

Analysis and Interpretation of Signals in Large Studies

Kenneth J. Rothman
February 16, 2011

Inferences From Large Data Sets

- **Studies can be conceptualized as exercises in measurement.**
- **Measurement error includes both random and systematic error. In large studies, random error is smaller, and most of the error is from bias.**
- **Statistical testing and interval estimation focuses on random error, and is therefore much less relevant, and even misleading, in large studies.**
- **Large datasets often use secondary data, which may lead to greater bias than in ad hoc research studies designed to reduce systematic error.**

Practical Implications

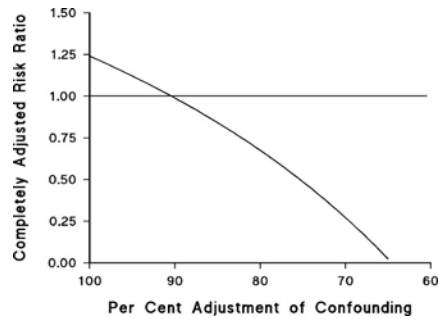
- **In large studies, statistical testing is even more misleading than usual.**
- **Confidence interval estimation is also problematic, because it measures only random error.**
- **Inference should incorporate biological plausibility.**
- **Bias analysis (sensitivity analysis) should be part of the analysis plan.**

Applying Quantitative Bias Analysis to Epidemiologic Data



**Lash T, Fox MP, Fink AK:
Applying Quantitative
Bias Analysis to
Epidemiologic Data,
Springer, 2009**

Example



**Rothman KJ, Wentworth
CE: Mortality of cystic
fibrosis patients receiving
tobramycin solution for
inhalation. *Epidemiology*
2003;14:55-59.**