Biostats for Winners!!!

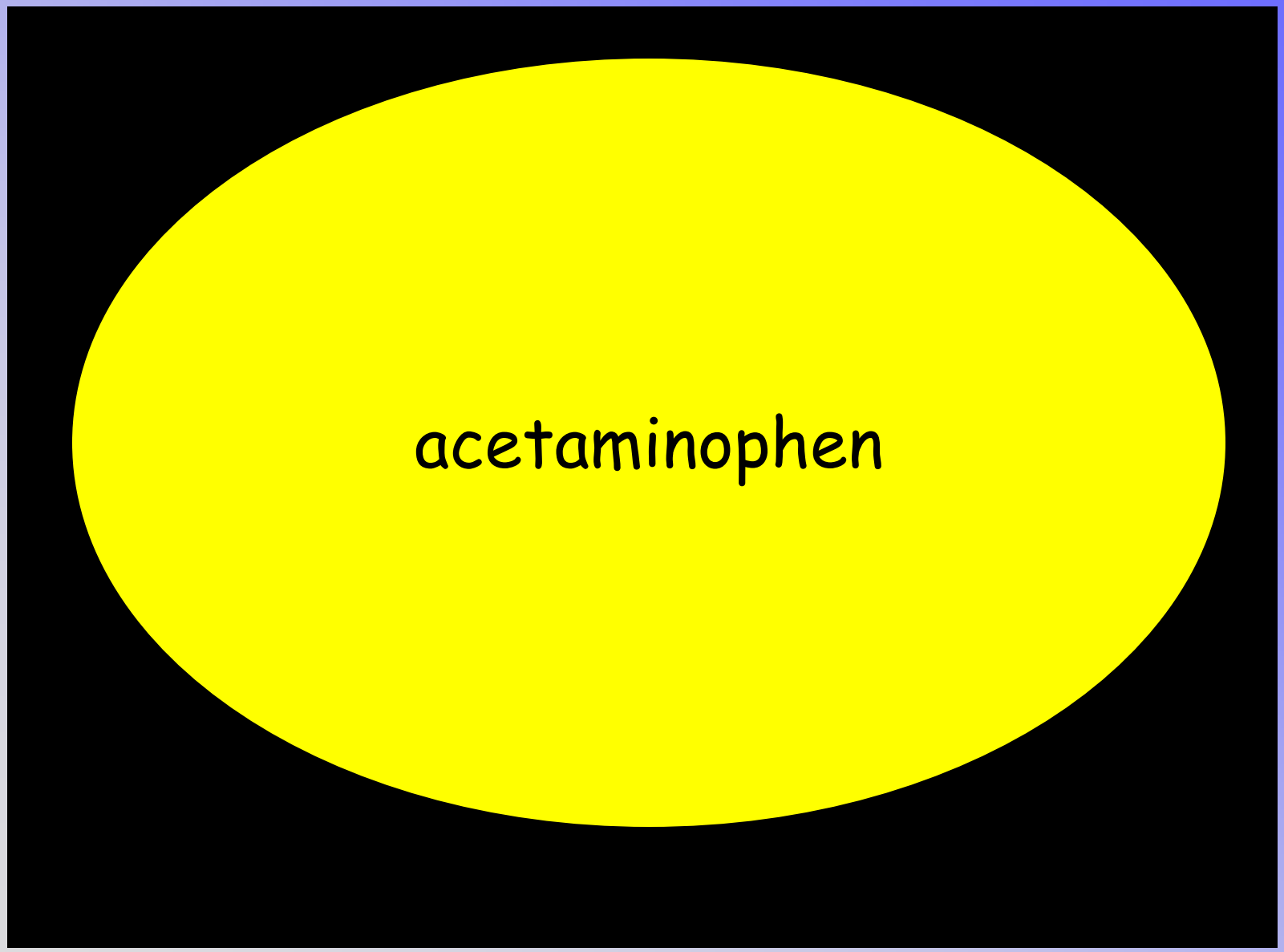
Jeffrey P Schaefer, MSc MD FRCPC

October 19, 2005

- No conflicts of interest to declare.

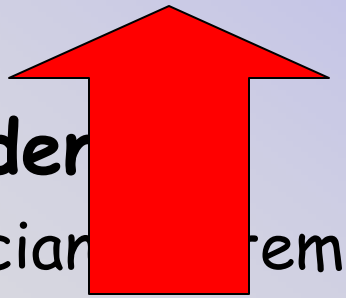
Objectives

- Know where Biostatistics fit into Evidence Based clinical practice.
- Define measures of treatment effect.
- Distinguish statistical from clinical significance.
- Common pitfalls.


acetaminophen

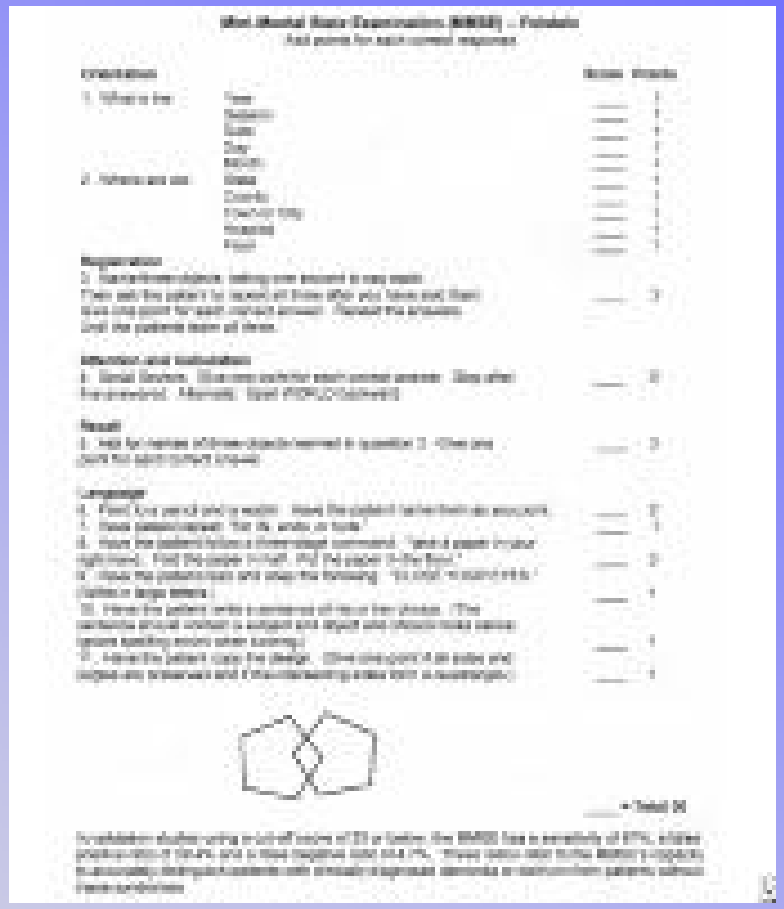
Evidence Based Medicine

- **Create a Clinical Question**
  - harm, diagnosis, therapy, prognosis, prevention
- **Find Answers**
  - unfiltered sources -> Medline
  - filtered sources -> Uptodate
- **Critically Appraise the Answers**
  - assess Validity, Results, Applicability
- **Apply the Evider**
  - patient, physician [illegible]em values



Clinical Scenario

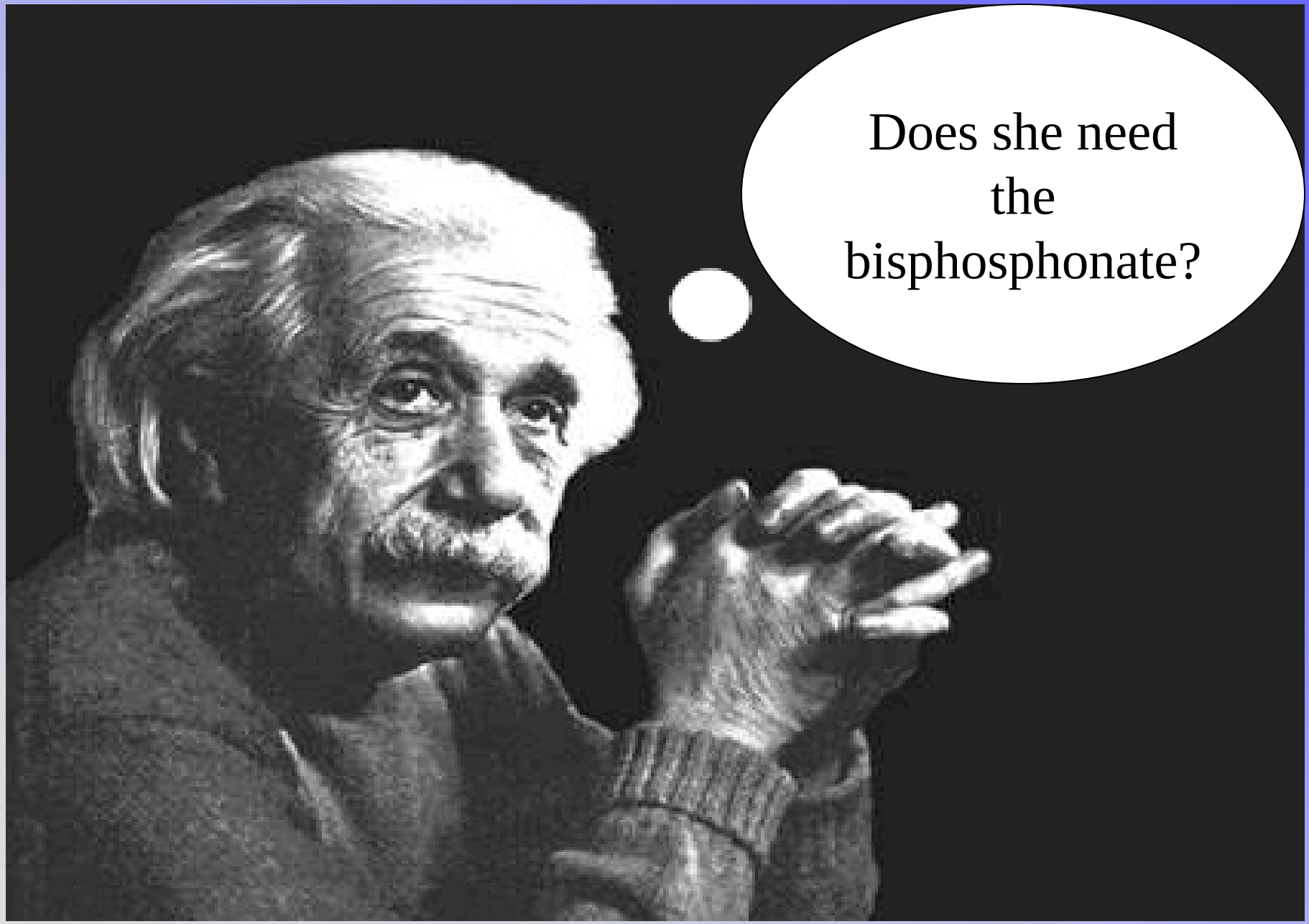
Margaret



Medication Review

- **Patient Profile**
  - 80 yr old female
  - Alzheimer's disease
  - hypertension
  - **osteoporosis**
  - osteoarthritis
- **Medications**

calcium 1,000 mg / day (elemental)	2 tab
vitamin D 1,000 IU / day	1 tab
risedronate 2.5 mg / day	**1 tab**
HCTZ 12.5 mg / day	1 tab
amolidpine 5 mg / day	1 tab
benazepril 10 mg / day	1 tab
ASA 81 mg / day	1 tab
colace 100 mg bid	2 tab
senokot 1 / day	1 tab
acetaminophen SR 650 mg tid	3 tab
multivitamin 1 / day	1 tab
resperidone 0.25 mg qhs	1 tab
	16 tabs/day


Does she need
the
bisphosphonate?

Clear Questions—Clear Answers

Compared to placebo, what is the effect of risedronate (Actonel) on fracture risk among elderly females with Alzheimer's Disease?

Patient

Outcome

Intervention & Comparator

ORIGINAL INVESTIGATION

The Prevention of Hip Fracture With Risedronate and Ergocalciferol Plus Calcium Supplementation in Elderly Women With Alzheimer Disease

A Randomized Controlled Trial

Yoshihiro Sato, MD; Tomohiro Kanoko, PhD; Kei Satoh, MD; Jun Iwamoto, MD

Background: A high incidence of fractures, particularly of the hip, represents an important problem in patients with Alzheimer disease (AD), who are prone to falls and have osteoporosis. We previously found that deficiency of 25-hydroxyvitamin D and compensatory hyperparathyroidism cause reduced bone mineral density in female patients with AD. We address the possibility that treatment with risedronate sodium and ergocalciferol plus calcium supplementation may reduce the incidence of nonvertebral fractures in elderly women with AD.

Methods: A total of 500 elderly women with AD were randomly assigned to daily treatment with 2.5 mg of risedronate sodium or a placebo, combined with 1000 IU of ergocalciferol and 1200 mg of elementary calcium, and followed up for 18 months.

Results: At baseline, patients of both groups showed 25-hydroxyvitamin D deficiency with compensatory hyperparathyroidism. During the study period, bone mineral density in the risedronate group increased by 4.1% and decreased by 0.9% in the control group. Vertebral fractures occurred in 29 patients (24 hip fractures) in the control group and 8 patients (5 hip fractures) in the risedronate group. The relative risk in the risedronate group compared with the control group was 0.28 (95% confidence interval, 0.13-0.59).

Conclusions: Elderly patients with AD hypovitaminosis D are at increased risk for hip fracture. Treatment with risedronate and ergocalciferol may be safe and effective in reducing the risk of a fracture in elderly patients with AD.

Arch Intern Med. 2005;165:1737-1742

Background

Among Elderly Patients

Fractures

7 : 1

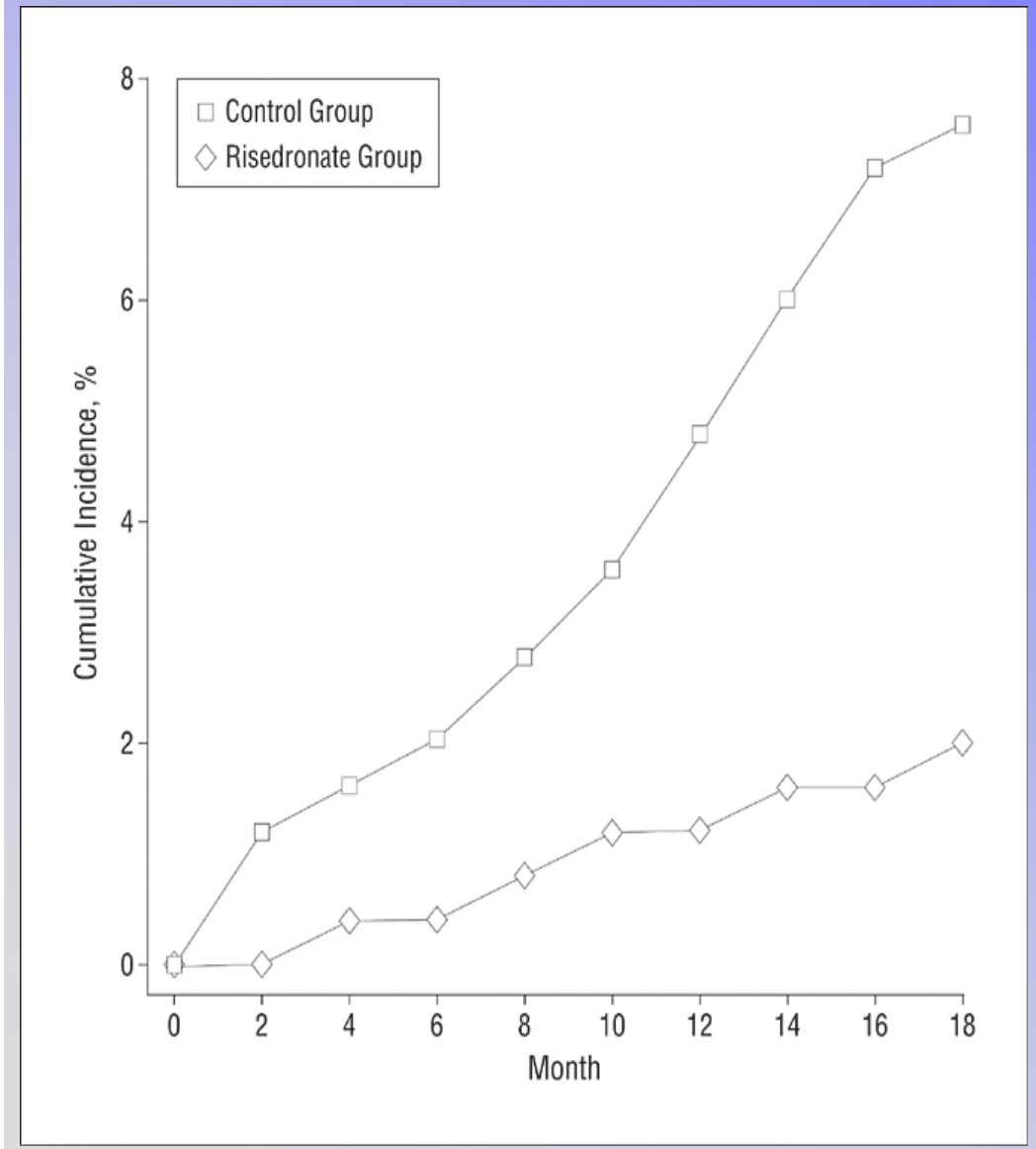
Alzheimer's Dx Controls

Objective

- What is the effect of
  - bisphosphonate (risedronate) on
  - frequency of fractures among
  - elderly females with AD?

Methods

- **Entry**
  - > 70 yr female with Alzheimer's Disease
- **Intervention (random allocation)**
  - 250 allocated to Risedronate 2.5 mg
  - 250 allocated to placebo
  - all get Calcium and Vit D
- **Observation**
  - 18 months
  - fractures, bone density, adverse fx



Incidence of All Fractures

control = 29/250

treatment = 8/250

log rank test = p <0.01

RR = 0.28

95%CI (0.13-0.59)

ARR= 0.084

NNT ~ 12

How good is this article?

CENTRE FOR HEALTH EVIDENCE

Home Site Map About Us Contact

You are here: CHE Home » Users' Guides » Therapy

- Home
- CHE News
- About Us
  - A Brief History
  - What We Do & Why
  - Where We're Going
- Users' Guides to EBP
- Projects
  - ABACUS
  - CHEDESK
  - EXTRA/FORCES
  - Nemours
  - NPLT
  - SEARCH
  - SGIM-EBM
  - TORCH
  - UGI
- Opportunities
- Communications
  - Annual Progress Report
  - Articles
  - Brochure
- Site Map
- Contact Us

How to Use an Article About Therapy or Prevention

Gordon H. Guyatt, David Sackett, Deborah J. Cook, for the Evidence Based Medicine Working Group

Based on the Users' Guides to Evidence-based Medicine and reproduced with permission from JAMA. (1993;270(21):2598-2601) and (1994;271(1):59-63). Copyright 1995, American Medical Association.

- Clinical Scenario
- The Search
- Introduction
- I. Are the results of this article valid
- II. What were the results
- III. Will the results help me in caring for my patients
- Resolution of the Scenario
- Conclusion
- References

Critical Appraisal
Therapy / Prevention

- **Validity**
  - random allocation?
  - complete follow-up?
  - intention to treat analysis?
  - were subjects and outcome assessment blinded?
  - were groups similar at baseline?
- **Results**
  - magnitude of effect?
  - how precise was the estimate of treatment fx?
- **Applicability**
  - apply to my patient?
  - all important outcomes assessed?
  - risks worth benefits and costs?

Article

- Validity?
  - random allocation? ☑
  - complete follow-up? ☑
  - intention to treat analysis? ☑
  - were subjects and outcome assessment blinded? ☑
  - were groups similar at baseline? ☑
- Results?
  - magnitude of effect? ?
  - how precise was the estimate?
- Applicability?
  - apply to my patient? ☑
  - all important outcomes assessed? ☑
  - risk / benefit assessment? ?

Primary Outcome
Non-vertebral fractures

	raw	probability	%
Treatment	8/250	0.032	3.2
Control	29/250	0.116	11.6

How do the groups compare?

Measures of Treatment Effect

1. Relative Risk Reduction
2. Relative Risk
3. Odd's Ratio
4. Absolute Risk Reduction
5. Number Needed to Treat

Relative Risk Reduction

Pr = probability of outcome

$$\frac{Pr_{control} - Pr_{treatment}}{Pr_{control}}$$

(0.116-0.032) / 0.116 = 0.724

Relative Risk Reduction = 72%

"Compared to controls, the treatment group had a 72% reduction in fracture risk."

Relative Risk

$$\frac{Pr_{treatment}}{Pr_{control}}$$

0.032 / 0.116 = 0.276

Relative Risk = 28%

"Compared to controls, the treatment group was 28% as likely to sustain fractures."

Relative Risk and Relative Risk Reduction

RR + RRR = 1.0 (100%)

0.72 + 0.28 = 1.0

72% + 28% = 100%

Odd's Ratio

"Ratio of the Odds"

$$\frac{\text{Odds of Fracture among Treatment}}{\text{Odds of Fracture among Control}}$$

What are Odds and Probabilities?

- Probability
  - how likely it is that some event will occur
  - heads in coin toss = $\frac{1}{2}$ = 0.5
  - '5' on single die roll = 1/6 = 0.17

- Odds
  - probability of event / (1 – probability of event)
  - heads in coin toss = 0.5 / (1-0.5) = 1
  - '5' on single die roll = 0.17 / (1-0.17) = 0.20

Odds of 1 = 1:1

Odds of 0.2 = 1:5

Odds Ratio

$$\frac{\text{Odds treatment}}{\text{Odds control}}$$

$$\frac{0.032 / (1-0.032)}{0.116 / (1-0.116)}$$

0.251

"Odds of fracture was 0.25 among the treatment group when compared to controls."

Odds Ratios and Relative Risks

- 'What if' treatment outcomes differed?

Tx	Cntl	RR	OR
0.001%	11.6%	0.0001	0.000076
1.0%	11.6%	0.09	0.08
3.2%	**11.6%**	**0.28**	**0.25**
11.6%	11.6%	1.00	1.00
50.0%	11.6%	4.31	7.62
75.0%	11.6%	6.47	22.86
99.0%	11.6%	8.53	754.45

Absolute Risk Reduction

$$Pr_{control} - Pr_{treatment}$$

0.116 - 0.032 = 0.084

'Compared to controls, there was an 8.4% reduction in fracture risk among those in the treatment group."

Number Needed to Treat

$$\frac{1}{ARR}$$

1 / (0.116-0.032)

11.9

"The number of elderly AD patients needed to treat with risedronate for 18 month to save one fracture is 12."

Measures of Treatment Effect

- RRR 72% or 0.72
- RR 28% or 0.28
- OR 0.25
- ARR 8% or 0.08
- NNT 12

- What numerical answer works best for you?.... for others?

Is there a best measure?

- Number Needed to Treat
  - hailed by EBM experts as best
  - reasonably intuitive
  - least likely to magnify effect
  - can associate with a confidence interval

Measures of Effect

Therapy	Endpoint	NNT (5yr)
stepped care for diastolic BP of 115 - 129	death, stroke, myocardial infarction	3
ASA for transient ischemic attack	death, stroke	6
stepped care for diastolic BP 90 - 109	death, stroke, myocardial infarction	141

What do patients understand?

- 350 patients
- Given a disease scenario
- Randomized to four treatment effect scenarios.

Relative Risk Reduction

'Treatment A reduces the chance that you will develop Disease Y by 25%.'

'Treatment B reduces the chance that you will develop Disease Y by 10%.'

Absolute Risk Reduction

'Treatment A reduces the chance that you will develop Disease Y by 10 per 1000 persons.'

'Treatment B reduces the chance that you will develop Disease Y by 4 per 1000 persons.'

Number Needed to Treat

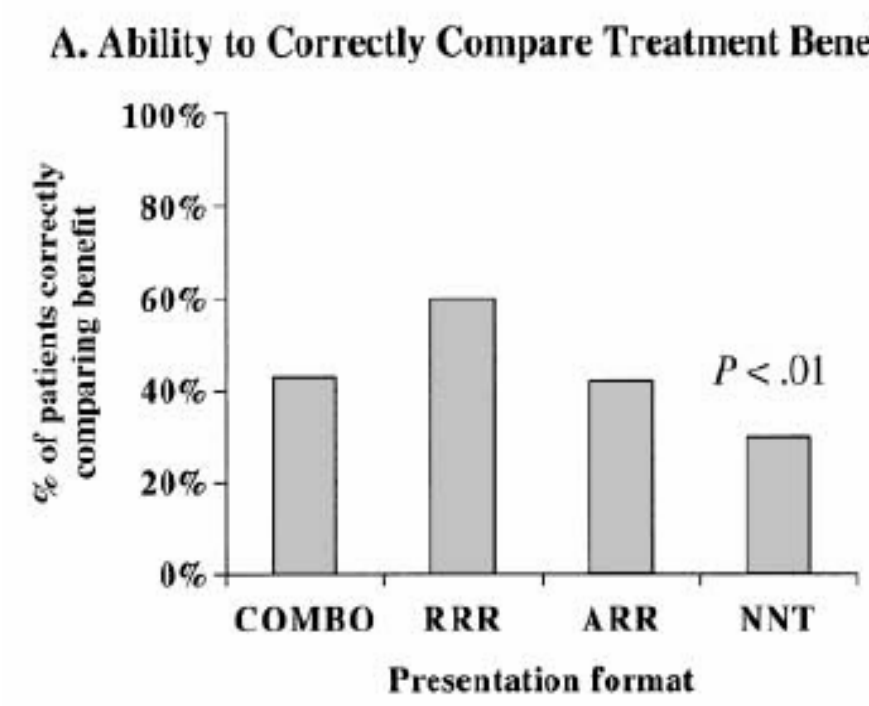
'100 persons just like you would have to be treated with Treatment A for 5 years for a benefit against Disease Y to be evident in one of you.'

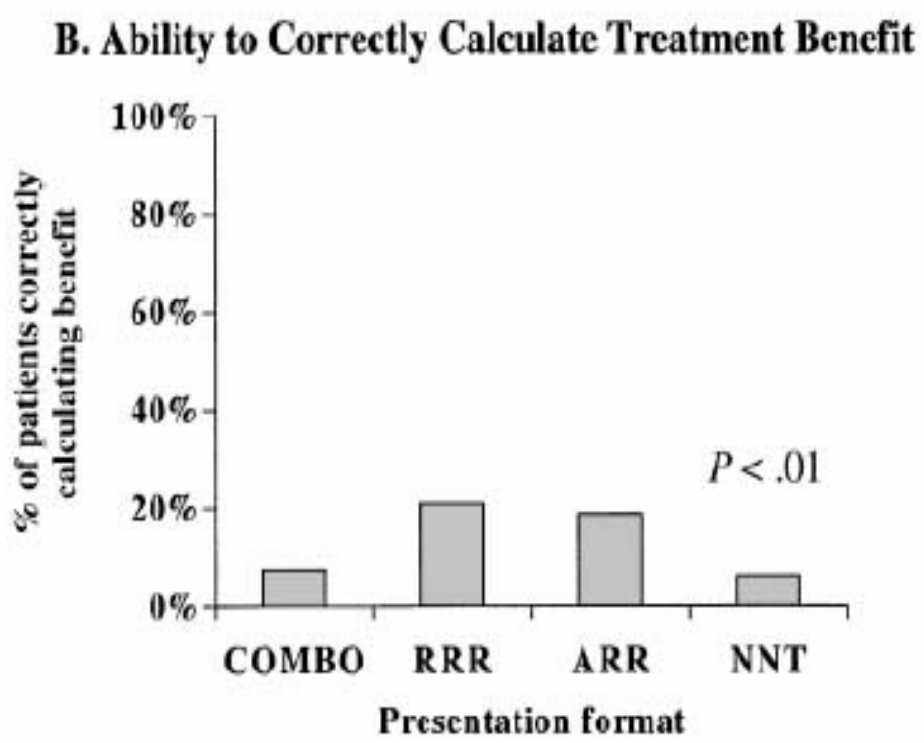
'250 persons just like you would have to be treated with Treatment B for 5 years for a benefit to be evident in one of you.'

Combination

'Treatment A reduces the chance that you will develop Disease Y by 25% or 10 per 1000 persons. This means that 100 persons just like you would have to be treated….'

'Treatment B reduces the chance that you'll develop Disease Y by 10% or 4 per 1000 persons. This means that 250 persons just like you would have to be treated….'


A. Ability to Correctly Compare Treatment Benefits
% of patients correctly comparing benefit
100%
80%
60%
40%
20%
0%
P < .01
COMBO
RRR
ARR
NNT
Presentation format


B. Ability to Correctly Calculate Treatment Benefit
% of patients correctly calculating benefit
100%
80%
60%
40%
20%
0%
P < .01
COMBO
RRR
ARR
NNT
Presentation format

Statistical Significance (p-values)

Common Sense

- Jack and Jill are competing.
- They take turns flipping a fair coin.
- Each HEADS pays $10

- What is likely to happen?

Jack	$10 / 2 tosses
Jill	$10 / 2 tosses

Statistics Language

- P(Jack $) - P(Jill $) = 0

- **Null Hypothesis**
  - no difference between Jack and Jill

Statistics Language

- After the trial EITHER

we REJECT the Null Hypothesis

or

we DO NOT REJECT the Null Hypothesis

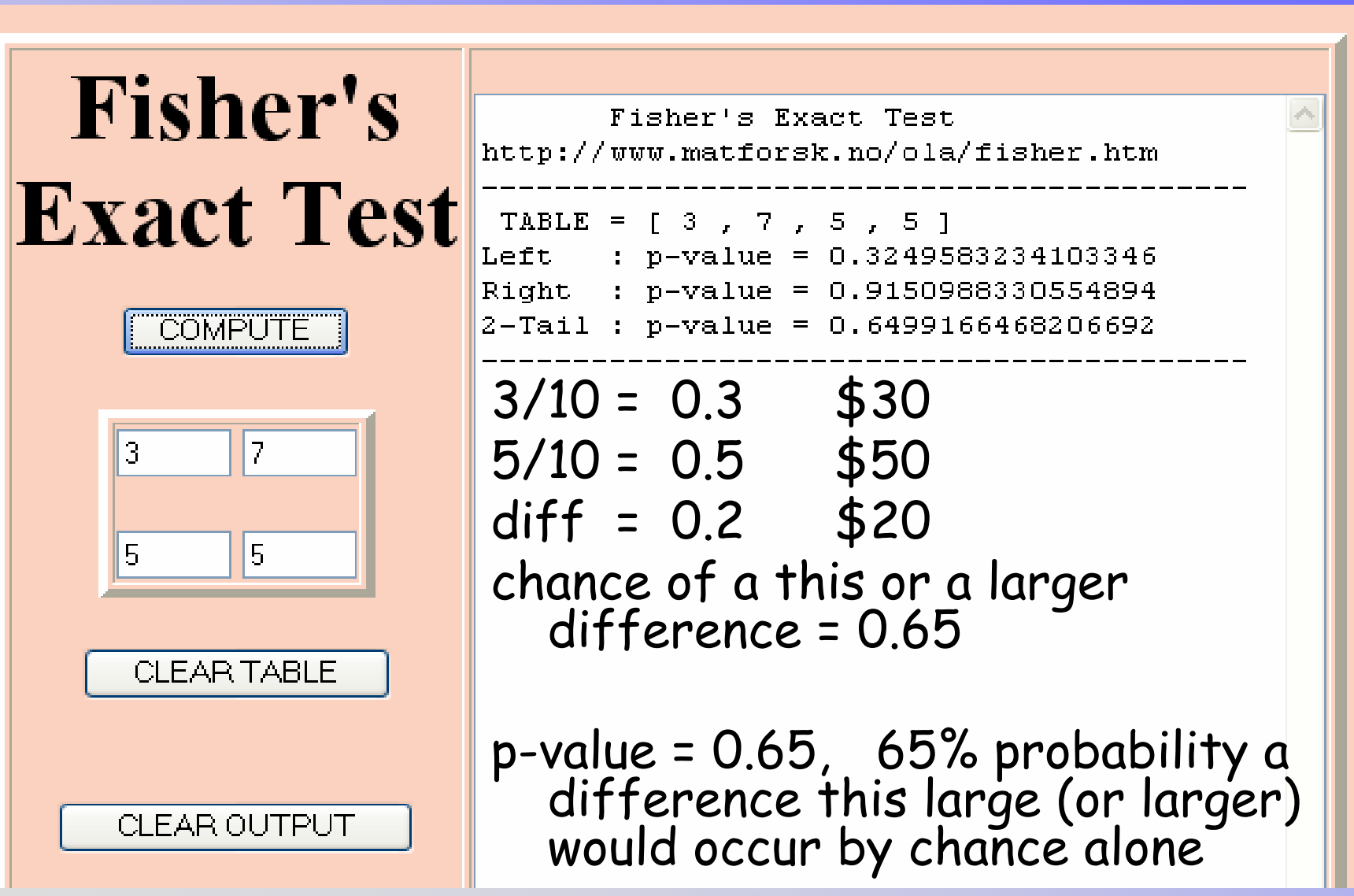
Play x 10 flips each...

- JACK
  - 10 flips -> 3 heads -> $30
- Jill
  - 10 flips -> 5 heads -> $50

- Is this enough evidence to conclude that Jack is a bad coin tosser (reject the null)?

No, this was ‘bad luck’.

How bad was the luck?



Play x 100 flips each...

- Jack flips a coin 100 times -> 30 heads
- Jill flips same coin 100 times -> 50 heads

- Do you reject the Null Hypothesis?

Sufficient Evidence to Reject the Null

Fisher's Exact Test

COMPUTE

30	70
50	50

CLEAR TABLE

```
        Fisher's Exact Test
http://www.matforsk.no/ola/fisher.htm
----------------------------------------
 TABLE = [ 30 , 70 , 50 , 50 ]
Left    : p-value = 0.002968670583006797
Right   : p-value = 0.9988303054752575
2-Tail  : p-value = 0.005937341166013594
----------------------------------------
```

3/10 = 0.3 $300

5/10 = 0.5 $500

diff = 0.2 $200

chance of a this or a larger difference = 0.0059

p-value = 0.0059, 0.6% probability a difference this large (or larger) would occur by chance alone

Test is significant -> this is unlikely to be bad luck alone!

Significance

- **Statistical significance**
  - when there is less than a 5% chance that the difference was due to luck alone
  - $p < 0.05$
  - says nothing about validity
  - says nothing about clinical benefit
- **Clinical significance**
  - requires statistical significance
  - is a judgment based on patient, physician, and system values

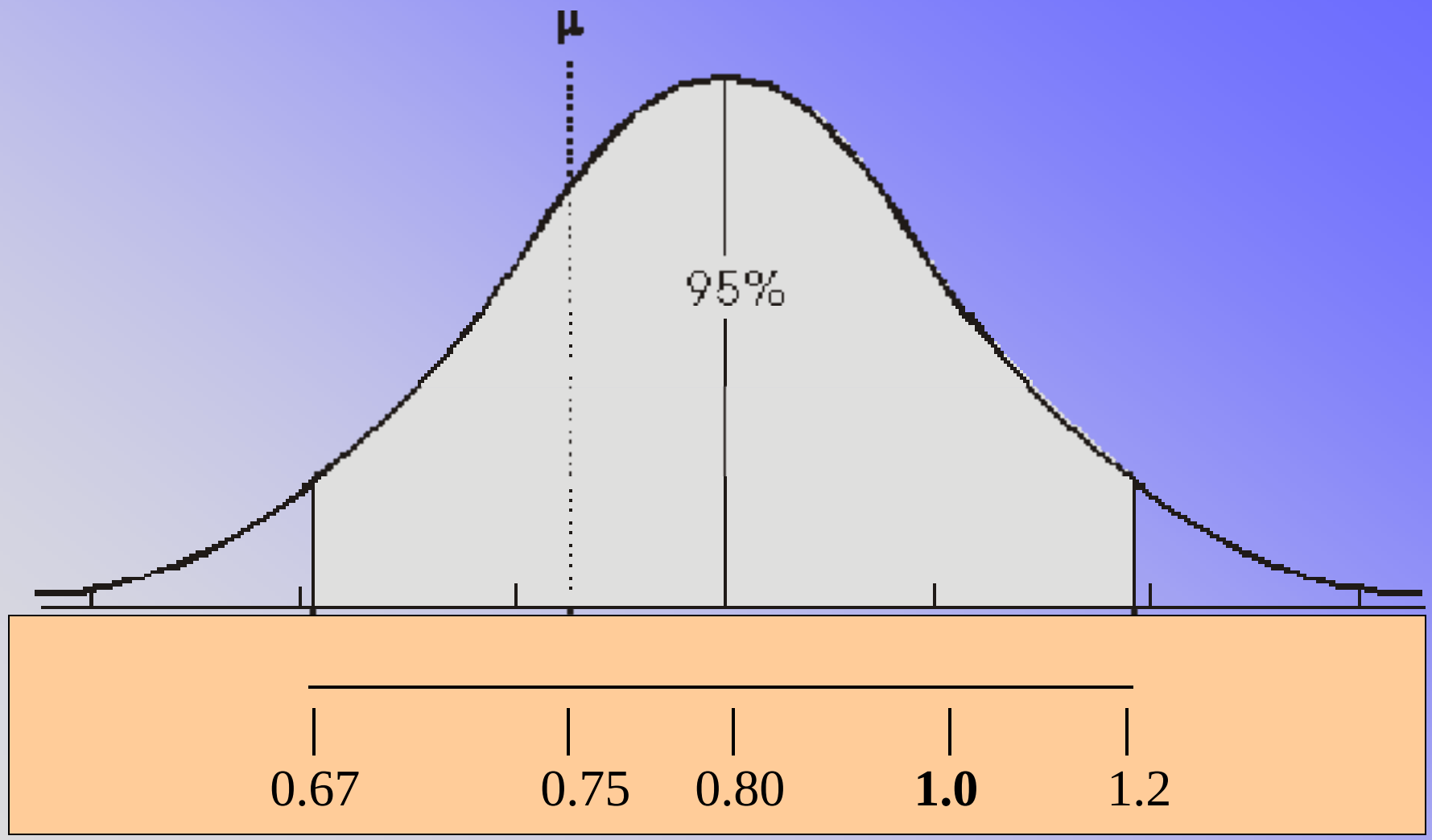
Confidence Intervals

The relative risk in the risedronate group compared to the control group was 0.28 (95% confidence interval, 0.13 to 0.59).

What's a Confidence Interval?

- Confidence Interval (95%)
  - measure of precision
  - is the range over which the TRUE population mean falls 95% of the time
  - If we repeated the risedronate study many times, we would expect the true Relative Risk to fall within the 95% CI, 95% of the time.

RR of 1 is the same as No Effect



What's the right test?

- Test selection depends on the study.
  - how many groups?
  - events versus measured quantities?
  - is the data normal?
  - is there a survival like function?
  - do we need to 'adjust' for variables other than the intervention?
  - are there repeated measures?

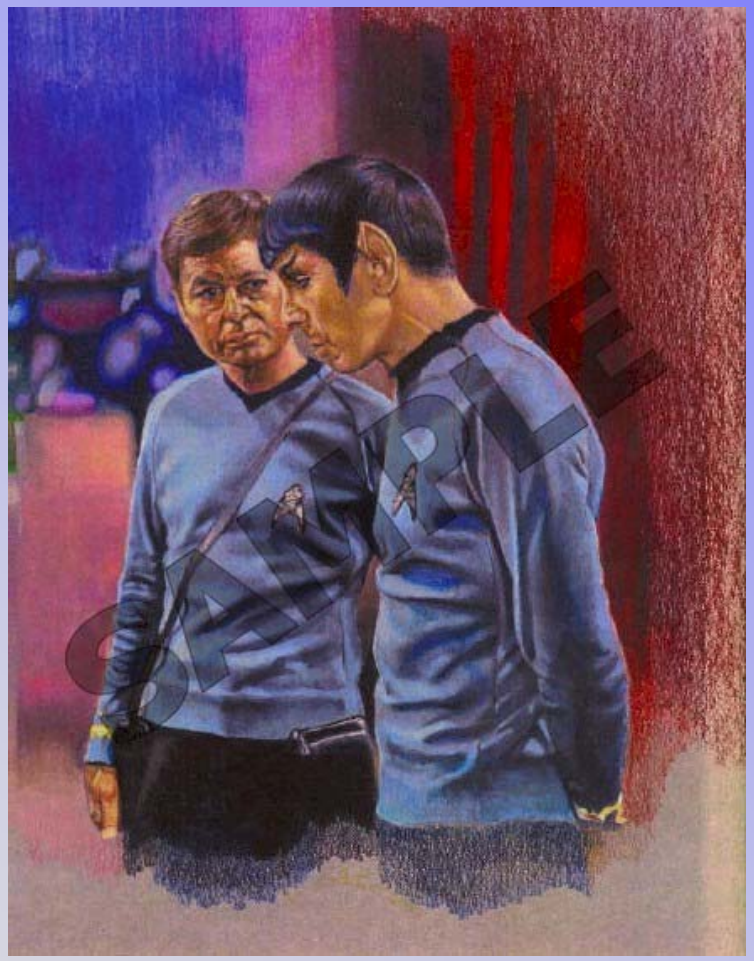
So many study designs, so many tests!

Number of Dependent* Variables	Number of Independent** Variables	Type of Dependent Variable(s)	Type of Independent Variable(s)	Measure	Test(s)
1	0 (1 population)	continuous normal	not applicable (none)	mean	one-sample t-test
		continuous non-normal		median	one-sample median
		categorical		proportions	Chi Square goodness-of-fit, binomial test
	1 (2 independent populations)	normal	2 categories	mean	2 independent sample t-test
		non-normal		medians	Mann Whitney, Wilcoxon rank sum test
		categorical		proportions	Chi square test Fisher's Exact test
	0 (1 population measured twice) ***or*** 1 (2 matched populations)	normal	not applicable/ categorical	means	paired t-test
		non-normal		medians	Wilcoxon signed ranks test
		categorical		proportions	McNemar, Chi-square test
	1 (3 or more populations)	normal	categorical	means	one-way ANOVA
		non-normal		medians	Kruskal Wallis
		categorical		proportions	Chi square test
	2 or more (e.g., 2-way ANOVA)	normal	categorical	means	Factorial ANOVA
		non-normal		medians	Friedman test
		categorical		proportions	log-linear, logistic regression
	0 (1 population measured 3 or more times)	normal	not applicable	means	Repeated measures ANOVA
	1	normal	continuous		correlation simple linear regression
		non-normal			non-parametric correlation
		categorical	categorical or continuous		logistic regression
			continuous		discriminant analysis

More tests...

Number of Dependent* Variables	Number of Independent** Variables	Type of Dependent Variable(s)	Type of Independent Variable(s)	Measure	Test(s)
	2 or more	normal	continuous		multiple linear regression
		non-normal			
		categorical			logistic regression
		normal	mixed categorical and continuous		Analysis of Covariance General Linear Models (regression)
		non-normal			
		categorical			logistic regression
2	2 or more	normal	categorical		MANOVA
2 or more	2 or more	normal	continuous		multivariate multiple linear regression
2 sets of 2 or more	0	normal	not applicable		canonical correlation
2 or more	0	normal	not applicable		factor analysis

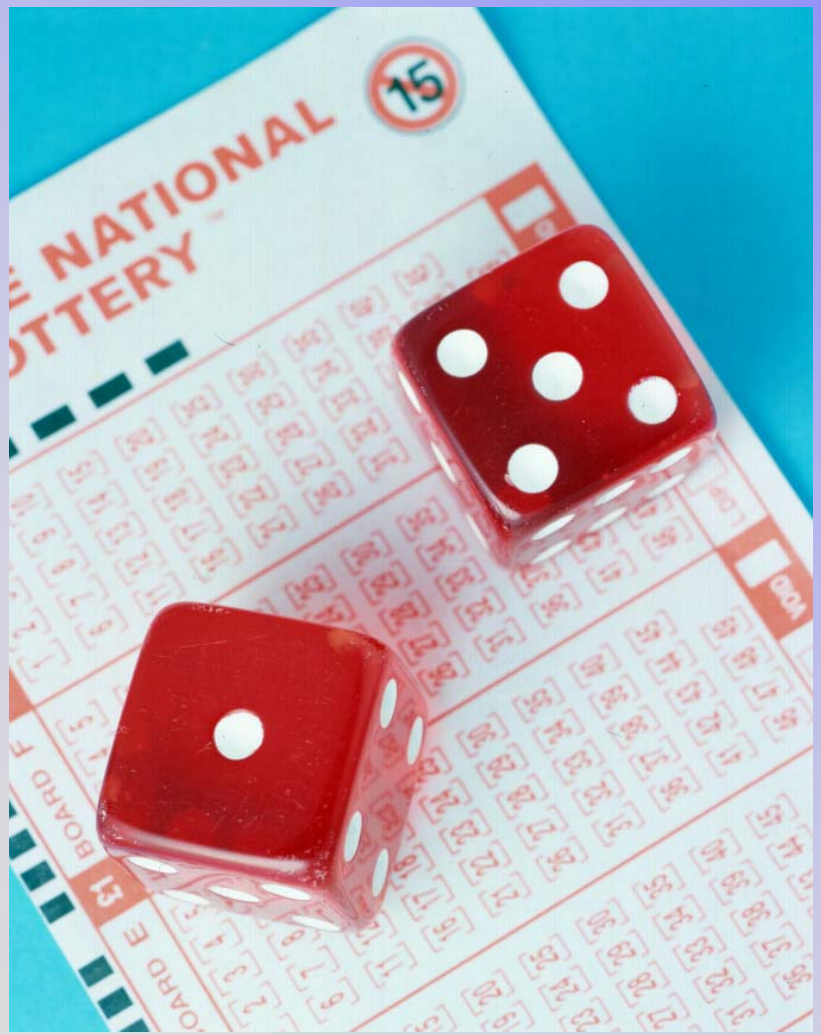
When to be concerned...



Large Study Samples

- More subjects
  - more likely that any difference will be statistically significant
  - tend to use Relative Measures
  - clinical significance needs careful review

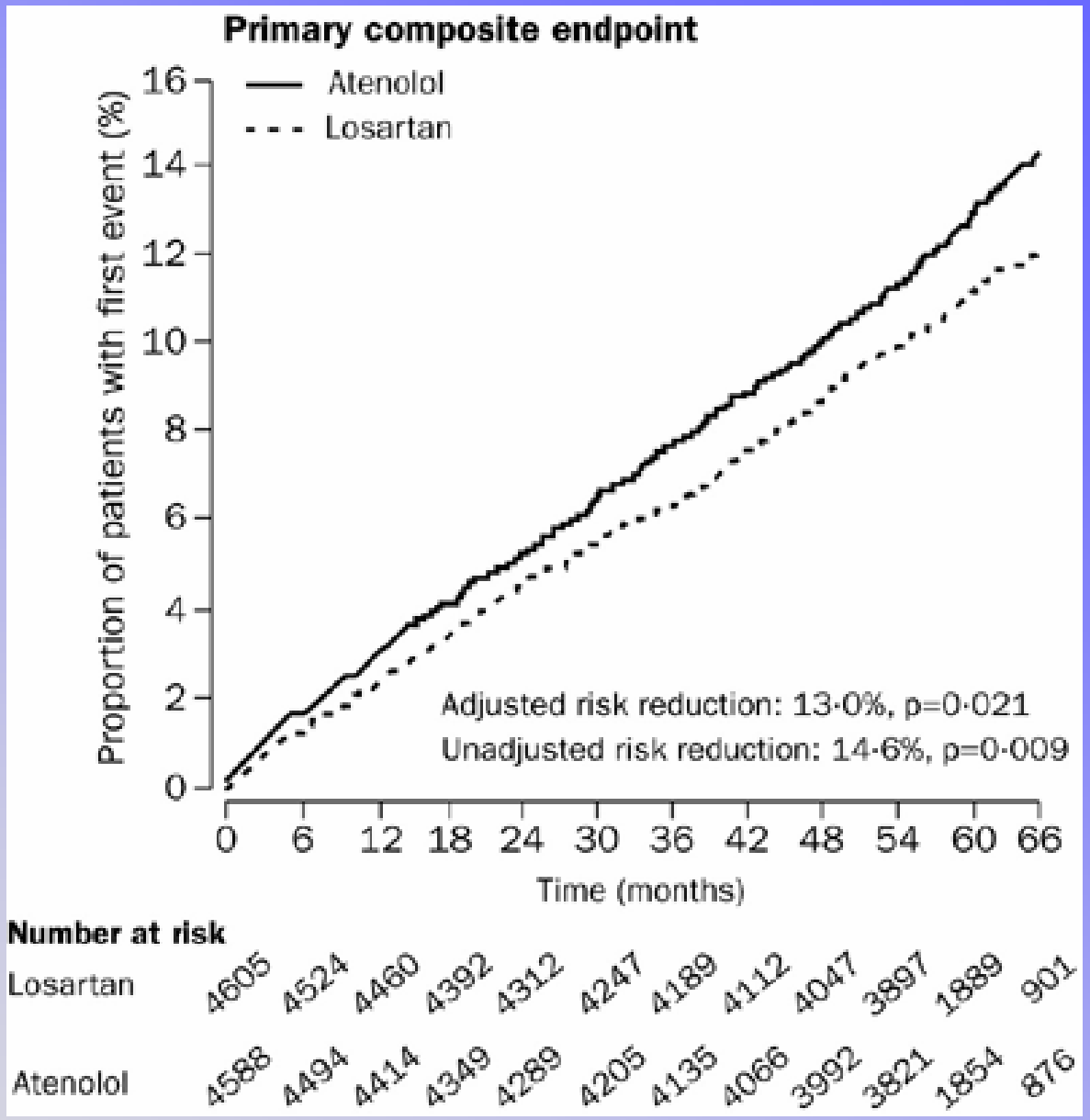
Lottery Example



- Lotto 649
- Background
  - 1 ticket
  - 1 / 13,983,816
- Intervention
  - buy another ticket
- Results
  - RR = 2.0
  - 2 /13,983,816
  - NNT = 13,983,815

LIFE trial
Lancet 2002

9,222 Subjects



Large Trials - Small Effects

- Raw data:
  - losartan: 508 events / 4605 subjects = 11.0%
  - atenolol: 588 events / 4588 subjects = 12.8%

- ARR:	0.018	1.8%
- RRR:	0.14	14.0%
- RR:	0.86	86.0%

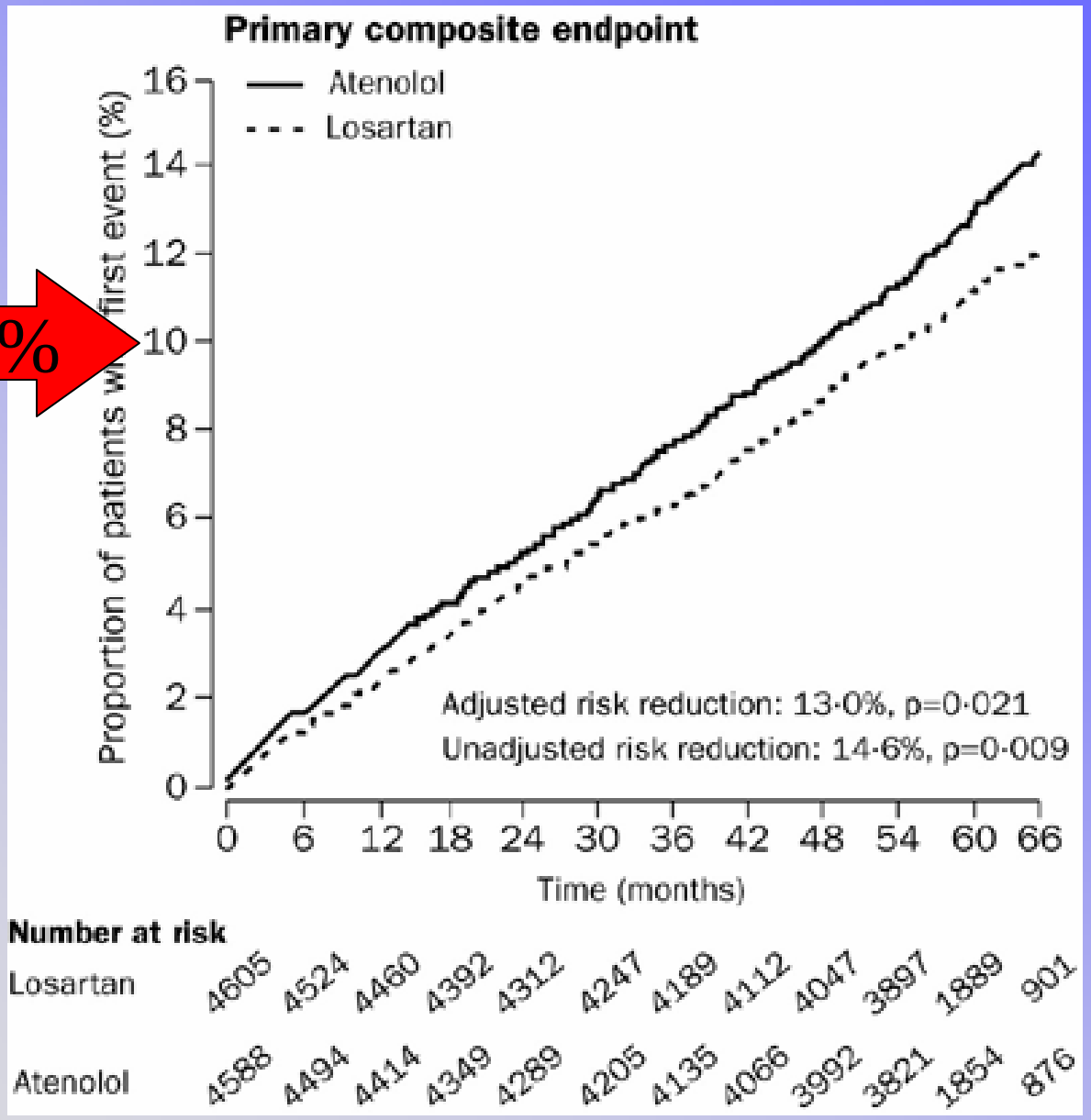
Which number would you like?

NNT = 55.5

Misleading Graphs

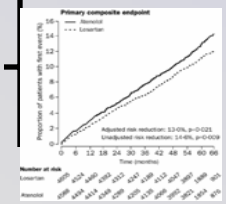
- Small effect sizes magnified…

scale 10%

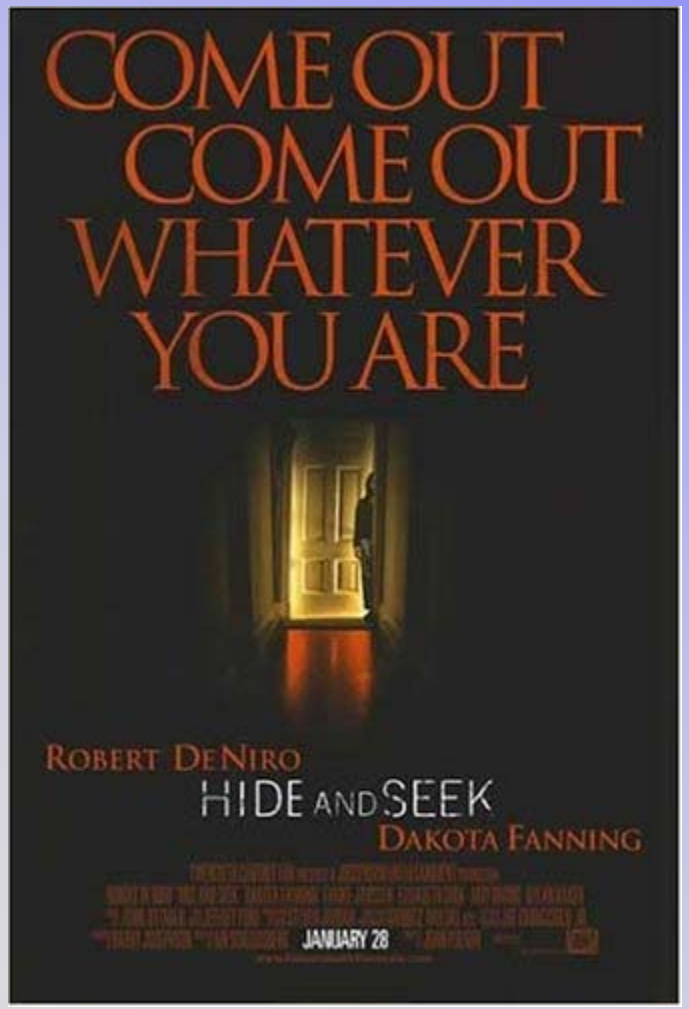

Primary composite endpoint
Atenolol
Losartan
Proportion of patients with first event (%)
16
14
12
10
8
6
4
2
0
Adjusted risk reduction: 13·0%, p=0·021
Unadjusted risk reduction: 14·6%, p=0·009
0 6 12 18 24 30 36 42 48 54 60 66
Time (months)
Number at risk
Losartan 4605 4524 4460 4392 4312 4247 4189 4112 4047 3897 1889 901
Atenolol 4588 4494 4414 4349 4289 4205 4135 4066 3992 3821 1854 876

Measures of Treatment & Prevention Effect

100% scale on y - axis



Subgroup Analyses



- Too many peaks…
- chances of finding a 'significant difference' increases with every look

Too many tries…

13 looks

if we accept $p<0.05$

1/20 will be ‘statistically significant’ by chance alone

Risk and Treatment Effect

- Highest risk individuals are most likely to benefit…
- Easiest to show benefit among patients most likely to respond.
- This leads to 'restricting entry' to high risk patients.

Measures of Treatment & Prevention Effect (examples)

Therapy	Endpoint	NNT (5yr)
stepped care for diastolic BP of 115 - 129	death, stroke, myocardial infarction	3
stepped care for diastolic BP 90 - 109	death, stroke, myocardial infarction	141

Biostats over Biosense

Aspirin use	No. of cases	Total person-years	Age-adjusted RR (95% CI)	Multivariable RR (95% CI)
Use				
Non-regular user†	96	969 187	1.0 (referent)	1.0 (referent)
Regular use (≥2 tablets per wk)	65	506 075	1.26 (0.92 to 1.73)	1.20 (0.87 to 1.65)
Current aspirin use per wk				
0 tablets	57	642 553	1.0 (referent)	1.0 (referent)
1–3 tablets	49	451 610	1.26 (0.86 to 1.85)	1.26 (0.85 to 1.85)
4–6 tablets	18	135 816	1.50 (0.88 to 2.54)	1.41 (0.82 to 2.40)
7–13 tablets	30	160 661	1.71 (1.10 to 2.65)	1.65 (1.05 to 2.59)
≥14 tablets	7	78 441	0.99 (0.45 to 2.20)	0.86 (0.39 to 1.89)
P_{trend}‡			.17	.41
Duration of regular use (≥2 tablets per wk):				
0 y	57	679 689	1.0 (referent)	1.0 (referent)
1–5 y	24	260 997	0.95 (0.59 to 1.54)	0.94 (0.58 to 1.53)
6–10 y	23	174 630	1.21 (0.75 to 1.96)	1.19 (0.73 to 1.95)
11–20 y	23	143 104	1.49 (0.91 to 2.45)	1.51 (0.92 to 2.48)
>20 y	34	214 650	1.58 (1.03 to 2.42)	1.58 (1.03 to 2.43)
P_{trend}‡			.01	.01

CASES	PT-YRS	PERCENT	RR
57	679689	0.0084%	
34	214650	0.0158%	1.89
	difference ->	0.0075%	

Tired Yet?

Sources

- Centre for Health Evidence
  - www.cche.net/che/home.asp
- Selecting Tests
  - www.whichtest.info/
- J Schaefer Presentations
  - www.ucalgary.ca/~jpschaef