

Using bias analysis to address misclassification in epidemiology

20 June 2011

SPER Workshop

Timothy Lash

University of Aarhus

Denmark

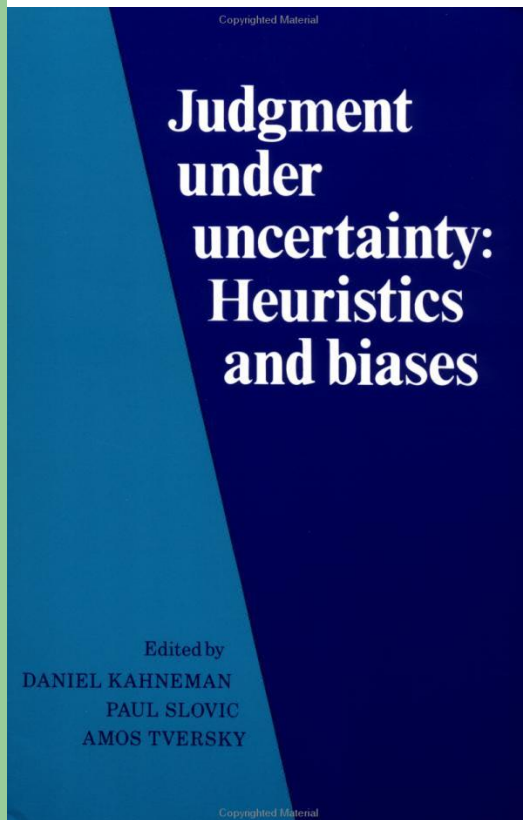
Premise

- If
The objective of etiologic epidemiologic research is to obtain a valid and precise measure of the effect of an exposure on the occurrence of a disease.
- Then
Epidemiologists have an obligation to quantify how far from the objective their estimate might fall.

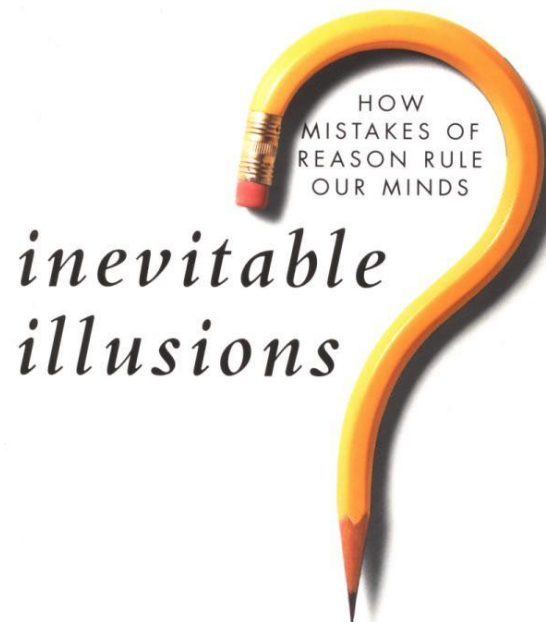
Conventionally

- Epidemiologists quantify how far from the mark they may have landed due to chance (confidence intervals, p-values, *etc*).
- All of these require an assumption of no bias acting on the estimate of effect
- Epidemiologist do NOT quantify how far from the mark they may have landed due to bias, although these errors often dominate the uncertainty
- Some argue that we make the bias assessment intuitively

Reasoning under uncertainty: heuristics and biases

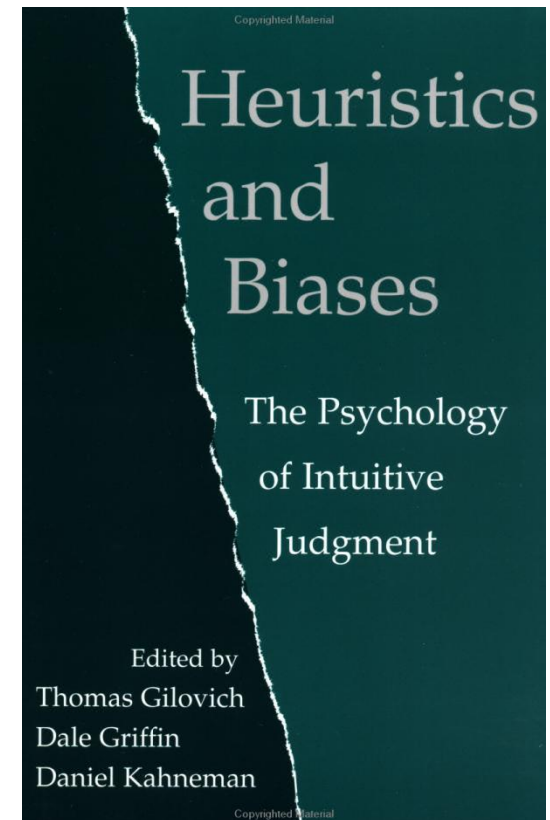


Copyrighted Material
"Delightful, informal . . . the best popular book in this field." —*Nature*



Massimo Piattelli-Palmarini

TRANSLATED BY
MASSIMO PIATTELLI-PALMARINI
AND KEITH BOTSFORD

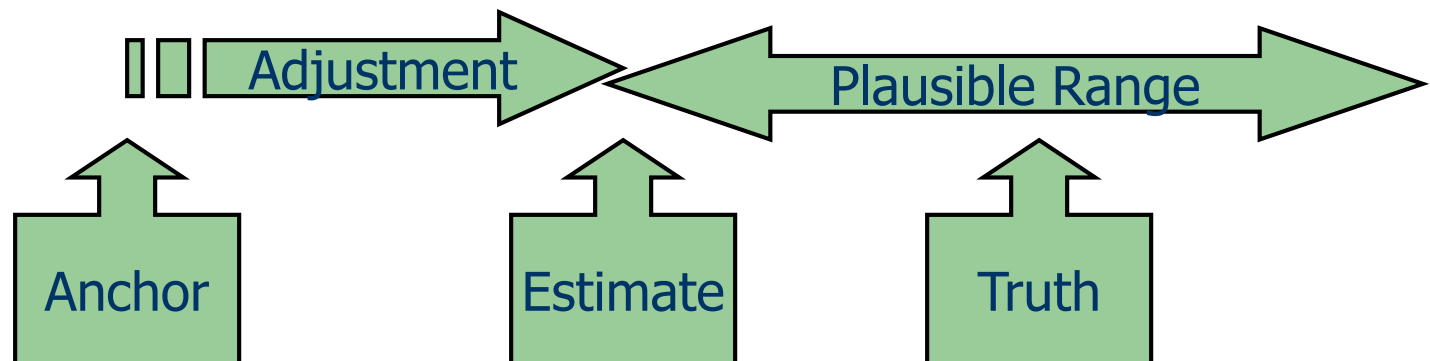


Heuristics

- Definition: Strategies that people use deliberately in order to simplify judgmental tasks that would otherwise be too difficult (or too resource intensive) to solve rationally.
- Example: Non-differential misclassification biases toward the null.
- Problem: While they often serve us well, heuristics can cause systematic errors in judgment, which in turn bias the judgments we reach.

Anchoring and adjustment (1)

- Estimates of uncertain quantities begin with an anchor.
- Adjust away from the anchor toward the plausible range.
- Adjustments are insufficient because we only adjust until we enter the plausible range.



Anchoring and adjustment (2)

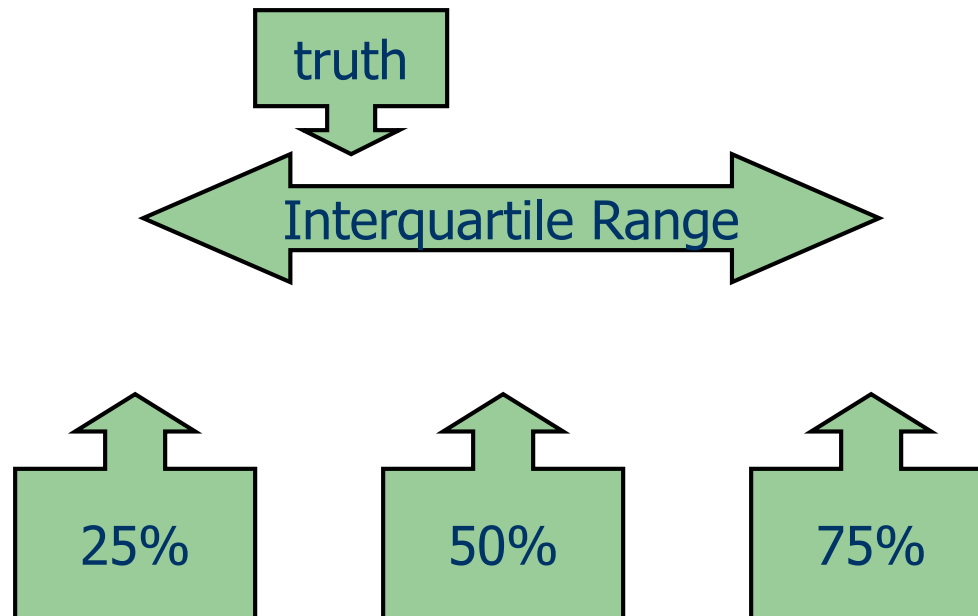
- Example #1: In what year was George Washington first elected president?
 - Anchor = 1776
 - Mean answer = 1779
 - Truth = 1788
- Example #2: At what temperature does vodka freeze?
 - Anchor = 0 C
 - Mean answer = -17 C
 - Truth = -29 C

Application to epidemiology

- Given a point estimate of effect, and
- Knowledge that the point estimate is subject to bias, along with the direction of that bias
- Should we expect that adjustment from the point estimate will be sufficient?

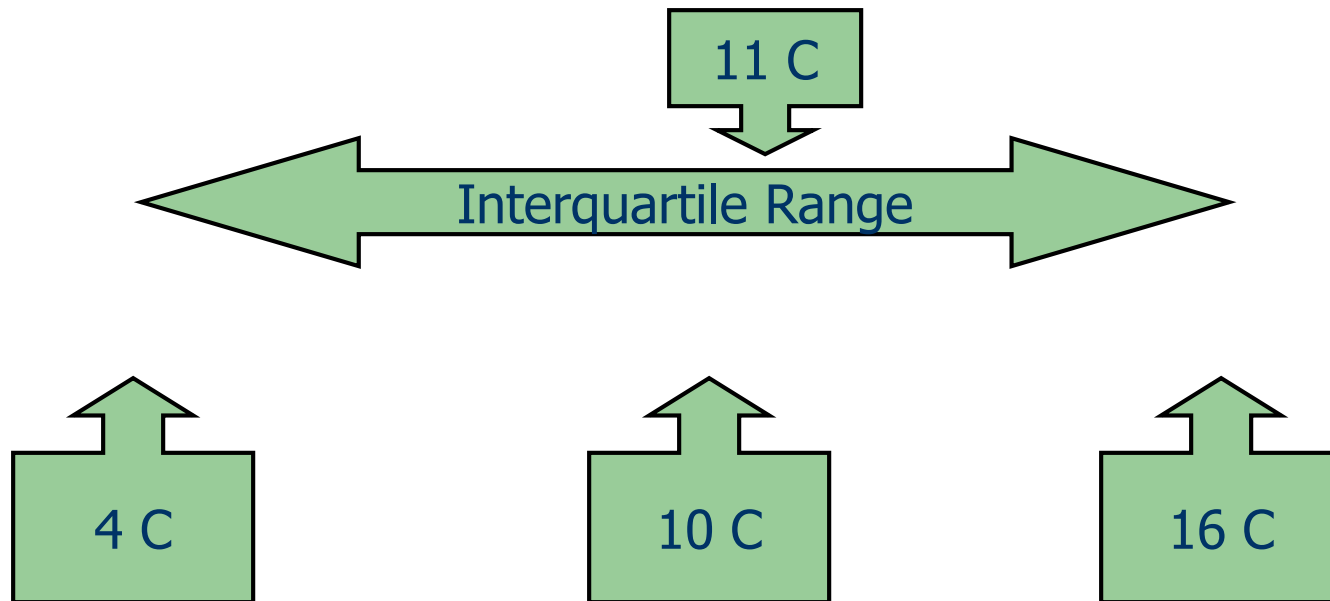
Overconfidence (1)

- Consistent tendency to overstate our certainty about unknown quantities.



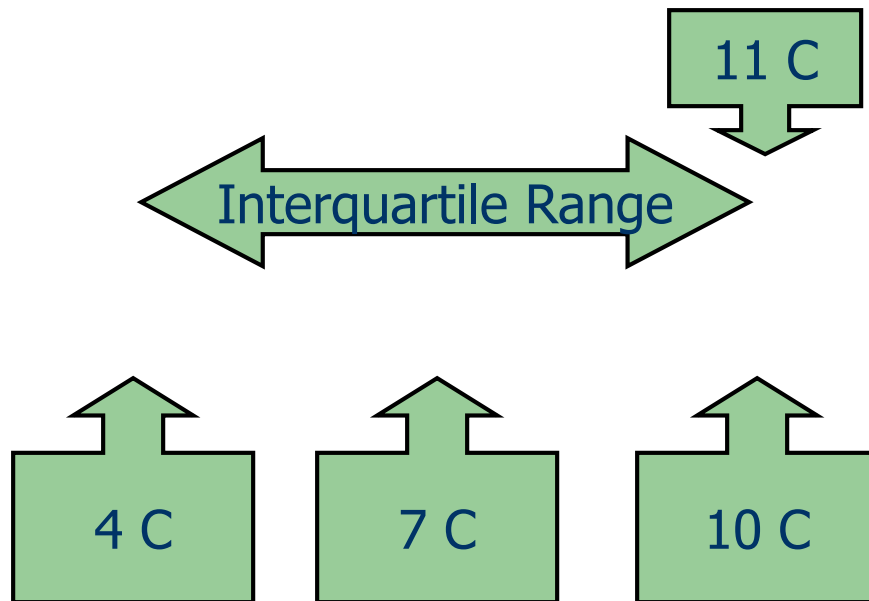
Overconfidence (2)

- What is the average annual temperature in Boston?



Overconfidence (3)

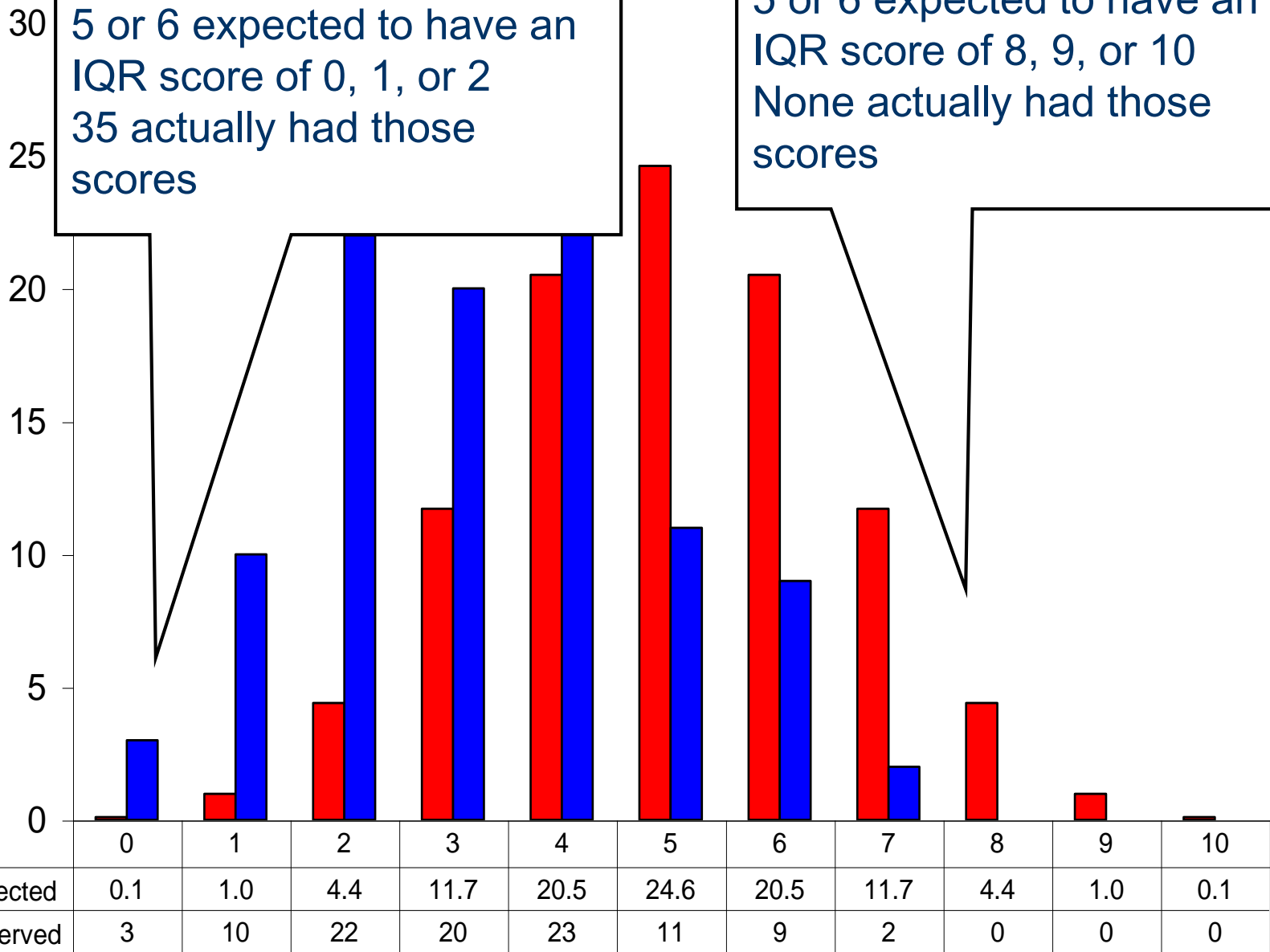
- What is the average annual temperature in Boston?



Overconfidence (4)

- Experiment
 - Ask 100 students to give numeric estimates for 10 uncertain quantities. For each give 25%, 50%, 75%.
- For a well calibrated estimator, truth should fall inside the IQR (25%–75%) as often as outside.
- IQR score equals the number (of 10) for which truth fell in the IQR (expect 5).

Frequency



IQR Score (# of 25%-75% ranges that included true value)

Application to epidemiology

- Given a confidence interval around a point estimate of effect, and
- Knowledge that the interval is too narrow
- Should we expect that inflation of the interval to account for additional error will be sufficient?

Common elements

- General \Rightarrow majority of humans.
- Systematic \Rightarrow tendency always in the same direction.
- Independent of intelligence and education \Rightarrow experts make the same mistakes as novices with only slightly harder problems.
- In fact, accuracy increases with expertise, but overconfidence increases faster.

**Simple, multidimensional, and
probabilistic bias analysis to
address exposure
misclassification**



Background

- Research question:
 - evaluate the association between use of antidepressants and breast cancer risk
- Chien *et al* (*Br Cancer Res Trt* 2006;95:131–40)
 - Population-based case-control study
 - Source population: Women 65-79 resident in 1 of 3 Western Washington State counties 1997-99
 - Cases from SEER registry
 - Controls from CMS records matched on age, yr, county
 - 20 year history of antidepressant use by in-person interview
 - Conventional result: “no association between ever use of antidepressants and breast cancer risk”
(OR=1.2; 95% CI 0.9,1.6)

Conventional Results

	Ever AD	Never AD
Cases	118	832
Controls	103	884
Crude OR	1.21 (0.92, 1.61)	
Adjusted OR	1.2 (0.9, 1.6)	

Could the null result be due to misclassification of self-reported antidepressant use?



Terminology (1)

- Sensitivity (SE) = the probability of being correctly classified as exposed. $\Pr(E+|T+)$
- Specificity (SP) = the probability of being correctly classified as unexposed $\Pr(E-|T-)$
- Positive predictive value (PPV) = the probability of truly being exposed if classified as exposed $\Pr(T+|E+)$
- Negative predictive value (NPV) = the probability of truly being unexposed if classified as unexposed $\Pr(T-|E-)$

Equationst

	Truly Exposed	Truly Unexposed
Classified as Exposed	A	B
Classified as Unexposed	C	D

$$\text{Sensitivity} = \frac{A}{(A+C)}$$

$$\text{Positive predictive value} = \frac{A}{(A+B)}$$

$$\text{Specificity} = \frac{D}{(B+D)}$$

$$\text{Negative predictive value} = \frac{D}{(C+D)}$$

† Assumes random sample of source population

Calculations (Boudreau *et al* Am J Epidemiol 2004;159:308–17)

	2 year Pharmacy Record (cases)		
		Ever AD	Never AD
Interview	Ever AD	24	2
	Never AD	19	144

$$\text{sensitivity} = \frac{24}{24 + 19} = 56\%$$

$$\text{specificity} = \frac{144}{144 + 2} = 99\%$$

Terminology (2)

- Non-differential exposure misclassification
 - the rates at which exposure status is misclassified are not expected to depend on disease status
- Differential exposure misclassification
 - the rates at which exposure status is misclassified are expected to depend on disease status

Calculations (Boudreau *et al.*, 2004)

	2 year Pharmacy Record (controls)		
		Ever AD	Never AD
Interview	Ever AD	18	4
	Never AD	13	130

$$\text{sensitivity} = \frac{18}{18 + 13} = 58\%$$

$$\text{specificity} = \frac{130}{130 + 4} = 97\%$$

Calculating expected observed from the truth

	Truth		Expected Observed	
	E+	E-	E+	E-
D+	A	B	$A(SE_{D+})+B(1-SP_{D+})$	$A(1-SE_{D+})+B(SP_{D+})$
D-	C	D	$C(SE_{D-})+D(1-SP_{D-})$	$C(1-SE_{D-})+D(SP_{D-})$

Example

SE=85% SP=95%	Truth		Observed	
	E+	E-	E+	E-
D+	200	100	$200 \times 0.85 + 100 \times 0.05$ 175	$200 \times 0.15 + 100 \times 0.95$ 25
D-				
OR				

$$A(SE_{D+}) + B(1 - SP_{D+})$$

Example (non-differential)

SE=85% SP=95%	Truth		Expected Observed	
	E+	E-	E+	E-
D+	200	100	$200 \times 0.85 + 100 \times 0.05$ $= 175$	$200 \times 0.15 + 100 \times 0.95$ $= 125$
D-	800	900	$800 \times 0.85 + 900 \times 0.05$ $= 725$	$800 \times 0.15 + 900 \times 0.95$ $= 975$
OR	2.25		1.88	

Observed to Expected Truth

	Observed		Expected Truth	
	E+	E-	E+	E-
D+	a	b	$\frac{[a - (1 - SP_{D+})D+]}{[SE_{D+} - (1 - SP_{D+})]}$	D+ - A
D-	c	d	$\frac{[c - (1 - SP_{D-})D-]}{[SE_{D-} - (1 - SP_{D-})]}$	D- - C

Example (antidepressant data)

SE =57% SP =98%	Observed		Expected Truth	
	Ever AD	Never AD	Ever AD	Never AD
Cases	118	832	183.3	766.7
Controls	103	884	154.2	832.8
OR	1.2		1.3	

$$A = \frac{[a - (1 - SP_{D+})D+]}{[SE_{D+} - (1 - SP_{D+})]} = \frac{[118 - (1 - 0.98)950]}{[0.57 - (1 - 0.98)]} = 183.3$$

Multidimensional approach

- Choose reasonable range of bias parameters
 - Create combinations of bias parameters
 - Calculate estimate of effect for each combination
- Choose combinations of bias parameters until an estimate of effect within a desired range is calculated (e.g., the null)
 - Present combinations until the desired estimate is observed
 - Judge the likelihood that the combinations required to achieve the desired estimate of effect is near the true combination

Multidimensional advantages

- More information than simple sensitivity analysis
 - More bias parameters
 - More combinations of bias parameters
- Provides bounding information
 - What combinations of bias parameters would be required to yield a null result?
 - What is the minimum and maximum result with combinations of reasonably chosen bias parameters?
- Useful to make strong statements about the potential impact of particular biases

Multidimensional Example

	Controls				
Cases	Sensitivity	0.6	0.5	0.6	0.5
Sensitivity	Specificity	1	1	0.95	0.95
0.6	1	1.24	0.99	1.11	0.86
0.5	1	1.57	1.25	1.41	1.09
0.6	0.95	1.57	1.28	1.42	1.8
0.5	0.95	1.98	1.62	1.8	1.44

- Under non-differential misclassification the corrected odds-ratio estimates are always further from the null than the uncorrected estimate (1.2) computed directly from the data

Multidimensional Example

	Controls				
Cases	Sensitivity	0.6	0.5	0.6	0.5
Sensitivity	Specificity	1	1	0.95	0.95
0.6	1	1.24	0.99	1.11	0.86
0.5	1	1.57	1.25	1.41	1.09
0.6	0.95	1.57	1.28	1.42	1.8
0.5	0.95	1.98	1.62	1.8	1.44

- The uncertainty in results due to the uncertainty about the classification probabilities can be much greater than the uncertainty conveyed by standard confidence intervals (95% CI 0.92, 1.61).

Multidimensional disadvantages

- Resource intensive to present
 - Text space
 - Table space
- No sense of the likelihood that any combination of bias parameters is the true set
 - No estimate of central tendency
 - No picture of the frequency distribution of the estimates of effect

Probabilistic extension

- Rather than one bias parameter, or one set of bias parameters, draw bias parameter from a distribution.
- Repeat draw over many iterations
- Accumulate results to generate a frequency distribution
 - Central tendency
 - Interval
- Simultaneously incorporate random error with resampling

Spreadsheet Example

Steps:

- 1) Input observed data (and create individual records)
- 2) Input probability distributions for sensitivity and specificity
- 3) Randomly choose a sensitivity and specificity from the specified distributions
- 4) “Correct” the observed data
- 5) Calculate PPV and NPV for cases and controls
- 6) Use these probabilities (PPV and NPV) to choose randomly who is reassigned
- 7) Recalculate measure of effect
- 8) Repeat
- 9) Create intervals from 2.5th to 97.5th percentile

Resampling

for $i = 1$ to n , let

$$\ln R_{b,i} = \ln \hat{R}_s - z_i \sqrt{V_s}$$

Accumulate over n iterations

Select 50% for point estimate

Select 2.5% & 95% for simulation interval

Calculations (Boudreau et al., 2004)

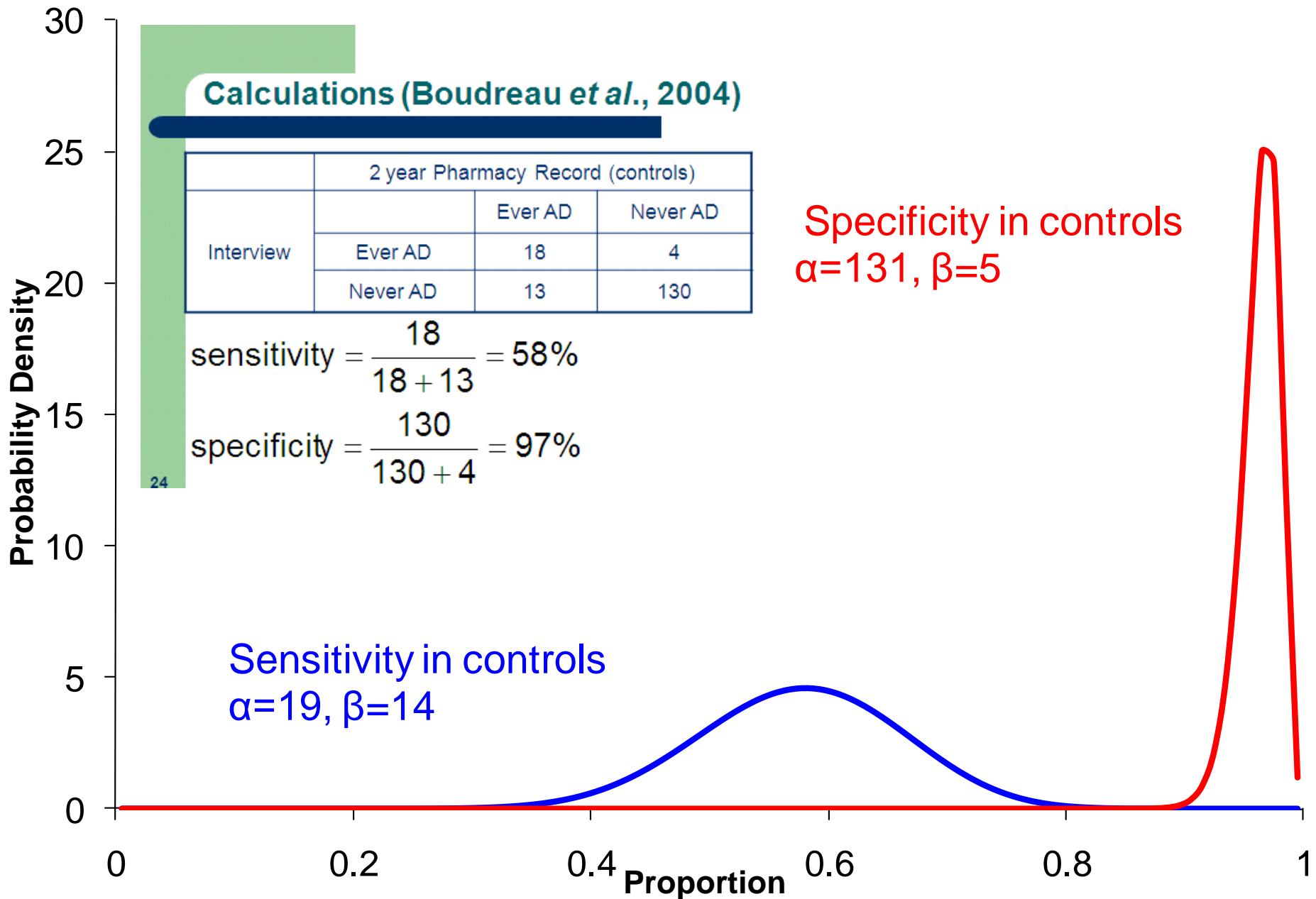
		2 year Pharmacy Record (controls)	
		Ever AD	Never AD
Interview	Ever AD	18	4
	Never AD	13	130

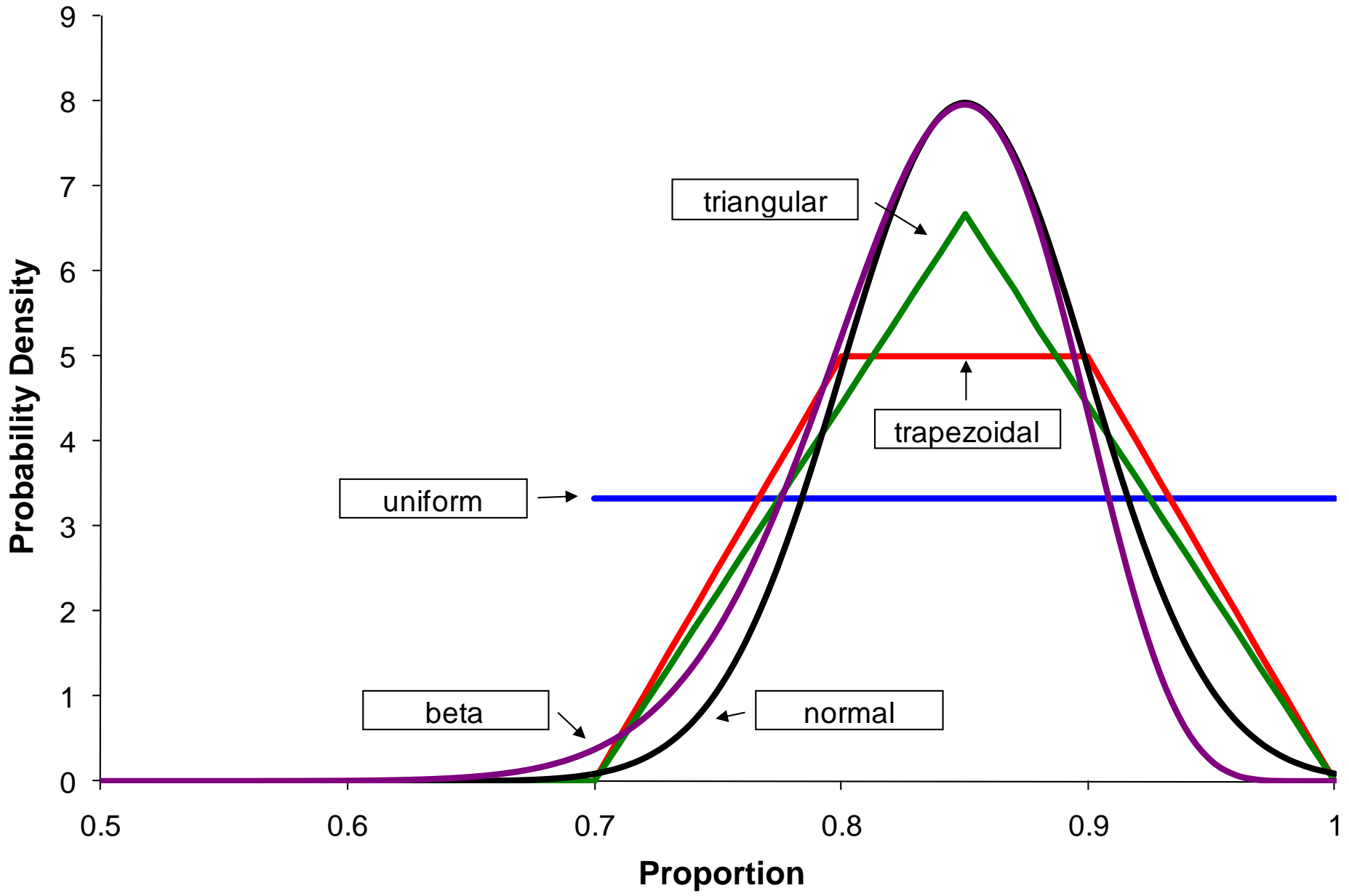
$$\text{sensitivity} = \frac{18}{18 + 13} = 58\%$$

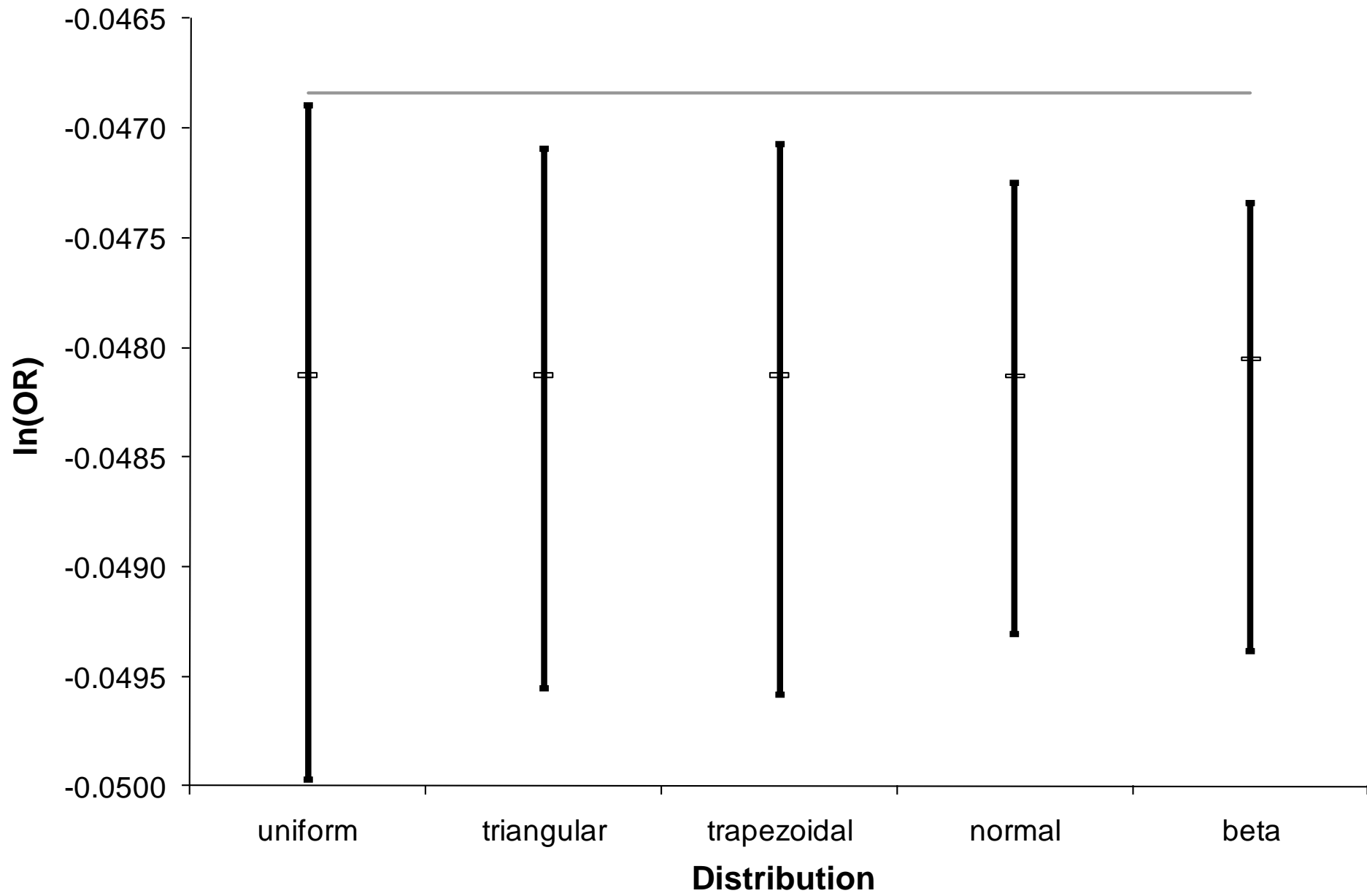
$$\text{specificity} = \frac{130}{130 + 4} = 97\%$$

Specificity in controls
 $\alpha=131, \beta=5$

Sensitivity in controls
 $\alpha=19, \beta=14$







Observed Data

	AntiDep +	AntiDep -	Total
BC +	118 ^a	832 ^b	950
BC -	103 ^c	884 ^d	987
Total	221 ^m	1716 ⁿ	

Corrected Data

	AntiDep +	AntiDep -	Total	Chosen Values
BC +	171.0 A_1	779.0 B_1	950	Se(D+) 57.6% Se(D-) 57.6%
BC -	142.1 C_1	844.9 D_1	987	Sp(D+) 97.5% Sp(D-) 97.5%
Total	313.1 M_1	1623.9 N_1		

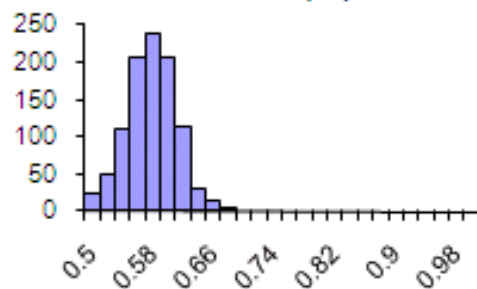
Observed	Measure (95% CI)
RR (AntiDep-BC)	1.1 (0.96 - 1.26)
OR (AntiDep-BC)	1.22 (0.92 - 1.61)

Single Corrected Estimate	Measure
RR (AntiDep-BC)	1.14
OR (AntiDep-BC)	1.31

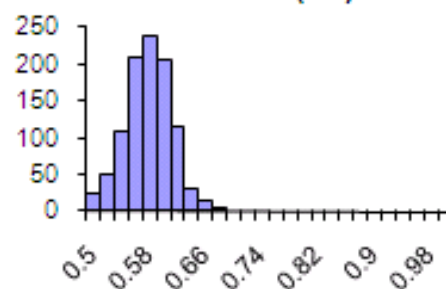
RR Simulation Results (N=1000)	
Analysis	Median (2.5 th -97.5 th percentile)
Conventional	1.1 (0.96 - 1.26)
Systematic	1.15 (1.13 - 1.19)
Total Error	1.16 (1.01 - 1.33)

OR Simulation Results (N=1000)		Illegal Values
Analysis	Median (2.5 th -97.5 th percentile)	
Conventional	1.22 (0.92 - 1.61)	0
Systematic	1.33 (1.29 - 1.44)	
Total Error	1.34 (1.01 - 1.81)	

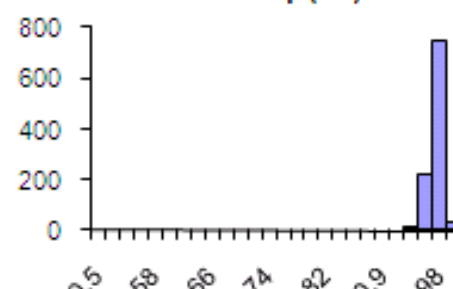
Chosen Se(D-)



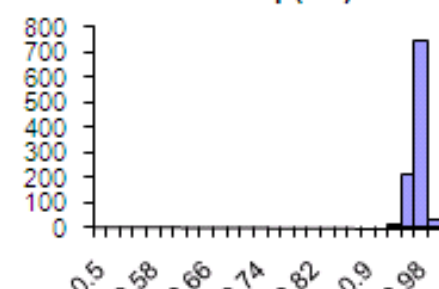
Chosen Se(D+)



Chosen Sp(D-)



Chosen Sp(D+)



Objections and answers

- Quantitative bias analysis will be used nefariously to annihilate important positive associations.
- Uncertainty quantified by bias analysis will be used to impede public health action.
- Quantitative bias analysis will supplant collection of internal validation studies, which provide superior analytic solutions.
- Results from bias analysis subject to over-interpretation.

Overconfidence and hypotheses (1)

- Experiment to test confidence about the accuracy of hypotheses
- Before nominations are announced:
 - Group 1 (Hypothesis generators): Students asked to name 3 candidates for Best Film, Best Actor, and Best Actress. Also give percent likely that the ultimate winner is among the 3.
 - Group 2 (Hypothesis evaluators): Students given 3 candidates suggested by a randomly paired student from group 1, give percent likely that the ultimate winner is among the 3.

Overconfidence and hypotheses (2)

	Generators	Evaluators
Best film	64%	74%
Best actress	28%	50%
Best actor	49%	75%

- True average accuracy equaled 33%, so both groups were generally overconfident
- Consideration of alternatives by generators, but not evaluators, led to lower estimate of confidence by hypothesis generators

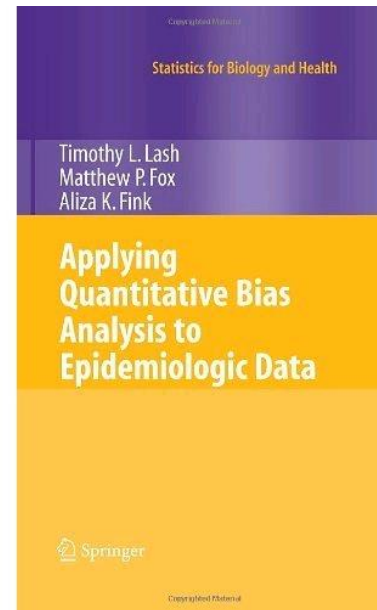
Application to epidemiology

- This study, and others, suggest that the poor correspondence between confidence and accuracy arises from the tendency to neglect alternatives to the focal hypothesis.
- In epidemiology, the focal hypothesis is usually the causal hypothesis, and alternatives are sources of error (chance, selection bias, measurement error, and confounding)
- Quantify total error rather than relying on intuition (anchoring & adjustment and overconfidence) to prevent misjudgment.

Acknowledgements & Resources

- Colleagues who have worked with me on these methods

- Mathew Fox
- Aliza Fink
- Sander Greenland
- Tony Blakely



- To download spreadsheets and software
 - <http://sites.google.com/site/biasanalysis/>

In case you think you're special

