Greatest Hits:

Top Ten Cognitive Optical Illusions in Clinical Research

Richard Chin, M.D.
richardchin@clinicaltrialist.com

Examples of Everyday Intellectual Illusions

- You're in a footrace and you pass the person in second place. What place are you in?
- A pencil and an eraser together cost $1.10. The pencil is $1 more than the eraser. How much is the eraser?

richardchin@clinicaltrialist.com

Examples of Everyday Intellectual Illusions

- You're in a footrace and you pass the person in second place. What place are you in?
- A pencil and an eraser together cost $1.10. The pencil is $1 more than the eraser. How much is the eraser?

- Answers:
  - Second place
  - 5 cents

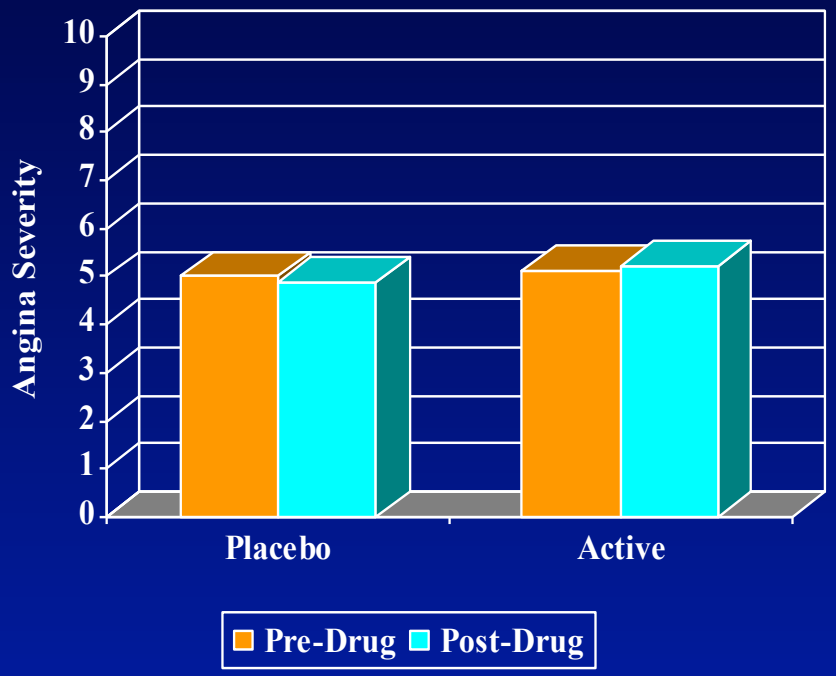
richardchin@clinicaltrialist.com

Intellectual Optical Illusions

- Primate brains are not hardwired to process aggregate data properly
- There is natural tendency to use heuristic processing, which is usually adequate for anecdotal data encountered in everyday life
- But this can lead to wrong conclusions when processing statistical information

Illusion 1: Regression to the Mean

- A promising drug for patients with depression is being developed
- Unfortunately, it fails to meet the primary endpoint (symptom severity) in the Phase 2 study

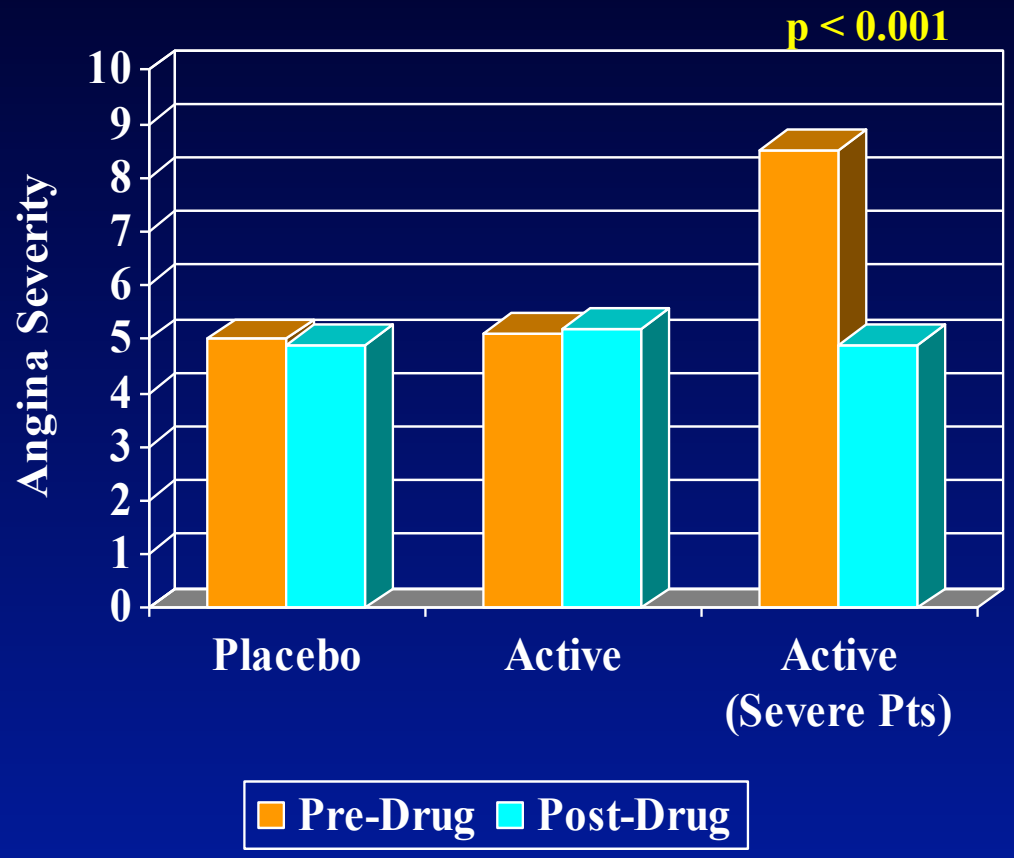


Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Illusion 1: Regression to the Mean

- But, preclinical data suggest that only the more severe patients would benefit because the drug affects receptors that are the most upregulated in severe disease
- So, a subgroup analysis is performed on the 50% most severe patients

In severe patients, the drug shows clear benefit



Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

The drug is advanced into Phase 3 for the more severe patients and it fails

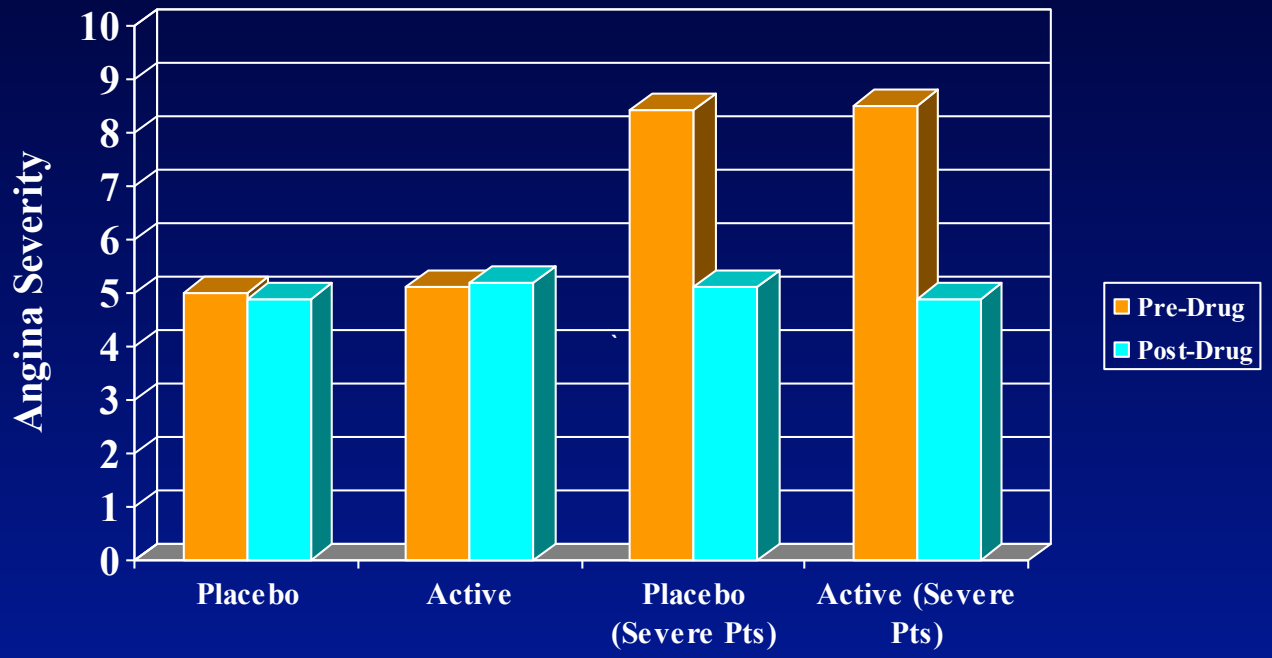
richardchin@clinicaltrialist.com

Illusion Explained

- In a waxing and waning diseases, all patients will have good periods and bad
- Taking only the patients who are having worse than usual days will result in patients appearing to improve on repeat measurement

Illusion Explained

- In the Phase 2 study, even the placebo patients appear to do well if only the severe patients are considered



Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Solution

- Never compare subgroup in one arm against the entire group from the other arm
- Stratification by severity at time of randomization can protect against regression to the mean

Other Common Instance of Regression to the Mean

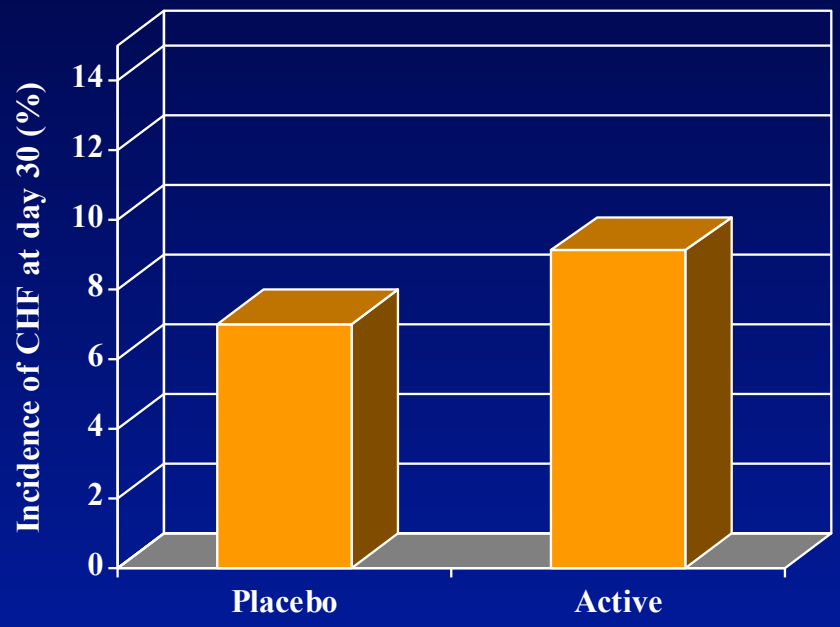
- Often, a Phase 2 study will yield spectacular results, and the subsequent Phase 3 will be less impressive. Given the number of Phase 2 studies that are conducted and given that only a subset proceed into Phase 3, it would be expected that in general, Phase 3 results will be less impressive than Phase 2.
- Often, a trial will fail due to "higher than expected placebo arm response." This is in some cases because patients who were having flares of disease or having a particularly bad day were enrolled.

Illusion 2: Survivor Bias

- A new anticoagulant is being developed for heart attack (MI) patients
- It is expected to decrease damage to the heart, lowering deaths and severity of the MI
- Two most common sequelae of MI are deaths and congestive heart failure (CHF)
- CHF is more common than death, so in order to increase the power of the study, CHF is selected as the primary endpoint

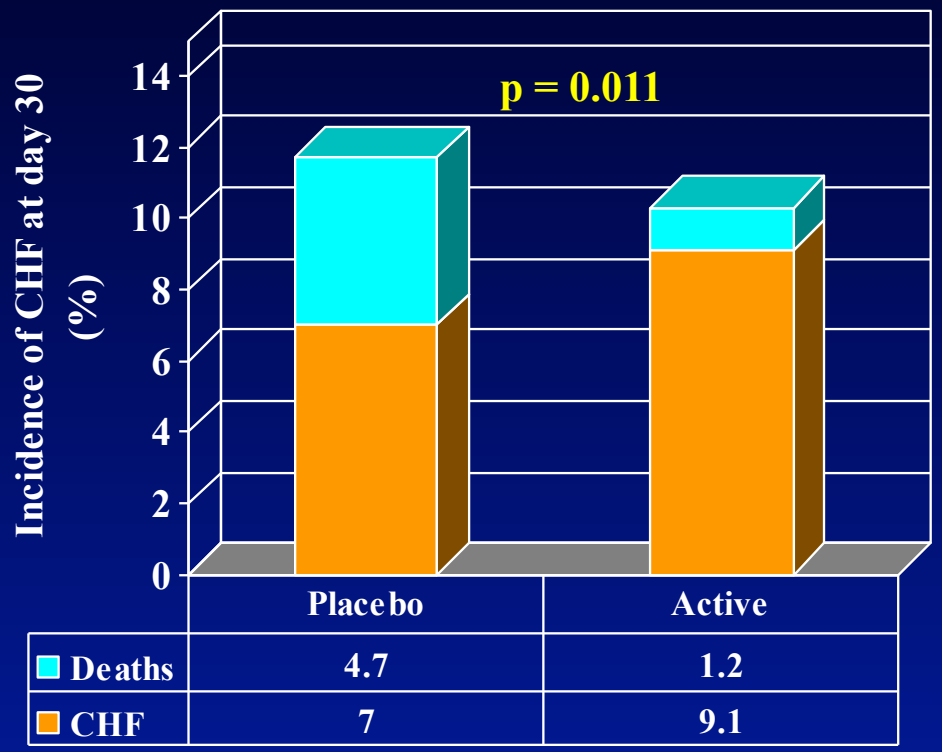
Illusion 2: Survivor Bias

- Contrary to expectation, CHF is increased rather than decreased in the treated group



Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

However, Death + CHF was decreased in the active group

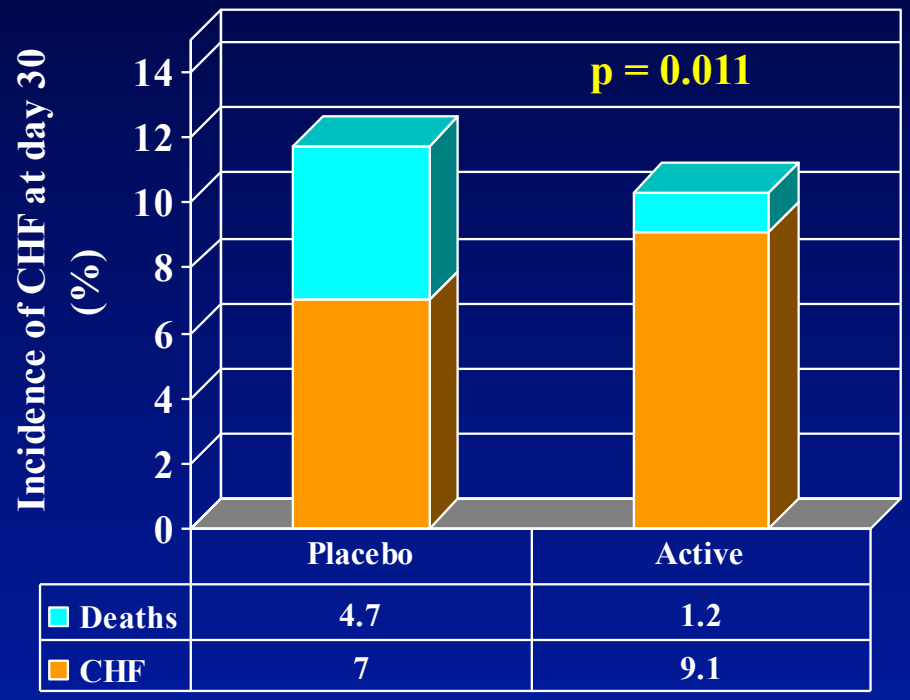


Death + CHF is significantly decreased in the treated group.

Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Illusion Explained

- The drug was very effective in preventing deaths
- The patients who would have died on placebo survived, but with enough damage that they developed CHF



Death + CHF is significantly decreased in the treated group.

Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Solution

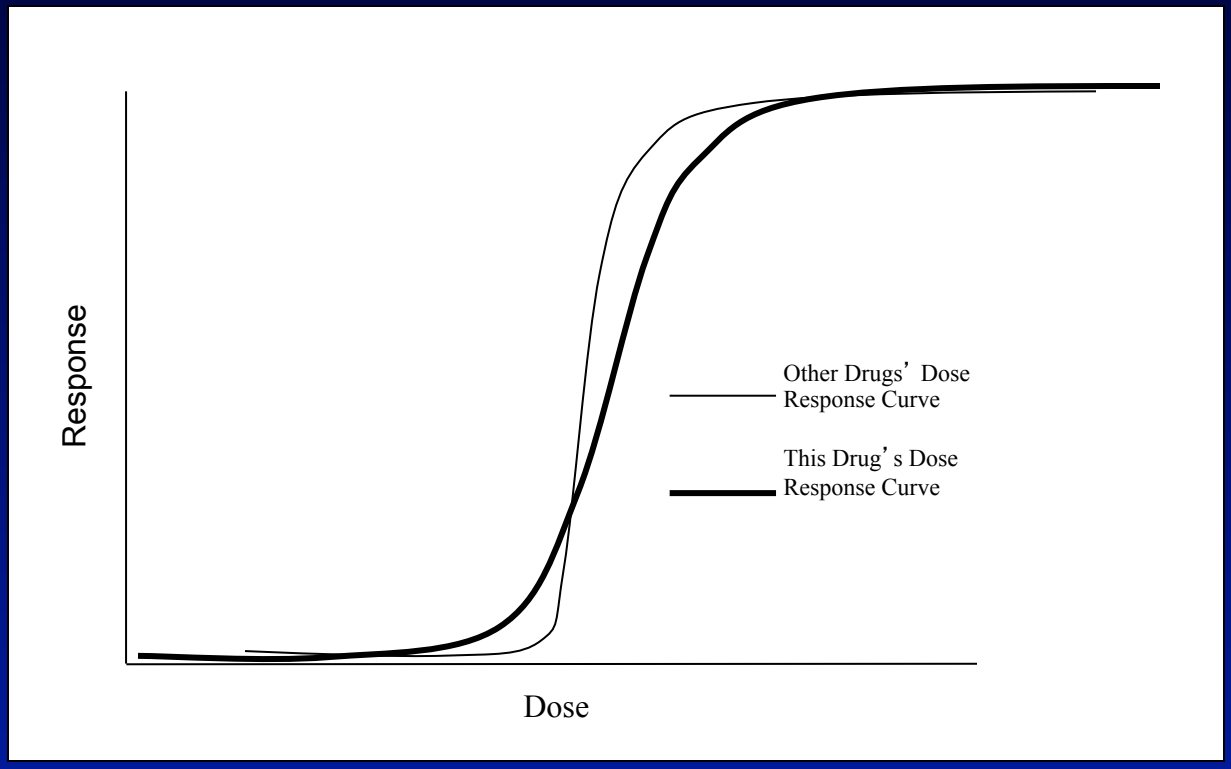
- When selecting endpoints, make sure that all critical endpoints are included
- Consider using composite endpoints instead of single endpoints if power needs to be increased

Illusion 3: Doses in Groups vs. Individuals

- A novel drug is being developed for seizures
- Unlike many other drugs, this drug appears to possibly have an extremely gradual dose-response curve that could lead to a wide therapeutic index
- A phase 2 study is conducted to test this hypothesis

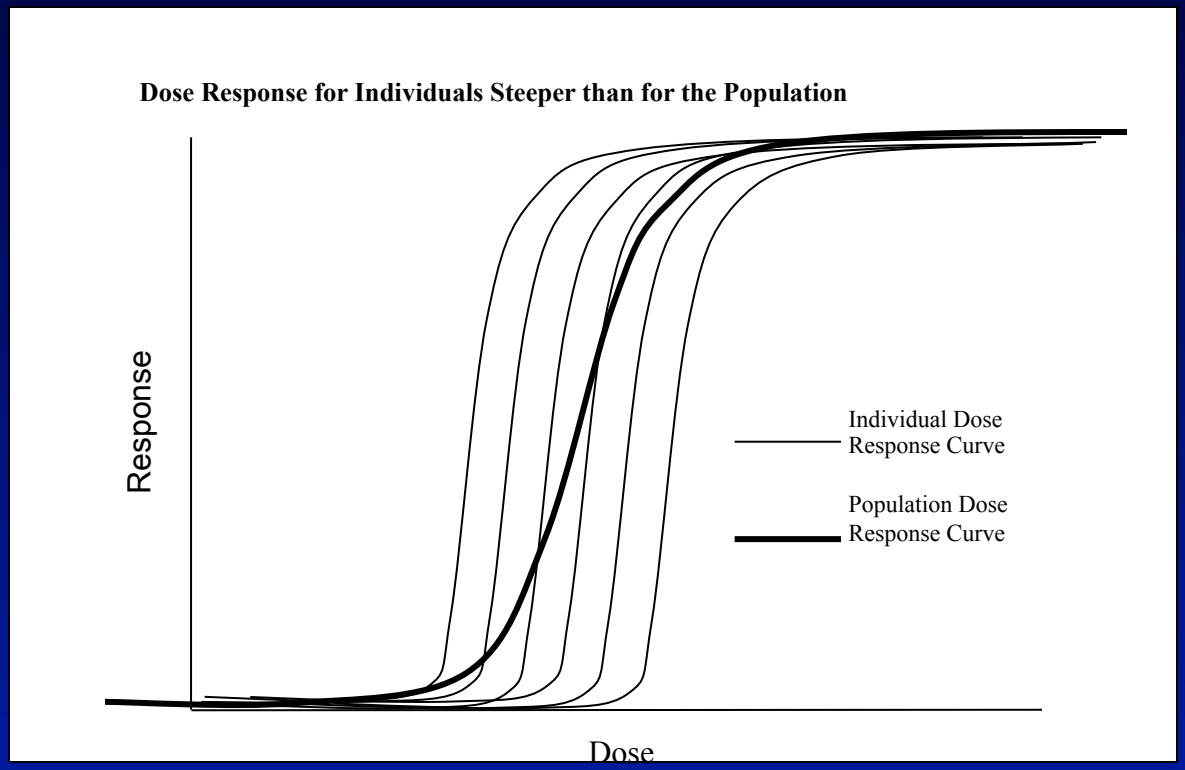
Illusion

- As expected, the dose response curve is very gradual
- This seems to be great news until a pharmacokineticist explains the data



Illusion: Doses in Groups vs. Individuals

- The gradual population dose-response curve does not necessarily translate into gradual individual dose-response curve



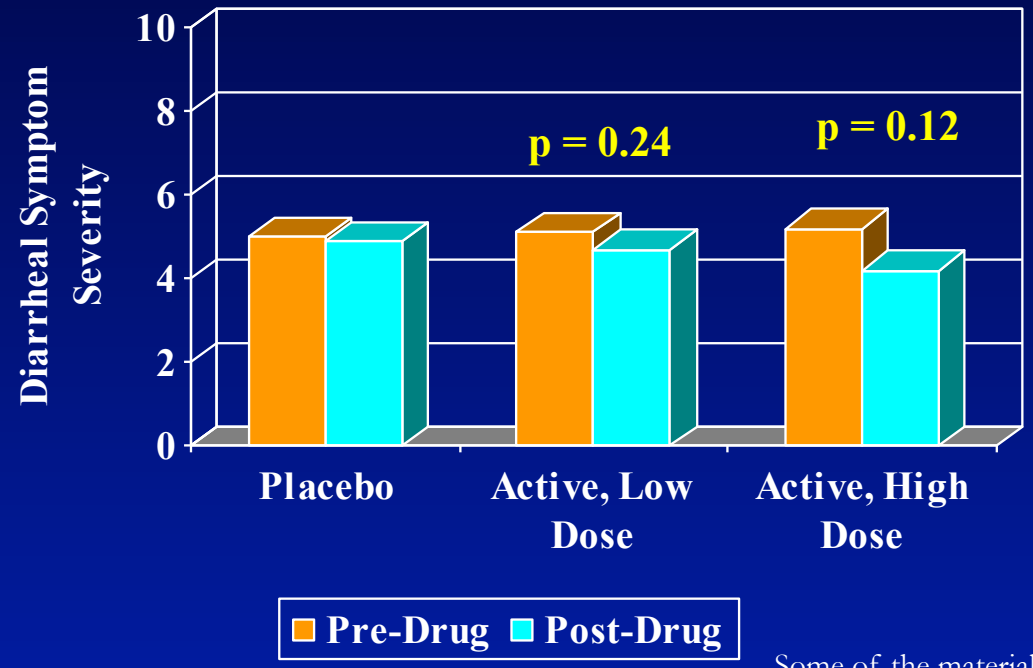
Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Illusion 4: Shifty Subgroups

- Subgroup analysis can often be useful, but unless interpreted correctly, it is a minefield of intellectual optical illusions
- Three most common illusions are:
  - Misuse of dependent variables
  - Completer analysis
  - Responder analysis (Mercedes in Chile phenomenon)

Confounded Dependent Variable

- A promising drug for gastroenteritis is being developed
- The study misses primary endpoint of diarrheal symptom severity, but there does seem to be some effect at the higher dose



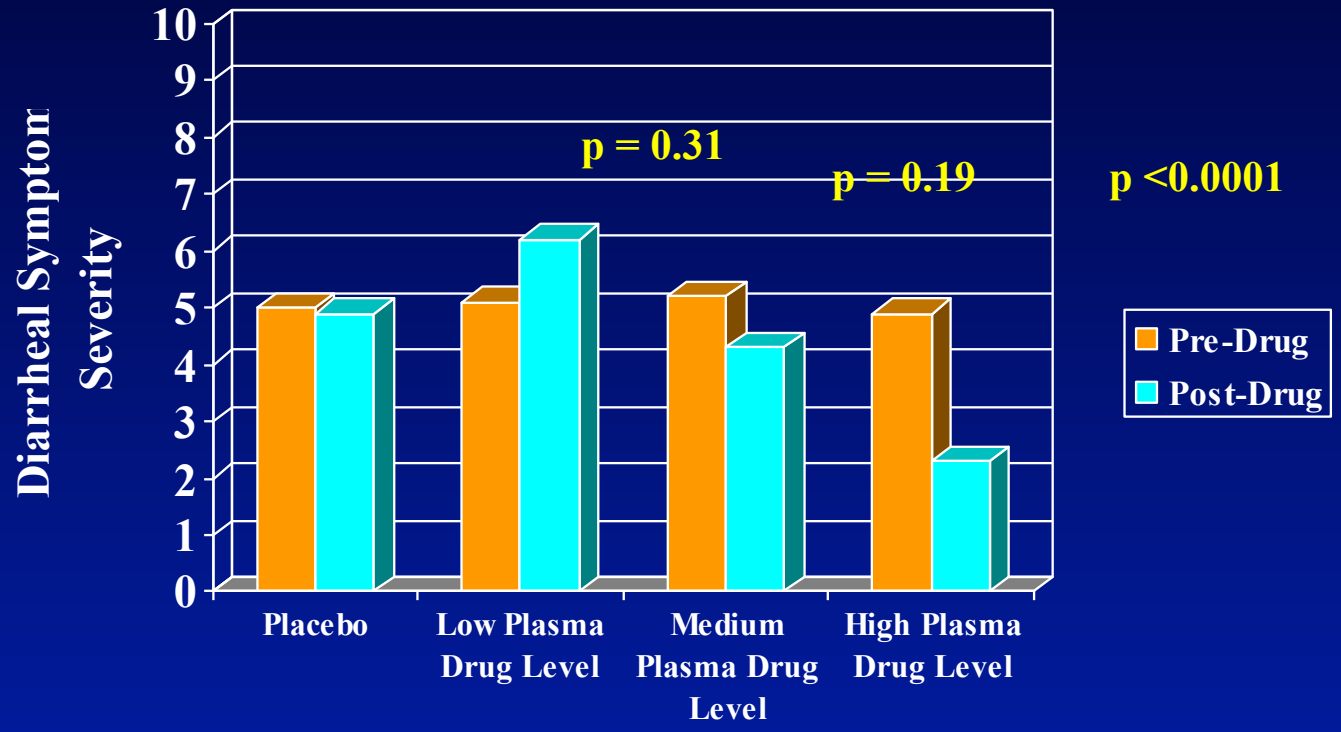
richardchin@clinicaltrialist.com

Confounded Dependent Variable

- One of the formulation experts thinks that the drug may not have been absorbed well in these patients with diarrhea
- If drug is not absorbed, then it can't be expected to work
- So, perhaps patients who absorbed the drug did well

Confounded Dependent Variable

- So a new analysis is performed. Patients are divided into 3 groups based on the drug levels in plasma
- There is a profound decrease in symptom score in patients who absorbed the drug best



The drug is advanced into large Phase 3 study with high dose, and the study is negative

Illusion Explained

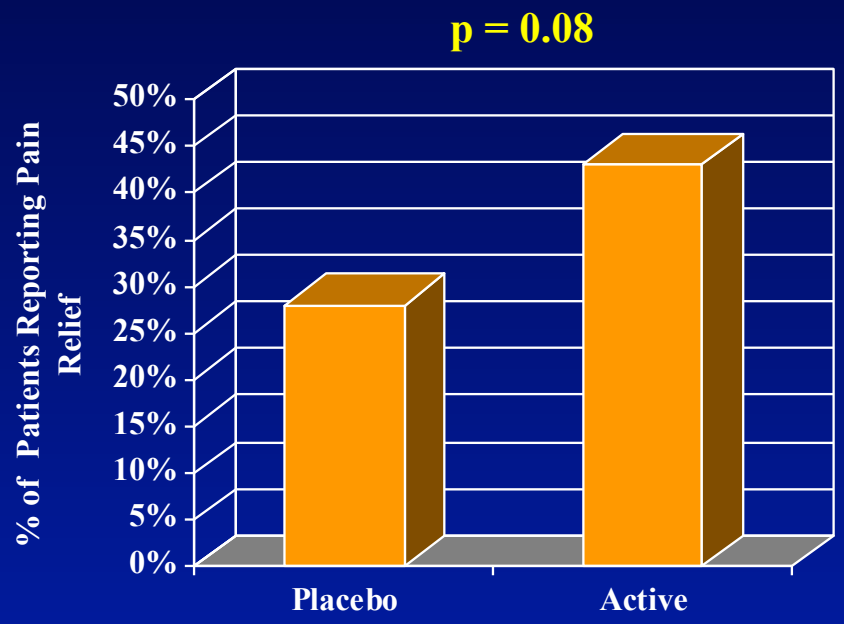
- Patients who were destined to improve had less severe form of gastroenteritis
- As a result, they were able to absorb drug better
- So, they absorbed drug better because they were in a subgroup with better prognosis, not the other way around
- Using any variable that changes during the course of the study (PK, PD, etc.) can lead to erroneous, confounded conclusions

Illusion Variation: Improper Imputation

- A promising new therapy for arthritis is being developed
- It appears to be highly effective in some patients
- Unfortunately, it causes severe itching in some patients that can lead to discontinuation
- Phase 2 is conducted, and as expected, dropouts are significant (about 30%)
- The results are therefore analyzed looking only at patients who completed the study (completers)

Completer Analysis

- The results among those patients who tolerated the drug and completed the study look promising, though statistical significance is not reached



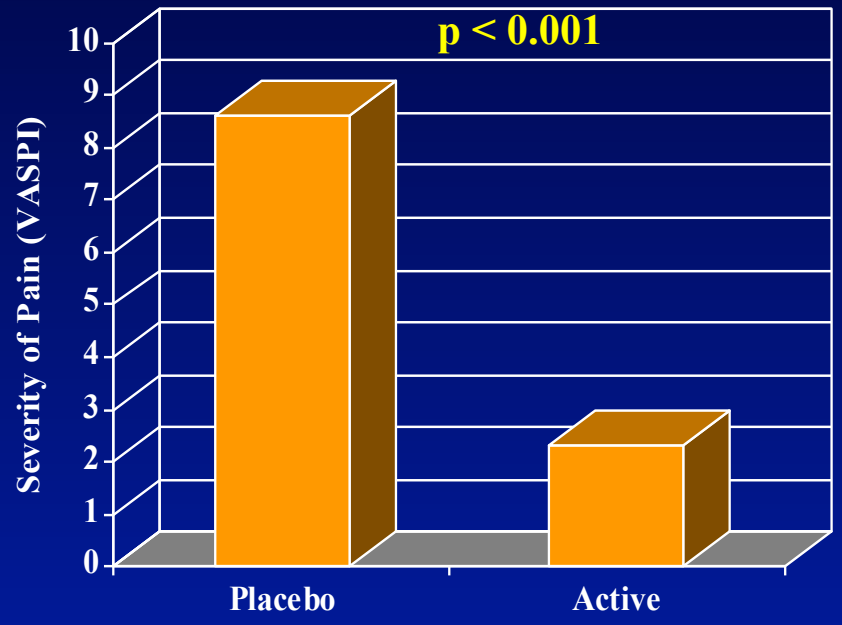
Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Responder Analysis

- Someone on the team recalls that comparing average scores (continuous endpoint) rather than looking just at success/failure (dichotomous endpoint) can increase the power of the study
- Also, the biology of the drug suggests that it will only work in some patients
- Therefore, the average pain score among those who reported pain relief is examined (responder analysis)

Responder Analysis

- In the responder population, the results are overwhelmingly positive
- In other words, it looks like the drug only works in some patients but in those where it works, it works astonishingly well



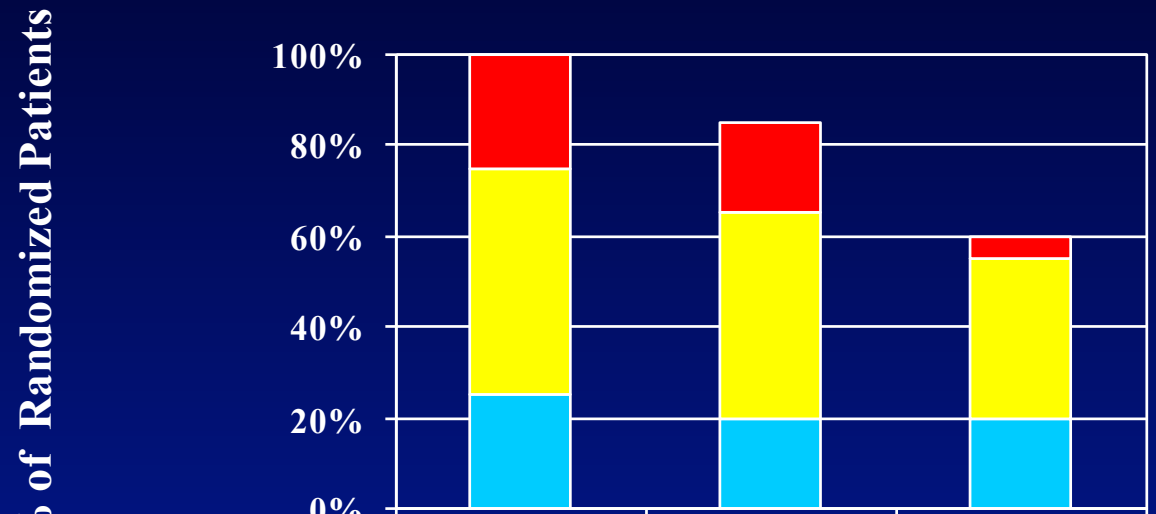
Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Bad Decision

- The drug is advanced into Phase 3
- It fails to show a benefit in Phase 3

Illusion

- The drug was most likely to cause itching in patients with worst arthritis
- Most patients in the active group who had the most severe arthritis dropped out



	All Randomized Patients	Placebo Completers	Active Completers
Severe Patients	25%	20%	5%
Moderate Patients	50%	45%	35%
Mild Patients	25%	20%	20%

Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Completer Illusion

- The numerator is similar between the groups, but denominators change, because more patients drop out in the active group
- The sicker patients are the ones who tend to drop out
- The healthier patients are the ones most likely to improve spontaneously

	Placebo	Active
Responder	27	26
Completer	82	61
Response Rate	33%	43%

Responder Illusion

- Among the responders in the placebo, some patients are from the severe group, who had higher pain scores to start with
- Among the responders in the active group, almost none are in the severe group because they nearly all dropped out
- The pain scores from the severe group skews the mean score in the placebo group

Mercedes in Chile

- Using only responders to gauge the efficacy of a drug is like trying to determine whether Chile or the U.S. is richer by using average income of Mercedes owners in each country
- Very small proportion of the people in Chile own a Mercedes but their average income is higher than Mercedes owners in the U.S.
- That doesn't mean that the average Chilean is richer than an average American
- Nor does it mean that the GNP of Chile is higher than U.S.'s

Solution

- Be very cautious with subgroup analysis
- Use intent-to-treat analysis whenever possible
- If subgroup analysis is important
  - use baseline variables, not variables that can change during the course of the study
  - stratify at randomization by the baseline variable
- Alternatively, titrate the drug to the dependent variable
  - For example, rather than a study with placebo/low dose/high dose, conduct a study with placebo/low target plasma level/ high plasma level

Illusion 5: Puzzling Proportions

- A company claims that its new drug reduces mortality by 90%
- Another drug is supposed to reduce mortality by 9%
- Yet another claims to reduce mortality from 10% to 1%
- Another company says that survival is increased from 90% to 99%.
- All of these claims are equivalent

Illusion

- The difference between 1% and 10% is the same as the difference between 90% and 99%
  - 1% mortality = 99% survival
  - 10% mortality = 90% survival
- The difference between 0% and 90% is comparable to the difference between 90% and 99%
  - 10% to 100% is tenfold difference
  - 1% to 10% is tenfold difference

Definitions

- Absolute difference: mathematical difference when one subtract one number from another
  - 100% - 90% = 10%
- Relative difference: absolute difference in relation to the baseline value
  - 25% - 20% = 5% -> 5%/25% = 20%
- Odds ratio: relative odds
  - 25%/75% ÷ 20%/80% = 1.33

Illusion 6: Post-hoc Analysis

"If you torture the data long enough, it will confess to anything"

- A Phase 3 trial of a drug for pulmonary fibrosis has completed
- It has missed the primary endpoint of walk distance
- But on a post-hoc exploratory analysis, a subgroup of patients with longstanding disease has demonstrated convincing improvement in mortality, from 10% to 2%, with $p < 0.001$
- Another phase 3 study is conducted, with all cause mortality as the endpoint
- The trial fails

Post-hoc Analysis

- Given enough time and enough analysis, a compelling subgroup and/or endpoint can be found for virtually any study
- In general, these findings are spurious and almost never repeatable.

Illusion 7: Surrogates

"I am not a doctor but I play one on TV."

- Many patients die of irregular heart rhythm (arrhythmias) after a heart attack (MI)
- The patients with the greatest number of premature ventricular contractions (PVC's) are clearly the highest risk of death
- Several drugs were developed to prevent PVC's, and they were believed to be effective in preventing life-threatening arrhythmias.

CAST

- CAST was a study of antiarrhythmic drugs designed to demonstrate that the drugs lowered mortality
- It was initiated amidst controversy, because many physicians thought it was unethical to randomize patients to placebo
- Unfortunately, the drugs did not prevent the arrhythmias
- They rather caused a proarrhythmic side effect that led to deaths
- The drugs increased mortality and CAST was terminated early by the DSMB

TNF Inhibitors

- There is overwhelming data proving that in congestive heart failure (CHF) patients, higher the level of TNF, more likely they are to die
- Because of this a large study was conducted to reduce mortality by administering TNF inhibitor
- The study showed that blocking TNF increased mortality

Illusion Explained

- Biomarkers that are correlated with a disease can sometimes be
  - a good surrogate if they are in the causal pathway
  - a good drug target if they are in the causal pathway between the drug action and clinical outcome
- However, they are more often
  - An epiphenomena – not in the causal pathway
  - Caused by the disease, rather than causing the disease
  - A protective counter-regulatory response to the disease
- A surrogate can be helpful if used correctly, but they must be validated first

Illusion 8: Statistical Flukes

- A new drug is being developed for wound healing
- The Phase 2 results are convincing, with 30% wound healing in placebo and 63% in active group (p = 0.002)
- However, one of the sites displays a worrisome effect. There, there were 60% wound healing in placebo and only 25% in active.
- An investigation is launched to determine what happened at the site, whether the randomization codes were mixed up, etc.

Illusion Explained

- With enough sites, one or more sites will show reversal of effect, just by chance
- The table below shows the probability of at least one site showing reversal of effect (alpha of 0.05, 80% power)

Probability of At Least One Center Showing Treatment Reversal								
Number of Centers	1	2	3	4	5	6	7	8
Probability of Treatment Reversal	.003	.05	.15	.29	.43	.56	.67	.75

Illusion Part 2

- The drug is taken into Phase 3
- Just to be sure, 4 identical Phase 3 studies are conducted rather than just 2. These studies have the exact same design as Phase 2.
- Despite being identical, 3 of the studies meet the primary endpoint, while the fourth one misses it with p=0.13
- Though pleased with the positive results, many people are puzzled why the fourth one was an anomaly.

Illusion Explained

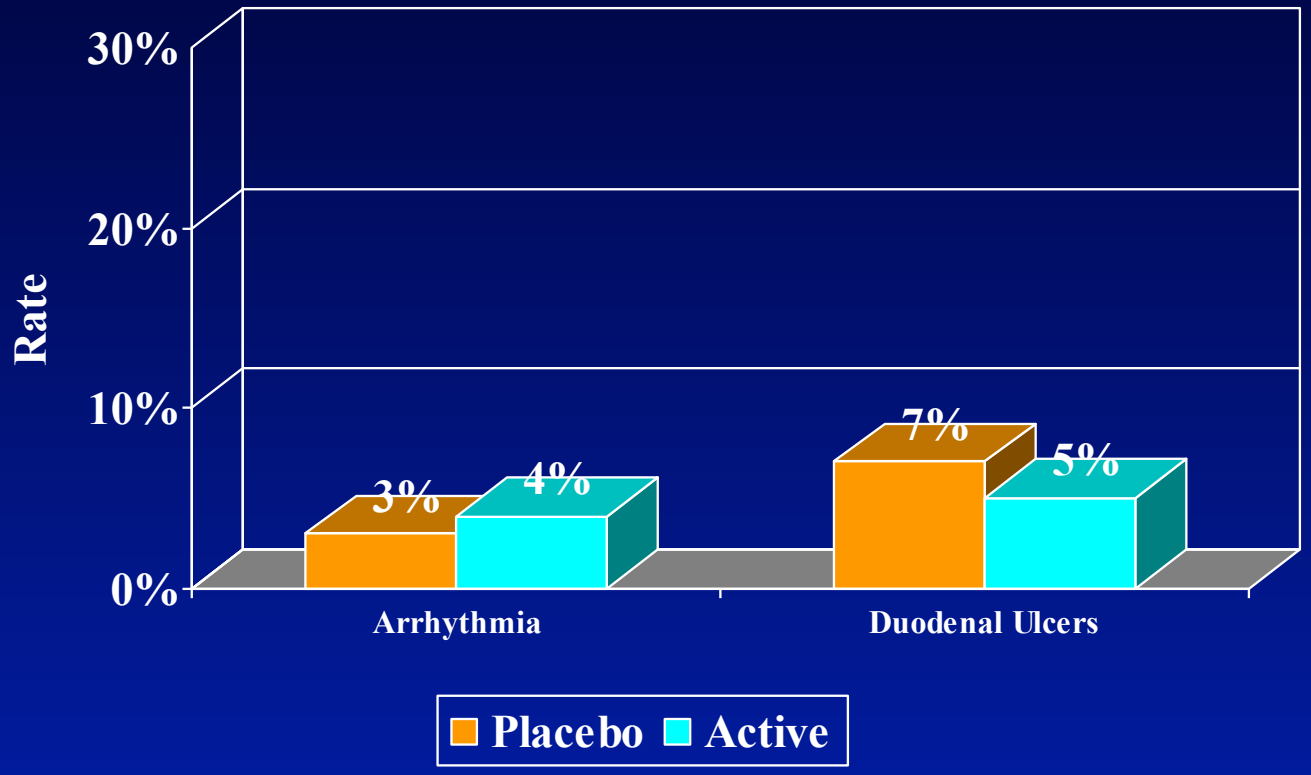
- Even with 90% power, and even with identical study designs, the results are not likely to be exactly the same
- The likelihood of 5 out of 5 studies meeting the primary endpoint is 0.9 x 0.9 x 0.9 x 0.9 x 0.9 = 0.59
- With 80% percent power, the likelihood is 0.8 x 0.8 x 0.8 x 0.8 x0.8 = 0.33
- Even with an active drug, if enough studies are conducted, one or more will eventually fail to show an effect

Illusion 9: Safety Shuffle

- A promising drug for refractory seizures is being developed
- Unfortunately, it appears to have two drawbacks
  - Potential to cause arrhythmias
  - Potential to cause duodenal ulcers
- Therefore, the Phase III safety data is carefully examined to assess potential signal in those two categories

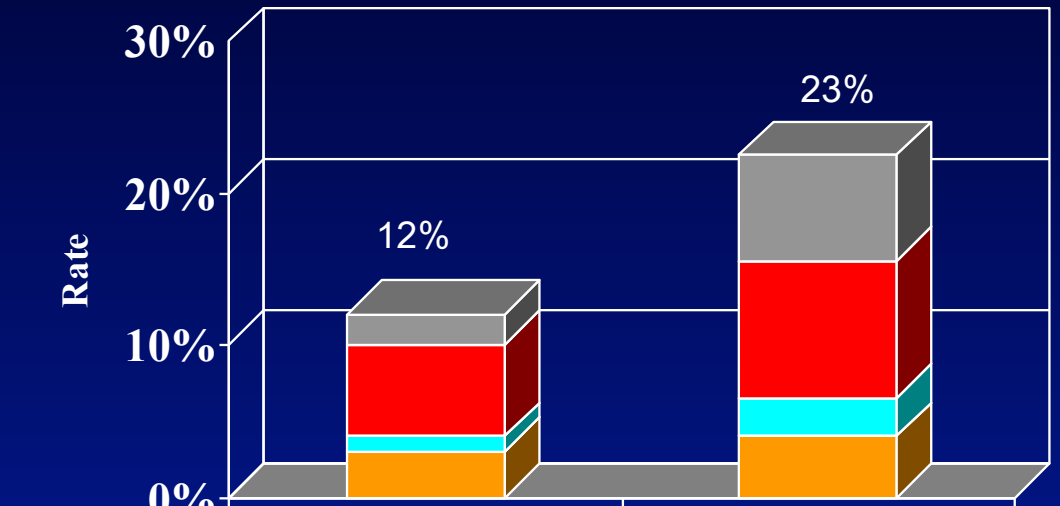
Initial reading

- Fortunately, neither concern seems to have been justified, as no signal is apparent



Illusion: Lumping

- Unfortunately, if the net for arrhythmias is cast more widely to include not just the term, “arrhythmia” but also “sudden death,” “palpitations,” “syncope,” (which can be other presentations of arrhythmia) then a signal becomes apparent

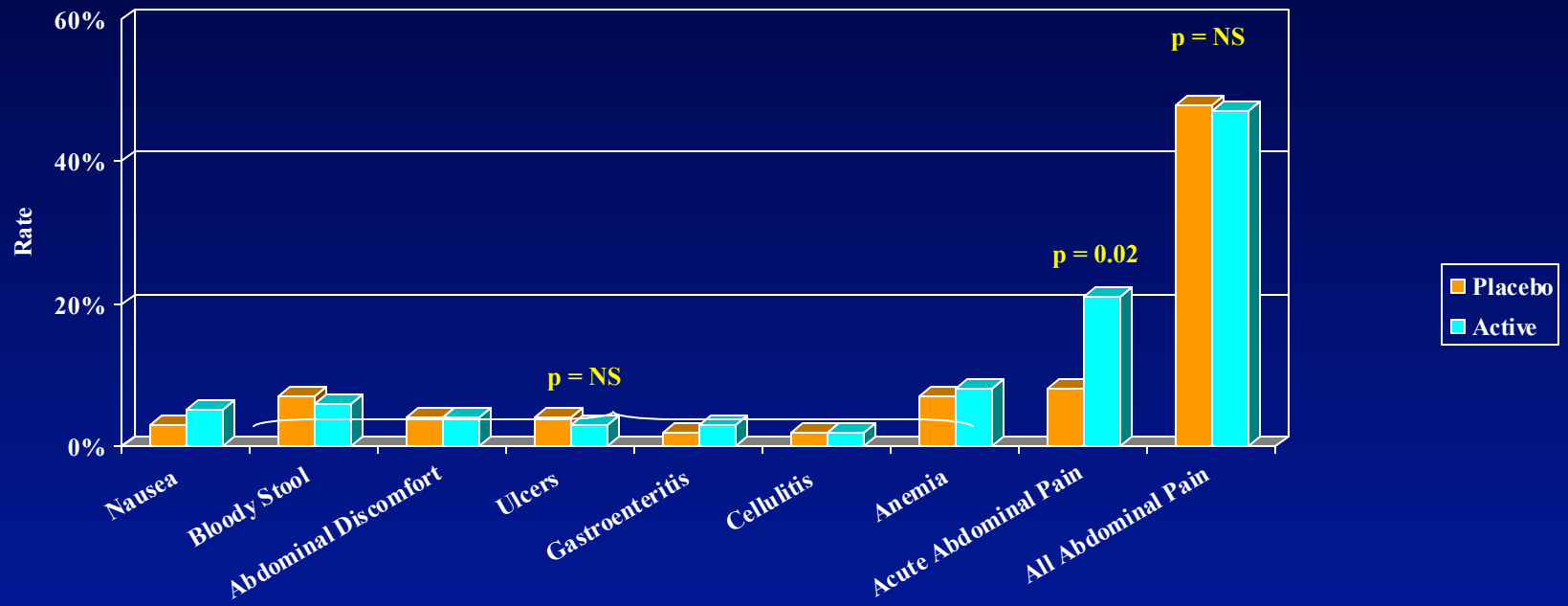


	Placebo	Active
Syncope	2%	7%
Palpitations	6%	9%
Sudden Death	1%	3%
Arrhythmia	3%	4%

Some of the material adapted with permission from
Chin, R. Principles and Practice of Clinical Trial Medicine
richardchin@clinicaltrialist.com

Lumping and Splitting

- For ulcers, no signal is apparent if the data is split too much or lumped too much
- But if we zoom to the right level, a clear signal comes into focus



Illusion 10: Playing with p Values

- Two companies are neck and neck in racing to develop a therapy for a congenital storage disease
- Their drugs are similar but when the two companies announce their Phase II results nearly simultaneously, the results appear quite different
  - The first company announces a successful study with a p=0.0001
  - The second announces a successful study with p=0.04
- What happened? Did the second company make a mistake in their trial design? Is the first company's drug more likely to work?

Illusion

- No, the drugs worked similarly, and to a similar magnitude in the two studies
- The first company's studies had 300 patients, the second had 80
- P values reflect 2 things: how well the drug works and how large the sample size is
- Impressive p values don't necessarily mean that a drug works better, if the studies are not equivalent in design and size

Other Things to Remember about p Values

- The bar is different for efficacy and safety
  - Bar for efficacy is $p<0.05$
  - Bar for safety is far less. Even a safety signal that does not come close to $p=0.05$ must be taken seriously.
- When looking at multiple endpoint, there must be adjustments for multiple comparisons.
- Any p values derived from post hoc analysis is nominal. It represents what would have been the p value, but it is not a real p value.

richardchin@clinicaltrialist.com