# STATISTICS 536B, Lecture #6

March 12, 2015

# Meta-Analysis - continued:

Selected comments prompted by the Algra and Rothwell paper

What's going on with statements like that in the abstract:

> *In case control studies, regular use of aspirin was associated with reduced risk of colorectal cancer (pooled odds ratio [OR] 0.62, 95% CI 0.58-0.67,* $p_{sig} < 0.0001$*, 17 studies), with little heterogeneity (*$p_{het} = 0.13$*) in effect between studies* ...

Relates to estimating $\tau^2$ in random effect meta-analysis

(Recall $Y_i|\theta_i \sim N(\theta_i, \sigma_i^2)$, $\theta_i \sim N(\mu, \tau^2)$)

# More thoughts from Algra and Rothwell

- Search strategy and selection criteria important (e.g., see Fig. 1)
- Note distinction between case-control studies, *standard cohort studies*, and *nested case-control* studies.
- Note emphasis on different *definitions* of exposure (e.g., Fig. 2). [And number of available studies depends on which definition is adopted.]

# Some general strengths of this work

- Thoughtful discussion/analysis of aspirin vs. colorectal cancer compared to aspirin vs. other cancers (Figs. 3, 4)
- Nicely aligned evidence:
  - association between aspirin and cancer incidence
  - association between aspirin and metastasis, given incidence (Fig. 5)
  - (lack of) association between aspirin and local spread, given incidence but no metastasis (Fig. 6)

# Congratulations: You've 'invented' a famous estimator!

# What is a "nested case-control" study???

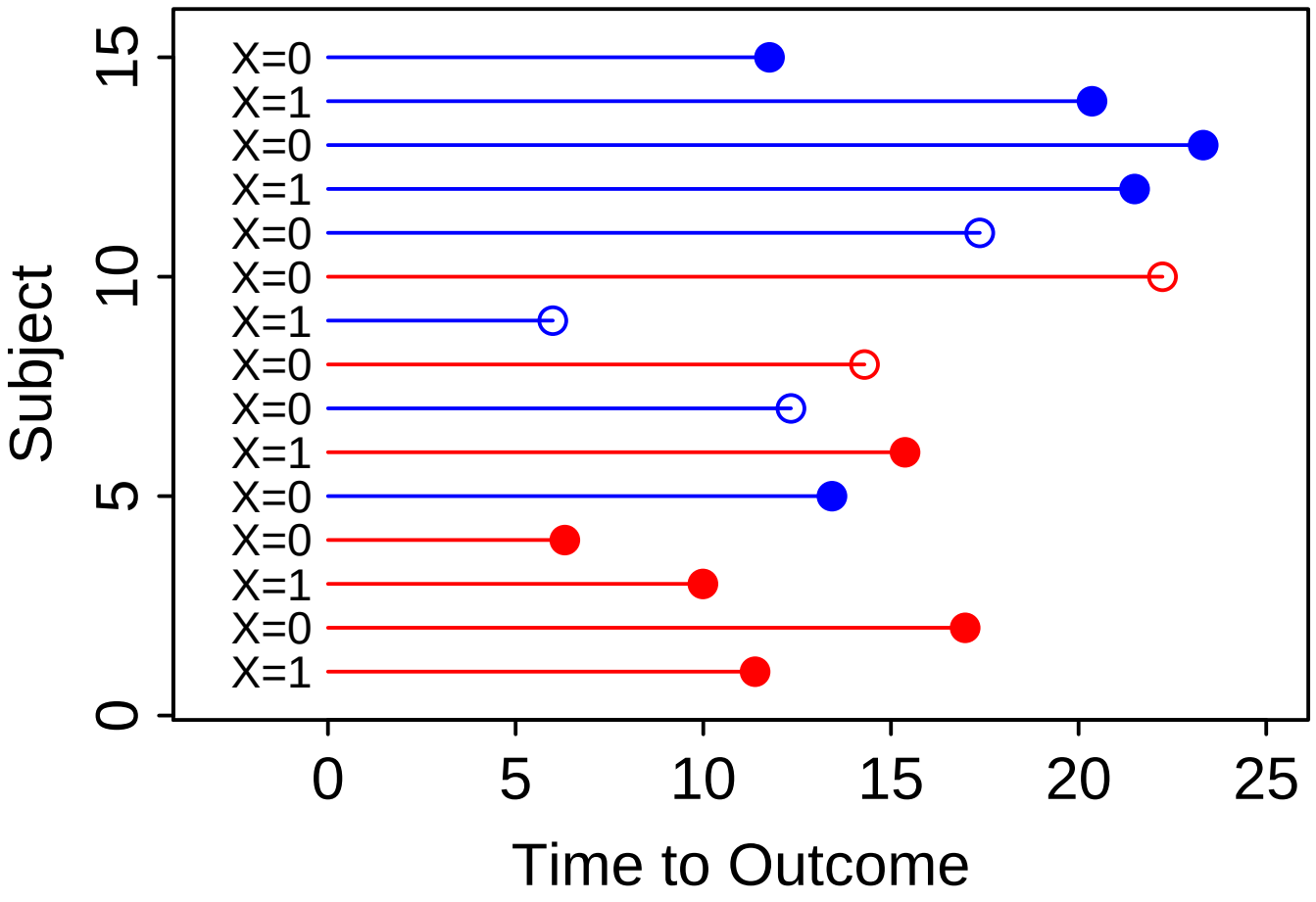
Think of a prospective cohort study

$T$ = time from "baseline" to bad outcome

$X$ = exposure (at baseline)

Could fit a survival analysis model for $(T|X)$. Or...

# Visualize the data



# Carry out a matched case-control study

For simplicity, think 1:1 matching as we considered before

For each case, randomly choose the control from amongst those subjects who:

- have matching covariate values
- are observed to be at risk at the case's failure time

So end up with matched case-control data with pairs in a 2 by 2 table, as before (recall, all the action is the discordant on X pairs, basically throw away the concordant on X pairs)

# In some specialized circumstances, nested case-control is wickedly good (compared to doing survival analysis)

Say $X$ is a binary genotype, say $Y$ is time to incident cancer

Maybe it's is cheap to freeze/store every subject's baseline blood sample

Maybe it's expensive to test the sample to determine if $X = 0$ or $X = 1$

Maybe reaching the disease outcome is quite rare

So if we only have to test the samples for the cases and their matched controls...