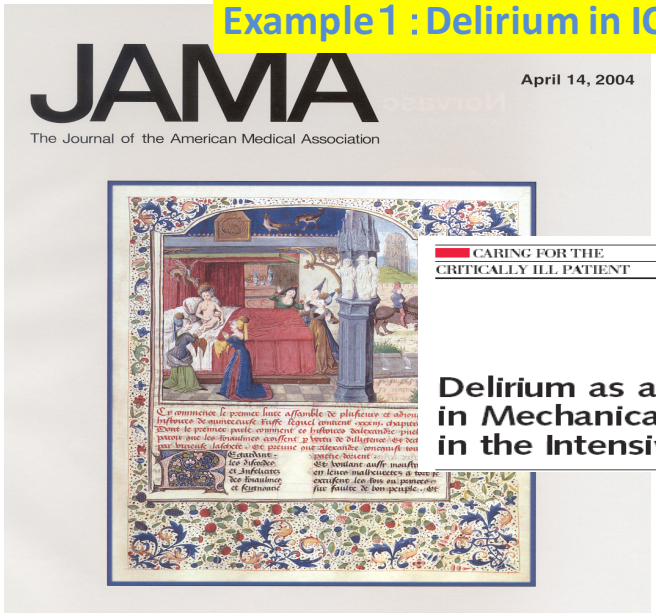
Basic Statistical Concept

Contents

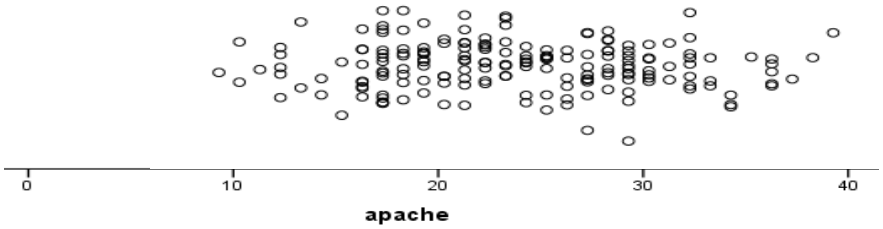
- Descriptive Statistics
  Mean and Standard Deviation (SD)
  Median and Inter-quartile-range (IQR)
- P-value and Hypothesis Testing
- 95% Confidence Interval
- Statistical graphs

Example 1 : Delirium in ICU and Mortality



Ely, Shintani, Truman et al, JAMA 2003;289:2983-91

Descriptive Statistics



Central Tendency — **Mean**

Data Variation — **Standard Deviation**

Example 1 : Delirium in ICU and Mortality

Table 1. Baseline Characteristics of the Patients*

Characteristic	No. (%)† No Delirium (n = 41)	Delirium (n = 183)
Age, mean (SD), y	54 (17)	56 (17)
Men	18 (44)	95 (52)
Race		
White	32 (78)	145 (79)
Black	9 (22)	38 (21)
Charlson Comorbidity Index, mean (SD)	3.2 (2.8)	3.2 (2.8)
Vision deficits, No./total (%)‡	23/33 (70)	104/153 (68)
Hearing deficits, No./total (%)‡	5/32 (16)	29/152 (19)
mBDRS score, mean (SD)	0.14 (0.6)	0.23 (0.8)
Activities of daily living, mean (SD)	0.81 (2.4)	0.91 (2.3)
APACHE II score, mean (SD)	23.2 (9.6)	25.6 (8.1)
SOFA score, mean (SD)	9.5 (2.9)	9.6 (3.4)
ICU admission diagnosis§		
Sepsis and/or acute respiratory distress syndrome	24 (59)	78 (43)
Pneumonia	6 (15)	35 (19)
Myocardial infarction/congestive heart failure	4 (10)	15 (8)
Hepatic or renal failure	0	11 (6)
Chronic obstructive pulmonary disease	2 (5)	18 (10)
Gastrointestinal bleeding	2 (5)	18 (10)
Malignancy	0	7 (4)
Drug overdose	3 (7)	8 (4)
Other	14 (34)	53 (29)

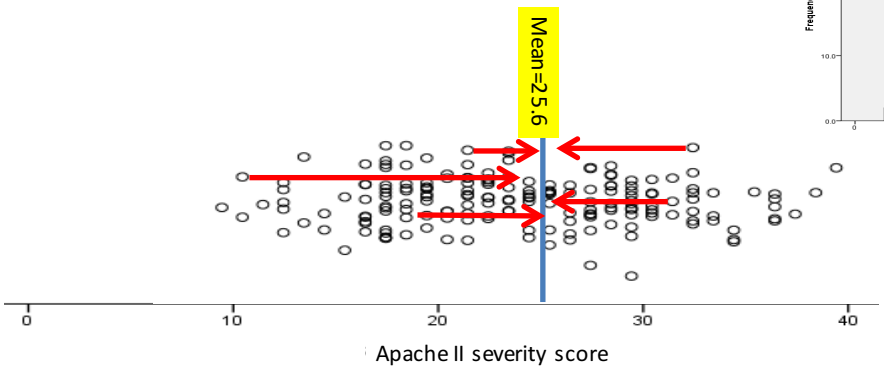
Abbreviations: APACHE II, Acute Physiology and Chronic Health Evaluation; ICU, intensive care unit; mBDRS, modified Blessed Dementia Rating Scale; SOFA, Sequential Organ Failure Assessment.
*All comparisons between the no delirium and delirium groups were nonsignificant ($P>.05$). See "Methods" section for descriptions of scales and for scale ranges.
†Except where noted otherwise.
‡Denominators indicate number of patients with available information.
§Recorded by the patients' medical team as the diagnoses most representative of the reason for admission to the ICU. Patients were sometimes given more than 1 admission diagnosis by the medical team, resulting in column totals >100%.

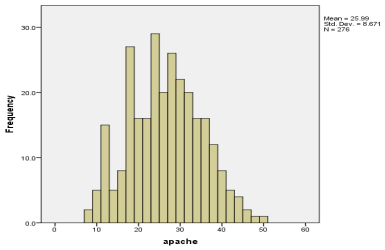
Mean (SD)

Standard Deviation (SD)

SD describes "variation" of data, which is approximately equivalent with average distance of each data point to their mean.

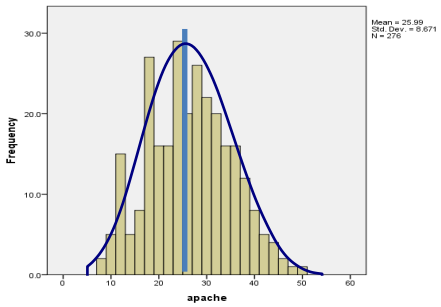
SD = 8.1





Standard Deviation (SD): Utility

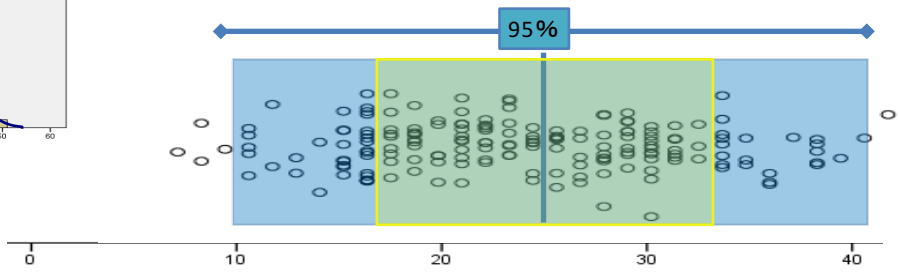
When data are normally distributed:



67%of patients' Apache score lie between a range defined by mean ±1SD , which is 25.6±8.1 =(17.5, 33.7)

67%

95%of patients' Apache score lie between a range defined by mean ±2SD , which is 25.6±2x8.1 =(9.4, 41.8)



Apache II severity score

Example1 : Delirium in ICU and Mortality

Table 1. Baseline Characteristics of the Patients*

Characteristic	No. (%)† No Delirium (n = 41)	Delirium (n = 183)
Age, mean (SD), y	54 (17)	56 (17)
Men	18 (44)	95 (52)
Race		
White	32 (78)	145 (79)
Black	9 (22)	38 (21)
Charlson Comorbidity Index, mean (SD)	3.2 (2.8)	3.2 (2.8)
Vision deficits, No./total (%)‡	23/33 (70)	104/153 (68)
Hearing deficits, No./total (%)‡	5/32 (16)	29/152 (19)
mBDRS score, mean (SD)	0.14 (0.6)	0.23 (0.8)
Activities of daily living, mean (SD)	0.81 (2.4)	0.91 (2.3)
APACHE II score, mean (SD)	23.2 (9.6)	25.6 (8.1)
SOFA score, mean (SD)	9.5 (2.9)	9.6 (3.4)
ICU admission diagnosis§		
Sepsis and/or acute respiratory distress syndrome	24 (59)	78 (43)
Pneumonia	6 (15)	35 (19)
Myocardial infarction/congestive heart failure	4 (10)	15 (8)
Hepatic or renal failure	0	11 (6)
Chronic obstructive pulmonary disease	2 (5)	18 (10)
tinal bleeding	2 (5)	18 (10)
	0	7 (4)
Drug overdose	3 (7)	8 (4)
		53 (29)

tied Blessed Dementia Rating Scale; SOFA, Sequential Organ Failure Assessment; mBDRS, modi-
*All comparisons between the no delirium and delirium groups were nonsignificant
descriptions of scales and for scale ranges.
†Except where noted otherwise.
‡Denominators indicate number of patients with available information.
§Recorded by the patients' medical team as the diagnoses most representative of the reason for admission to the ICU.
Patients were sometimes given more than 1 admission diagnosis by the medical team, resulting in column totals

Mean ± SD

Let's use it !

95%of patients' Apache score lie between a range defined by mean ±2SD

25.6±2 x 8.1 =(9.4, 41.8)

Example1 : Delirium in ICU and Mortality

Table 2. Daily and Cumulative Doses of Sedative and Analgesic Medications

Drug	Daily ICU Dose, Mean (SD), mg: No Delirium (n = 41)	Delirium (n = 183)	*P* Value†
Lorazepam	1.12 (2.2)	4.8 (12.8)	.01
Propofol	36.6 (258.6)	48.4 (172.9)	.19
Morphine	5.8 (17.0)	17.3 (163.8)	.79
Fentanyl	0.53 (1.7)	0.78 (1.7)	.22

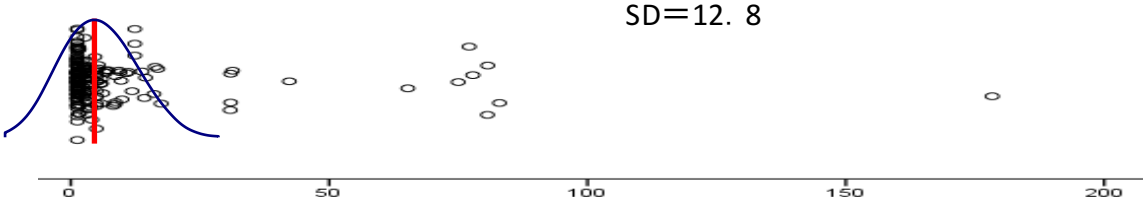
Abbreviation: ICU, intensive care unit.
*In the persistently comatose patients, the mean (SD) cumulative doses of these medications were fentanyl, 3 (12) mg.
†By Wilcoxon rank sum test for no delirium vs delirium.
‡Fentanyl is commonly reported to be 100 times more potent than morphine.[54] Therefore, using a dose of fentanyl given to patients in the no delirium and delirium groups would equate to 310 m confounded in vivo, such large values are plausible considering fentanyl's initially short durati ion of fentanyl as a continuous infusion ra

Let's use it !

4.8±2 x 12.8 =(-20.8, 30.4)

95%of patients' daily ICU dose of lorazepam is -20.8mg to 30.4 mg?????

95% of patients' daily ICU dose of lorazepam is -20.8mg to 30.4 mg???



Mean=4.8
SD=12. 8

Median [Inter-quartile range, IQR] = 1 [0, 4.25]
25% 0 mg
50% 1 mg
75% 4.25mg

THE LANCET
Volume 371 · Number 9607 · Pages 89–176 · January 12–18, 2008 www.thelancet.com

Articles

Efficacy and safety of a paired sedation and ventilator weaning protocol for mechanically ventilated patients in intensive care (Awakening and Breathing Controlled trial): a randomised controlled trial

Lancet 2008; 371: 126–34
See Comment page 95

Timothy D Girard, John P Kress, Barry D Fuchs, Jason W W Thomason, William D Schweickert, Brenda T Pun, Darren B Taichman, Jan G Dunn, Anne S Pohlman, Paul A Kinniry, James C Jackson, Angelo E Canonico, Richard W Light, Ayumi K Shintani, Jennifer L Thompson, Sharon M Gordon, Jesse B Hall, Robert S Dittus, Gordon R Bernard, E Wesley Ely

The Lancet (ISSN 0099-5355) is published weekly, except for the last issue in December which is a double issue, by Elsevier Ltd. Periodical postage paid at New York NY and additional mailing offices. # 585480 USPS CDN PM#0905372
POSTMASTER: Send address changes to The Lancet, Elsevier, Subscription Customer Service, 6277 Sea Harbour Drive, Orlando, FL 32887-4800, USA.
 Printed in USA.
Founded 1823 · Published weekly

Patients' characteristics

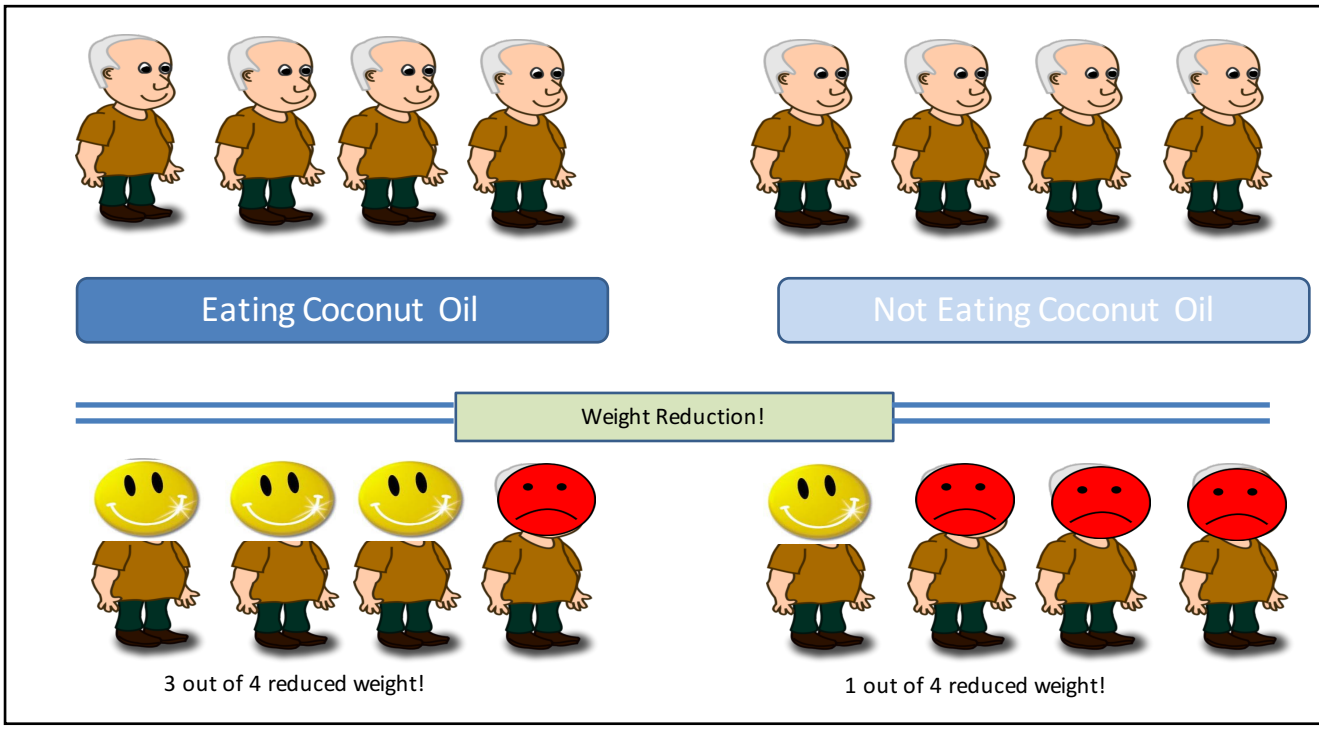
	Intervention group (n=167)	Control group (n=168)
Age (years)	60 (48 to 71)	64 (51 to 75)
Sex (female)	77 (46%)	83 (49%)
APACHE II score	26 (21 to 33)	26·5 (21 to 31)
SOFA score	9 (6 to 11)	8 (6 to 11·5)
Diagnosis on admission to intensive care		
Sepsis/acute respiratory distress syndrome	79 (47%)	87 (52%)
Myocardial infarction/congestive heart failure	22 (13%)	29 (17%)
Chronic obstructive pulmonary disease/asthma	17 (10%)	12 (7%)
Altered mental status	18 (11%)	12 (7%)
Hepatic or renal failure	9 (5%)	5 (3%)
Malignancy	3 (2%)	2 (1%)
Alcohol withdrawal	1 (1%)	1 (1%)
Other*	18 (11%)	20 (12%)
RASS on first study day	−4 (−5 to −2)	−4 (−5 to −2)
Sedation before enrolment		
Benzodiazepines (mg)†	8 (4 to 34)	10 (2 to 41)
Opiates (μg)‡	815 (184 to 4380)	850 (142 to 4685)
Propofol (mg)	5102 (2340 to 9720)	3248 (1455 to 7420)
Time from admission to enrolment (days)	2·2 (1·1 to 3·9)	2·2 (1·1 to 3·9)

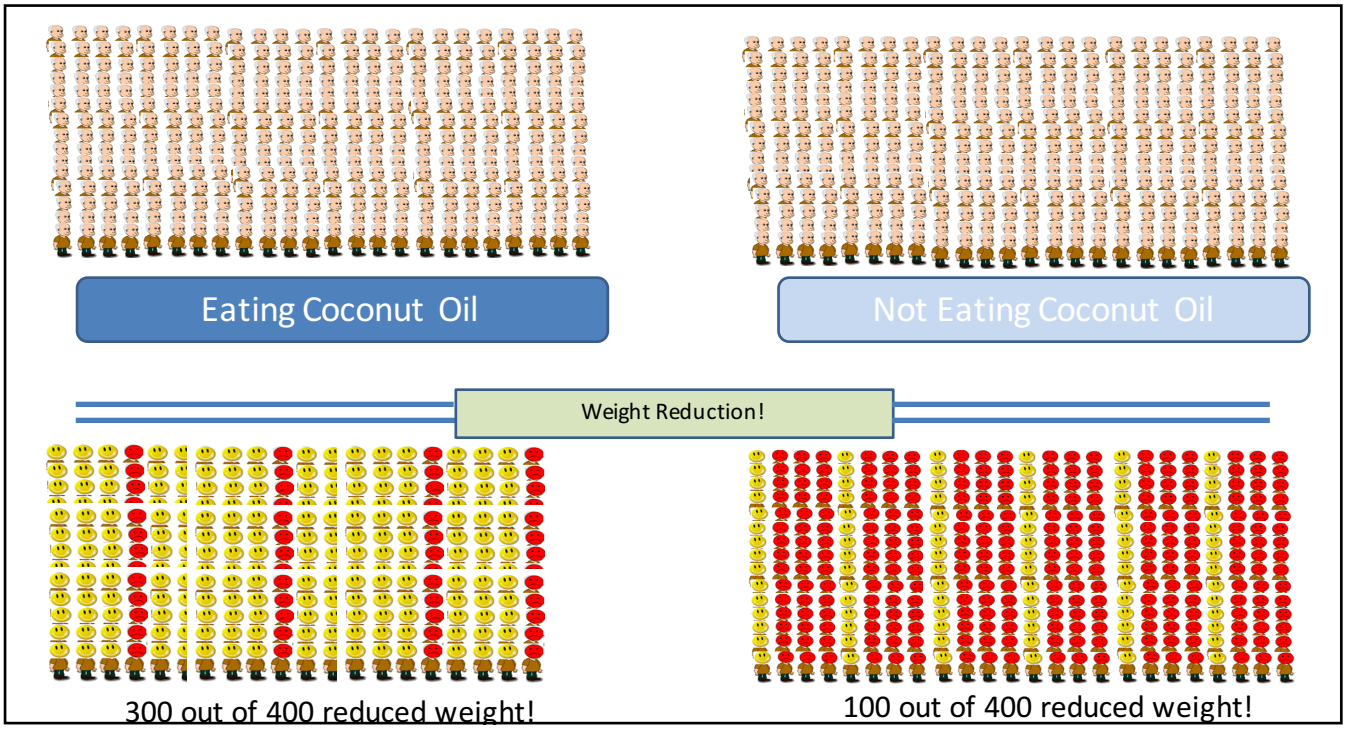
Data are n (%) or median (IQR). APACHE II=acute physiology and chronic health evaluation II. RASS=Richmond agitation-sedation scale. SAT=spontaneous awakening trial. SBT=spontaneous breathing trial. SOFA=sequential organ failure assessment. *Including gastrointestinal bleeding, metabolic disarray, haemoptysis, pulmonary embolism, and status epilepticus. †Expressed in lorazepam equivalents.[34] ‡Expressed in fentanyl equivalents.[34]

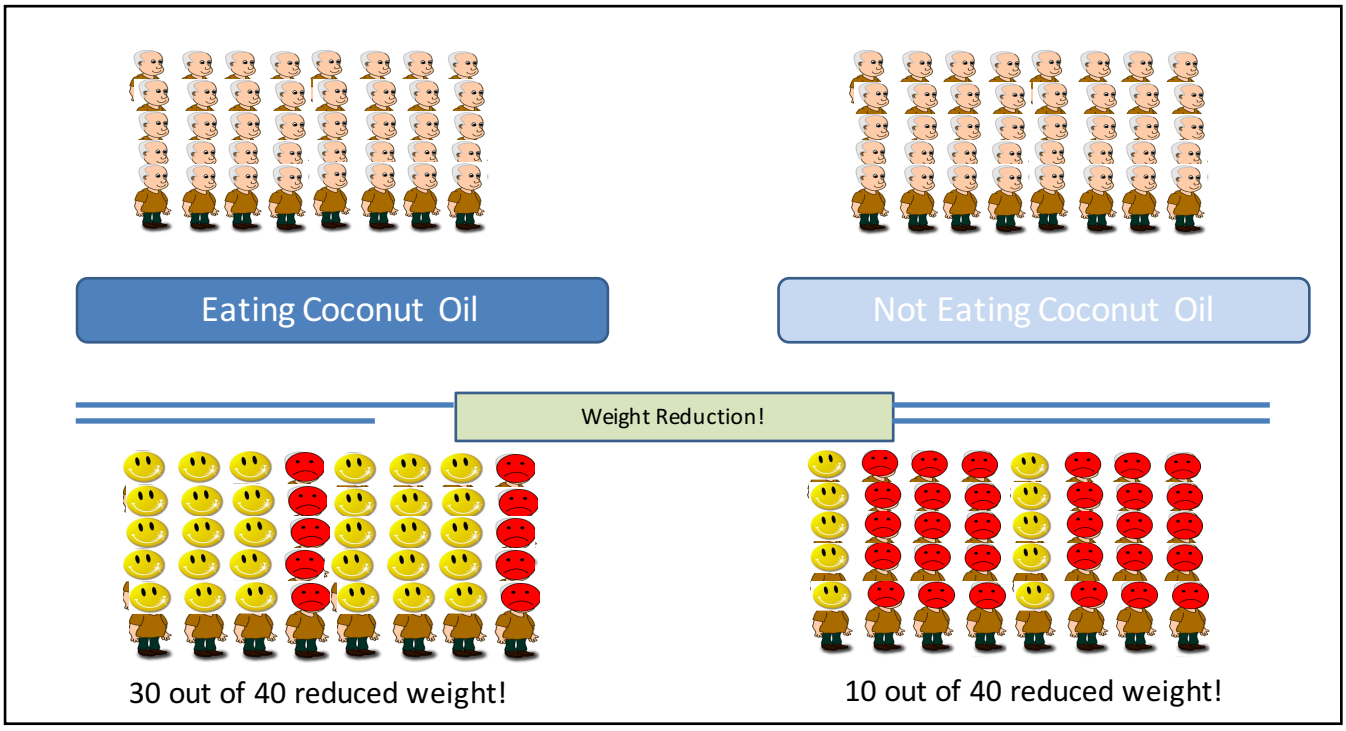
50% 10 mg
25% 2 mg
75% 41 mg

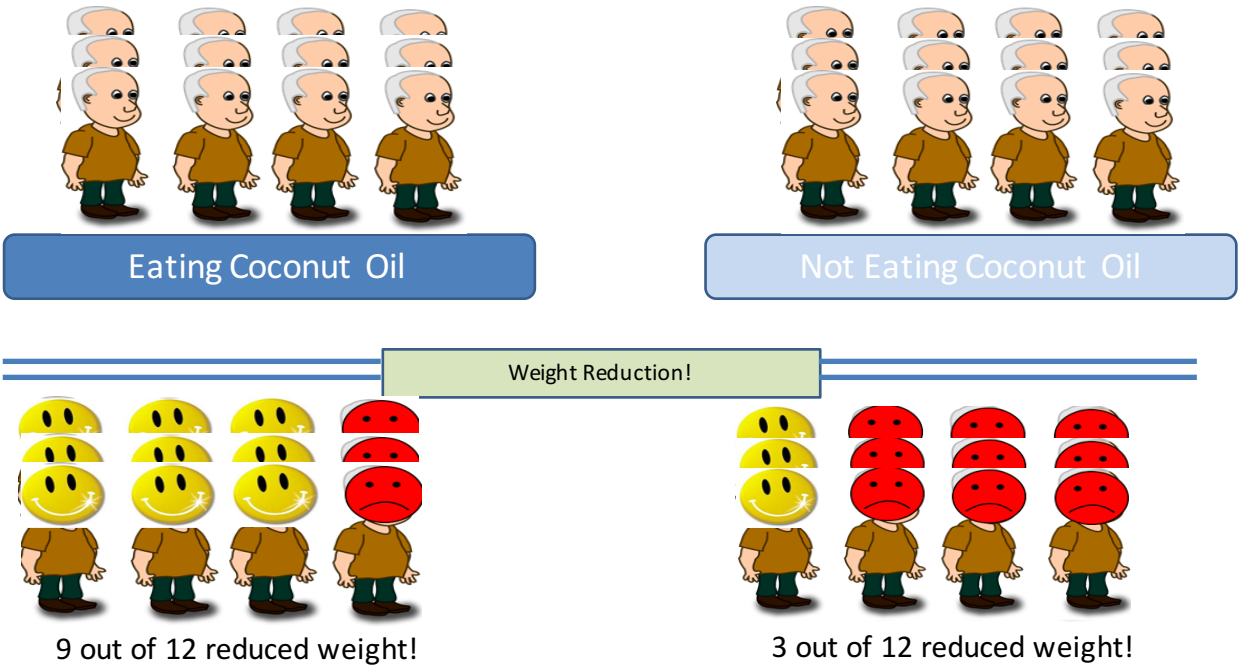
Contents

- Descriptive Statistics
  Mean and Standard Deviation (SD)
  Median and Inter-quartile-range (IQR)
- P-value and Hypothesis Testing
- 95% Confidence Interval
- Statistical graphs




Eating Coconut Oil
Not Eating Coconut Oil
Weight Reduction!
300 out of 400 reduced weight!
100 out of 400 reduced weight!


Eating Coconut Oil
Not Eating Coconut Oil
Weight Reduction!
30 out of 40 reduced weight!
10 out of 40 reduced weight!



What is a scientific evidence ?

When you want to prove that a new drug works, which approach do you want to take?

Alternative hypothesis (H_a)

A. Give evidences to support that the drug works

B. Give an evidence to against that the drug does not work

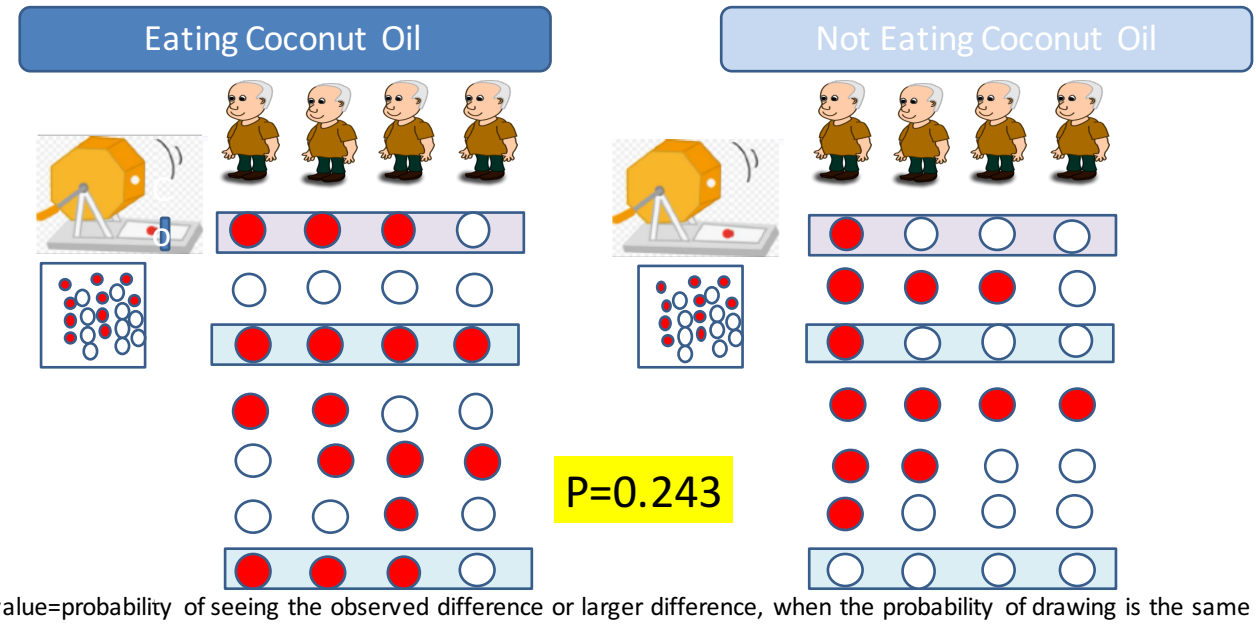
Which approach do you think more convincing (or easier to collect evidence)?

Null hypothesis (H_o)

Disproving Hypothesis in a Evidence Based Medicine

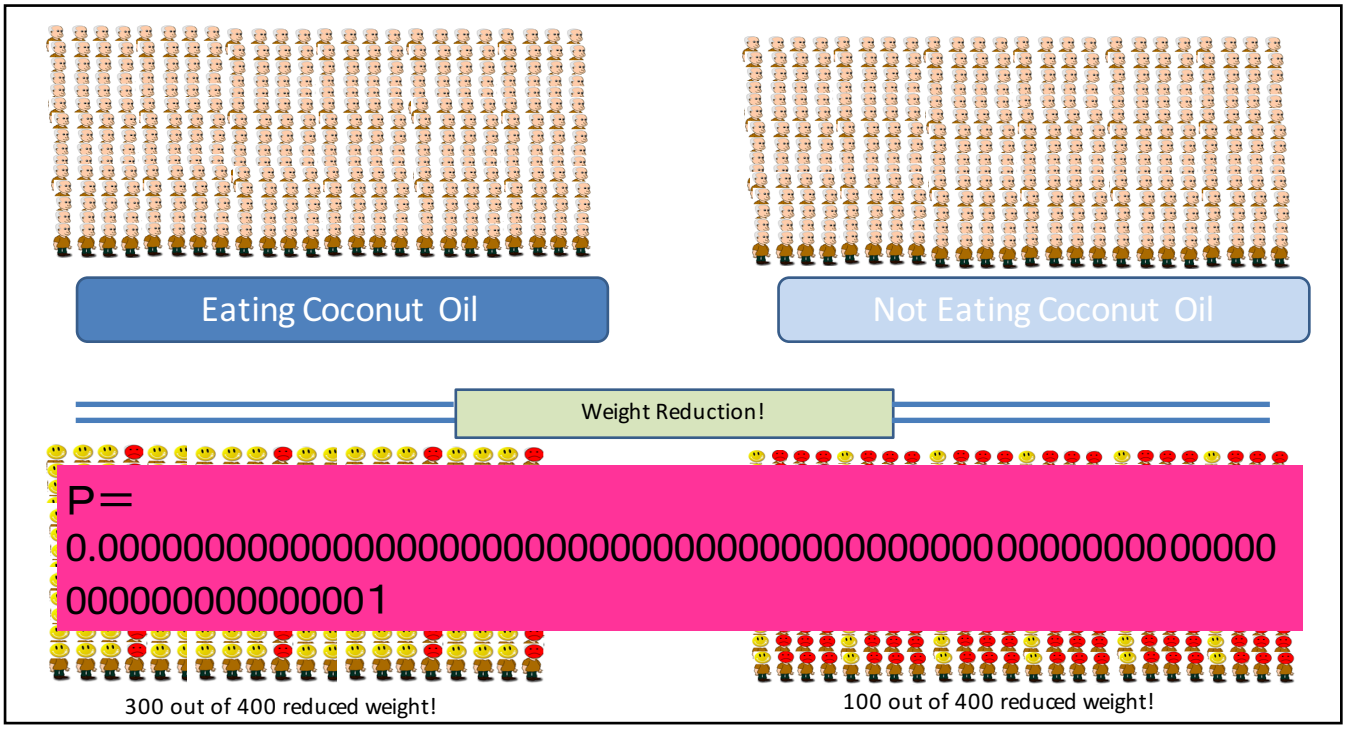
In general, it is much easier to find evidence which *againsts* a hypothesis than an evidence which proves that it is correct. In fact, one view of science is a process of disproving hypothesis. Statistical methods formalize this idea by looking for evidence against a **null hypothesis (H_0)**: that there is no difference between groups or no association between variable.

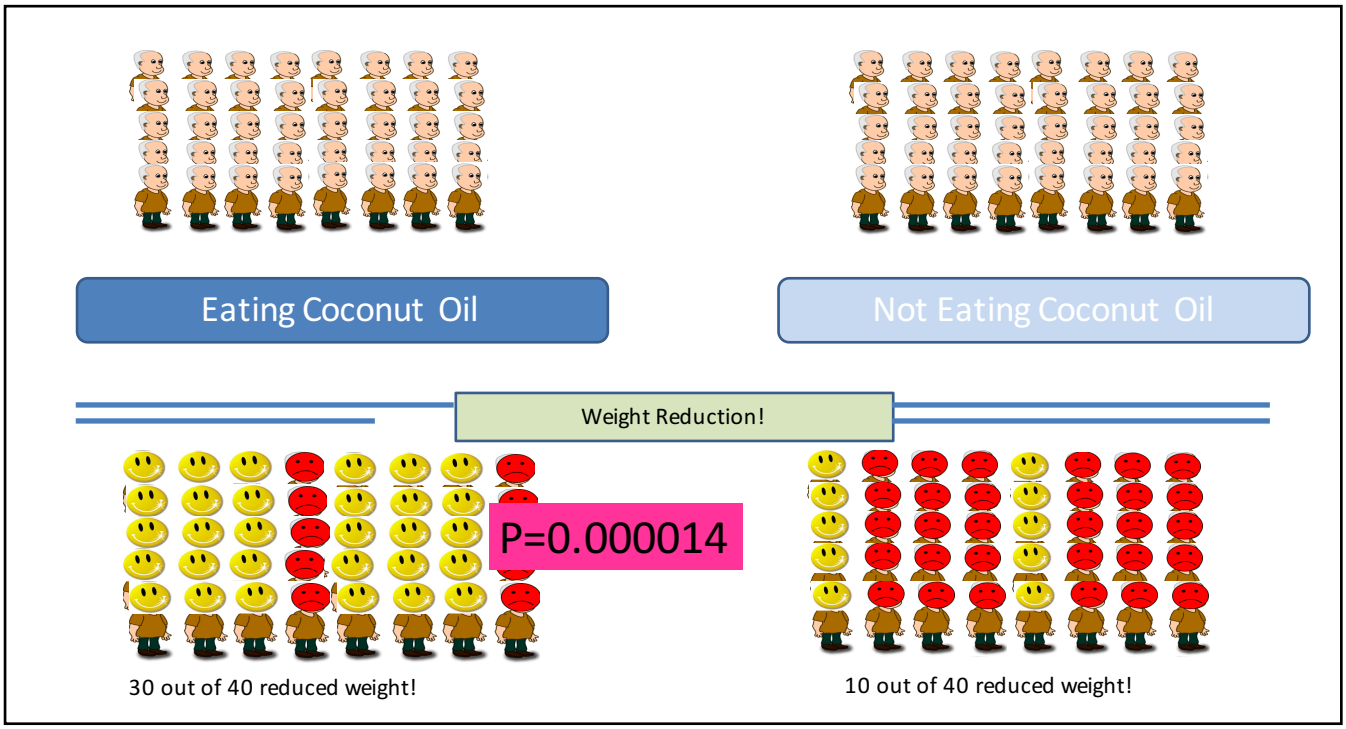
P-value is used as an evidence **to against the null hypothesis**, it is defined as "probability of observing the observed difference or greater difference when the null hypothesis is true". (**So smaller is better**)

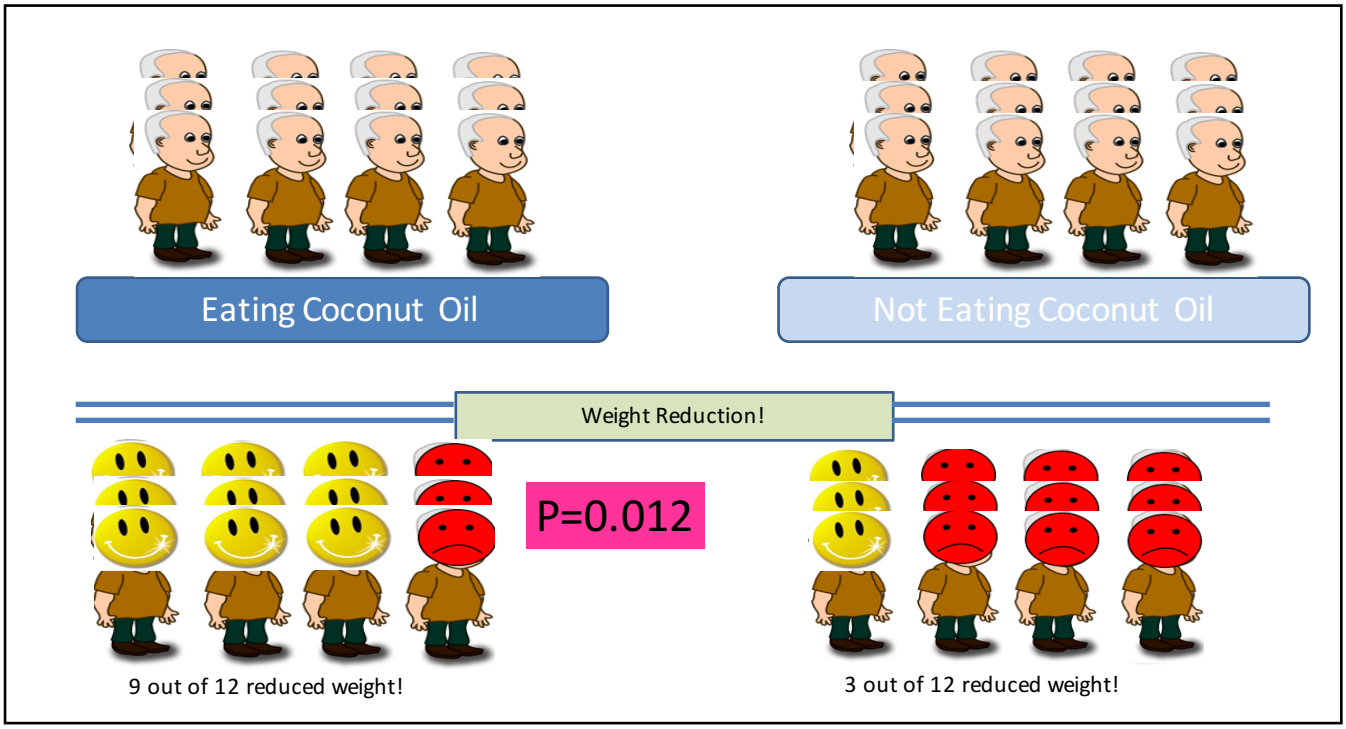


P-value=probability of seeing the observed difference or larger difference, when the probability of drawing is the same (i.e., no effect of coconut oil)

P-value: Probability of observing the difference by chance alone.


Eating Coconut Oil
Not Eating Coconut Oil
Weight Reduction!
P=
0.000000000000000000000000000000000000000000000000000000000001
300 out of 400 reduced weight!
100 out of 400 reduced weight!


Eating Coconut Oil
Not Eating Coconut Oil
Weight Reduction!
P=0.000014
30 out of 40 reduced weight!
10 out of 40 reduced weight!



Pit falls of P-value

- P-value becomes larger with a smaller difference.
- P-value becomes larger with a smaller sample size.

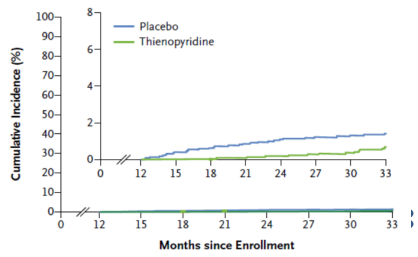
Thus, we cannot really tell why statistical significance is absent, due to small effect, or small sample size?

The NEW ENGLAND JOURNAL of MEDICINE
MARCH 24, 2011

Tiotropium versus Salmeterol for the Prevention of Exacerbations of COPD

Claus Vogelmeier, M.D., Bettina Hederer, M.D., Thomas Glaab, M.D., Hendrik Schmidt, Ph.D., Maureen P.M.H. Rutten-van Mölken, Ph.D., Kai M. Beeh, M.D., Klaus F. Rabe, M.D., and Leonardo M. Fabbri, M.D., for the POET-COPD Investigators*

Stent Thrombosis
12–30 mo Thienopyridine vs. placebo, 0.4% vs. 1.4%; hazard ratio, 0.29; P<0.001



No. at Risk

Thienopyridine	5020	4934	4870	4828	4765	4686	4642	3110
Placebo	4941	4845	4775	4721	4651	4603	4556	3105

Figure 2. Cumulative Incidence of Stent Thrombosis, According to Study

CONCLUSIONS

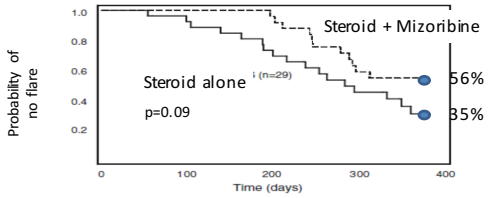
Dual antiplatelet therapy beyond 1 year after placement of a drug-eluting stent, as compared with aspirin therapy alone, significantly reduced the risks of stent thrombosis and major adverse cardiovascular and cerebrovascular events but was associated with an increased risk of bleeding.

Pediatr Nephrol (2010) 25:877–882
DOI 10.1007/s00467-009-1341-4

ORIGINAL ARTICLE

Combination therapy with steroids and mizoribine in juvenile SLE: a randomized controlled trial

Yuriko Tanaka · Norishige Yoshikawa · Shinzaburo Hattori · Satoshi Sasaki · Takashi Ando · Masahiro Ikeda · Masataka Honda · for Japanese Study Group for Renal Disease in Children



Number at risk	day 0	50	100	150	200	250	300	350
Group S+M	28	28	27	25	24	18	14	10
Group S	29	27	24	22	19	15	11	8

Conclusion: The time to the first flare from treatment initiation was not significantly different between the two groups (p=0.09)

Suggestion :
Clinically meaningful reduction of flare was observed, though it did not reach statistical significance due to a small sample size.

Pediatr Nephrol (2010) 25:877–882
DOI 10.1007/s00467-009-1341-4

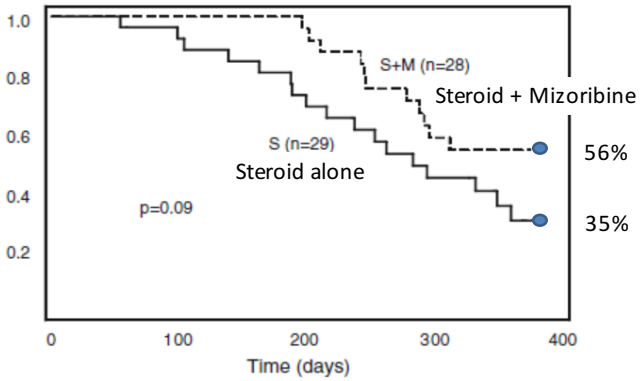
ORIGINAL ARTICLE

Combination therapy with steroids and mizoribine in juvenile SLE: a randomized controlled trial

Yuriko Tanaka · Norishige Yoshikawa · Shinzaburo Hattori · Satoshi Sasaki · Takashi Ando · Masahiro Ikeda · Masataka Honda · for Japanese Study Group for Renal Disease in Children

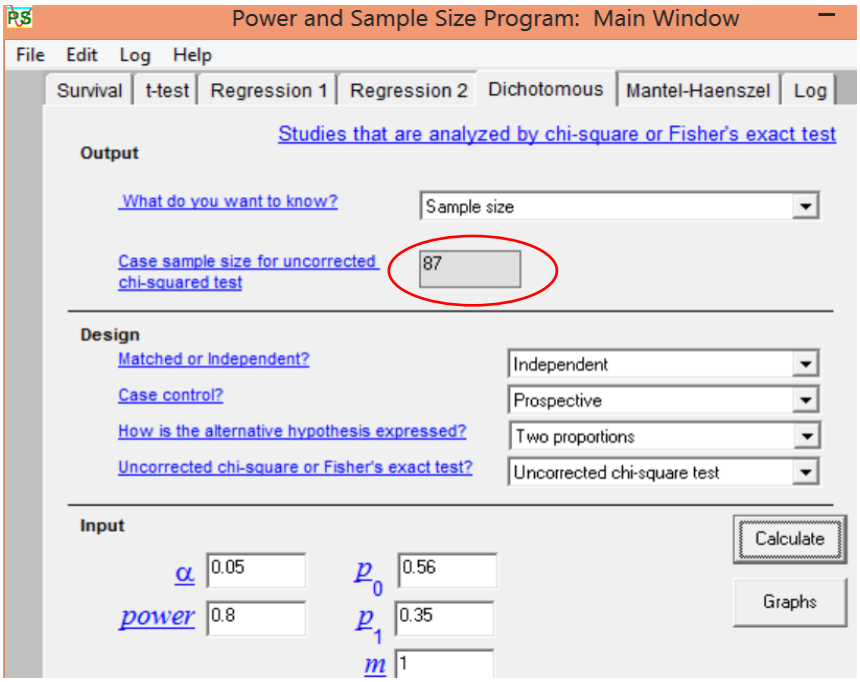
Sample size computation :

87 patients are required in order to reach a statistical significance with this magnitude of difference.



Number at risk	day 0	50	100	150	200	250	300	350
Group S+M	28	28	27	25	24	18	14	10
Group S	29	27	24	22	19	15	11	8

Fig. 2 Survival time without flare

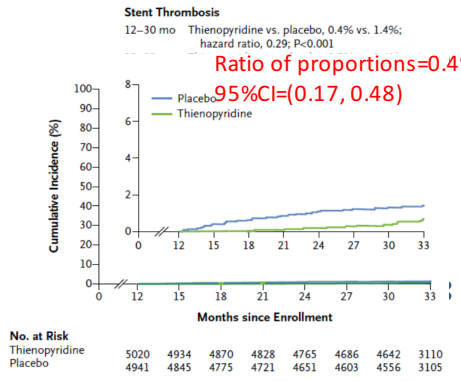


Contents

- Descriptive Statistics
  Mean and Standard Deviation (SD)
  Median and Inter-quartile-range (IQR)
- P-value and Hypothesis Testing
- 95% Confidence Interval
- Statistical graphs

The NEW ENGLAND JOURNAL of MEDICINE
ESTABLISHED IN 1812 DECEMBER 4, 2014 VOL. 371 NO. 23

Twelve or 30 Months of Dual Antiplatelet Therapy after Drug-Eluting Stents



Ratio of proportions=0.4%/1.4%=0.29
95%CI=(0.17, 0.48)

Figure 2. Cumulative Incidence of Stent Thrombosis, According to Study

CONCLUSIONS

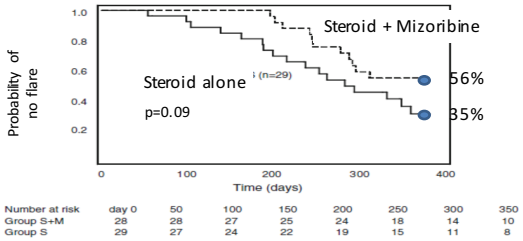
Dual antiplatelet therapy beyond 1 year after placement of a drug-eluting stent, as compared with aspirin therapy alone, significantly reduced the risks of stent thrombosis and major adverse cardiovascular and cerebrovascular events but was associated with an increased risk of bleeding.

Pediatr Nephrol (2010) 25:877–882
DOI 10.1007/s00467-009-1341-4

ORIGINAL ARTICLE

Combination therapy with steroids and mizoribine in juvenile SLE: a randomized controlled trial

Yuriko Tanaka · Norishige Yoshikawa · Shinzaburo Hattori · Satoshi Sasaki · Takashi Ando · Masahiro Ikeda · Masataka Honda · for Japanese Study Group for Renal Disease in Children



Conclusion: The time to the first flare from treatment initiation was not significantly different between the two groups (p=0.09)

Suggestion :
Clinically meaningful reduction of flare was observed, though it did not reach statistical significance due to a small sample size.

Relationship between p-value and 95% Confidence Interval, CI

95% CI including the null value ↔ P>0.05 No difference detected

95% CI excluding the null value ↔ P<=0.05 A difference detected

Null value = 1 when a ratio between two means (or proportions) is evaluated.

Null value = 0 when a difference between two means is evaluated.

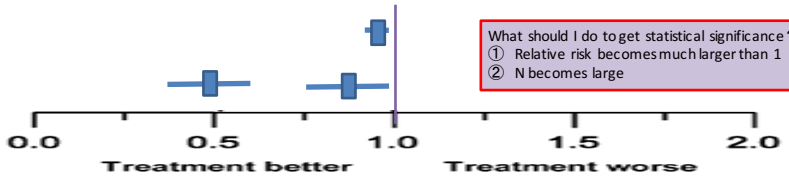
Table 2. Effect of n-3 PUFA in the primary and secondary analyses of GISSI-Prevenzione

	End-study events in the control group (%)	Relative risk (95% confidence interval)	
Primary endpoints			
Death, nonfatal MI, and nonfatal stroke	14.8	0.85 (0.74–0.98)	$p<0.05$
CV death, nonfatal MI, and nonfatal stroke	11.7	0.80 (0.68–0.94)	$p<0.05$
Secondary analyses			
All-cause mortality	10.5	0.80 (0.67–0.94)	$p<0.05$
Cardiovascular mortality	7.2	0.70 (0.56–0.86)	$p<0.05$
Coronary mortality	5.2	0.68 (0.53–0.88)	$p<0.05$
Sudden death	3.3	0.56 (0.40–0.79)	$p<0.05$
CHD death and nonfatal MI	8.9	0.78 (0.65–0.94)	$p<0.05$
Fatal and non-fatal stroke	1.4	1.24 (0.82–1.87)	$p\geq0.05$
Nonfatal MI	4.0	0.95 (0.79–1.14)	$p\geq0.05$
Nonfatal stroke	1.1	1.08 (0.75–1.55)	$p\geq0.05$

Significant

Not significant

95%CI = (mean − 2 x SE - mean + 2 x SE)
= (mean − 2 x SD/√N - mean + 2 x SD/√N)



What should I do to get statistical significance?
① Relative risk becomes much larger than 1
② N becomes large

CI includes 1 (Null value, i.e., no difference)
P ≥ 0.05 (Not significantly different)

CI does not include 1
P < 0.05 (Significantly different)

Let's use Confidence Interval!

The NEW ENGLAND JOURNAL of MEDICINE

ORIGINAL ARTICLE

Premature Coronary-Artery Atherosclerosis in Systemic Lupus Erythematosus

Yu Asanuma, M.D., Ph.D., Annette Oeser, B.S., Ayumi K. Shintani, Ph.D., M.P.H., Elizabeth Turner, M.D., Nancy Olsen, M.D., Sergio Fazio, M.D., Ph.D., MacRae F. Linton, M.D., Paolo Raggi, M.D., and C. Michael Stein, M.D.

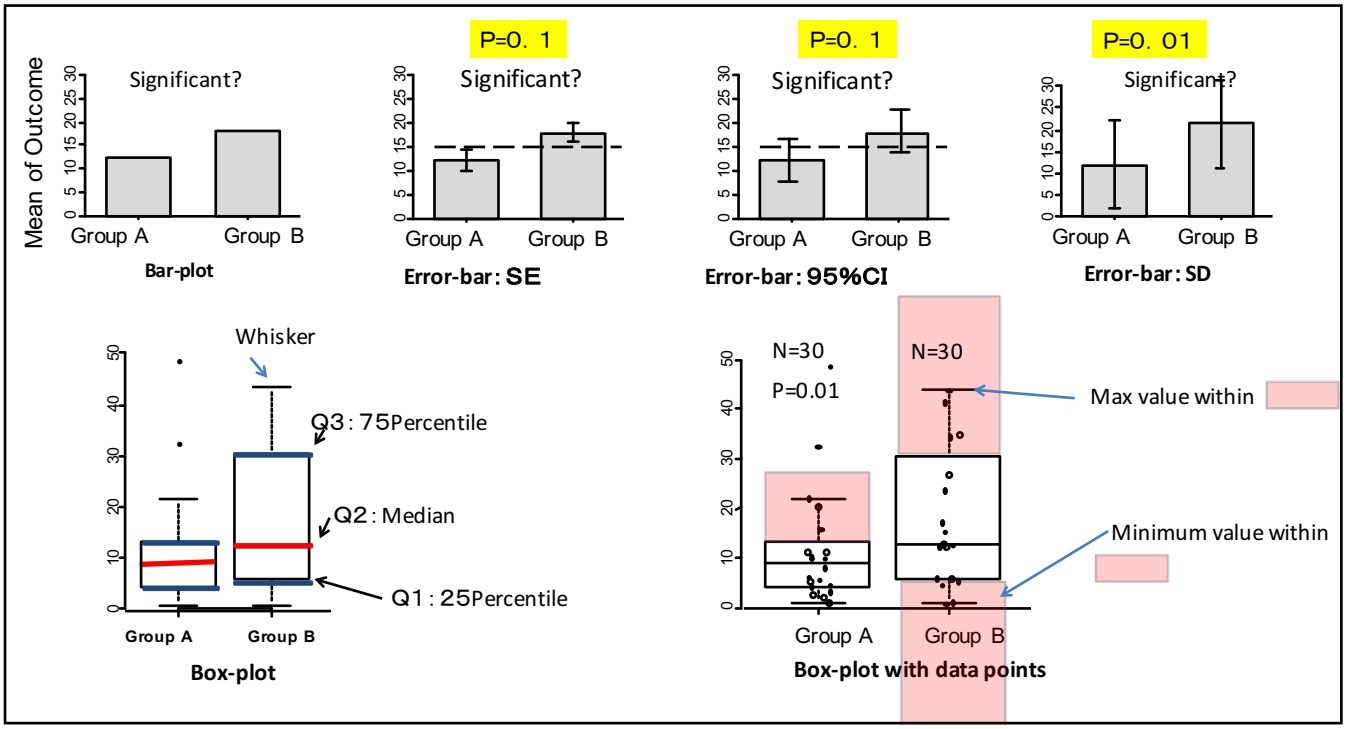
N Engl J Med 2003;349:2407-15.

Table 3. Characteristics of Patients with Lupus, According to the Presence or Absence of Coronary-Artery Calcification.*

Characteristic	Calcification (N=20)	No Calcification (N=45)	P Value†	Adjusted for Age and Sex	
				Odds Ratio (95% CI)‡	P Value
Age (yr)	49.9±10.9	36.0±9.2	<0.001	NA	NA
Female sex (%)	75	98	0.008§	NA	NA
White race (%)	70	73	0.88§	2.47 (0.5–11.4)	0.25
Duration of disease (yr)	12.1±10.3	8.9±7.9	0.16	0.96 (0.88–1.06)	0.41
Blood pressure (mm Hg)					
Systolic	127.0±22.0	119.4±18.0	0.12	1.01 (0.98–1.05)	0.58
Diastolic	78.0±15.2	77.0±14.6	0.66	1.00 (0.95–1.05)	0.99
Family history of coronary heart disease (%)	25	18	0.52§	1.80 (0.39–8.42)	0.45
Current smoking (%)	45	31	0.40§	1.19 (0.30–4.76)	0.80
Total pack-yr of smoking	11.5±13.3	5.7±10.3	0.18	1.00 (0.94–1.05)	0.91
Body-mass index	29.5±6.1	27.6±6.1	0.22	1.03 (0.92–1.15)	0.59
Creatinine (mg/dl)	0.9±0.1	0.8±0.4	<0.001	3.96 (0.52–30.4)	0.19
Albumin (g/dl)	3.6±0.4	3.7±0.6	0.49	0.43 (0.09–2.02)	0.28

Contents

- Descriptive Statistics
  - Mean and Standard Deviation (SD)
  - Median and Inter-quartile-range (IQR)
- P-value and Hypothesis Testing
- 95% Confidence Interval
- Statistical graphs



Summary of this lecture

1. Mean and Standard deviation are useful when data are normally distributed, otherwise, you may consider using Median and Inter-quartile range because they are more robust for data distribution.

2. P-value indicates the probability of falsely detecting a difference when there is no difference. Larger effect and larger sample size leads to more significant result (smaller p-value). So even clinically meaningless difference could reach statistical significance, so be careful.

3. Using confidence interval can refer statistical significance. When 95% CI does not include null value, which links with $P < 0.05$.

4. Use SD to describe sampled data, and use 95% CI to make statistical inference.

5. Don't forget to label error–bars, and use 95% CI for error bars comparing means. Also consider Box-Whisker's plot instead of error-bar plots.