

Subgroup Analyses in Clinical Trials

Exploratory Subgroup Analyses: Why Do We Need Particular Caution?

Thomas R. Fleming, Ph.D.
Professor, Dept. of Biostatistics
University of Washington

- * Fleming TR “Clinical Trials: Discerning Hype from Substance”
 - *Annals of Internal Medicine* 2010; 153:400-406

Interest in “*Positive*” Results in Clinical Trials

➤ Industry Sponsors

~ Company profits, ↑ value of stock options, promotion

➤ Government Sponsors

~ Claims of success in advancing health care

~ Leverage for ↑ in federal funding

➤ Journal Editors (Publication bias)

➤ Academic Investigators / Caregivers

~ Increased ability to publish results

↑ professional stature, earlier promotion, ↑ salary

~ Desire to offer more therapeutic options to patients

....Result: *Wide Spread & Significant Conflicts of Interest*

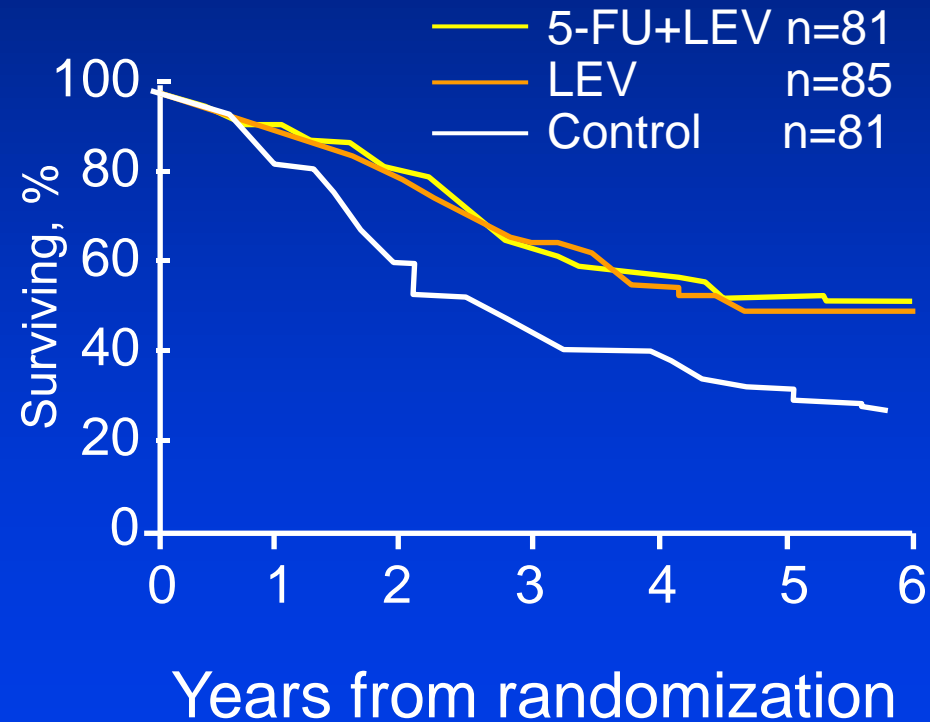
An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Surgical Adjuvant Therapy
of Colorectal Cancer



Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial



NORTH CENTRAL TREATMENT GROUP STUDY

Looking at Treatment Effect on Overall Survival

Females Only

Males Only

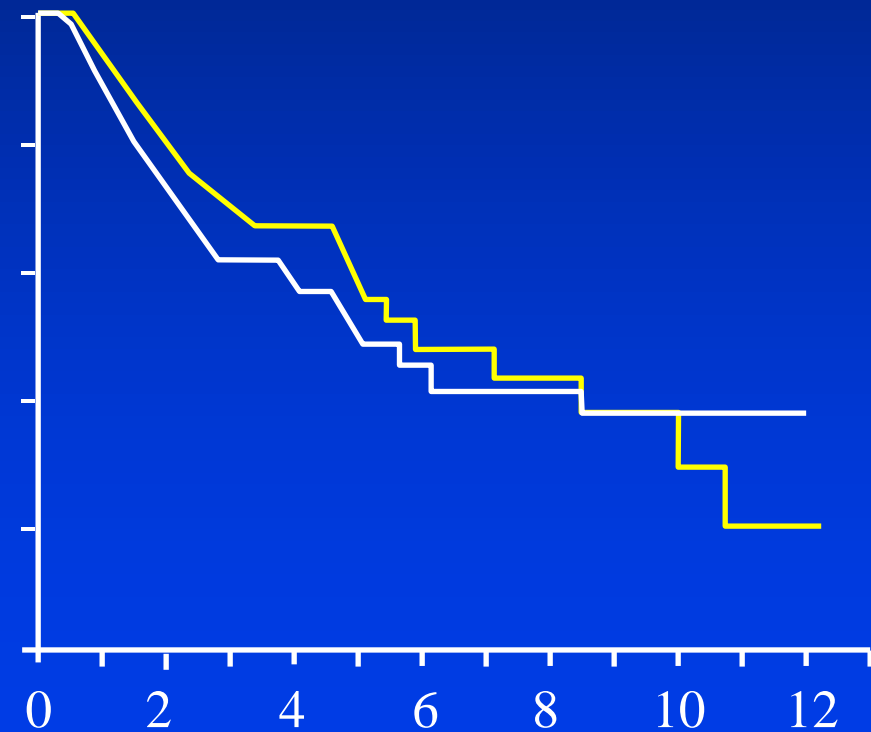
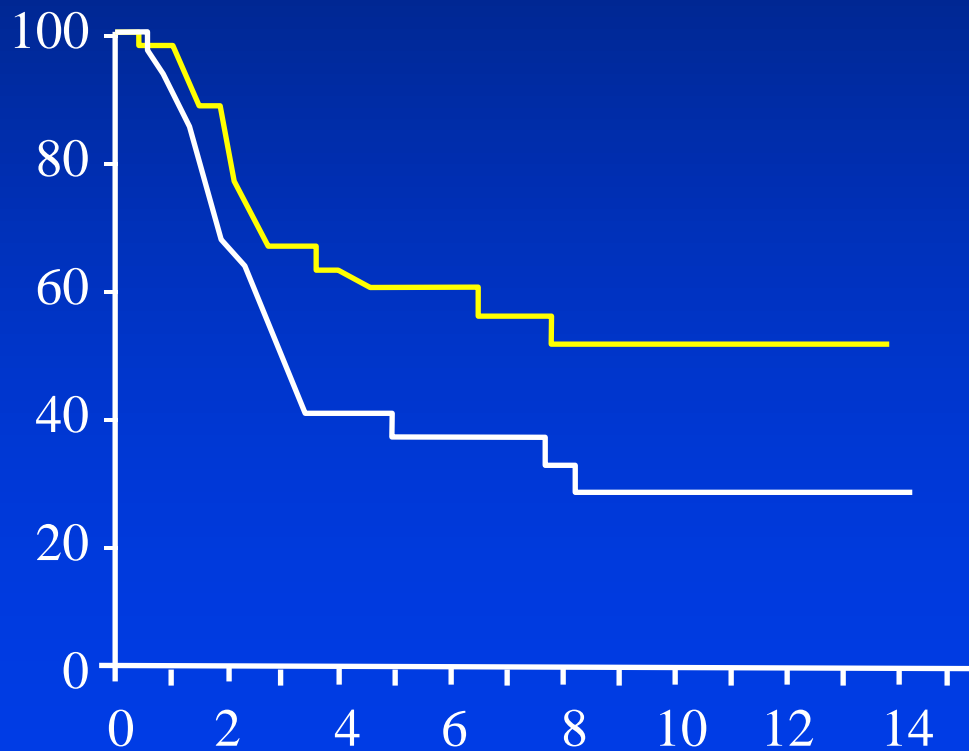
At Risk Death 5-Yr Estimate

At Risk Death 5-Yr Estimate

— 5-FU+Levamisole
— Follow-Up Only

| | | | |
|-----------------|----|----|-----|
| 5-FU+Levamisole | 44 | 21 | 57% |
| Follow-Up Only | 43 | 29 | 40% |

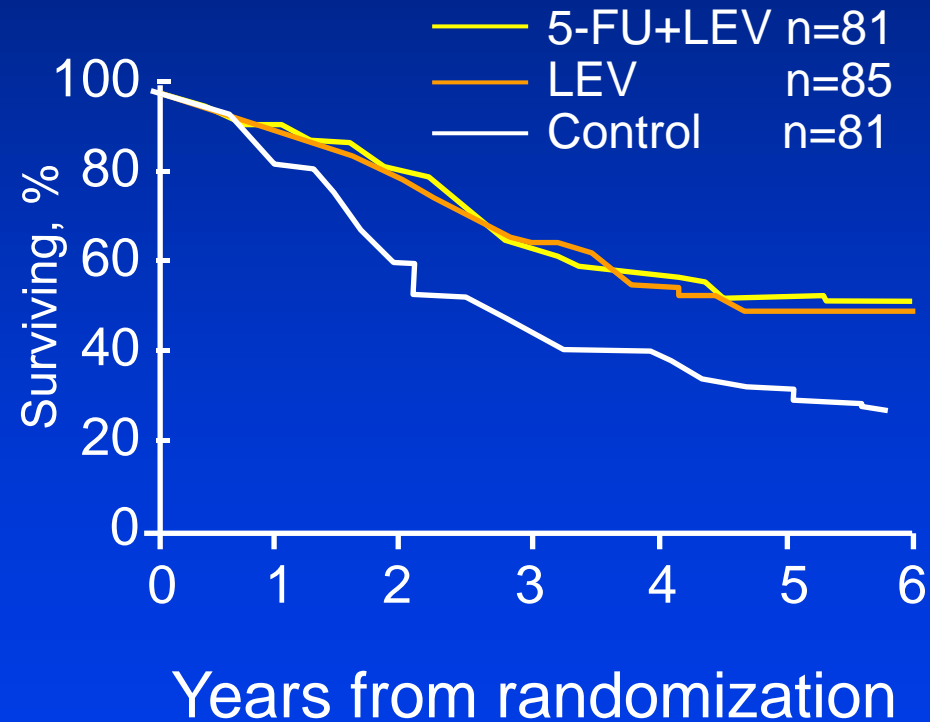
| | | | |
|-----------------|----|----|-----|
| 5-FU+Levamisole | 37 | 25 | 51% |
| Follow-Up Only | 38 | 24 | 47% |



Years from Registration

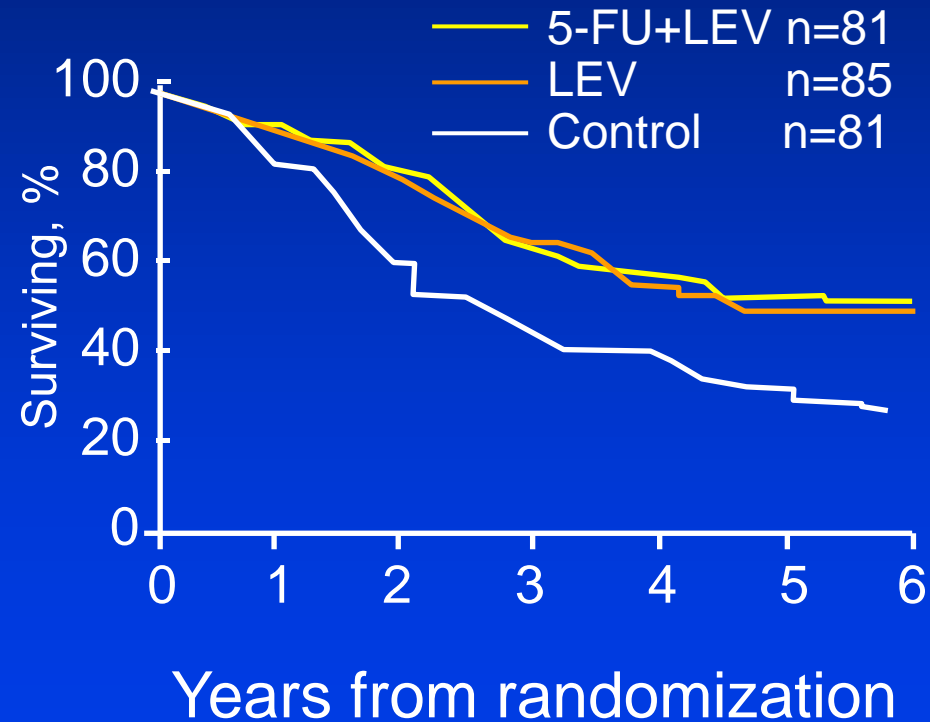
Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial

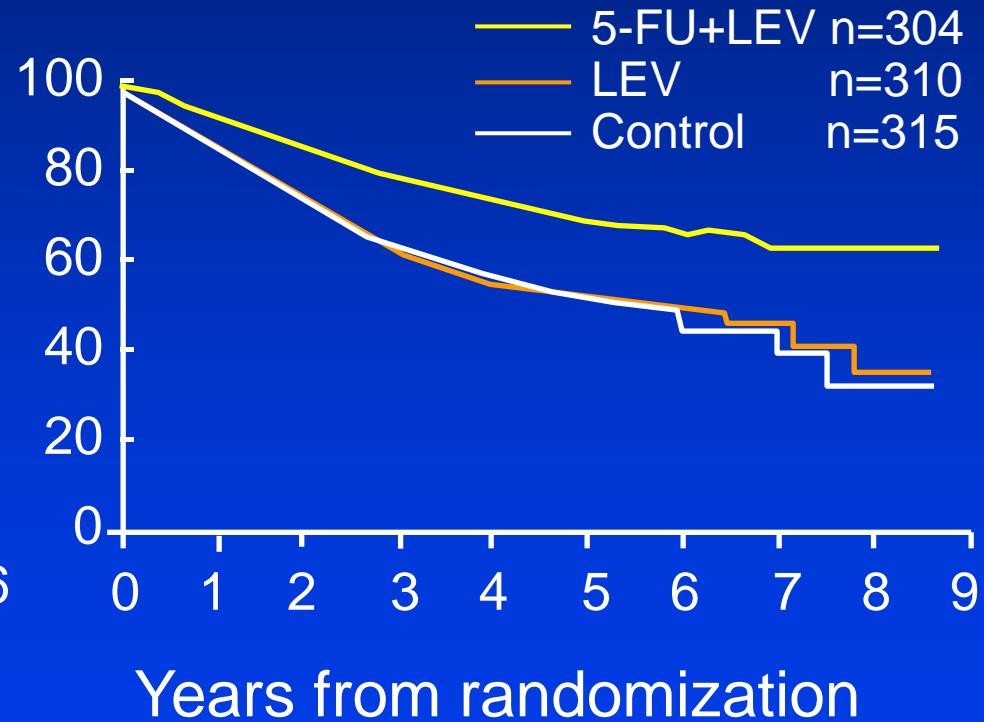


Surgical Adjuvant Therapy: Colorectal Cancer

NCCTG Trial



Cancer Intergroup Trial



INTERGROUP STUDY 0035

Looking at Treatment Effect on Overall Survival

Females Only

Males Only

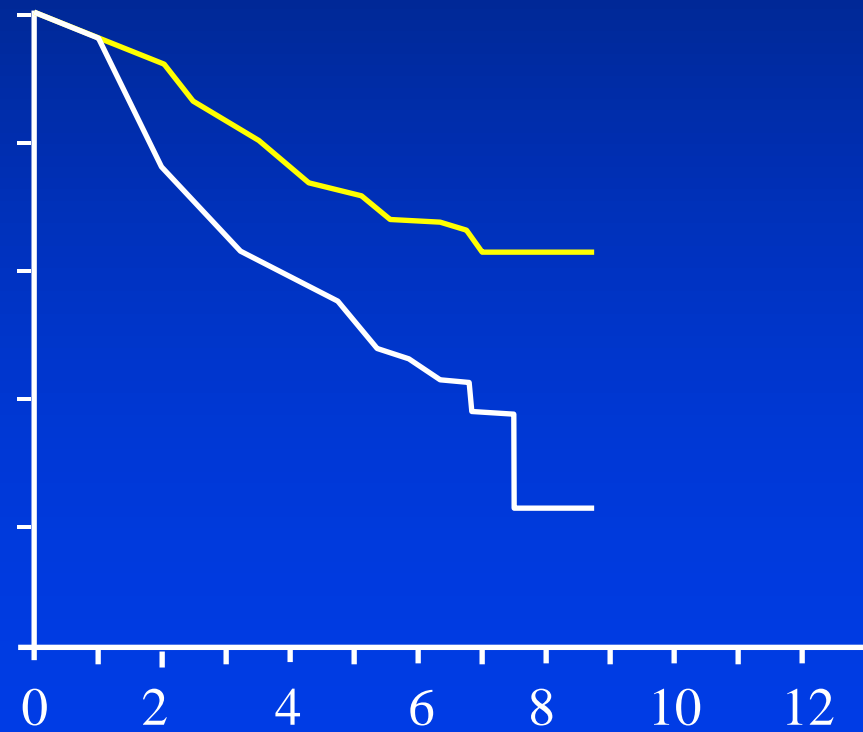
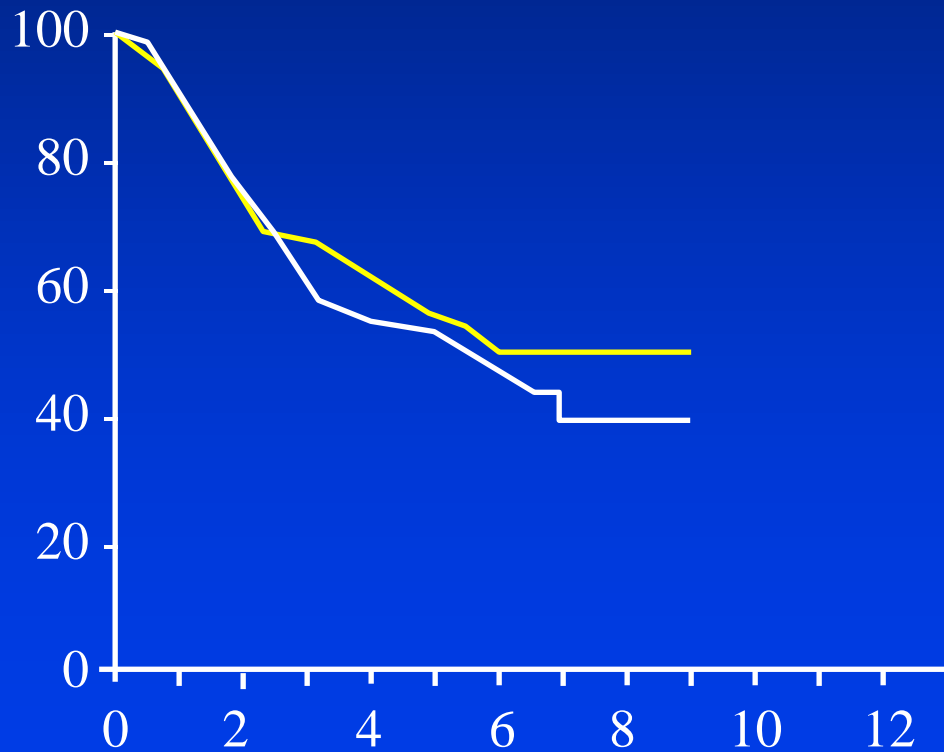
At Risk Death 5-Yr Estimate

At Risk Death 5-Yr Estimate

— 5-FU+Levamisole
— Follow-Up Only

| | | |
|-----|----|-----|
| 163 | 74 | 58% |
| 149 | 77 | 54% |

| | | |
|-----|----|-----|
| 141 | 47 | 70% |
| 166 | 91 | 51% |



Years from Registration

Duke's C Colon Cancer Adjuvant

Percent ↓ in Death Rate: 5-FU + Levamisole
Control

| Analysis Group | North Central Treatment Group Study (n = 162) | Intergroup Study # 0035 (n = 619) |
|----------------|---|-----------------------------------|
| All patients | 28% | 33% |
| Female | 43% | 15% |
| Male | 9% | 50% |
| Young | 40% | 23% |
| Old | 13% | 41% |

Confirmatory vs. Exploratory Analyses

- **Hyp. Confirmation vs. Hyp. Generation**
 - ~ **Post-hoc analyses & *Random High Bias***
(new endpoints, new analyses, interim analyses
subgroup analyses, covariate adjustments)

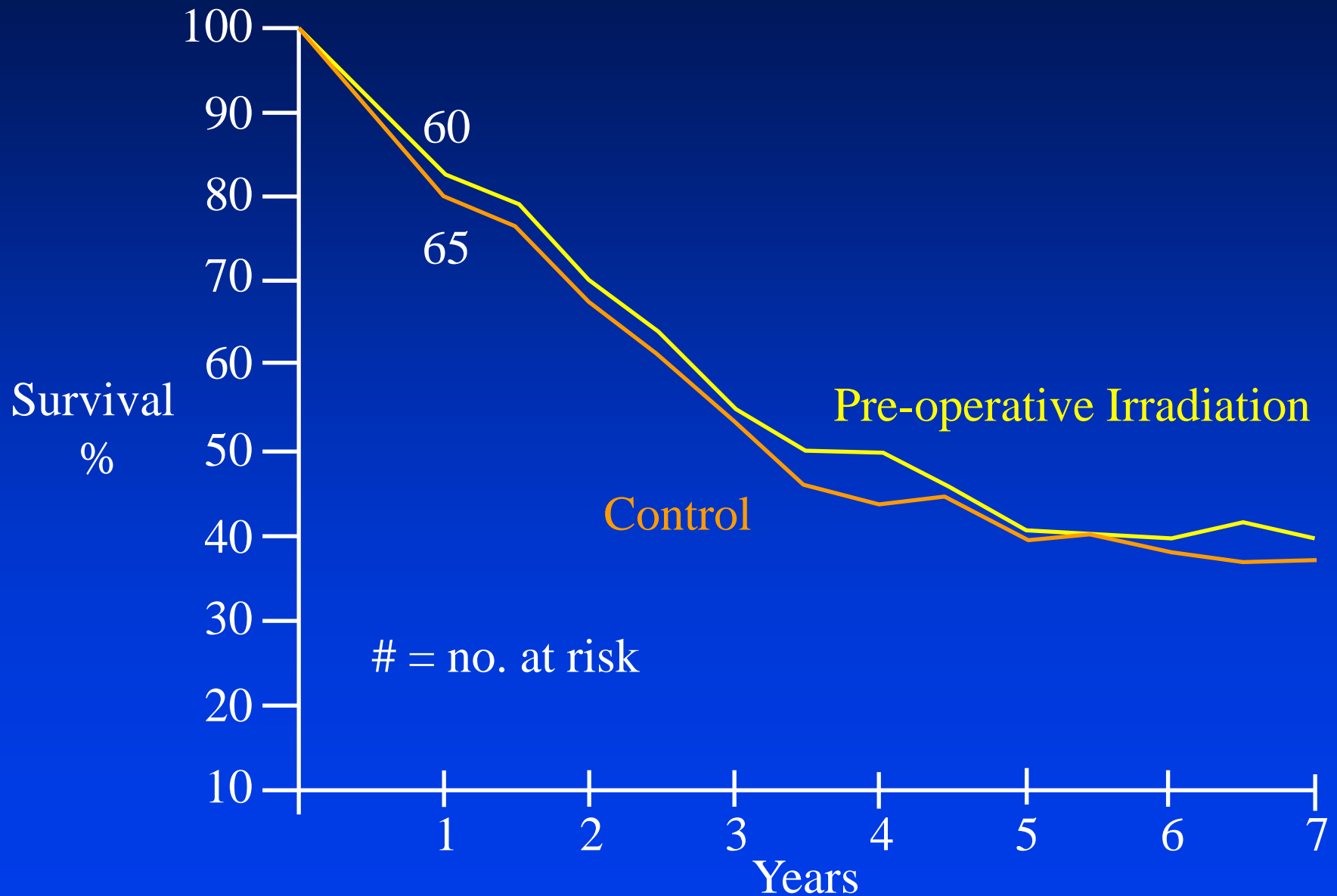
Illustrations and Motivation:
Baseball & Clinical Research

An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Radiation Treatment in Rectal Cancer
Princess Margaret Hospital

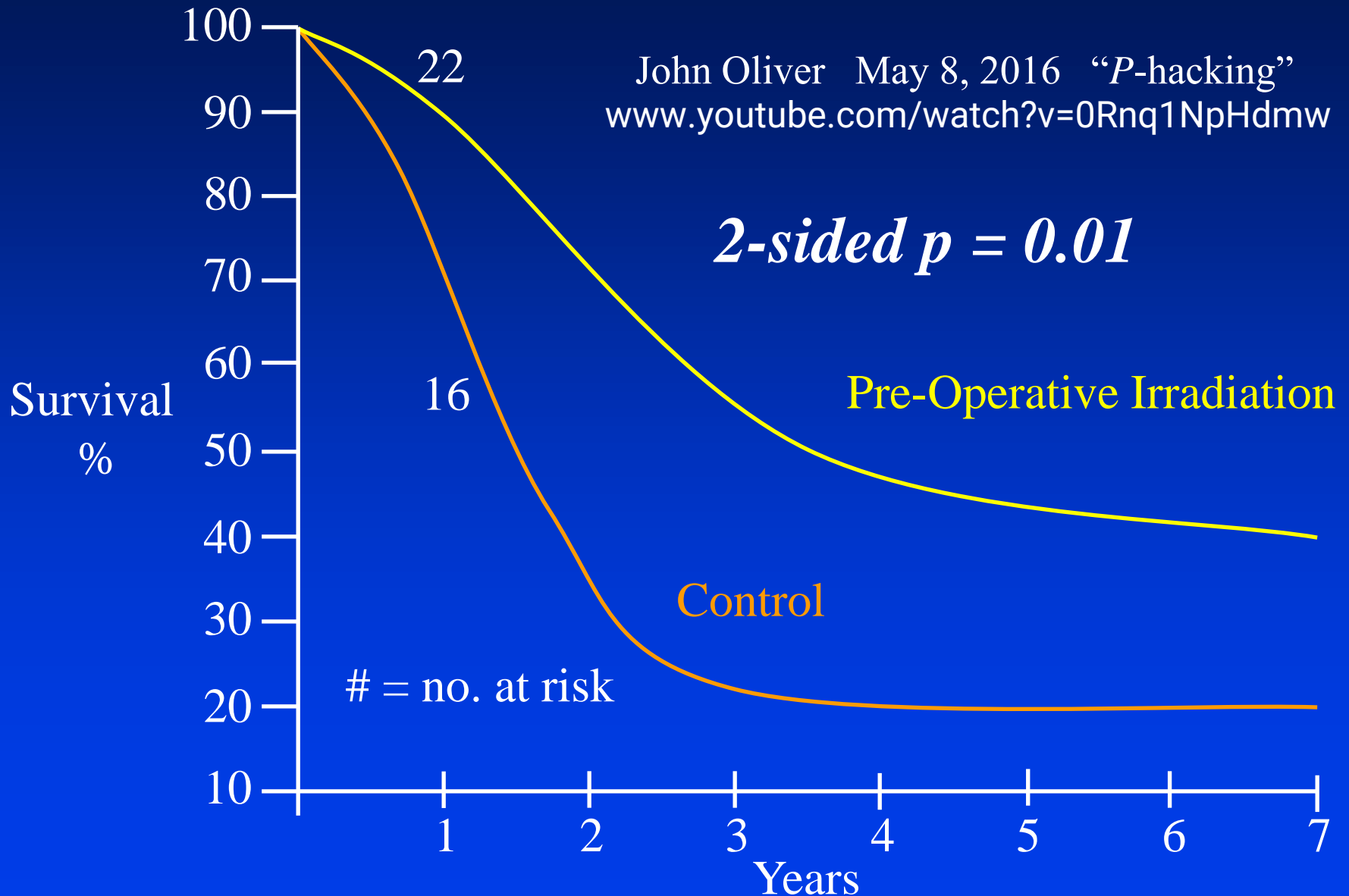


Survival of Patients with Rectal Carcinoma Princess Margaret Hospital, Toronto (1977)

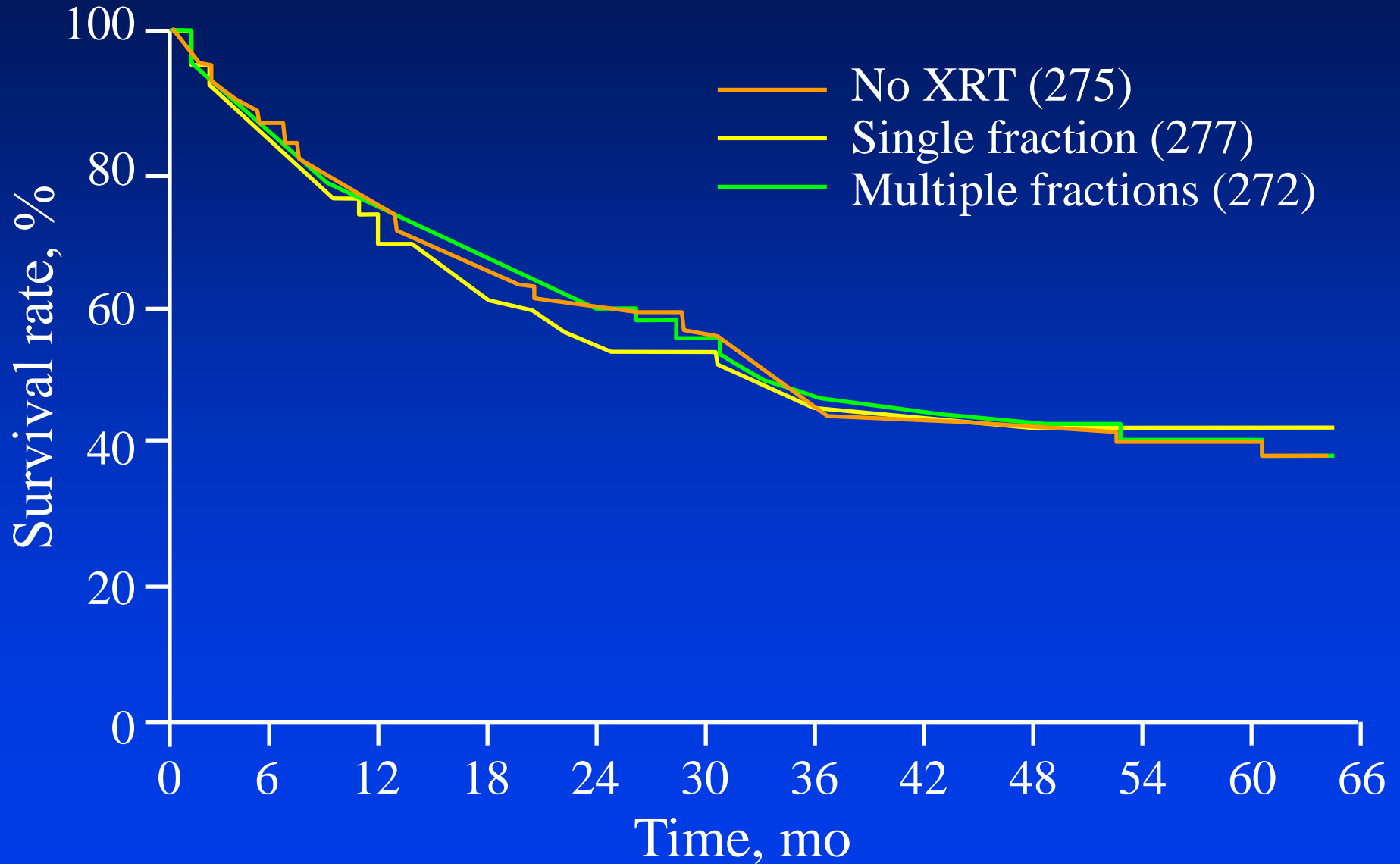


Survival of Patients with Rectal Carcinoma

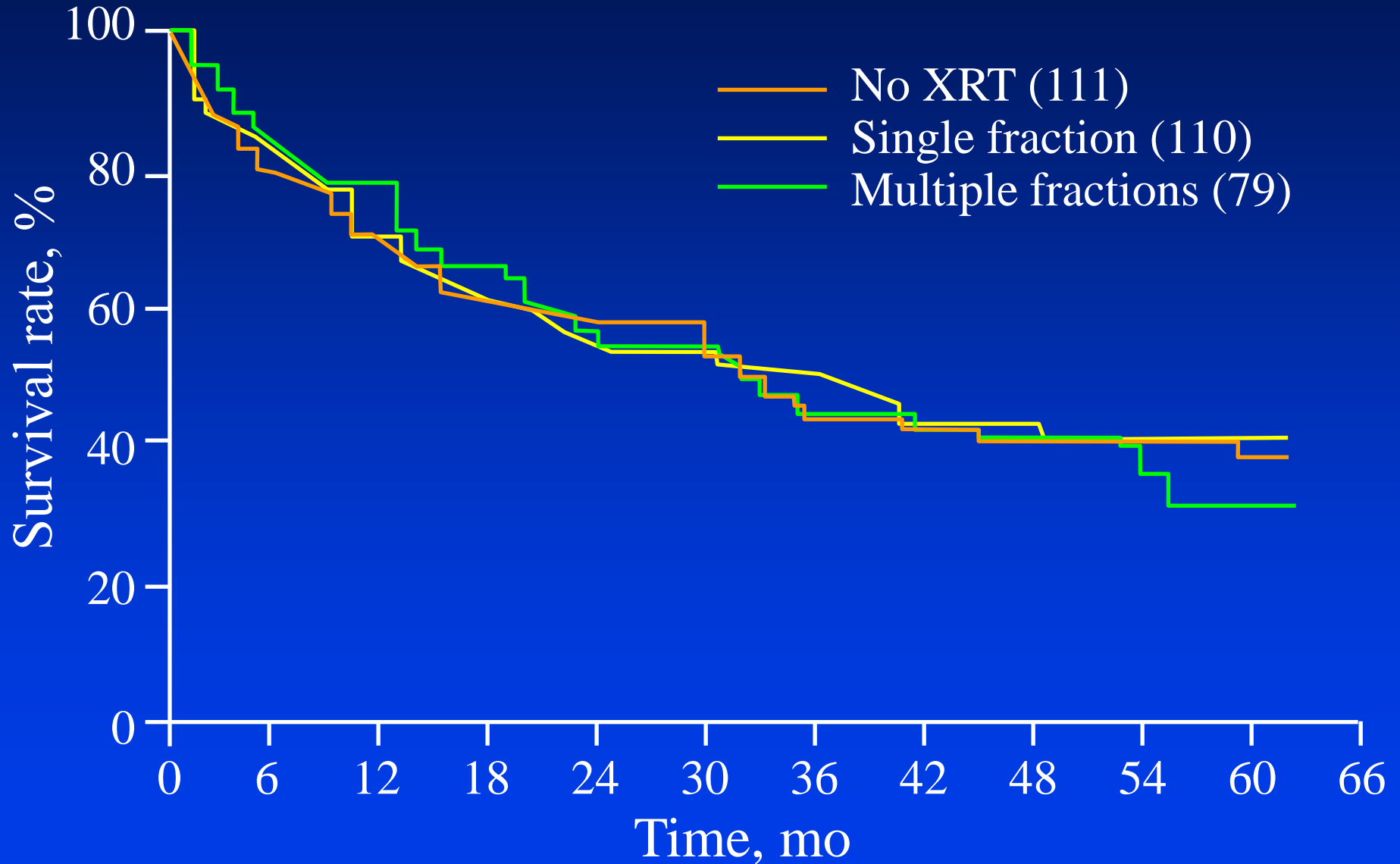
Exploratory Subgroup: **Dukes' Stage C Disease**



Medical Research Council (MRC) Confirmatory Trial



MRC Subgroup Analysis: **Dukes' C Cases**



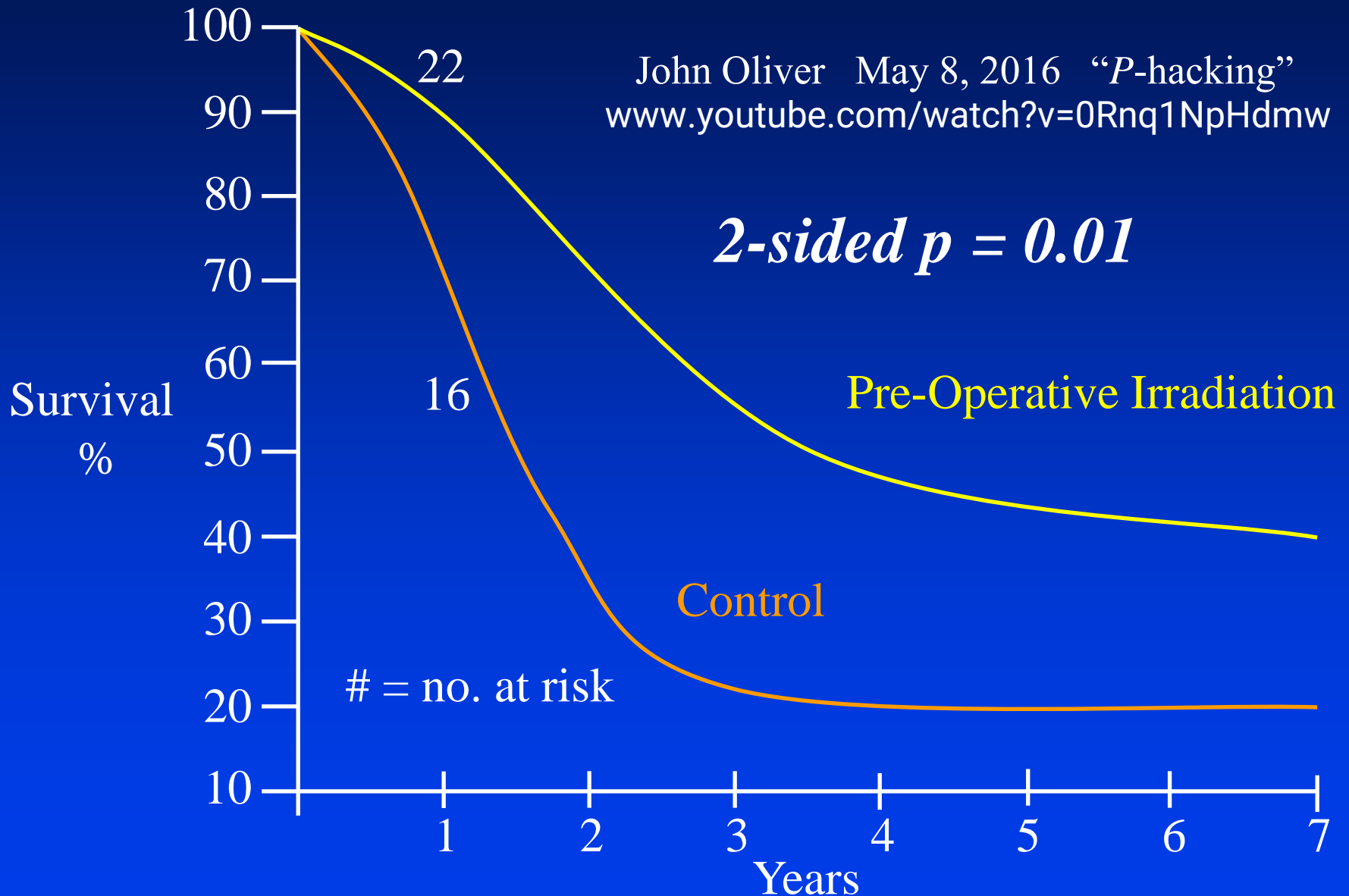
Some Important Observations

- P -values are only interpretable when you understand the sampling context from which they were derived
- Random High bias is real
- Exploratory Analyses usually should be viewed to be “Hypothesis Generating”
- Confirmatory Trials
greatly enhance the reliability of conclusions

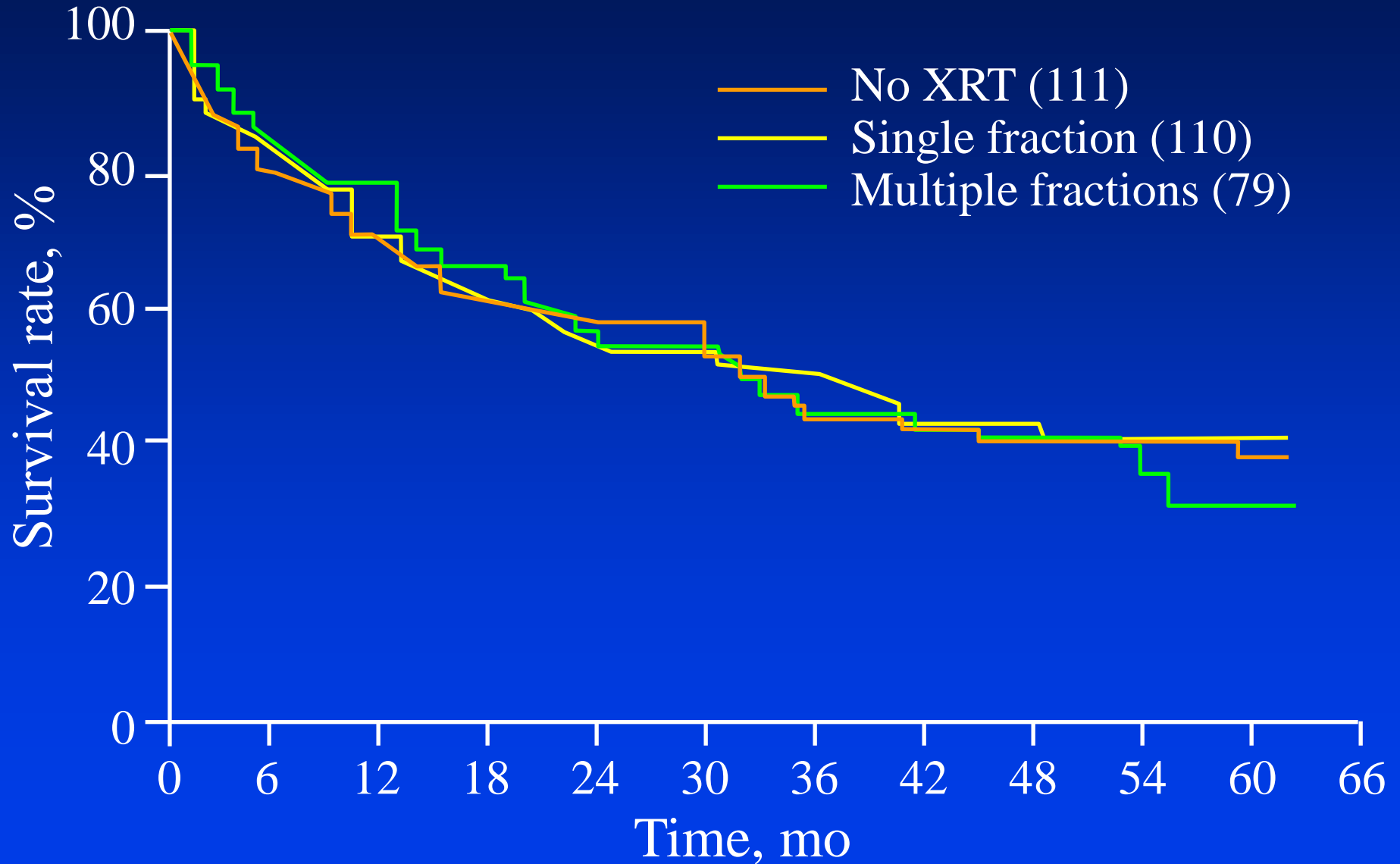
Fleming TR “Clinical Trials: Discerning Hype from Substance”
Annals of Internal Medicine 2010; 153:400-406

Survival of Patients with Rectal Carcinoma

Exploratory Subgroup: **Dukes' Stage C Disease**



MRC Subgroup Analysis: **Dukes' C Cases**



“It isn’t so much the things we *don’t know*
that get us in trouble.

It’s the things we *know* that aren’t so”.

—Artemus Ward (1834-1867)

Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet '86)
 - **SK** reduces mortality by 20%
- confined to:
- anterior MI**
 - < 65 years**
 - < 6 hours from symptom onset**

Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet '86)
 - **SK** reduces mortality by 20%
confined to:
 - anterior MI
 - < 65 years
 - < 6 hours from symptom onset
 - Subset restriction not confirmed by ISIS-2, ASSET, AIMS
 - While in ISIS-2:
 - Aspirin beneficial overall...**

Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet '86)
 - **SK** reduces mortality by 20%
confined to:
 - anterior MI
 - < 65 years
 - < 6 hours from symptom onset
 - Subset restriction not confirmed by ISIS-2, ASSET, AIMS
 - While in ISIS-2:
 - Aspirin** beneficial overall...
 - ... yet **harmful** to patients with
astrological signs **Libra** and **Gemini**

Bias for “Positive” Results in Clinical Trials

- Protocol Specified Clinical Trial “Primary Objective”
 - Very frequent wording:
 - ~ “To *establish* that the experimental regimen is safe and effective”

Bias for “Positive” Results in Clinical Trials

- Protocol Specified Clinical Trial “Primary Objective”
 - Very frequent wording:
 - ~ “To *establish* that the experimental regimen is safe and effective”
 - Scientifically unbiased wording:
 - ~ “To *determine whether* the experimental regimen is safe and effective”
- Regulatory Industry Statistics Workshop: (9/22/2011)
 - ...Credibility of exploratory analyses...
 - ...a Paradox...

Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- Trial #1: Time to renal flare: Minimal effect, ($2p = 0.51$)

Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- Trial #1: Time to renal flare: Minimal effect, ($2p = 0.51$)
...exploratory *high affinity* subgroup: $2p = 0.007$
- Trial #2 conducted in *high affinity* subgroup:
Time to renal flare:

Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- Trial #1: Time to renal flare: Minimal effect, ($2p = 0.51$)
...exploratory *high affinity* subgroup: $2p = 0.007$
- Trial #2 conducted in *high affinity* subgroup:
Time to renal flare: Minimal non-significant effect

Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- Trial #1: Time to renal flare: Minimal effect, ($2p = 0.51$)
...exploratory *high affinity* subgroup: $2p = 0.007$
- Trial #2 conducted in *high affinity* subgroup:
Time to renal flare: Minimal non-significant effect
...exploratory *truncation at 12 months* is favorable
- Trial #3 conducted in *high affinity* subgroup
with prespecified *truncation at 12 months follow-up*:

Interest in “Positive” Results in Clinical Trials

- Abetimus Sodium: Reducing Renal Flare Rate in Lupus
- Trial #1: Time to renal flare: Minimal effect, ($2p = 0.51$)
...exploratory *high affinity* subgroup: $2p = 0.007$
- Trial #2 conducted in *high affinity* subgroup:
Time to renal flare: Minimal non-significant effect
...exploratory *truncation at 12 months* is favorable
- Trial #3 conducted in *high affinity* subgroup
with prespecified *truncation at 12 months follow-up*:
...early termination by DMC for futility.

“If you Torture Data Long Enough,
They will Confess”

- * Fleming TR “Clinical Trials: Discerning Hype from Substance”
Annals of Internal Medicine 2010; 153:400-406

Some Conclusions

- Recognize bias resulting from strong interest to achieve “positive” results
- When refereeing journal publications, request:
 - *the clinical trial protocol*
 - *the statistical analysis plan (SAP)*
 - *the clinical study report (CSR)*
- The only P -values presented in CSRs & publications should be for α -spending analyses pre-specified in the SAP
- Recognize unreliability of Exploratory Analyses...
...generating hypotheses, but with “random high” bias
- Exploratory subgroup analyses should be presented descriptively, for example using forest plots

Some Conclusions

- For reliable evidence regarding effects by subgroups, such as evaluating effects in biomarker positive vs negative subgroups, it is important to have pre-specified hypotheses (potentially with alpha spending)
- Cautionary Note: *“When it is prespecified that biomarker-negative patients should not be included in the primary analysis of treatment effect in biomarker-positive patients because of the likelihood that treatment effects would differ between the 2 subgroups, it is logically inconsistent to include biomarker-positive patients in the primary analysis of treatment effect in biomarker-negative patients.” **
- * Rothmann et al. “Testing in a Pre-specified Subgroup and the Intent-to-treat Population. *Drug Information Journal* 46(2): 175-179, 2012.

