Canadian perspectives – TRIG
- Study aimed at validating the effectiveness of training to improve probabilistic belief revision.
- Study aimed at validating the quality of predictive assessments from an intel assessment division.
- Closing comments, questions and discussion



## Outline

- **Canadian perspectives – TRIG**
- Study aimed at validating the effectiveness of training to improve probabilistic belief revision.
- Study aimed at validating the quality of predictive assessments from an intel assessment division.
- Closing comments, questions and discussion



## Canadian perspectives - TRIG

- Thinking, Risk, and Intelligence Group (TRIG) is unique in Canada given its objective:
  - *To support Canada's defence and security community through applied behavioral science aimed at promoting human effectiveness in risk management and intelligence production.*
  - stood up Jan 2008.
  - has 4 behavioral scientists
  - 3-4 yr projects, each \$250K/yr



## Outline

- Canadian perspectives – TRIG
- **Study aimed at validating the effectiveness of training to improve probabilistic belief revision.**
- Study aimed at validating the quality of predictive assessments from an intel assessment division.
- Closing comments, questions and discussion



## Training development and validation

- In partnership with CFSMI, we are testing the effectiveness of new training modules developed at DRDC aimed at improving various aspects of analytic judgment.
- This study examines the effectiveness of a brief 20-min training protocol for improving Bayesian reasoning.
- Pretest – training – posttest design.



## Example 1: intelligence judgment

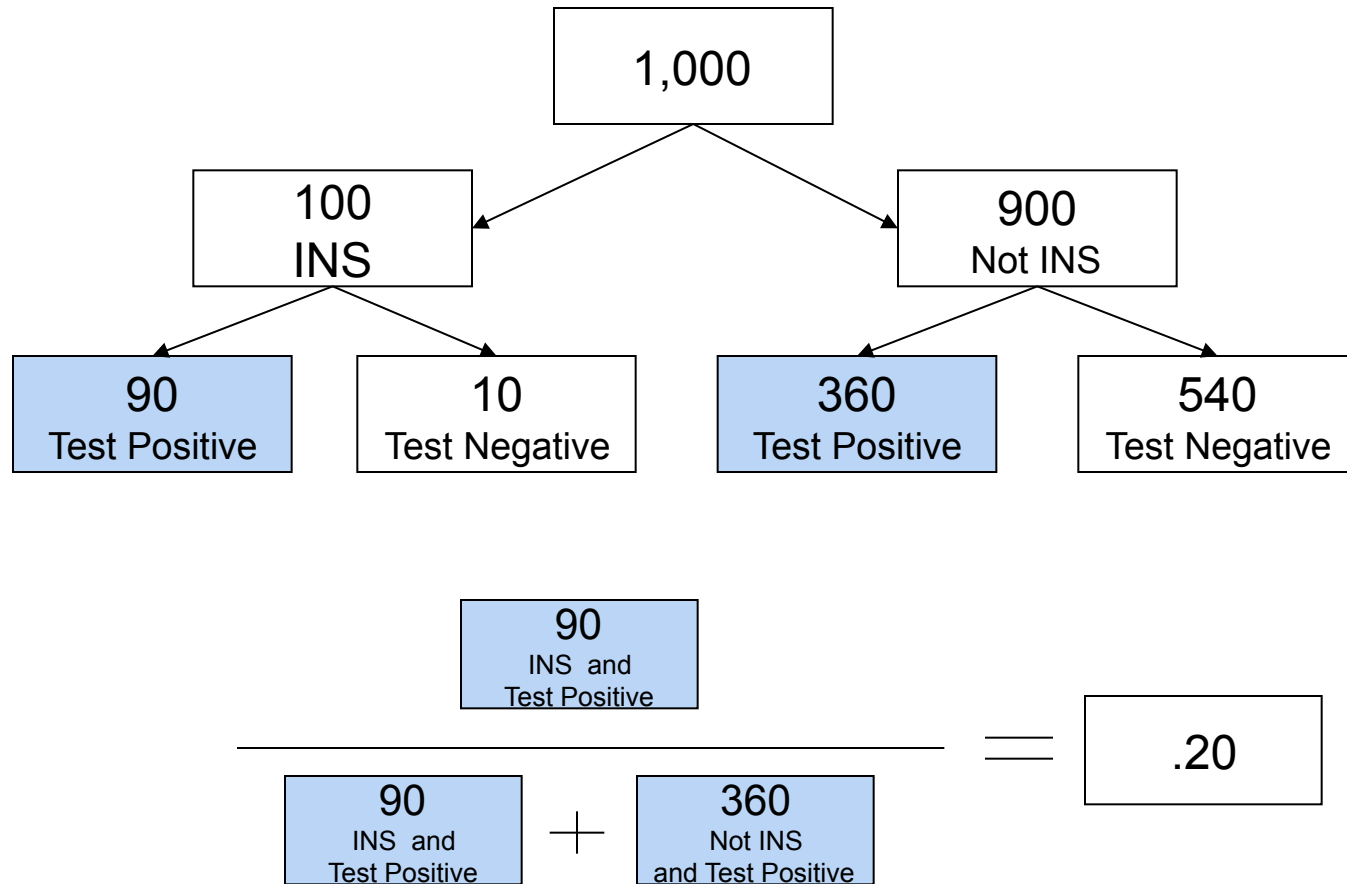
- Imagine that you are an analyst and you suspect that one of your targets (T) might be an insurgent (INS). T has been tested for explosive residue and the test result was positive. Now, you would like to know the probability of his actually being INS given that he tested positive. Assume the following statistics are accurate: 10% of the population in this area are INS, and 90% of people who are INS test positive on the same test you gave to T. As well, 40% of people who are not INS also test positive for explosive residue. Thus the test is not a 100% accurate diagnostic for INS. It makes an incorrect diagnosis 10% of time for people who are INS and 40% of time for people who are *not* INS.

If you wanted to assess the probability that T is INS given that he tested positive, how would you go about doing that?





# Graphical representation of the evidence





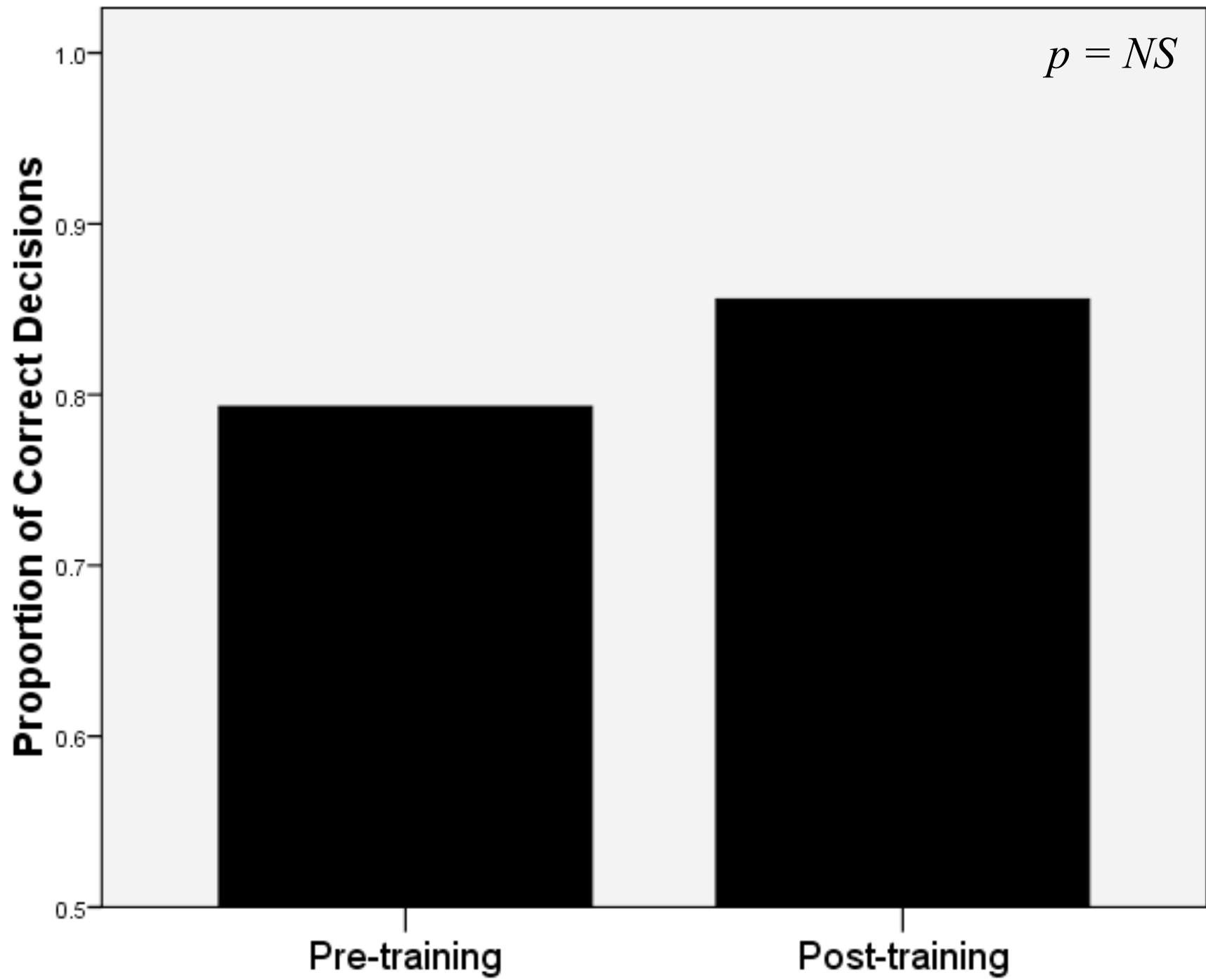
## Pre- and post-training task

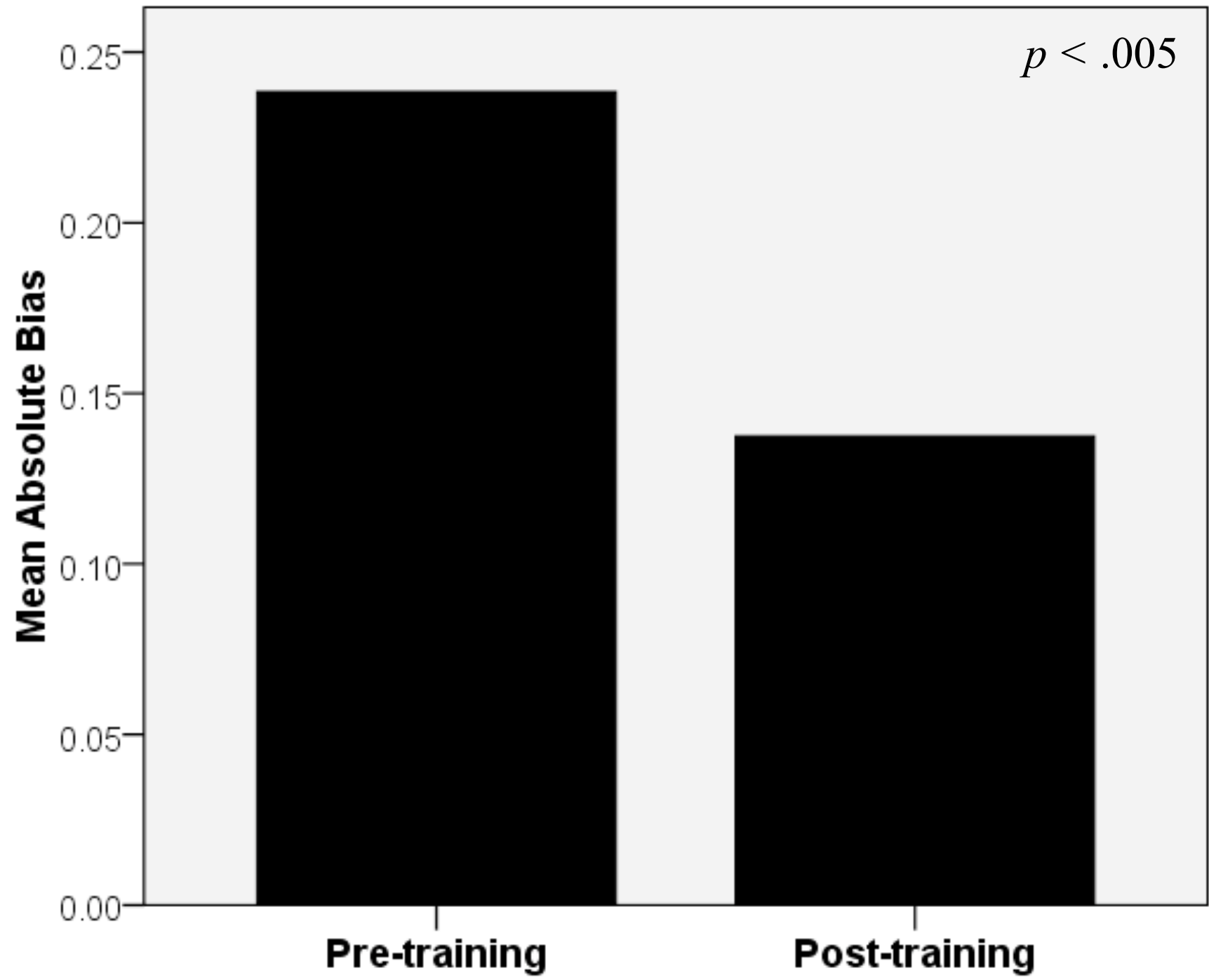
- As a contestant in this game show, you will meet eight “mystery people” one person at a time. In each case, your task is to classify the mystery person into one of two categories – namely, Group A or Group B.
- Each of these groups has 100 members in total. The eight mystery people that you will meet have been randomly selected from the total pool of members in Groups A and B.
- After a mystery person is introduced to you, the game-show host will ask the mystery person a question about a characteristic that he or she might possess. The mystery person will answer the question – *always honestly* – by either saying YES or NO. Then the host will tell you the % of people in Groups A and B who have that characteristic.
- Subjects judge  $P(A)$ ,  $P(B)$ , pick A or B.

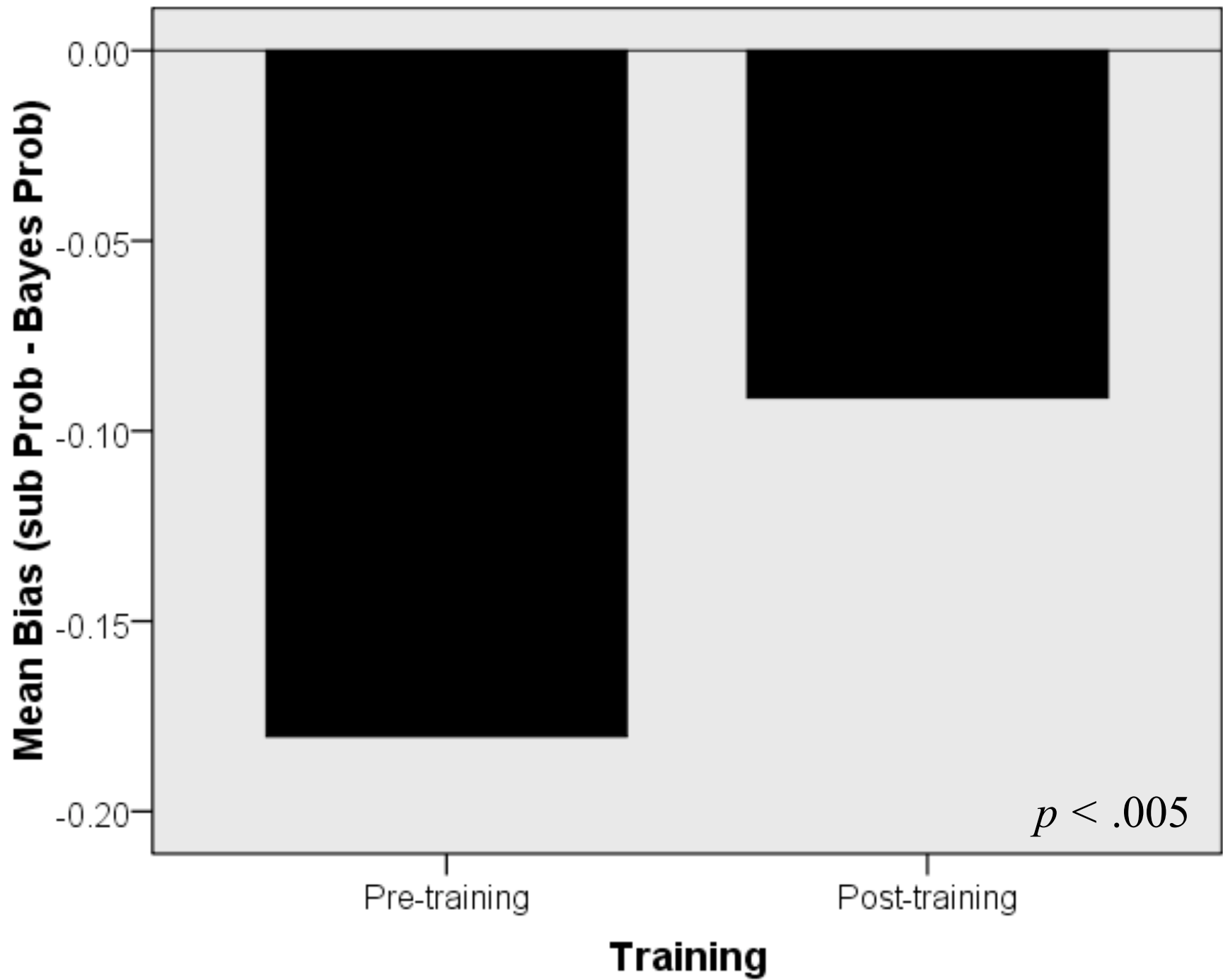


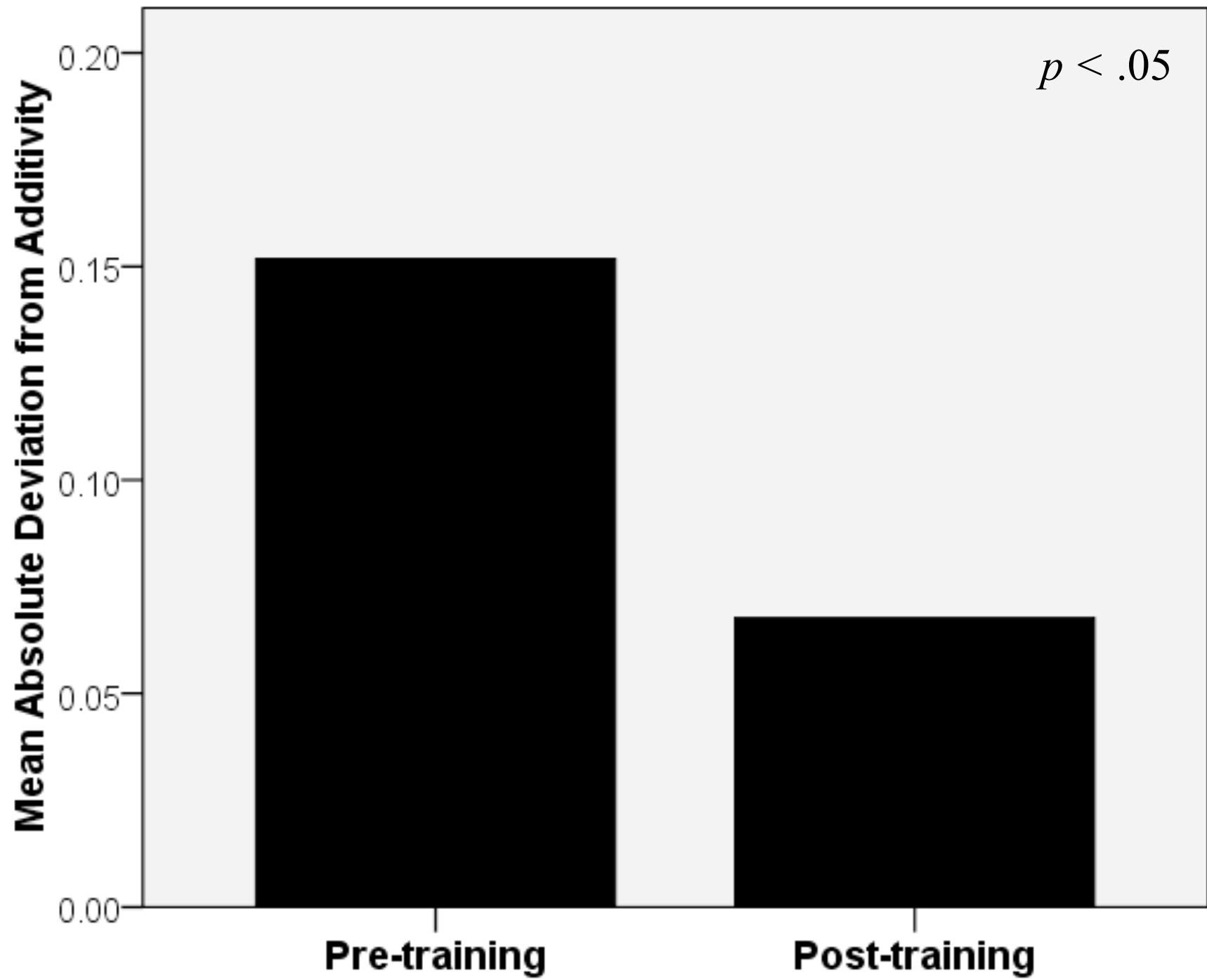
## Sample

- 32 trainees in military intelligence analysis training courses.
- 19 from junior course (entry course for INT branch) and 13 from senior course (strategic defence intelligence analysis course).
- Found no difference in either accuracy or coherence pre- or post-training.











## Summary

- Initial tests show promise of training to improve accuracy and coherence of analysts' judgments.
- Next step in June is to run study with one week post-training follow up to measure skill retention.
- One benefit: demonstrates to instructors value of validation research.





## Outline

- Canadian perspectives – TRIG
- Study aimed at validating the effectiveness of training to improve probabilistic belief revision.
- **Study aimed at validating the quality of predictive assessments from an intel assessment division.**
- Closing comments, questions and discussion



## Validation of predictive judgments

- Assessments of intelligence effectiveness often come in the form of autopsies of failures or near misses.
- Aim here was to examine the quality of predictive judgments of a strategic intelligence assessment division using
  - standard quantitative methods
  - in a situation where the assessment was not motivated by failure.



## Study data

- Judgment data: 649 predictive judgments abstracted from 51 Intelligence Memoranda produced by IAS Middle East and Africa Div from Mar 05 - Oct 06.
- Nine analysts produced these judgments, but the level of analysis here is at the division level.
- Outcome data: binary codes of whether or not the predicted event occurred in the relevant timeframe.
- Of the 649 judgments, 580 (89%) could be coded in terms of outcome (the remainder were ambiguous and were excluded from the analysis).



## Quantification of uncertainty in the division

- The IAS analysts are routinely directed to think first in terms of the numeric estimate on a 0/10 to 10/10 uncertainty scale.
- Only then for reporting purposes do they translate that estimate into a congruent verbal expression.
- An advantage of this approach is that accuracy of predictions can be assessed using precise measures of judgment quality.

**PROBABILITY TERMS**

Verbal Expression	Probability	Remarks
will is certain	[10/10]	where you can envisage no plausible scenario— however remote—where this event would not happen
almost certain extremely likely highly likely	[9/10]	there remains some conceivable scenario—albeit very remote—that this event would not happen
likely probable, probably	[7-8/10]	
slightly greater than even chance	[6/10]	use rarely, only when there is a specific reason to judge the probability at greater than even but cannot be categorized as "likely"
even chance	[5/10]	
slightly less than even chance	[4/10]	use rarely, only when there is a specific reason to judge the probability at less than even but cannot be categorized as "unlikely"
unlikely (only a) low probability probably not	[2-3/10]	
very unlikely highly unlikely extremely unlikely little prospect	[1/10]	there remains some conceivable scenario—albeit very remote—that this event could happen
no prospect will not	[0/10]	where you can envisage no plausible scenario— however remote—where this event could happen



## Measures of judgment quality

- **Calibration:** an analyst is well calibrated if his or her probability judgments of event occurrence match the relative frequencies of event occurrence.
- **Discrimination:** An analyst's discrimination skill is revealed in his or her ability to differentiate instances when an event is going to occur from those when it is not.
- In effect, discrimination is directly proportional to explained variance in the outcome variable.

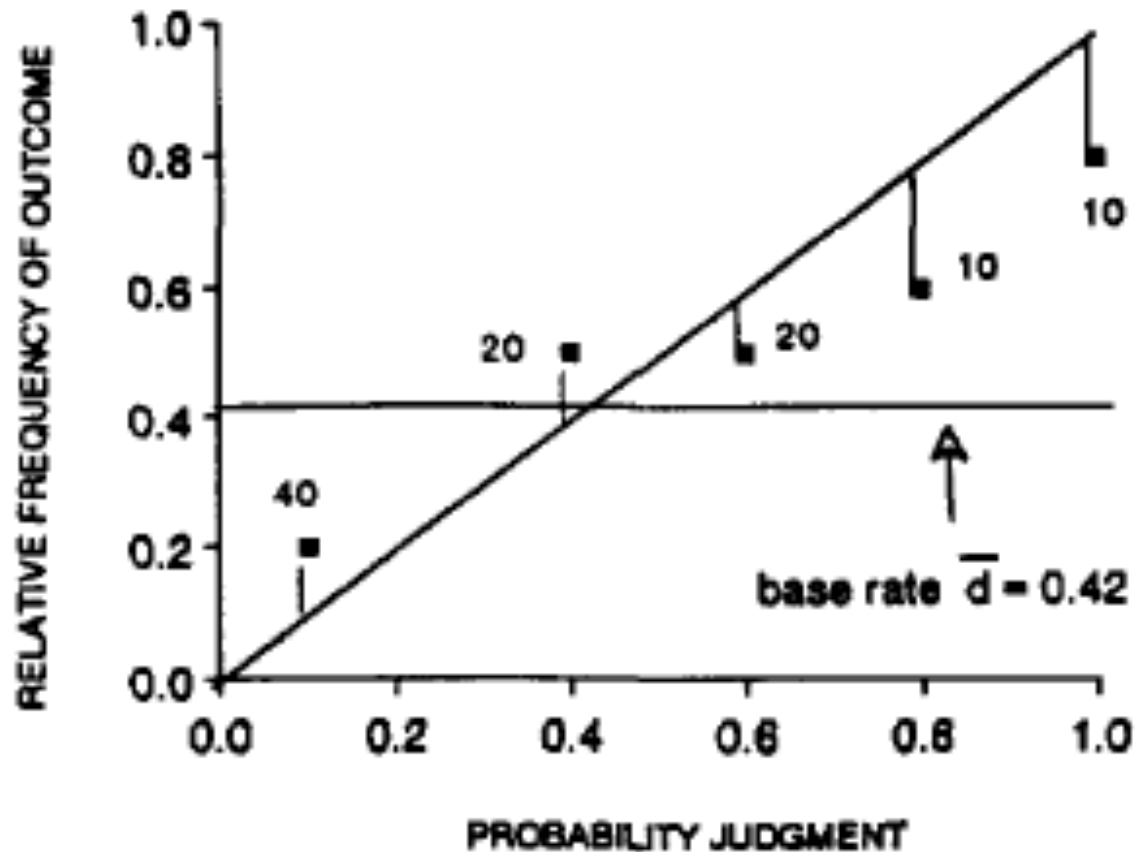


Figure 2. The relative frequencies or proportions,  $\bar{d}_1, \bar{d}_2, \dots, \bar{d}_j$ , plotted as a function of the judged probabilities,  $f_1, f_2, \dots, f_j$ . The numbers inside the graph indicate the number of observations contributing to each point. Calibration skill is a function of the distances of  $\bar{d}_1, \bar{d}_2, \dots, \bar{d}_j$  from the 45° perfect calibration line.

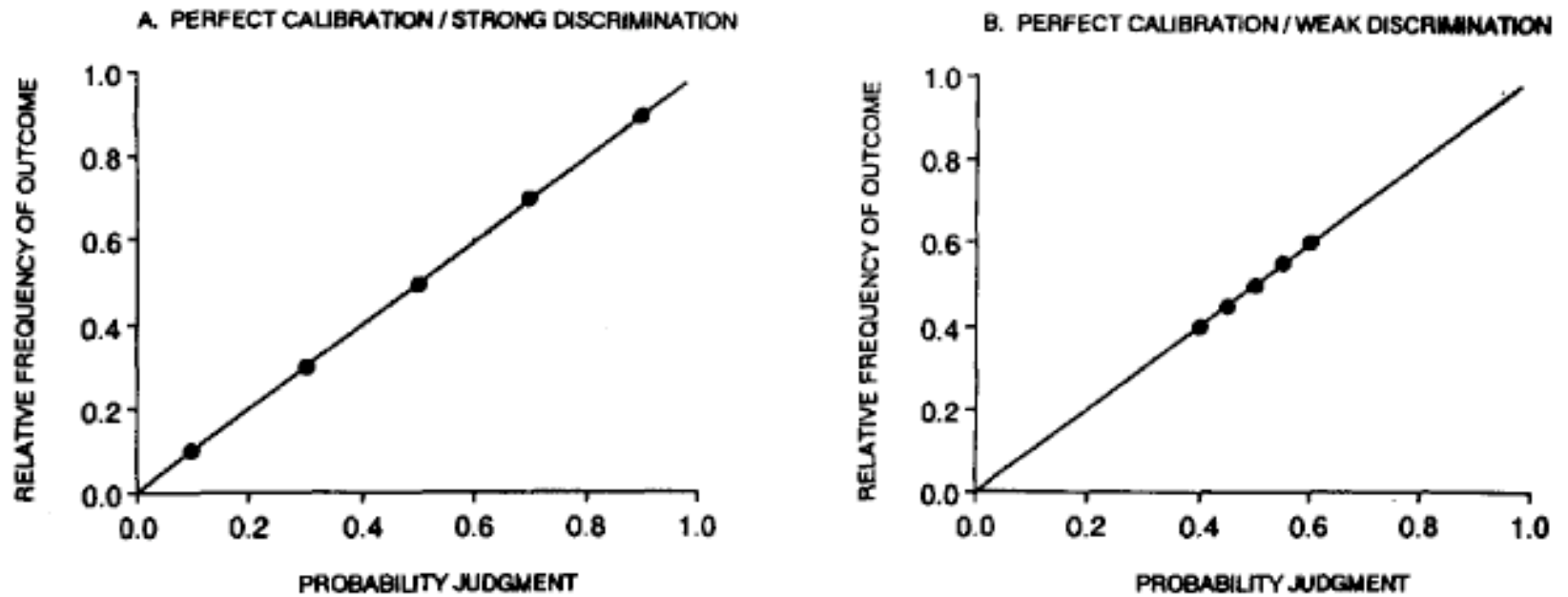
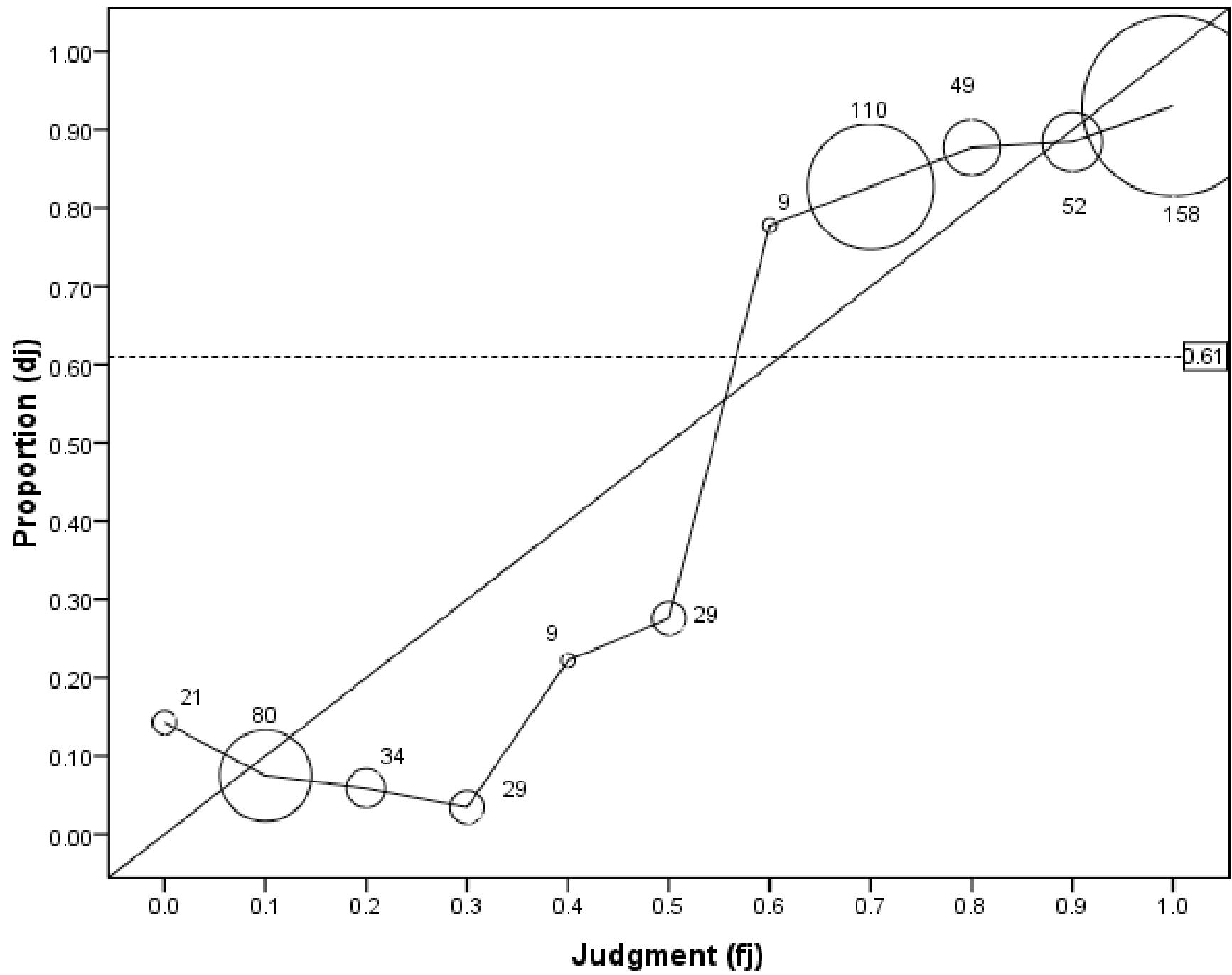


Figure 3. Representations of the assessments of Judges A and B who are equally (and perfectly) well calibrated. (Judge A's discrimination skill is better than that of Judge B)

From Yaniv, Yates, & Smith, 1991 *Psychol Bull* 110, 611.







## Quantitative index of calibration

- Calibration index,  $CI = \frac{1}{N} \sum_{j=1}^J N_j (f_j - \bar{d}_j)^2$ .
- CI ranges from 0 (perfect calibration) to 1.0 (worse possible).
- In the present research,  $CI = .014$
- This represents a very high degree of calibration.



## Quantitative index of discrimination

- Adjusted normalized discrimination index, ANDI

$$\text{NDI} = \frac{\frac{1}{N} \sum_{j=1}^J N_j (\bar{d}_j - \bar{d})^2}{\bar{d}(1 - \bar{d})} = \frac{\text{DI}}{\text{Var}(d)} = \eta^2.$$

$$\text{ANDI} = 1 - \frac{(1 - \text{NDI})}{(1 - \text{NDI}_0)} = \frac{N \cdot \text{NDI} - J + 1}{N - J + 1}$$

- ANDI ranges from 0 (no variance explained) to 1 (all variance explained).
- In the present research, ANDI = .583 (analysts are accounting for ~60% of the variance in outcomes).



Classification Table<sup>a</sup>

Observed			Predicted		
			Occurance: Partials Ignored		Percentage Correct
			No	Yes	
Step 1	Occurance: Partials Ignored	No	172	32	84.3
		Yes	22	324	93.6
	Overall Percentage				90.2

a. The cut value is .500



## Summary

- Quality of analytic judgments was “good to excellent” across the various indices measured.
- However, the causal bases for this high performance are currently unknown.
- Since calibration did not differ between junior and senior analysts, experience seems an unlikely explanation.



## Outline

- Canadian perspectives – TRIG
- Study aimed at validating the effectiveness of training to improve probabilistic belief revision.
- Study aimed at validating the quality of predictive assessments from an intel assessment division.
- **Closing comments, questions and discussion**



## Closing comments

- There is a need for applied behavioral science to improve intelligence analysis--
- *“Intelligence analysis--the assessment of complex national security issues shrouded by gaps in authentic and diagnostic information--is essentially a mental and social process.... Effective management of the impact of cognitive biases and other psychological challenges to the analytic process is at least as important in ensuring the soundness of assessments on complex issues as the degree of substantive expertise invested in the effort.”*

» Jack Davis (2008)



## Closing comments

- But there are also many challenges—
  - Need robust partnerships
    - Trust and effective communication
    - Shared understanding of roles
    - Institutional support and coordination
    - Flexibility and patience



DEFENCE



DÉFENSE