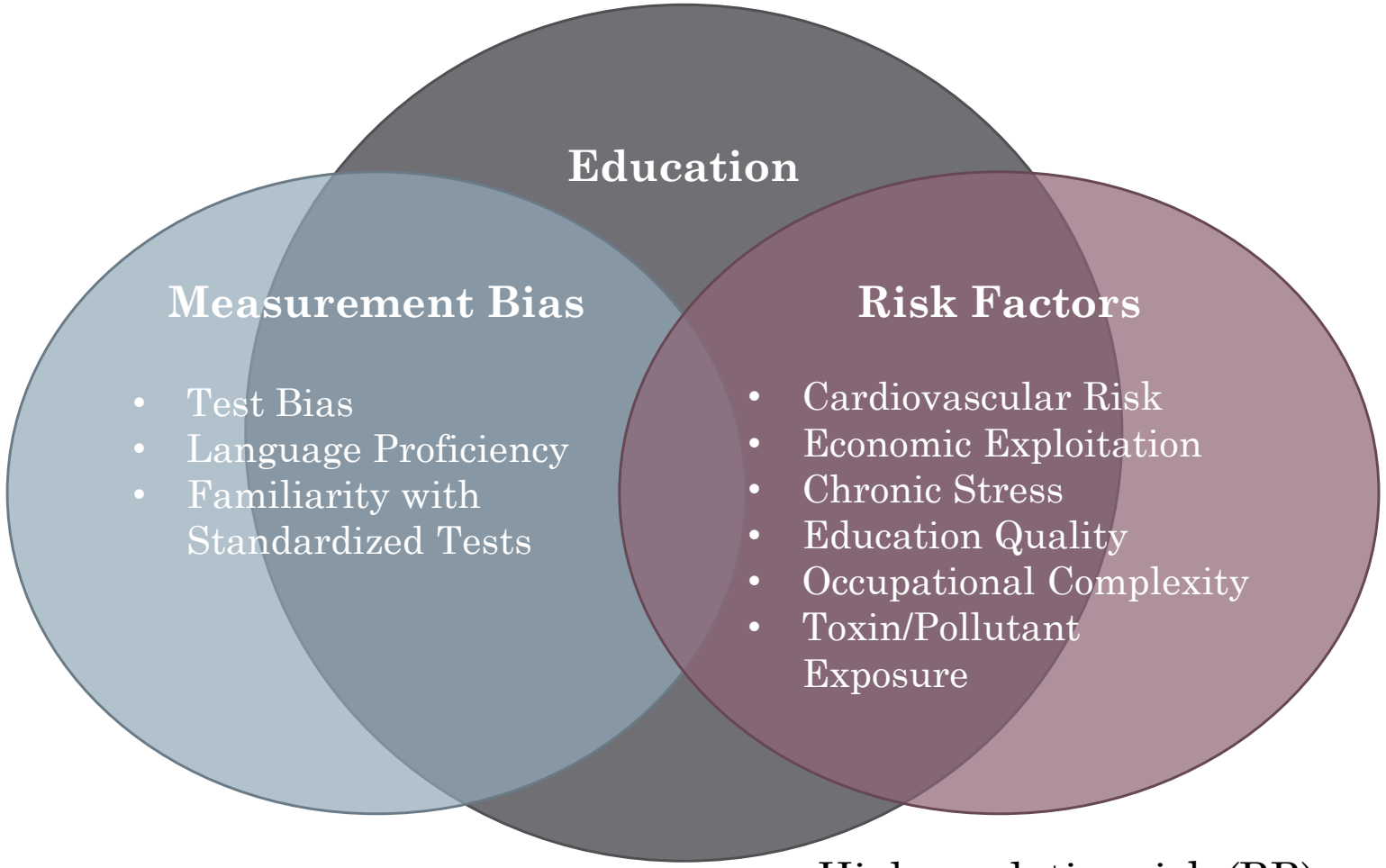
Demographic Adjustment is Not Demographic Correction: A Simulation Study

Jacob Fiala



Group differences caused by measurement bias are artificial

Higher relative risk (RR)
↓
Higher prevalence
↓
More people with low scores
↓
Lower group mean

Demographic group differences in cognitive test scores reflect a mixture of two categories of influences

The Problem

- It is likely that the demographic group differences we adjust out of our cognitive tests are due to both:
  - Measurement Bias Factors
  - Different level of exposure to risk factors for cognitive impairment
- However, when we adjust for a variable like education, we are not separating these two sources of variance, and thus we're discarding not only measurement bias, but also important true differences in cognitive risk.
- As this visual metaphor suggests, we're throwing the baby – true score differences – out with the bathwater – bias.

Purpose

1. Under what conditions does demographic adjustment of test scores *improve* diagnostic accuracy and under what conditions could it be *counter-productive*?
2. Is adjusting for risk factors themselves a viable alternative to adjusting for their demographic proxy variables?
3. Which test score adjustment methodology works the best across all conditions?

Method

The Need for a Simulation Study

- Simulations allow for varying parameters' values across iterations to make sure findings hold across a wide array of conditions
- A simulation can provide knowledge about:
  - The **true impairment status** of each participant
  - The **true prevalence of cognitive impairment** within my norming sample
  - The **true, exact amount of measurement bias** that my hypothetical cognitive test has
  - Each participant's individual "Risk Profile"

Key Parameters

1. Overall Prevalence/Risk: The prevalence of cognitive impairment and average level of risk in the High Education group was set to a predefined level (ranging from .05 to .30)

2. Relative Prevalence/Risk Difference: The prevalence of cognitive impairment and average level of risk in the Low Education group was determined by a predefined Relative Risk (ranging from 1 to 1.75)

3. Mean Measurement Bias: Raw scores in the Low Education group were reduced a predefined amount from their *actual*, true score (ranging from 0 *Z* to -1 *Z*)

4. Variance Measurement Bias: The variance of raw scores in the Low Education group was reduced or increased a predefined amount from the variance of their true scores (ranging from 50% to 150%)

5. Validity Measurement Bias: In the Low Education group, the correlation between raw score and impairment classification was reduced a predefined amount from the correlation between their true score and impairment classification (ranging from 0% to 50%)

Procedure

1. Randomly generate a normative sample including:
   - Predefined N in each demographic group
   - Predefined prevalence of cognitive impairment in each demographic group (corresponding to the mean level of risk for cognitive impairment in that group)
   - Cognitive Test True Score (i.e., No Measurement Bias present)
   - Cognitive Test Raw Score (i.e., May include Measurement Bias)
   - Individual Risk Profile score (with predefined validity, or correlation to true impairment status)
2. Adjust raw scores based on different variables including:
   - Demographic Group
   - Risk Profile
   - Both Demographic Group & Risk Profile
3. Calculate statistics for each score to evaluate the effectiveness of each adjustment method
   - Descriptives (Means, SDs, Correlations)
   - ROC Analyses (AUC, Ideal Cutoff, Sensitivity, Specificity, and Accuracy)

Results & Discussion

Under what conditions can demographic adjustment of test scores *improve* diagnostic accuracy and under what conditions could it be *counter-productive*?

Question 1

Overall, Demographic Adjustment Improved Diagnostic Accuracy Compared to Raw Scores, but Did Not Achieve Comparable Accuracy as True Scores

Accuracy	Full Sample	High Education	Low Education
True Score	80%	80%	80%
Raw Score	-3%	+3%	-9%
Demographically Adjusted	-2%	-2%	-3%

	Sensitivity			Specificity		
	Full Sample	High Education	Low Education	Full Sample	High Education	Low Education
True Score	81%	81%	81%	80%	80%	80%
Raw Score	-2%	-7%	+1%	-3%	+5%	-12%
Demographically Adjusted	-3%	+2%	-7%	-2%	-2%	-2%

Overall, Demographic Adjustment Reduced Demographic Bias Compared to Raw Scores, but Did Not Fully Remove Demographic Bias

	Full Sample Mean	High Education Mean	Low Education Mean
True Score	0.00	0.04	-0.04
Raw Score	0.00	0.26	-0.26
Demographically Adjusted	0.00	0.00	0.00

	Normal Cognition		Impaired Cognition	
	High Education Mean	Low Education Mean	High Education Mean	Low Education Mean
True Score	0.26	0.26	-1.16	-1.16
Raw Score	0.48	-0.02	-0.91	-1.17
Demographically Adjusted	0.22	0.25	-1.22	-0.96

When is demographic adjustment helpful?

Helpful
(improves diagnostic accuracy compared to raw, unadjusted score)

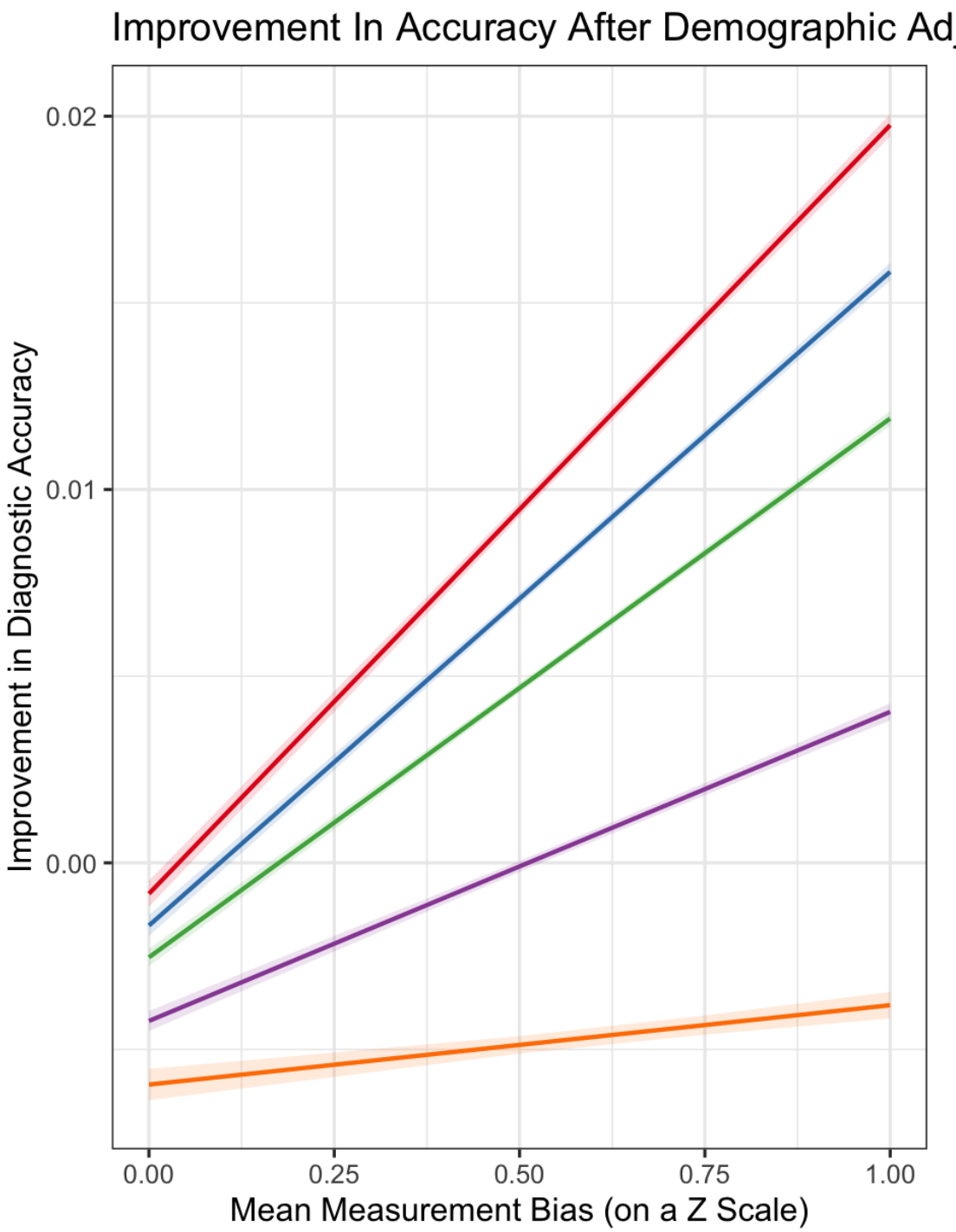
- Higher levels of measurement bias

	β
Mean Measurement Bias	0.11***
Variance Measurement Bias	0.10***
Validity Measurement Bias	0.07***

Counter-productive
(reduces diagnostic accuracy compared to raw, unadjusted score)

- Higher *relative* prevalence/risk difference between the groups

	β
Relative Prevalence/Risk	-0.14***



RR = 1	Normal Cognition		Impaired Cognition	
	High Education Mean	Low Education Mean	High Education Mean	Low Education Mean
True Score	0.22	0.22	-1.21	-1.21
Dem Adjusted	0.22	0.22	-1.22	-1.22

RR = 1.75	Normal Cognition		Impaired Cognition	
	High Education Mean	Low Education Mean	High Education Mean	Low Education Mean
True Score	0.30	0.30	-1.10	-1.10
Dem Adjusted	0.22	0.38	-1.18	-1.02

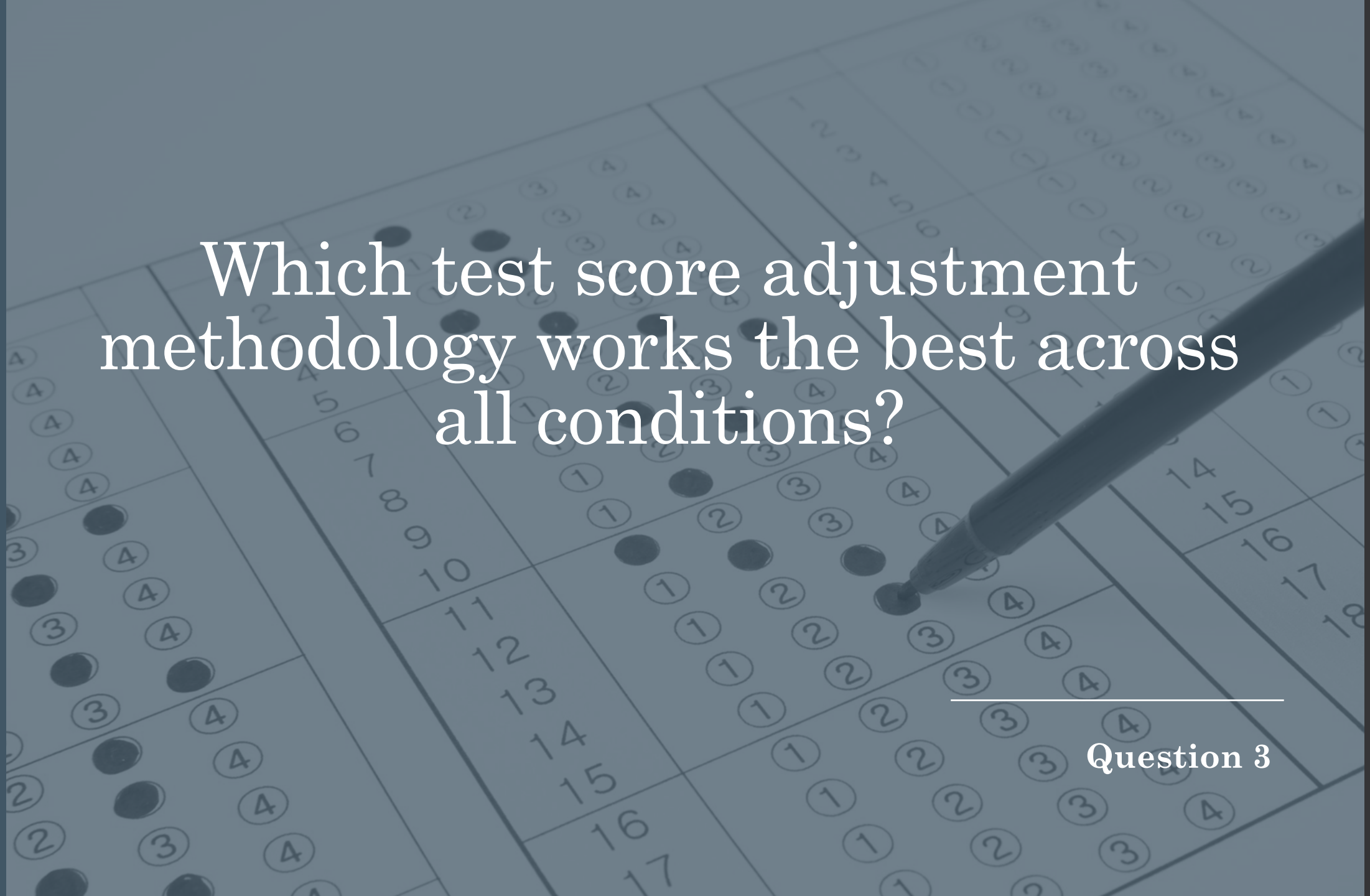
Is adjusting for risk factors themselves a viable alternative to adjusting for their demographic proxy variables?

Question 2

Risk Profile Adjustments Were Unambiguously Counter-Productive

Accuracy	Full Sample	High Education	Low Education
True Score	80%	80%	80%
Risk Adjusted	-14%	-13%	-16%
Risk & Demographically Adjusted	-15%	-14%	-15%

	Sensitivity			Specificity		
	Full Sample	High Education	Low Education	Full Sample	High Education	Low Education
True Score	81%	81%	81%	80%	80%	80%
Risk Adjusted	-13%	-9%	-17%	-14%	-14%	-15%
Risk & Demographically Adjusted	-14%	-12%	-16%	-15%	-15%	-15%



Which test score adjustment methodology works the best across all conditions?

Question 3

Implications & Future Research

Utility of Demographic Adjustment

- Accurate Demographic Adjustment requires knowledge of prevalence within the normative sample
- Future research is desperately needed to determine if current practices are on the whole constructive or counterproductive

Adjusting based on Risk Factors?

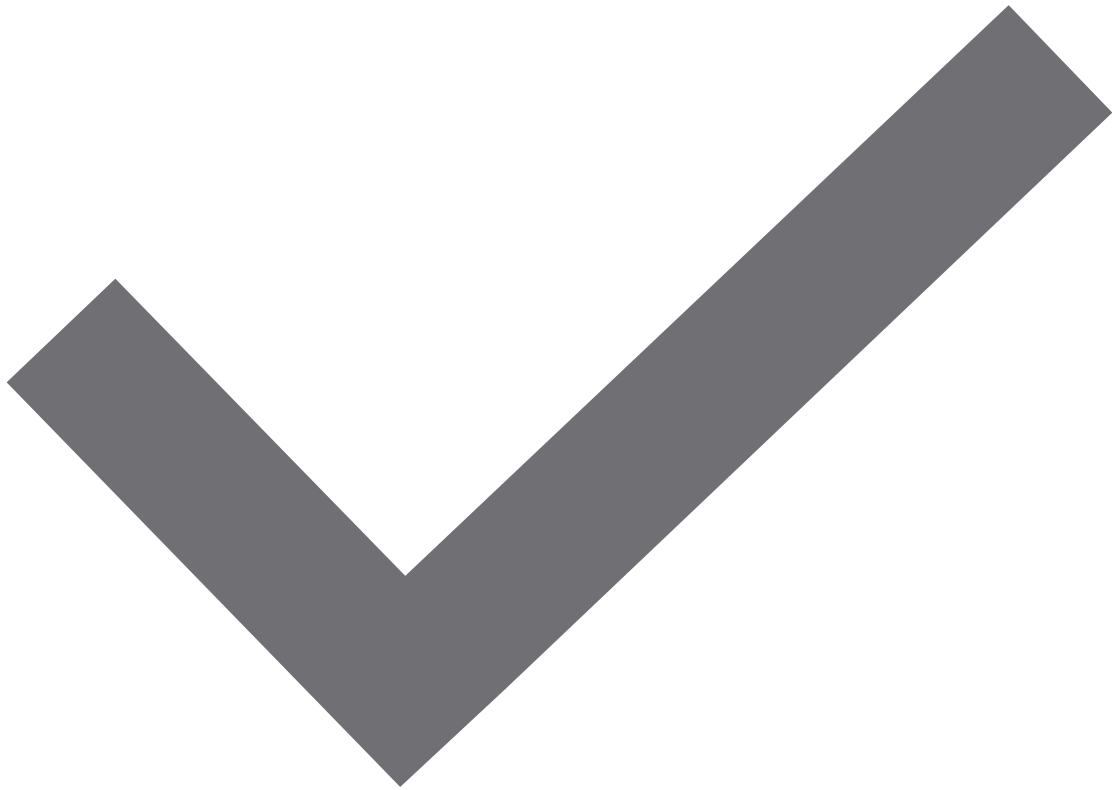
Across all simulated conditions, adjusting for risk profile, whether in isolation or in addition to Demographic Grouping, was counterproductive

What method works best?

Novel Method :

Weighted Mean Composite of Demographically Adjusted Score and Risk Profile

- Improved diagnostic accuracy even above than the accuracy of the true cognitive score
- Reduced demographic bias better than each of the other methods



Future Research

1. Construct a Risk Profile including many biomarkers and medical risk factor variables
2. Further refine my novel methodology
3. Cross-validate the new norming approach and compare the validity of this methodology to existing methodologies, by norming the UDS's Cognitive Batter, and comparing their relationships with biomarkers