Summer Institute in Statistics for Clinical Research

# Exploratory Analyses: Why Do We Need Particular Caution?

**July 28, 2017**

**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

# Data Driven Hypothesis for the Cancer Risk with Vytorin in Aortic-Valve Stenosis

- **SEAS Trial**

| | N | CA. Incidence | CA. Deaths |
|---|---|---|---|
| Vytorin | 944 | 101 | 37 |
| Placebo | 929 | 65 | 20 |
| Relative Risk: | | **1.55** | **1.78** |
| 95% C.I.: | | (1.13, 2.12) | (1.03. 3.11) |

- **IMPROVE-IT & SHARP Trials**

| | N | CA. Incidence | CA. Deaths |
|---|---|---|---|
| Vytorin | 10,391 | 313 | 97 |
| Control | 10,298 | 326 | 72 |
| Relative Risk: | | **0.96** | **1.34** |
| 95% C.I.: | | (0.82, 1.12) | (0.98, 1.84) |

# *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*

# Bias for "Positive" Results in Clinical Trials

~ What is the definition of a

*successful* clinical trial?

➢ A very common response:

"A clinical trial that achieves a *positive* result"

# Bias for "Positive" Results in Clinical Trials

~ What is the definition of a

*successful* clinical trial?

- A very common response:

  "A clinical trial that achieves a *positive* result"

- The proper scientific response:

  "A clinical trial that

  addresses a clinically important issue,

  and that *reliably answers* the questions

  it was designed to address"

# *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)

# *Confirmatory* vs. *Exploratory* Analyses

- Clinical Endpoints in Pulmonary Arterial Hypertension
  - ~ Overall survival
  - ~ Quality of Life: SF-36 (8 domains), Borg Dyspnea Score
- ~ NYHA Functional Class
  - ~ **6MWT**: @18 wk, **24 wk**, 48 wk, etc.
  - ~ Time to Clinical Worsening
    - ✓ Death, PAH Hosp, L.T., (NYHA↑ & 6MWT↓)
- Analysis Methods
  - ~ Normally distributed: **T-test**, ANCOVA, Wilcoxon
  - ~ Time to event: Log-rank, Cox Regression
  - ~ Dichotomous: Fisher's Exact Test, Pearson $\chi^2$

# *Confirmatory* vs. *Exploratory* Analyses

- Biomarker Endpoints (Hemodynamic parameters)
  - ~ Pulmonary Arterial Pressure
  - ~ Systolic & Diastolic Systemic Arterial Pressure
  - ~ Systemic & Pulmonary Vascular Resistance
  - ~ Heart Rate & Cardiac Output
- Analyses over Calendar Time
- ~ Normally distributed: T-test, ANOVA, Wilcoxon
- ~ Time to event: Log-rank, Cox Regression
  ~ Dichotomous: Fisher's Exact Test, Pearson $\chi 2$

# *Confirmatory* vs. *Exploratory* Analyses

- Subgroup Analysis & Prognostic Covariate Adjustment

  ~ WHO PAH Functional Class: I v II v III v IV

  ~ Etiology: Idiopathic PAH, Assoc w CTD, SLE, Other

  ~ Baseline Walking Distance: < 325 v > **325 meters**

  ~ Gender: male v female

  ~ Age: By decade

  ~ Ethnicity: White v Black v Asian v Other

  ~ mean PAP: < 50 v > 50

Epoprostenol +/– Sildenafil

# *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:

# *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:

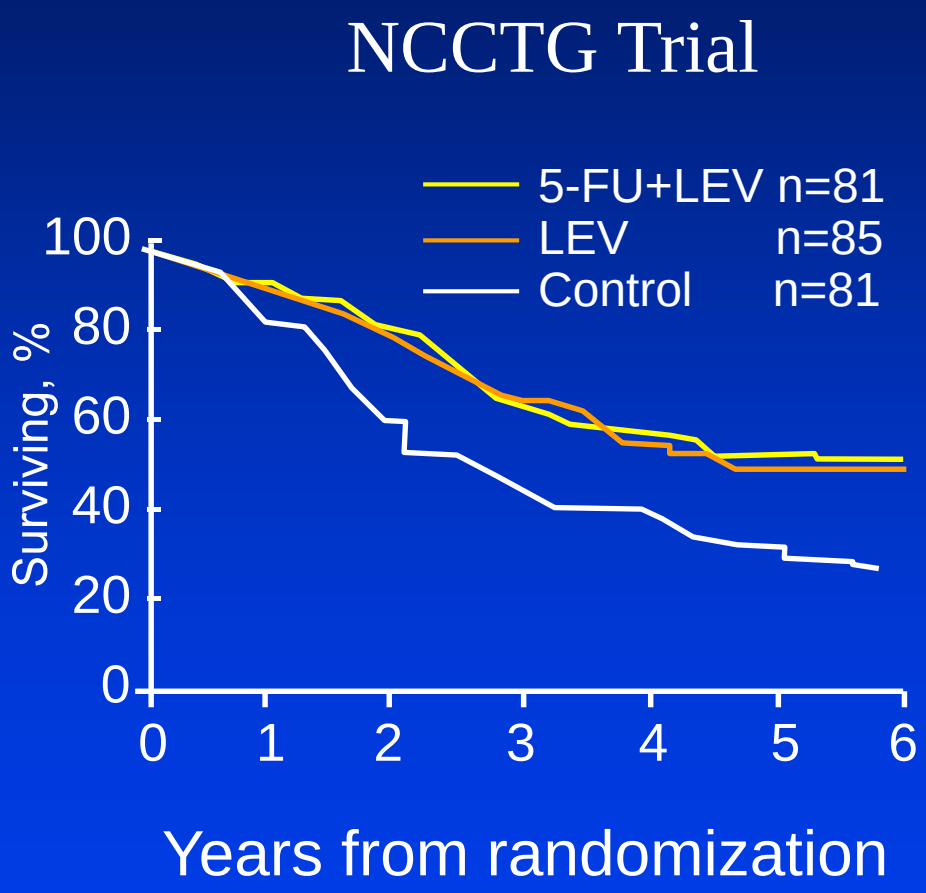
***Maternity Wards***, Baseball & Clinical Research

20 vs 2: (.71, .99), 2p = 0.0001

# An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Surgical Adjuvant Therapy of Colorectal Cancer

# Surgical Adjuvant Therapy: Colorectal Cancer



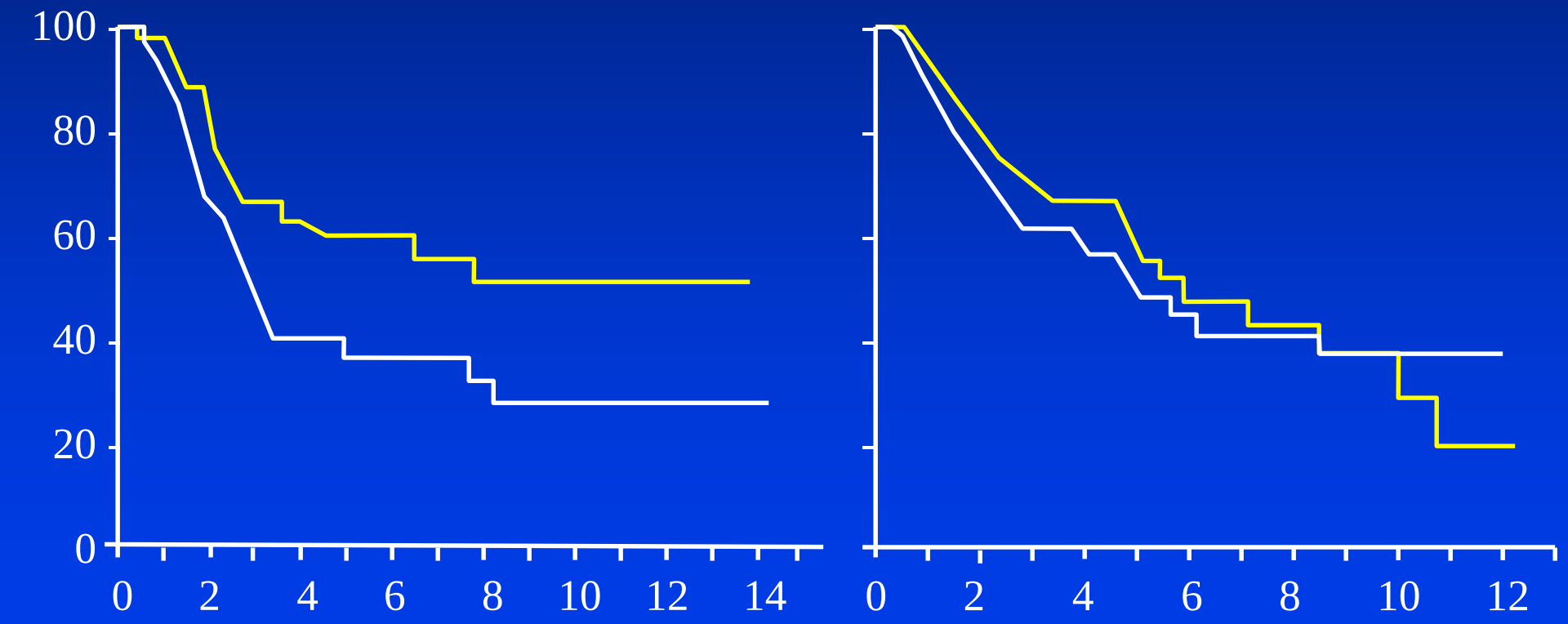
# NORTH CENTRAL TREATMENT GROUP STUDY
## Looking at Treatment Effect on Overall Survival

### Females Only

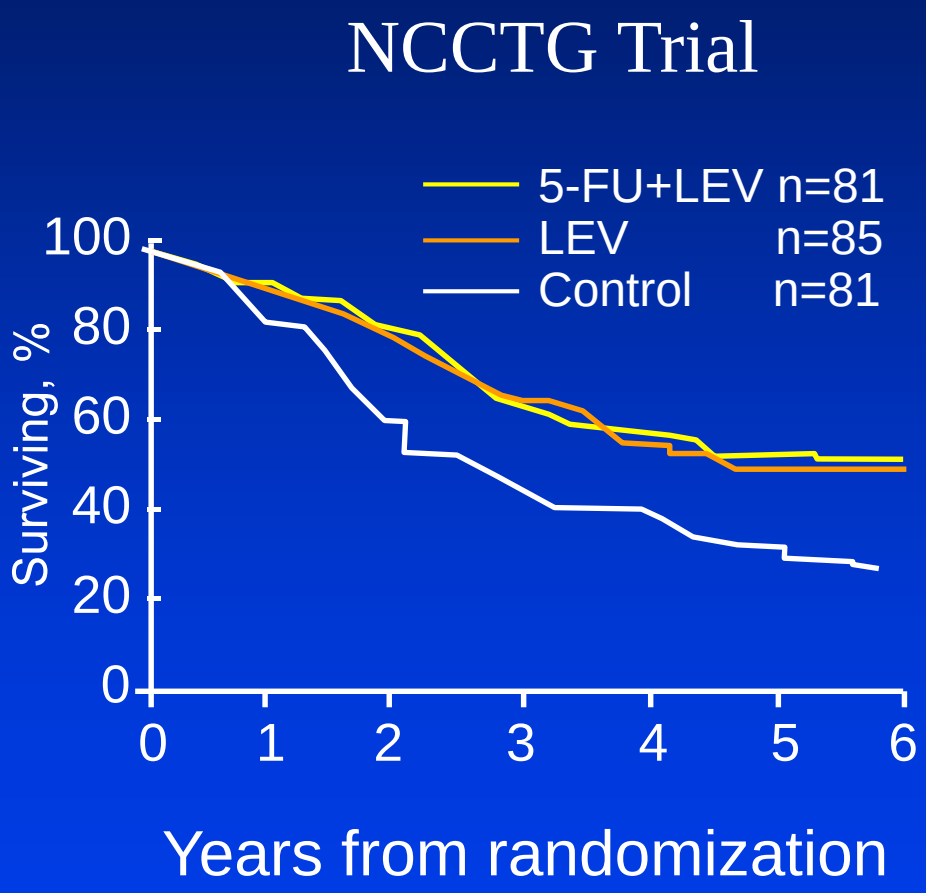
| | At Risk | Death | 5-Yr Estimate |
|---|---|---|---|
| 5-FU+Levamisole | 44 | 21 | 57% |
| Follow-Up Only | 43 | 29 | 40% |

### Males Only

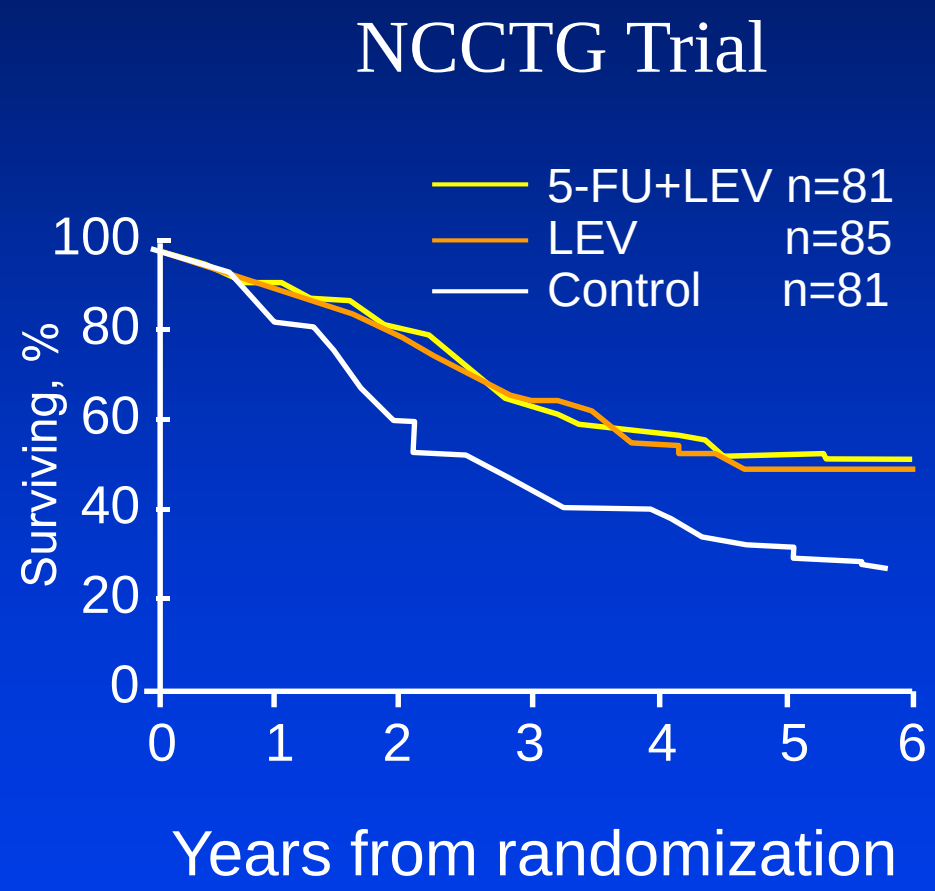
| | At Risk | Death | 5-Yr Estimate |
|---|---|---|---|
| 5-FU+Levamisole | 37 | 25 | 51% |
| Follow-Up Only | 38 | 24 | 47% |

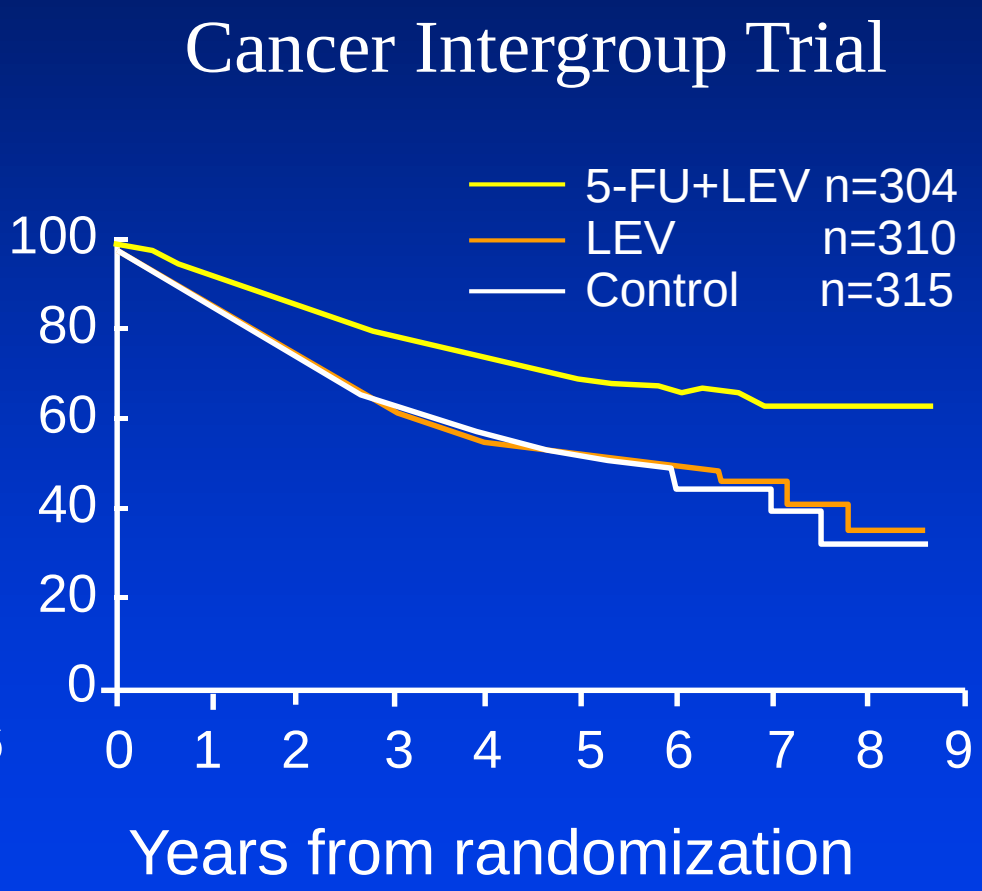


# Surgical Adjuvant Therapy: Colorectal Cancer



# Surgical Adjuvant Therapy: Colorectal Cancer



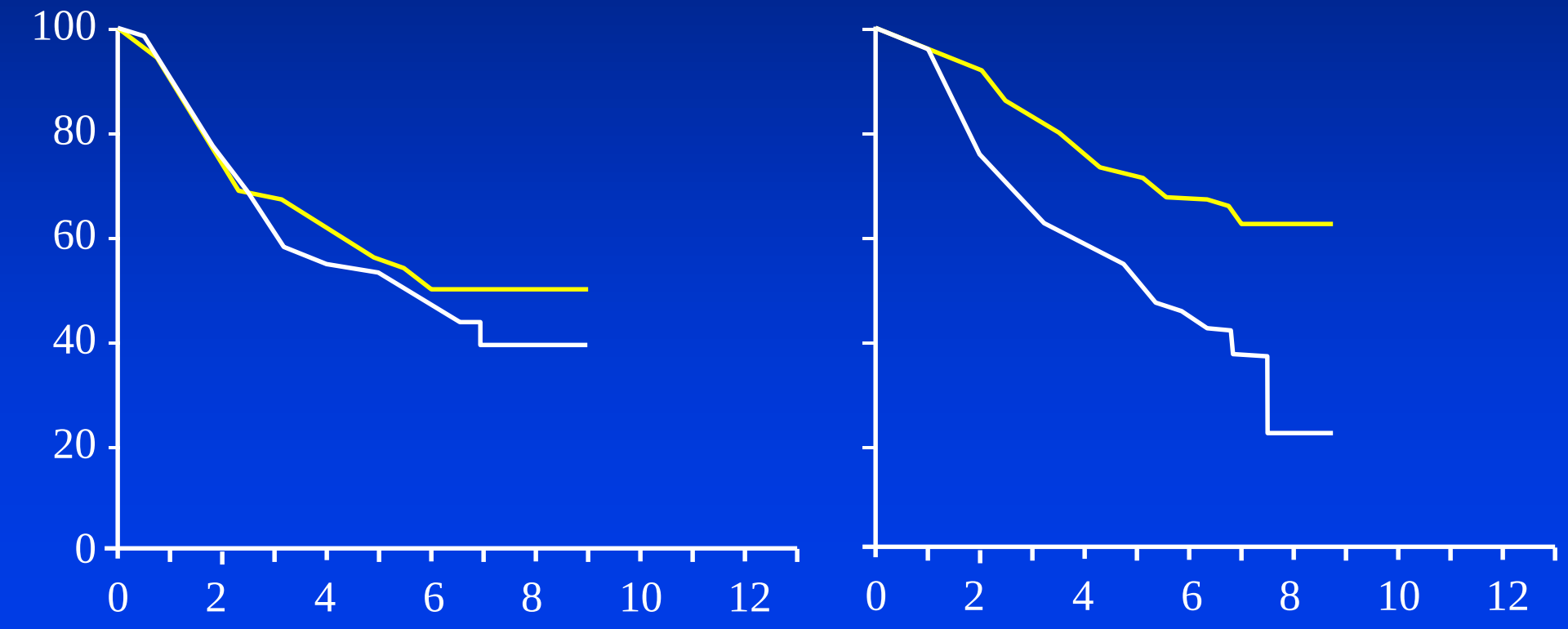
# INTERGROUP STUDY 0035
## Looking at Treatment Effect on Overall Survival

### Females Only

| | At Risk | Death | 5-Yr Estimate |
|---|---|---|---|
| 5-FU+Levamisole | 163 | 74 | 58% |
| Follow-Up Only | 149 | 77 | 54% |

### Males Only

| | At Risk | Death | 5-Yr Estimate |
|---|---|---|---|
| 5-FU+Levamisole | 141 | 47 | 70% |
| Follow-Up Only | 166 | 91 | 51% |



# Duke's C Colon Cancer Adjuvant

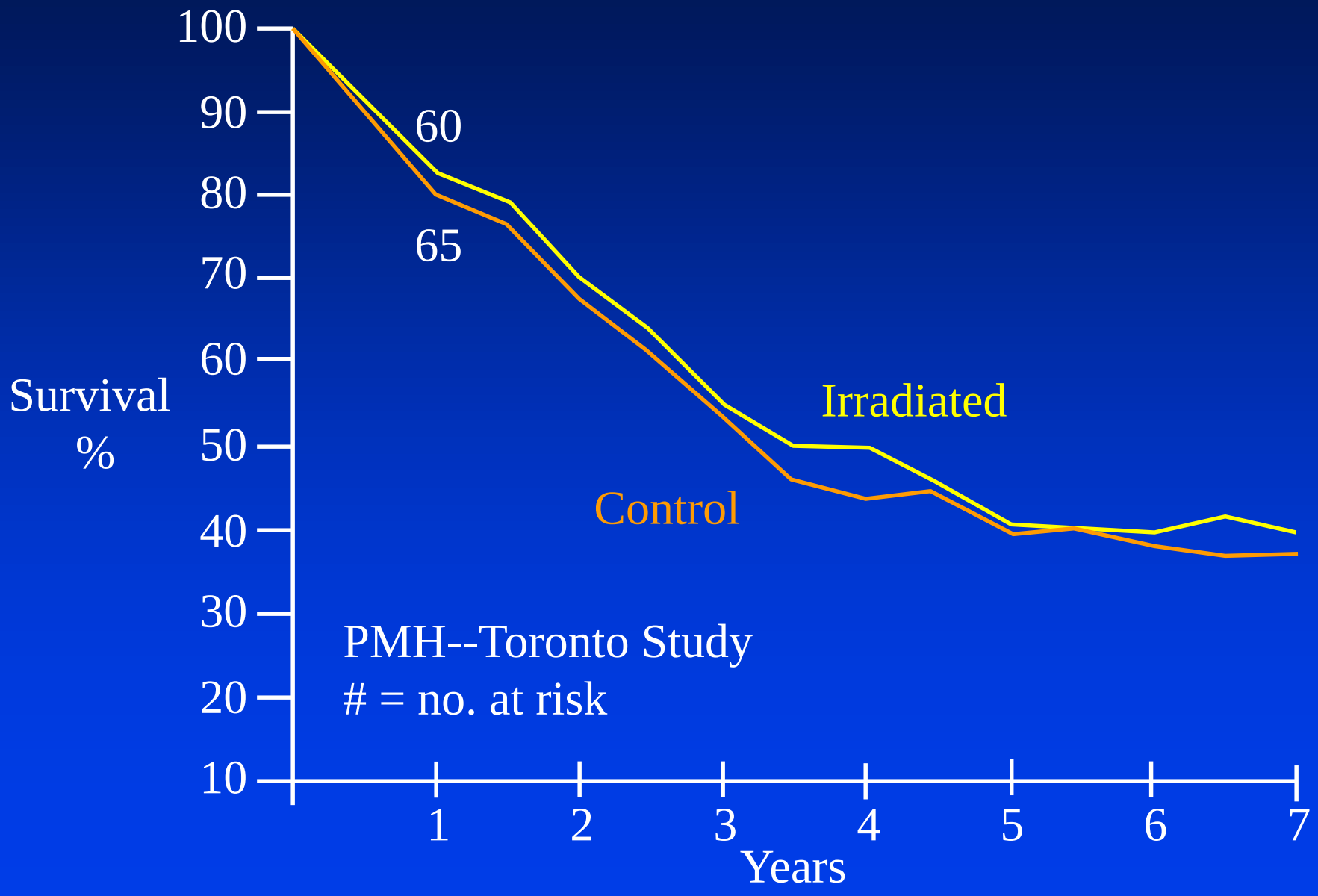
Percent ↓ in Death Rate: 5-FU + Levamisole / Control

| 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%** |

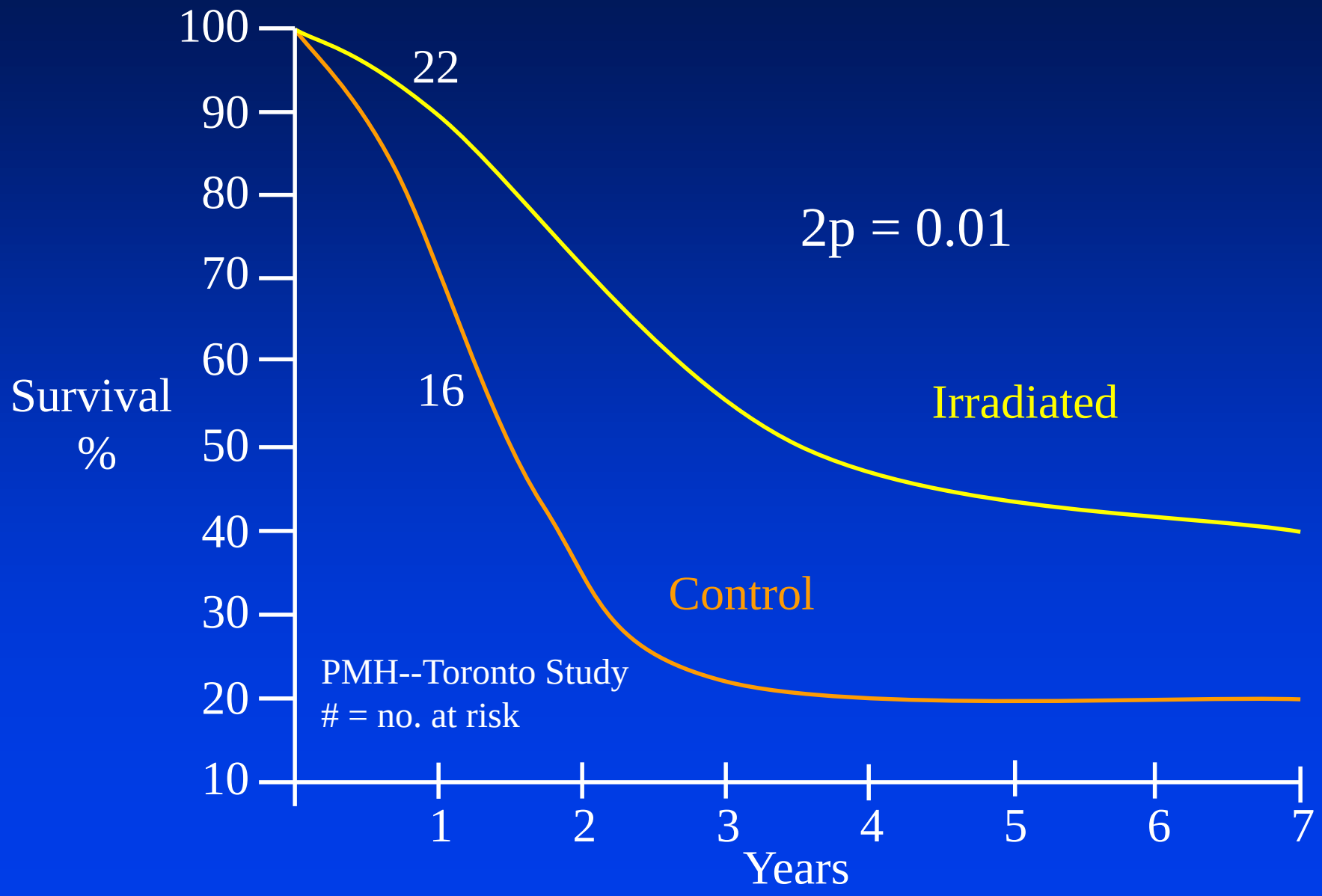
# An Illustration of Exploratory Analyses: Post-hoc Subgroup Analyses

Radiation Treatment in Rectal Cancer
Princess Margaret Hospital

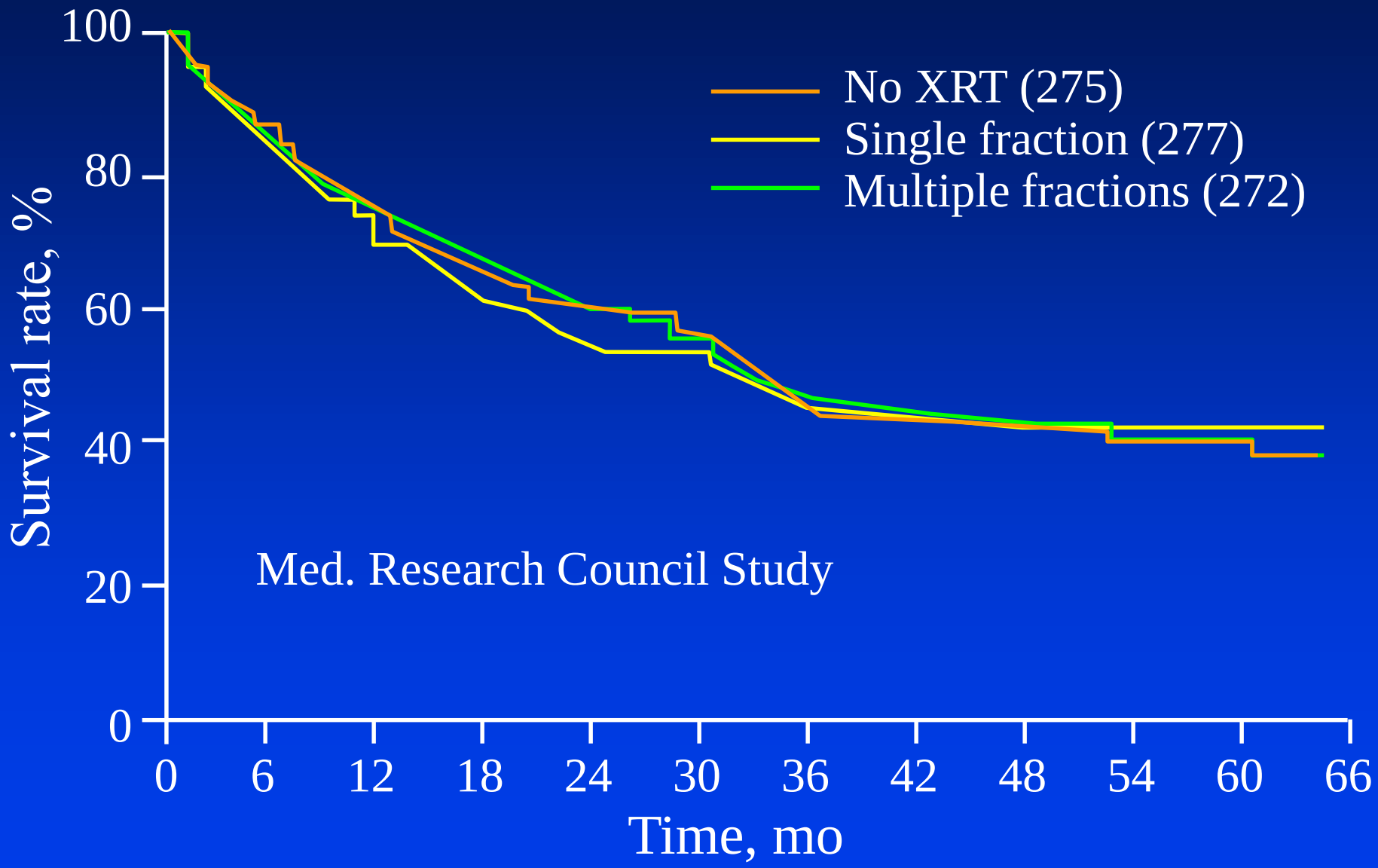
# Survival of Patients with Rectal Carcinoma in Control and Irradiated Groups



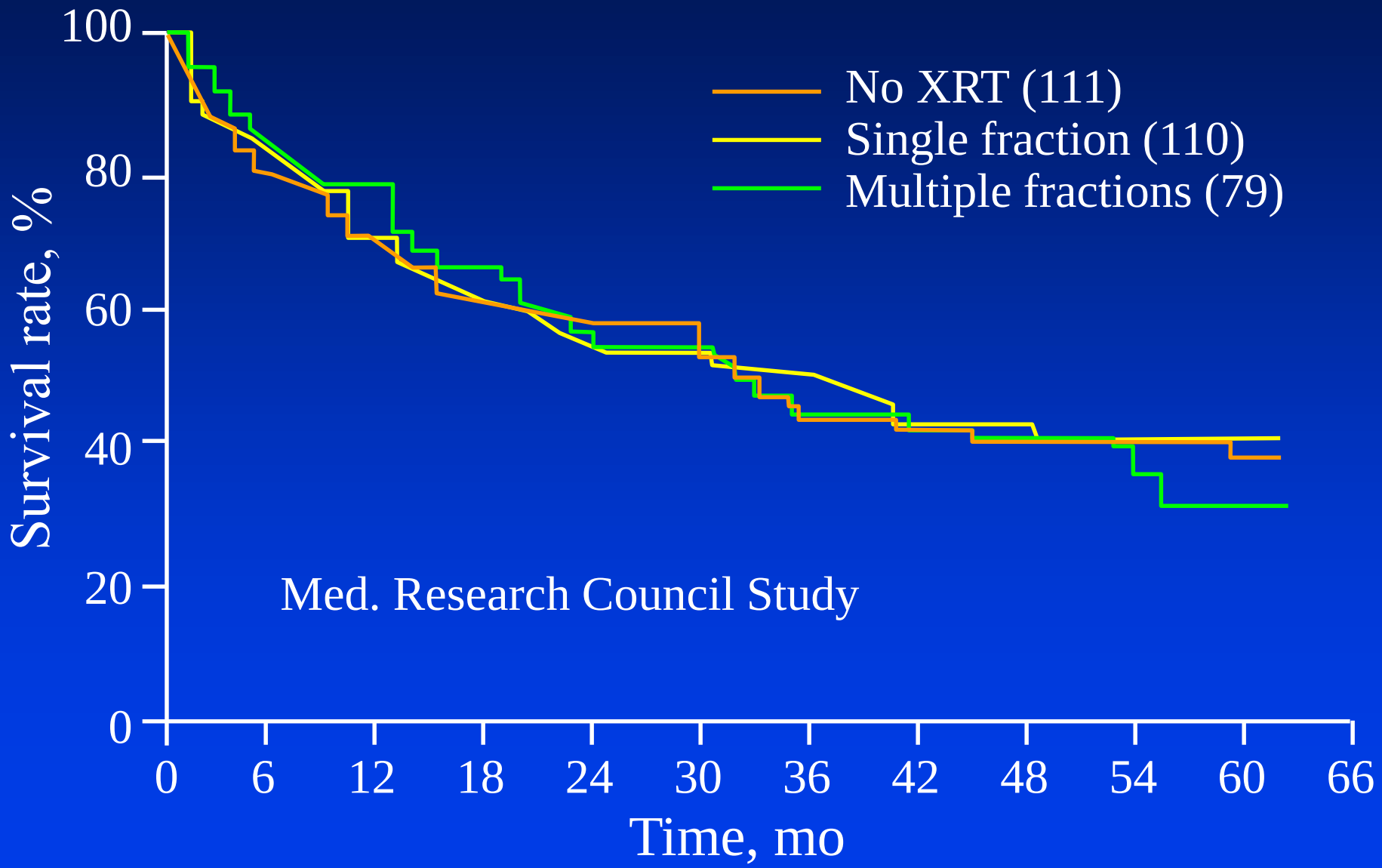
# Survival of Patients with Dukes' Stage C Rectal Carcinoma in Control and Irradiated Groups



# Survival by Treatment Allocated



# Survival by Treatment for Dukes' C Cases



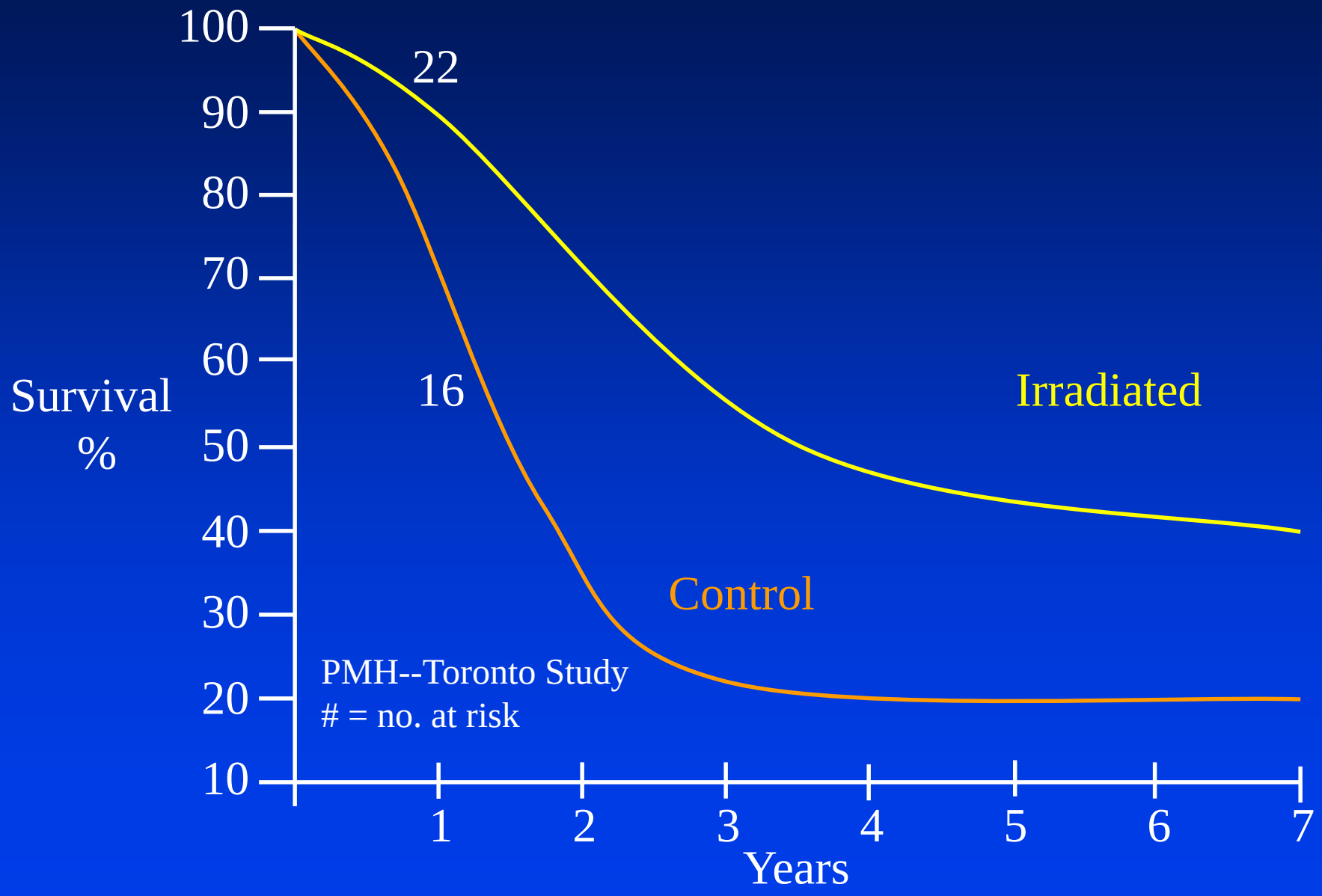
# *Confirmatory* vs. *Exploratory* Analyses

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

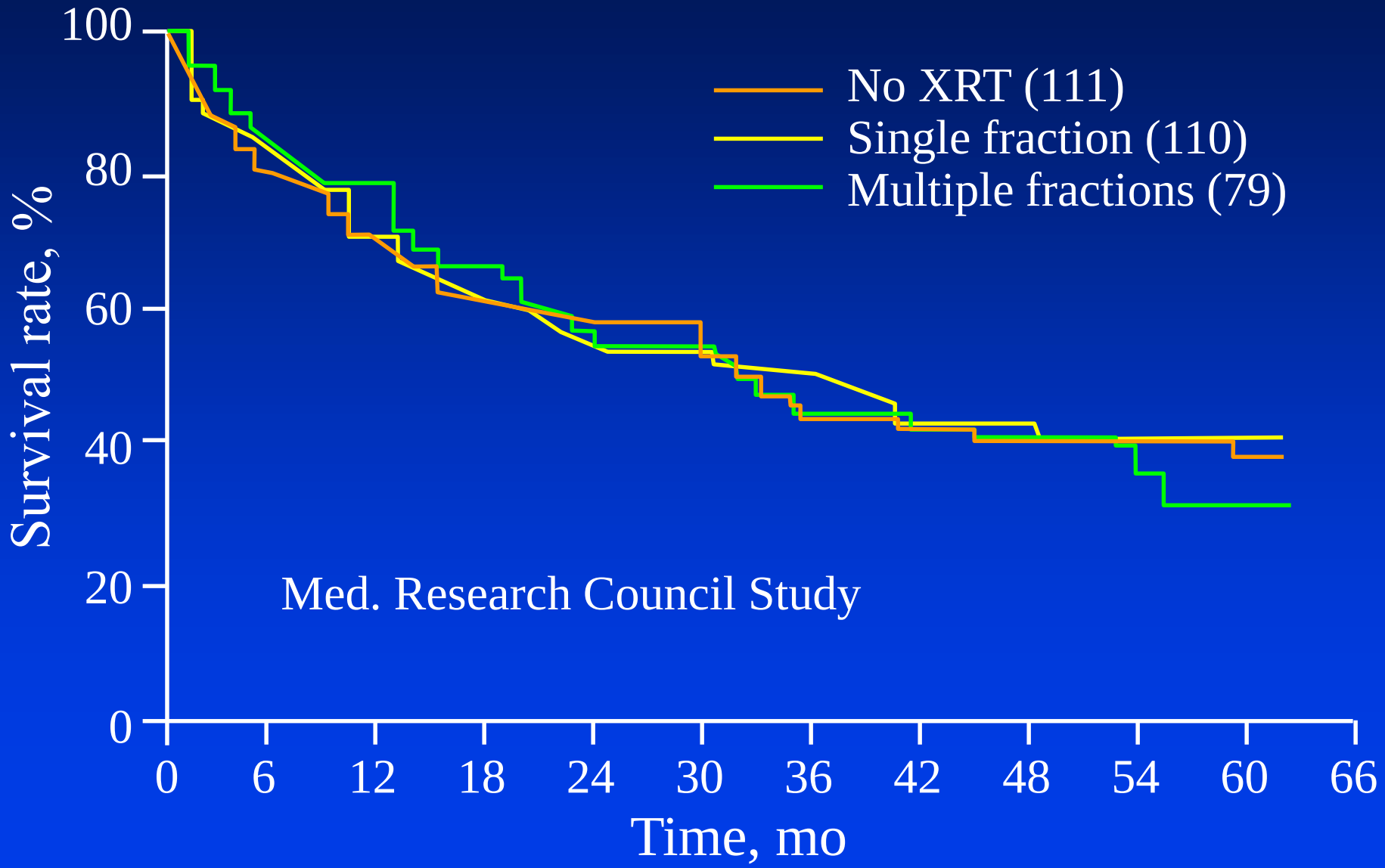
Illustrations and Motivation:

Maternity Wards, ***Baseball*** & Clinical Research

# Survival of Patients with Dukes' Stage C Rectal Carcinoma in Control and Irradiated Groups



# Survival by Treatment for Dukes' C Cases



# Thrombolytics in Acute Myocardial Infarction

- GISSI (Lancet ’86)
  - **SK reduces mortality by 20%**

# 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

# 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**

# Can Efficacy or Safety Signals Discovered in Exploratory Analyses Be Viewed to be Reliable Results?

- Criteria to be simultaneously satisfied:

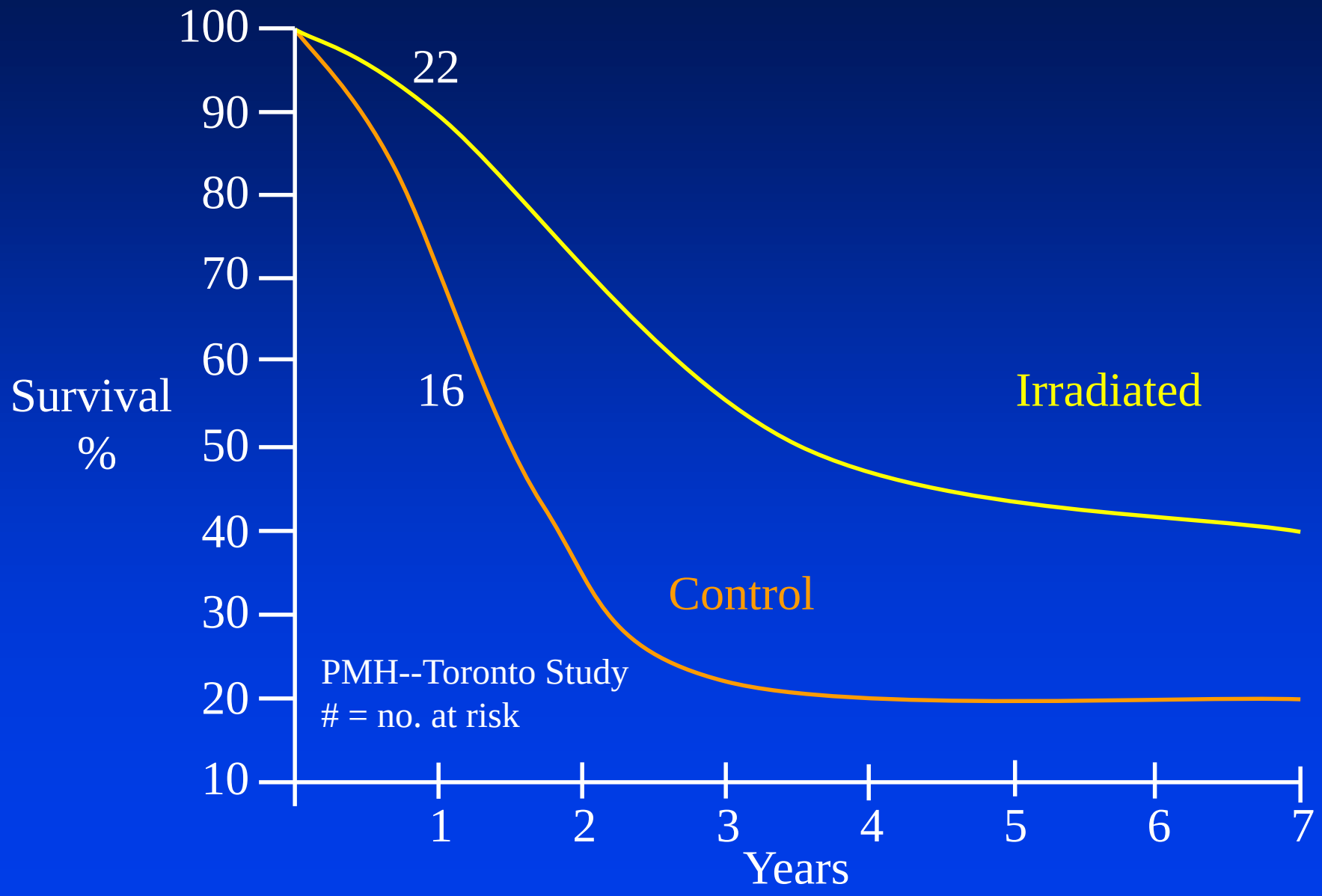
✓ < < P-values *(e.g., Natalizumab & PML & Carvedilol in Heart Failure)*

✓ Biologically plausible effect

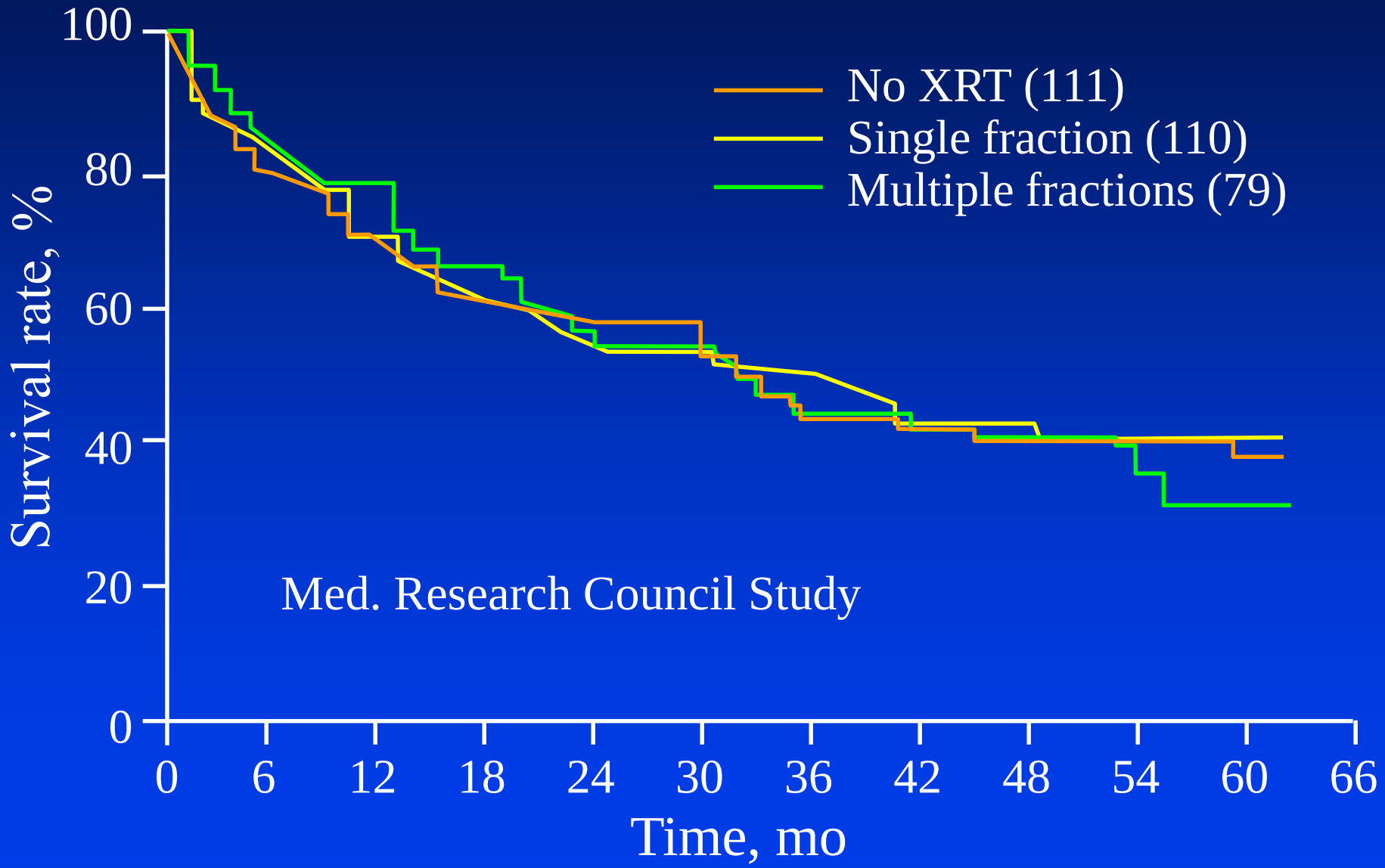
- *White Paper Illustration*

✓ Confirmed by external results

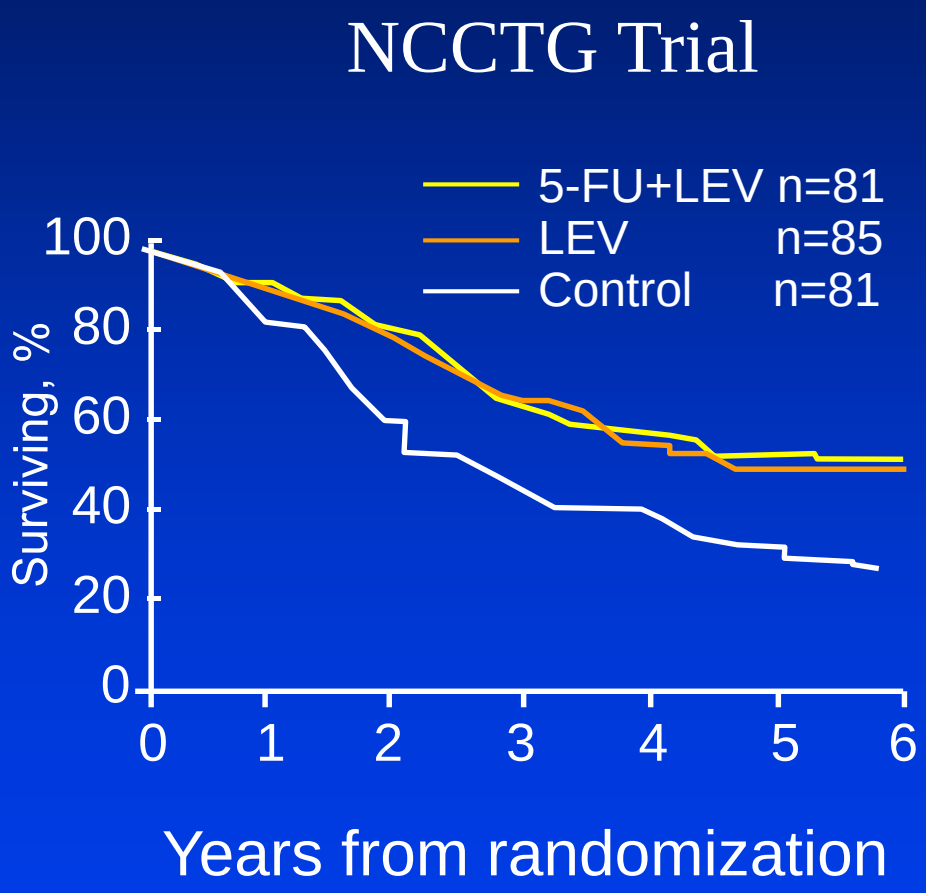
# Survival of Patients with Dukes' Stage C Rectal Carcinoma in Control and Irradiated Groups



# Survival by Treatment for Dukes' C Cases

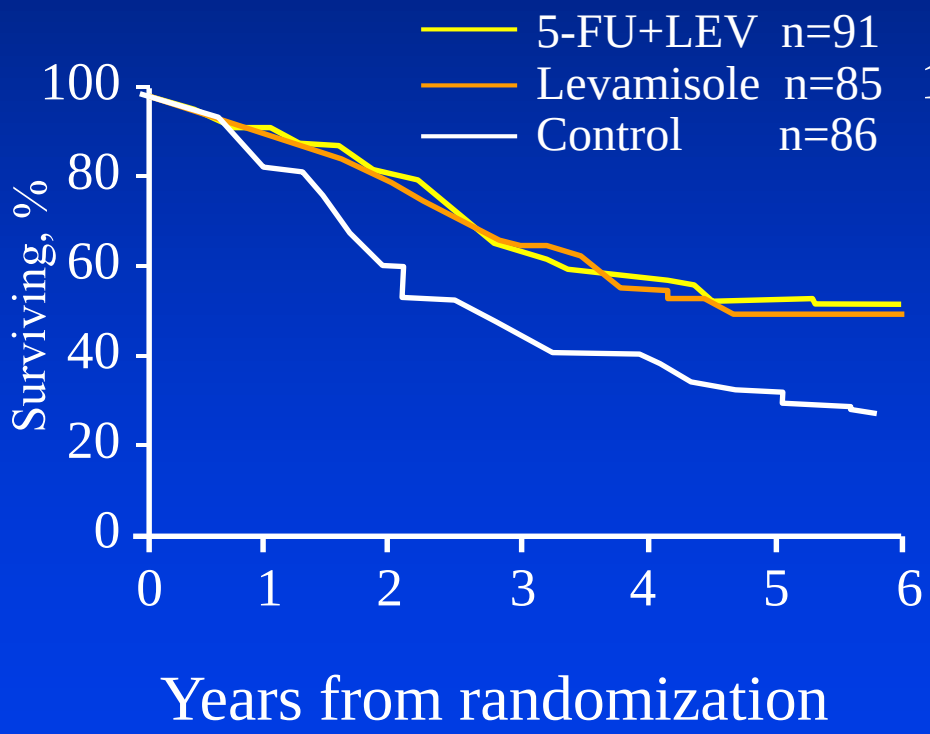


# Surgical Adjuvant Therapy: Colorectal Cancer

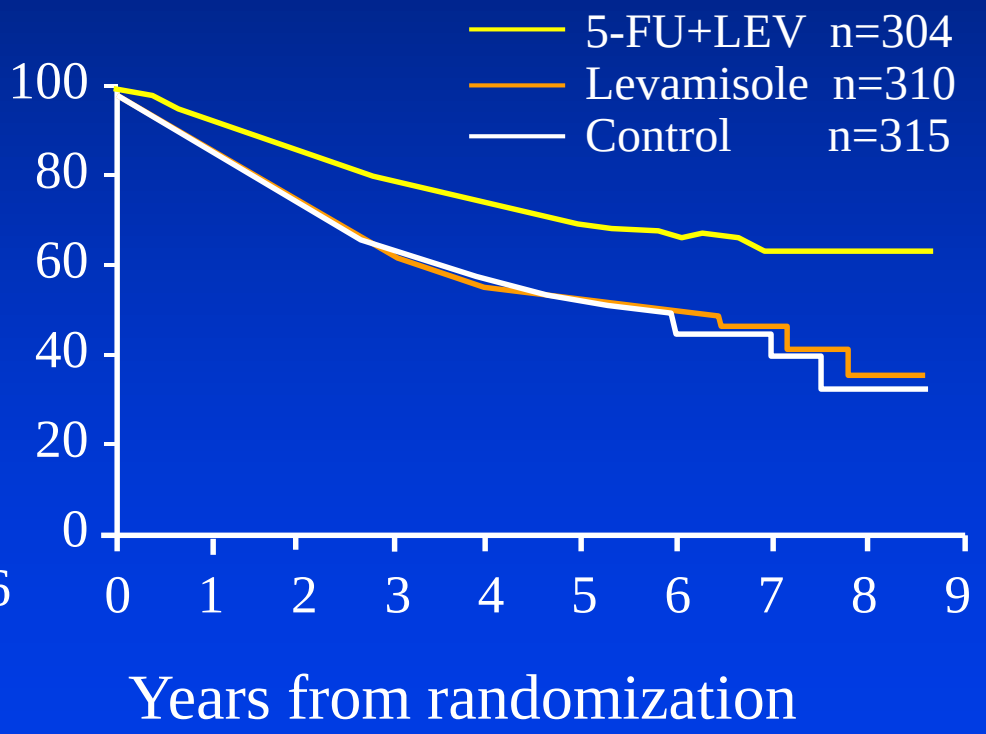


# Surgical Adjuvant Therapy Of Colorectal Cancer

## NCCTG Trial




## Cancer Intergroup Trial



Of all experimental interventions studied in colon adjuvant, suppose only 4% are truly positive & 96% are truly negative.

Suppose the "*false negative error rate*" is $\beta = 0.10$
(so the "*statistical power*" is $1-\beta = 0.90$ )
& Suppose the "*false positive error rate*" is $\alpha = 0.025$

Then, the probability a trial positive will be a true positive is $36 / 60 = 0.60$

| RESULT OF EXPERIMENT | TRUTH Positive | Negative | |
|---|---|---|---|
| Positive | 36 | 24 | 60 |
| Negative | 4 | 936 | 940 |
| | 40 | 960 | 1000 |

Of all experimental interventions studied,
suppose 60% are truly positive & 40% are truly negative

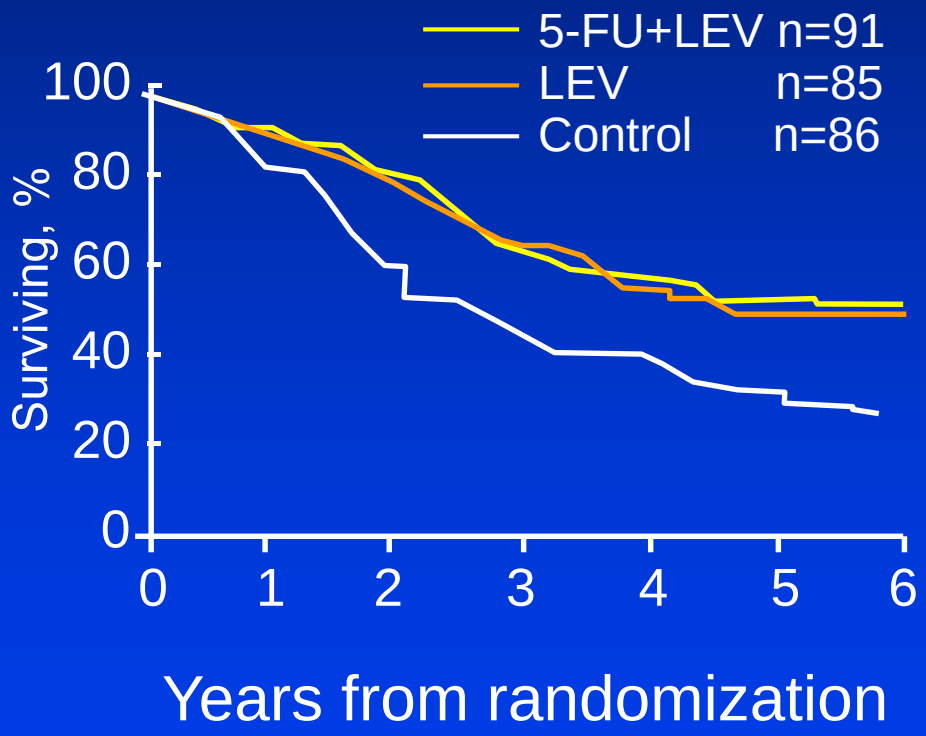
Suppose the "*false negative error rate*" is $\beta = 0.10$
(so the "*statistical power*" is $1-\beta = 0.90$ )
& Suppose the "*false positive error rate*" is $\alpha = 0.025$

Then, the probability a trial positive will be
a true positive is $540 / 550 = 0.98$

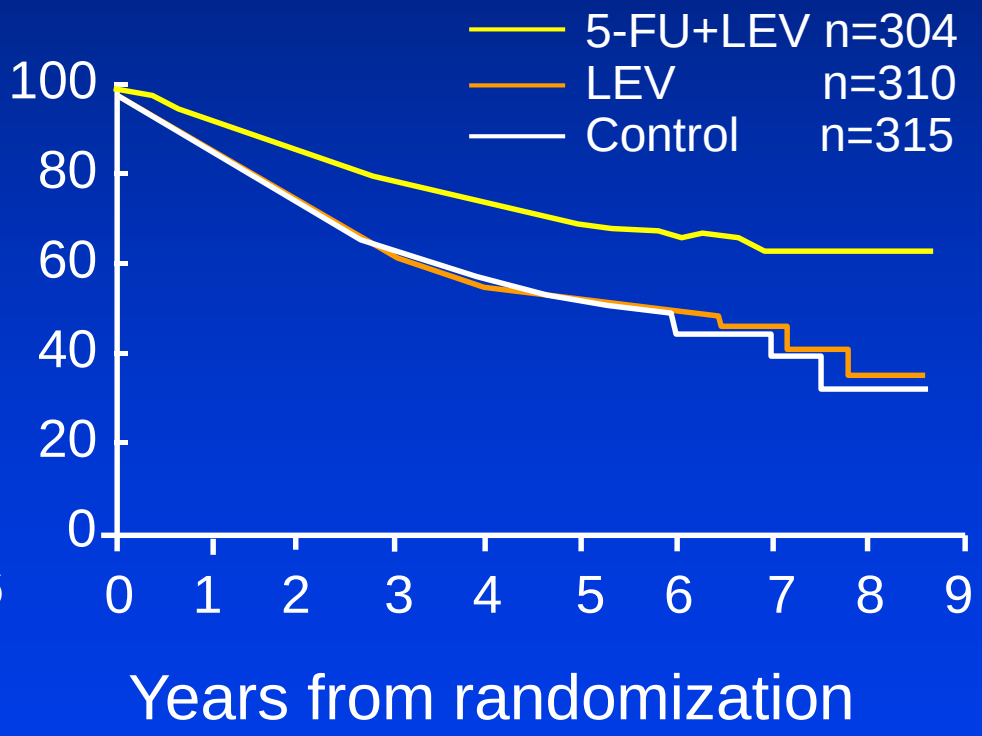
| RESULT OF EXPERIMENT | TRUTH Positive | Negative | |
|---|---|---|---|
| Positive | 540 | 10 | 550 |
| Negative | 60 | 390 | 450 |
| | 600 | 400 | 1000 |

# Surgical Adjuvant Therapy Of Colorectal Cancer

## NCCTG Trial




## Cancer Intergroup Trial



"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)

# Some Conclusions

- 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

# *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:

***Maternity Wards***, Baseball & Clinical Research

20 vs 2: (.71, .99), 2p = 0.0001

Meta-Analysis: 31 vs 13: (.55, .83), 2p = 0.0096

# Bias for "Positive" Results in Clinical Trials

- Protocol Specified Primary Objective of the Clinical trial:

- Very frequent wording:
  - ~ " To *establish* that the experimental regimen is safe and effective"

# Bias for "Positive" Results in Clinical Trials

- Protocol Specified Primary Objective of the Clinical trial:

- 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"

…building a story with supportive analyses…

# Bias for "Positive" Results in Clinical Trials

…Andrew Fleming's insight from Psychology…

"*Cognitive Dissonance*"

…The Harvard Professor's Course…

…The Apparent Lack of Benefit in Males…

# 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

# 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

# Principles & Insights

"The Goal of Clinical Research:

# Principles & Insights

"The Goal of Clinical Research:
To *Determine Whether*,
Not to *Establish*,
the Experimental Regimen
Is Safe and Effective"

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