Abstract

Medical observational studies are complex with multiple endpoints and statistical analysis is essentially exploratory. There is a need to have principled evaluation strategies not only within a particular study, but also over multiple studies. Our idea is to borrow techniques from the quality and statistical literatures and benchmark evaluation strategies against randomized clinical trials. The benefits of this approach should lead to a logical framework for evaluation of claims coming from observational studies.

Understanding the Role of Cohort Study Analysis, Goal: Valid Claims

S. Stanley Young

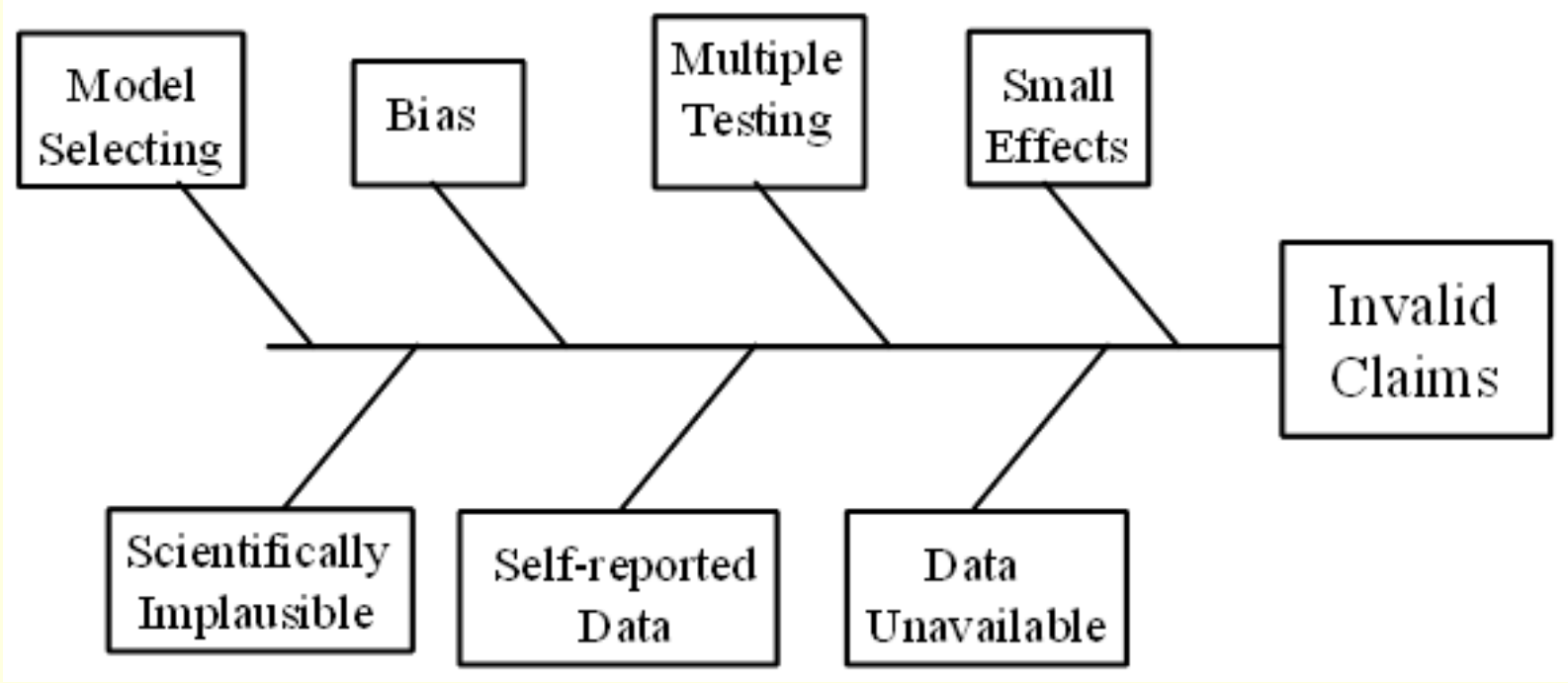
National Institute of Statistical Sciences

young@niss.org, 919 685 9328

Claims based on observational studies, and tested in RCT

ID#	Pos	Neg	#Claims	Treatment(s)
				Claims based on observational study
1	0	1	3	Vit E, beta-carotene
3	0	3	4	Hormone Replacement Ther.
5	0	1	2	Vit E, beta-carotene
6	0	0	3	Vit E
10	0	0	3	Low Fat
11	0	0	3	Vit D, Calcium
12	0	0	2	Folic acid, Vit B6, B12
13	0	0	2	Low Fat
14	0	0	12	Vit C, Vit E, beta-carotene
17	0	0	12	Vit C, Vit E
18	0	0	3	Vit E, Selenium
new	0	0	3	HRT+antioxidant vits**
	0	**5**	**52**	

Fish-Bone Diagram



The big three factors

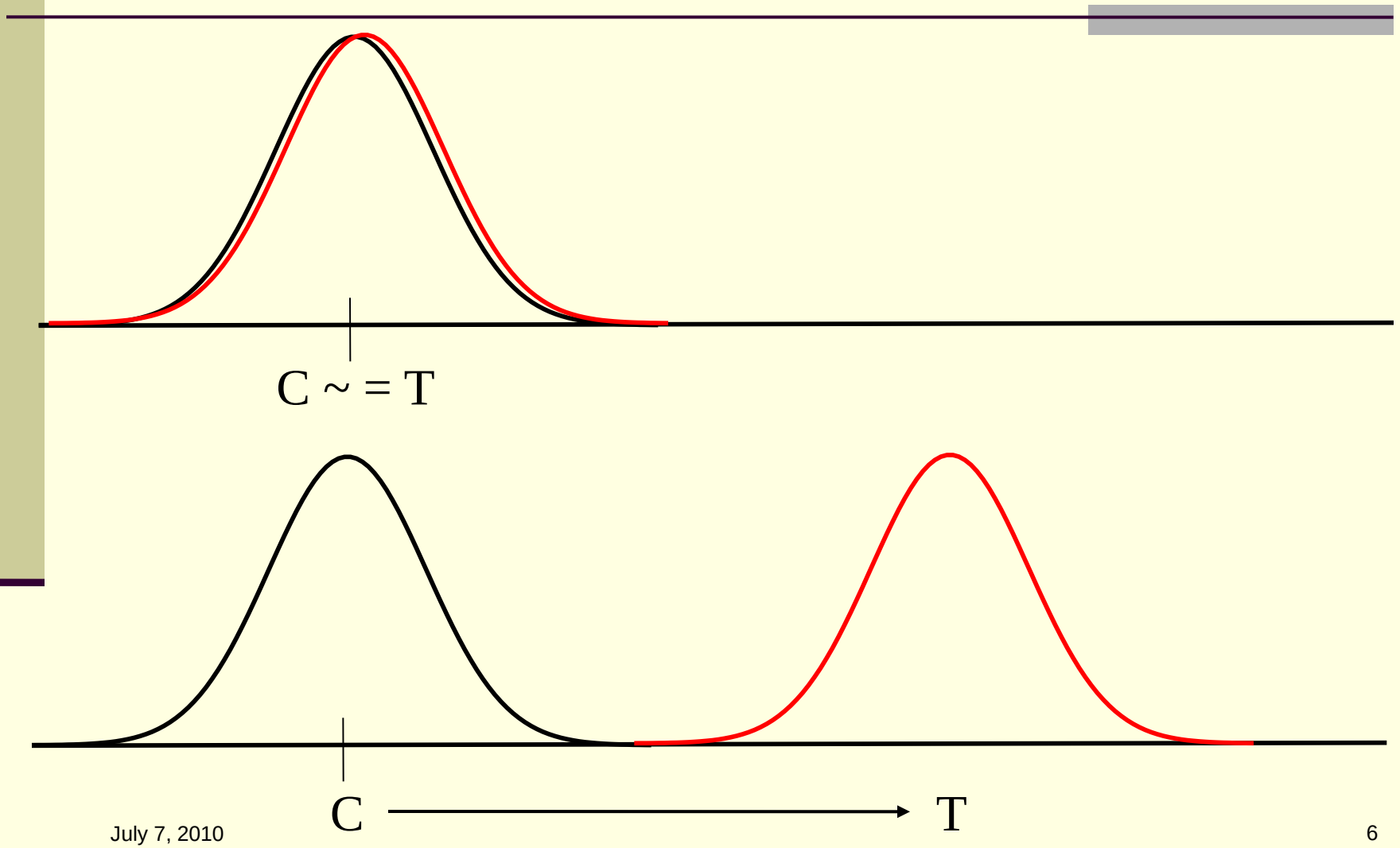
1. Bias (and small effects)

2. Multiple testing

3. Multiple model searching

Any or all can lead to false claims.

No bias: Randomized Clinical Trial



First, Bias

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \ldots + \beta_p X_{pt} + \varepsilon$$

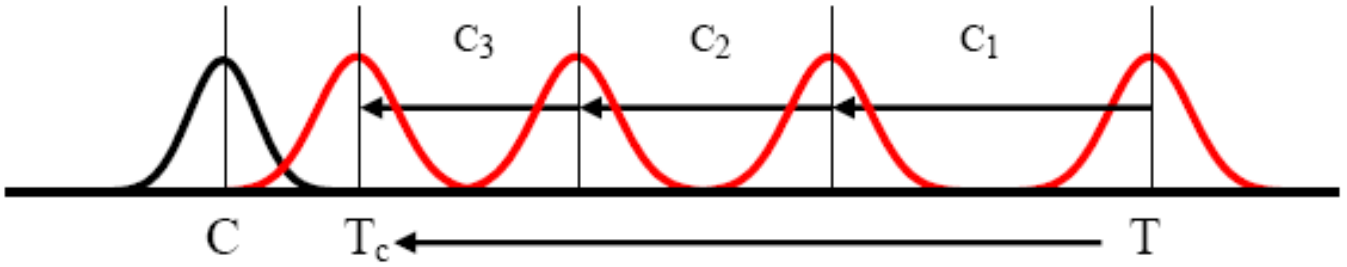
$$Y_c = \beta_0 + \beta_1 X_{1c} + \beta_2 X_{2c} + \beta_3 X_{3c} + \beta_4 X_{4c} + \ldots + \beta_p X_{pc} + \varepsilon$$

$$\Delta_{t\text{-}c} = \left(\bar{Y}_t - \bar{Y}_c\right) = \beta_1\left(\bar{X}_{1t} - \bar{X}_{1c}\right) + \beta_2\left(\bar{X}_{2t} - \bar{X}_{2c}\right) + \ldots + \beta_p\left(\bar{X}_{pt} - \bar{X}_{pc}\right) + \left(\bar{\varepsilon}_t - \bar{\varepsilon}_c\right)$$

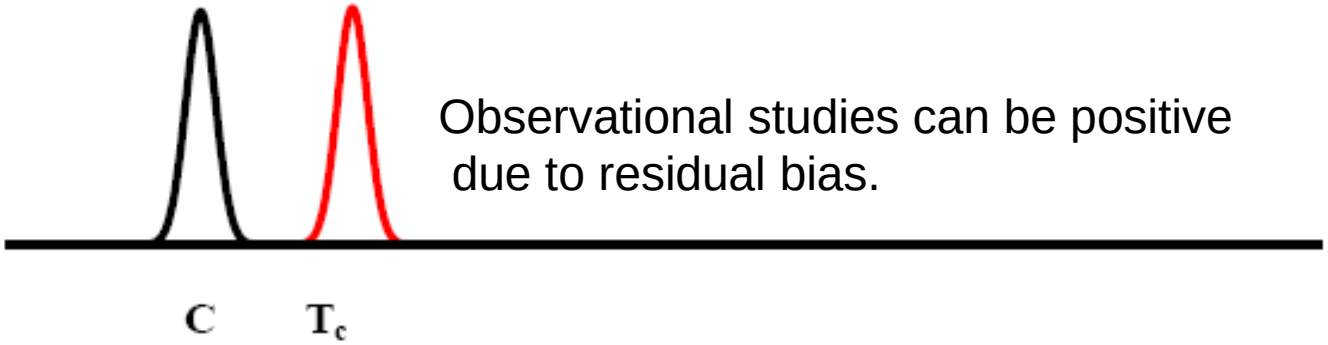
$$\Delta_{t\text{-}c} - [\text{known confounders}] = \beta_1 + [\text{unknown confounders}]$$

Residual bias: observational studies

(a) Use confounding variables to reduce bias.



(b) As n get large the standard error of the mean gets small.



Observational studies can be positive due to residual bias.

Possible Solution: Local Treatment Differences*

1. Cluster people base on covariates.

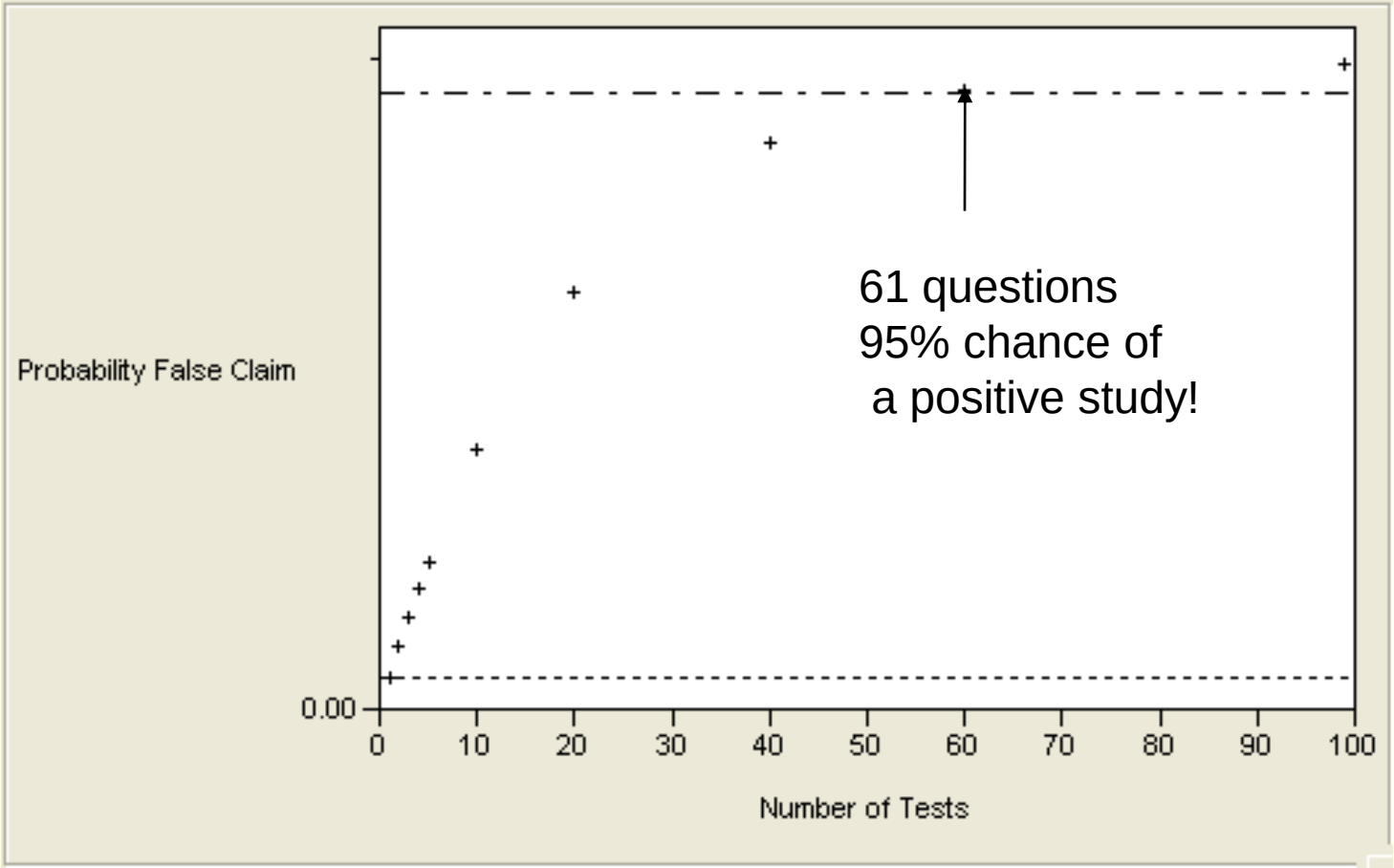
2. Compute treatment differences within clusters.

3. Examine LTDs over different cluster sizes.

4. Let the analysis unit be the cluster and use recursive partitioning to examine covariates.

* Bob Obenchain method.

Multiple testing will produce multiple "p-values < 0.05"



HIV Drug Classes (~864 combinations) (28 "main" effects)

NRTIs (12/8) (nucleoside or nucleotide reverse transcriptase inhibitors)

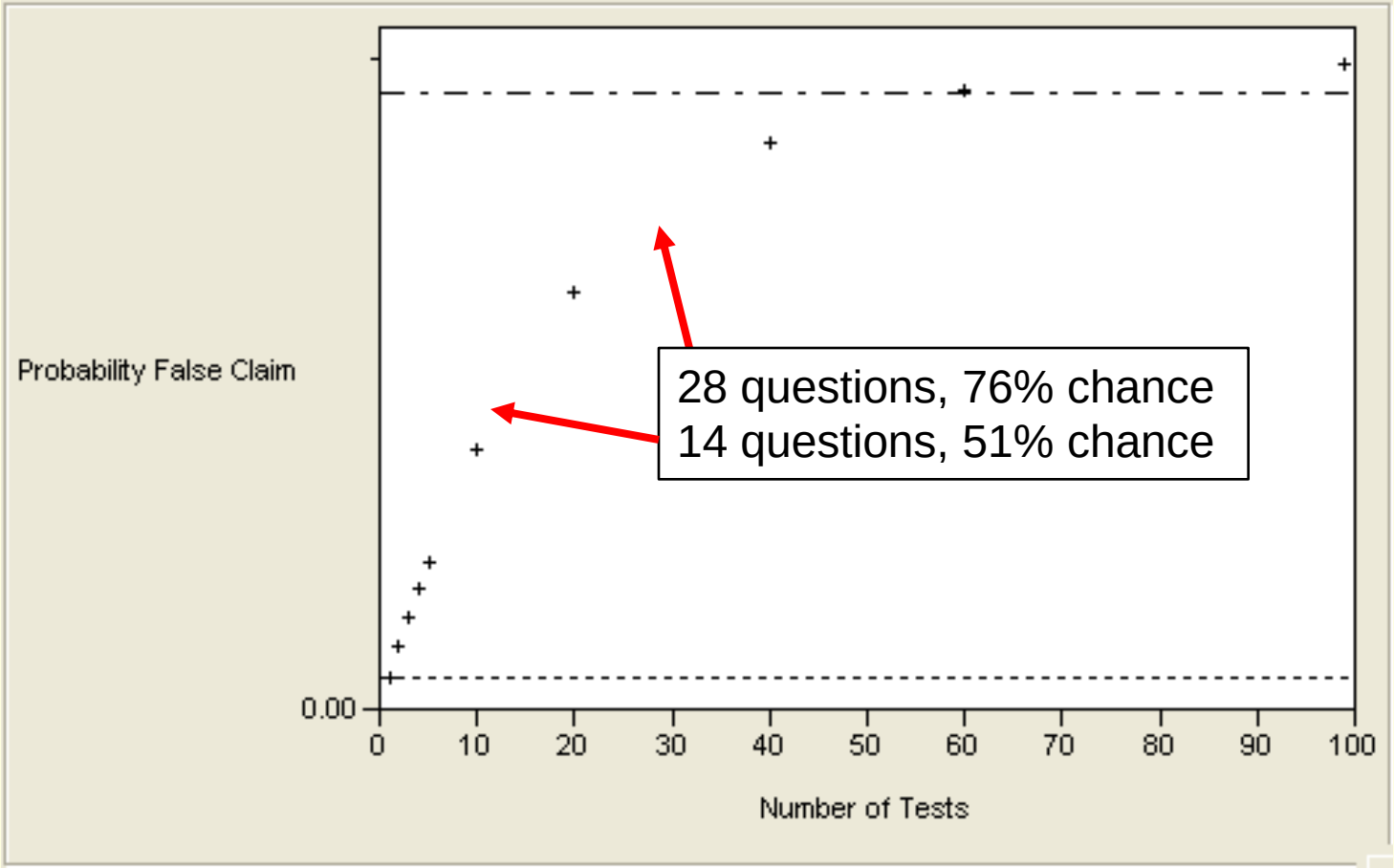
NNRTIs (4/4) (non-nucleoside reverse transcriptase inhibitors)

PIs (9/8)(protease inhibitors)

Entry inhibitors (2)

Integrase inhibitors (1)

Multiple testing will produce multiple "p-values < 0.05"



It is not that easy to count

Quantitative Evaluation of Multiplicity in Epidemiology and Public Health Research*.

173 articles examined, ~20 questions/article.

Attempted to count the questions at issue.

“The reporting style in some of the articles made the determination of the exact number of statistical tests conducted and the number found statistically significant a difficult task.”

*Ottenbacher, 1998, AJE

Potential solutions

0. Depend on others to replicate findings.

1. Run hypothesis generating study, followed by a focused study.

2. Define a family of tests, multiplicity adjust.

3. Test and hold out data sets.

Example of multiple testing/modeling

Association of Urinary Bisphenol A Concentration With Medical Disorders and Laboratory Abnormalities in Adults

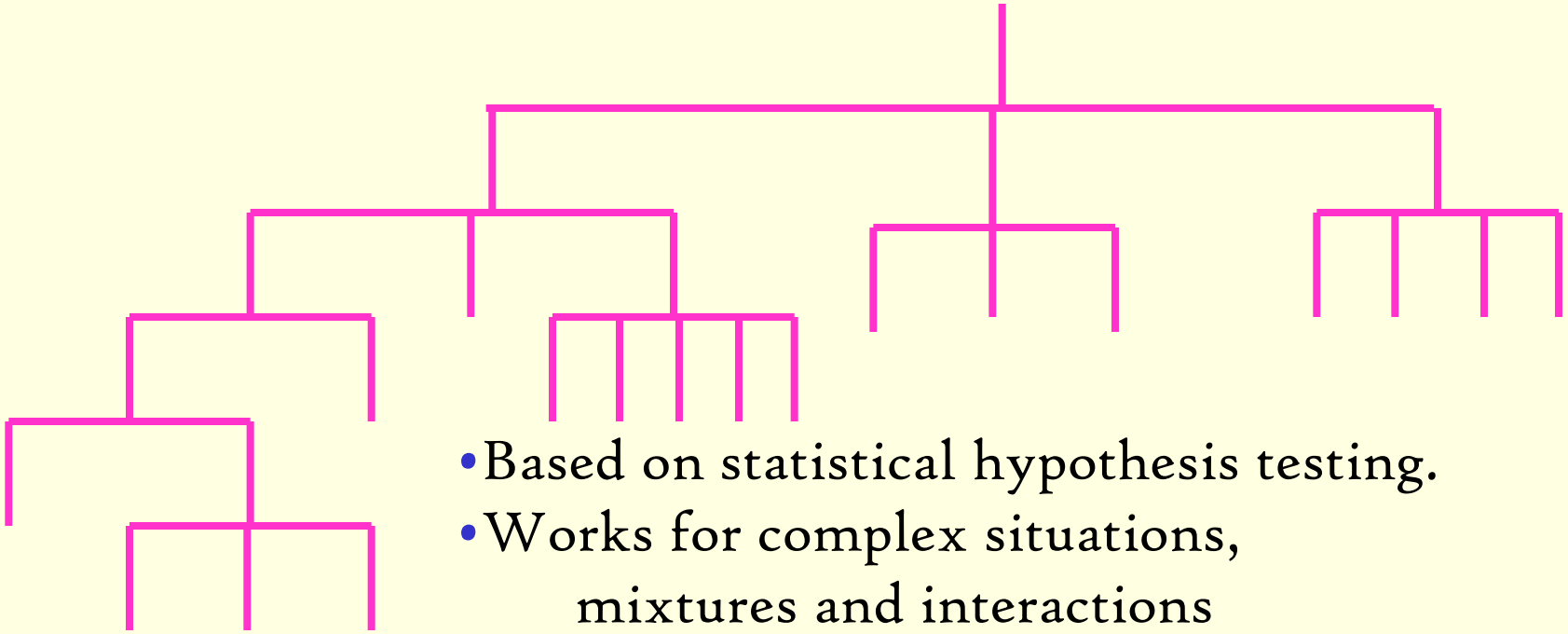
JAMA. 2008;300(11):1303-1310

1. 275 chemicals
2. 32 medical outcomes
3. 10 demographic covariates

$275 \times 32 = 8800 \times 2^{10} =$ **~9 million**

Claims: diabetes and CVD

Recursive Partitioning: Finding Sub-Groups



- Based on statistical hypothesis testing.
- Works for complex situations, mixtures and interactions
- Statistical method easy to understand.
- Excellent for subgroup analysis.
- Handles more predictors than observations.

CV Risk Factors

1. Age
2. Gender
3. BMI
4. LDL/HDL
5. BP (systolic and diastolic)
6. Diabetes
7. Statins
8. Family history
9. Personal history
10. Smoking

Etc.

Things to consider

residual bias

multiple testing

multiple modeling

small effects

Without considerable care, every study will have positive effects.

Follow up causes worry and is costly.

References

Austin PC, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. J Clin Epidemiol 2006;59:964 – 969.

Ioannidis JPA. Contradicted and initially stronger effects in highly cited clinical research. JAMA. 2005;294:218-228

Ottenbacher KJ. Quantitative evaluation of multiplicity in epidemiology and public health research. Amer J Epi 1998;147:615-619.

Peto, R, Emberson, J. Landray, et al. Analysis of cancer data from three Ezetimibe trials. NEJM. 2008;358, 1357-1366.

Pocock SJ, Collier TJ, Dandreo KJ, et al. Issues in the reporting of epidemiological studies: a survey of recent practice. BMJ. 2004;329:883-888.