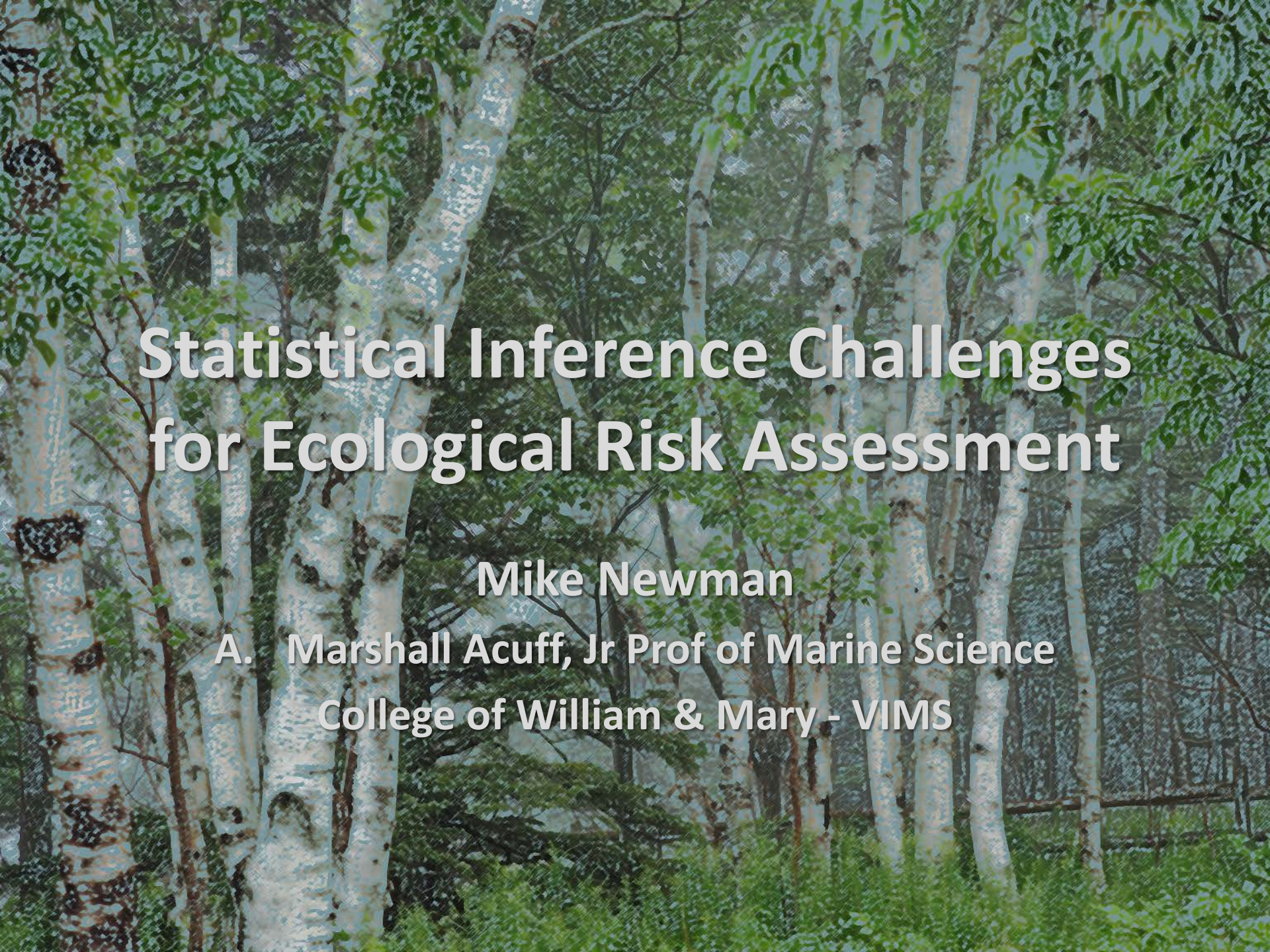


Statistical Inference Challenges for Ecological Risk Assessment

Mike Newman

A. Marshall Acuff, Jr Prof of Marine Science

College of William & Mary - VIMS

1. Overview
2. Reasoning with Probabilities
3. Null Hypothesis Significance Testing

1. Overview

Mimicry Heuristic: People evolved to imitate successful individuals instead of conducting individual trial-and-error learning. This strategy is fast and efficient …

EXCEPT in highly variable environments due to the high risk of using inappropriate or out-of-date information. Can result in informational cascades. Then it is better to rely more on independent individual learning.

2. Reasoning with Probabilities

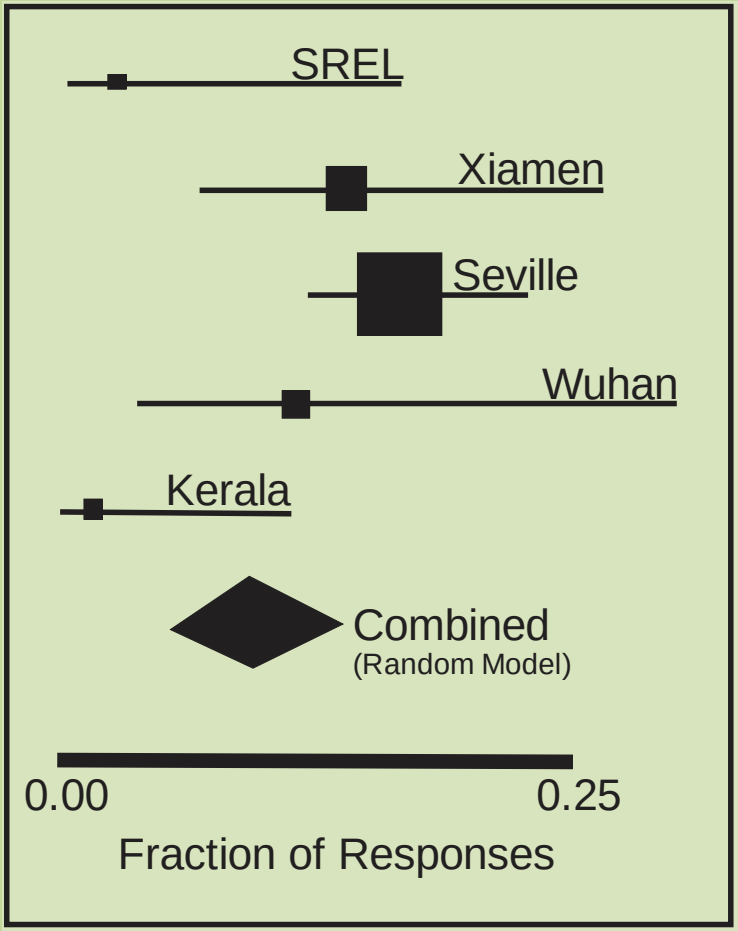
Analysis of Drinking Water for Contaminant Scenario
(From Rizak & Hrudey. 2006. *EST* 40:5244)

"Monitoring evidence for [City X] has indicated that in treated drinking water, [pesticide Y] is truly present above the recognized methods detection limit once every 1000 water samples from consumers' taps.

- *95% of tests will be positive for detection when [Y] is truly present*
- *98% of tests will be negative for detection when [Y] is truly absent*

Given a positive test on the analytical test for [Y] in the [City X] drinking water system, how likely do you think this positive test reflected the true presence of [Y]?"

Typically Involves Global Introspection



Fraction Picking Correct Answer:
10% 95% CI: 5 to 19%

How likely is it that the pesticide is actually present given the test result?

★ 80-100%	20-39%
60-79%	10-19%
40-59%	★ 0-9%

$$PPV = \frac{(1-\beta)R}{R - \beta R + \alpha}$$

Base Rate Error

Would it be acceptable to use an analytical method that was always off by so much if a better method were as easily applied? Reporting an effective dose equivalent of 5,000 instead of the correct 220 mrem.

2. Reasoning with Probabilities – Conclusion

Global introspection is unreliable. Cultivate a working knowledge of probability calculus and apply it whenever possible.

Natural frequencies, BBNetworks

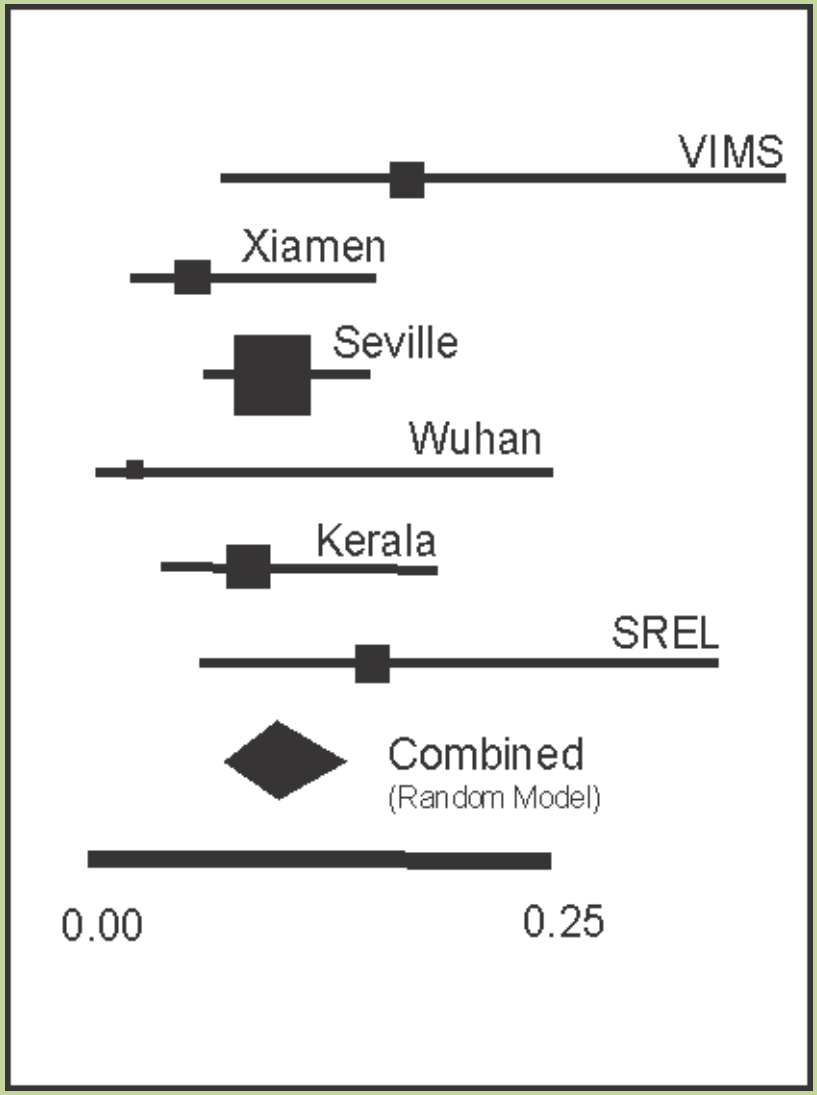
…we are instinctively very poor evaluators of probability and equally poor at choosing between alternative possibilities…

Piatelli-Palmarini (1994)

3. Null Hypothesis Significance Testing

When you took statistics, or when first exposed to significance testing, you were taught that a "statistically significant" outcome with a p-value of 0.01 means …

a. The null hypothesis is disproven.
b. The p-value is the estimated probability of the null hypothesis being true.
c. You proved that an effect is present.
d. You can deduce the probability of the effect actually being present.
e. You know the probability that you would make a wrong decision if you rejected the null hypothesis.
f. You know the probability of getting these, or more extreme, data if the null hypothesis were true.
g. If you repeated the test many times, you would obtain a significant effect in 99 out of 100 trials



Fraction Providing Correct Answer:
9.6% 95% CI: 6.9 to 13.2%

Consistent with Oakes (1986) and Haller and Krauss (2002) surveys of German research and teaching faculty, and advanced students

Only 9.6% of those surveyed chose the correct answer (Answer f).

Randomly picking any answer would have resulted in 14.2% correct answers. This 14.2% is outside of the confidence interval for the observed data, suggesting that the surveyed environmental science faculty and students were not simply poorly informed: they were misinformed during their training about the interpretation of p-values.

Fisher – Significance Testing

Influenced by Popper's partial solution to the problem of induction. You can never know that something is true but you can prove that it is false.

Practical Falsification – something is false if sufficiently improbable (set α).

P-value used to gauge "sufficiently improbable" data if H_0 true

Focused on **falsifying a single H_0**.

Neyman-Pearson Hypothesis Testing

- H_0 is stated and given a (Type I) decision error rate (α).
- H_A is stated and given a (Type II) decision error rate (β).
- Effect size (ES) is selected for a meaningful difference or effect.

Does ***NOT*** falsify the null hypothesis.

Decision is correct more often than not if based on the test outcome.

Current NHST Convention

"mishmash of Fisher and Neyman-Pearson, with invalid Bayesian interpretation" (Cohen, 1994)

"a hybridization of Fisher's significance testing and Neyman and Pearson's hypothesis testing (Anderson et al., 2000)

3. NHST - Conclusion

NHST is an illogical mishmash that is currently being reevaluated. Be very thoughtful if you use it.

Other approaches exist.

- Estimation with CI or HDI
- Information-theoretic
- Bayesian

It ain't so much the things we don't know that get us in trouble. It's the things we know for sure that just ain't so.

Artemus Ward


Questions or Comments?