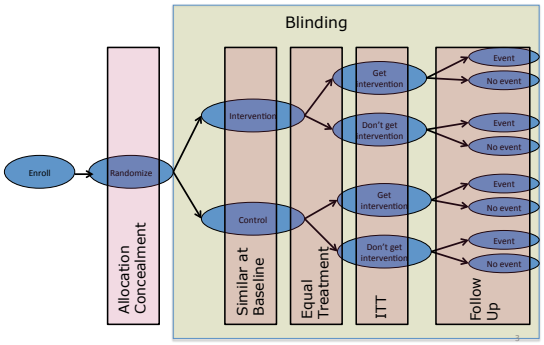
Beyond the Basics: A Few More Facts on Critical Appraisal, Stats and other Highly Confusing Concepts

Elizabeth Allen, MD
8-26-14

- 75 % of the research literature involves therapeutic intervention
- Really understanding the nuances of appraisal and interpretation of the RCT is key

Anatomy of an RCT



Some things that confuse or mislead…

- Choice of suboptimal control
- Choice of outcome - surrogate, composite
- Choice of measure of association - relative vs. absolute
- Clinical vs. statistical significance
- P-values, confidence intervals and all that jazz
- Special case: Negative studies – what about power?

Choice of Control

- Examples of suboptimal controls:
  - Placebo, if active therapy is standard
  - Standard therapy in lower doses, dosing intervals, or potencies
  - Sub-standard therapies
  - Placebo isn't really a placebo

Choice of Outcome

- Ideal: The most relevant clinical endpoint
- Watch for **surrogates (substitute) outcomes**
  - **Suboptimal surrogate: no clear correlation with relevant clinical outcome**
    - **#premature ventricular contractions on ECG as a measure of anti-arrhythmic medication effectiveness**
    - **thickness of plaque in carotid arteries as an outcome of cholesterol lowering therapy**
- Watch for **composite outcomes**

Vascular Medicine

Oral Buflomedil in the Prevention of Cardiovascular Events in Patients With Peripheral Arterial Obstructive Disease

A Randomized, Placebo-Controlled, 4-Year Study

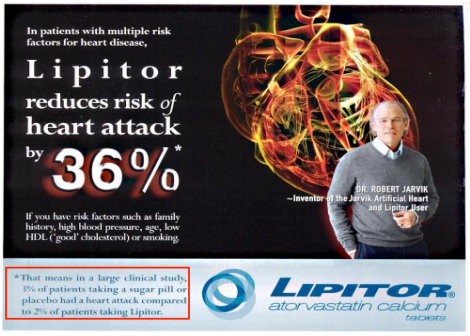
Limbs International Medicinal Buflomedil (LIMB) Study Group

Background—Cardiovascular-related morbidity and mortality in patients with peripheral arterial obstructive disease remain high. We performed an international, multicenter, randomized, double-blind, placebo-controlled trial to investigate whether long-term administration of oral buflomedil could reduce the rate of cardiovascular events in patients with intermittent claudication.

Methods and Results—Patients >40 years of age with documented peripheral arterial obstructive disease, intermittent claudication, and an ankle-brachial index between 0.30 and 0.80 were eligible for inclusion and were randomized to receive orally either buflomedil or placebo for 2 to 4 years. Aspirin was recommended for all patients (unless they were receiving other antithrombotic treatments at inclusion). The primary efficacy outcome was critical cardiovascular events, defined as the composite of cardiovascular death, nonfatal myocardial infarction, nonfatal stroke, symptomatic deterioration of peripheral arterial obstructive disease, or leg amputation. A total of 2078 patients were recruited. Mean treatment duration was 33 months. The rate of critical cardiovascular events was significantly lower in buflomedil-randomized patients than in placebo-randomized patients (9.1% versus 12.4%; hazard ratio, 0.742; 95% confidence interval, 0.603 to 0.915; $P=0.0163$). Ankle-brachial index increased by 9.2% in buflomedil-randomized patients and decreased by 3.6% in placebo-randomized patients ($P<0.001$). Tolerance of buflomedil and placebo was comparable.

Conclusions—Compared with placebo, buflomedil administered for 3 years reduced the occurrence of symptomatic cardiovascular events by 26%. The main contributor to the difference in the composite outcome was the reduction in symptomatic deterioration of peripheral arterial disease. The use of buflomedil should be considered in addition to an antiplatelet agent in patients with peripheral arterial obstructive disease and intermittent claudication. (***Circulation*. 2008;117:816-822.**)

Relative vs. Absolute Risk

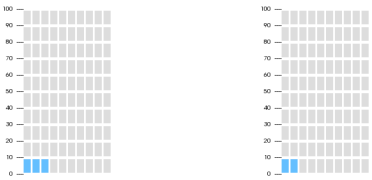


Relative vs. Absolute Risk

- Treatment is associated with MI event rate of **2%,** Placebo group event rate is **3%**
- Absolute Risk Reduction = Difference between experimental and control event rates
  = EER – CER = (3%-2%) = 1%
- Relative Risk = EER/CER = 2%/3% = 0.67
- Relative Risk Reduction = (1 – RR)100 = 33%
- **Absolute Risk Reduction keeps in context baseline rate of events in control group.**
- **When baseline event rate is low, ARR and RRR will diverge**

ARR Demonstrated in an Icon Array

MI Rate on Placebo **MI Rate on Lipitor**



Icon array courtesy of Picto-Generator at http://cbssm.org/reasearch_tools/pictographs

Statistical vs. Clinical Significance

- **Is the clinical benefit of a magnitude you care about?**
  - **Treatment A decreases mortality by 1% compared with placebo**
  - **P value for this difference is 0.05**
  - **Do you care?**

1.2 million Americans will have an MI this year. Potentially, 12,000 less people would have an event.

What does the p-value actually mean?

When doing a study:

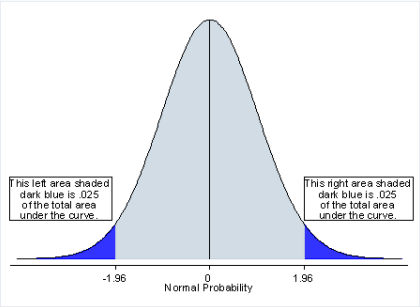
- No way to be sure that the difference you are seeing is real or just a coincidence of random sampling
  - Maybe, the 1% ARR is a coincidence
- We use statistical calculations to help determine the probability of seeing an ARR this size if the treatment and control groups are really the same
  → This is the p-value
- A p-value can range from 0 – 1
- If the p-value is small, you conclude that the risk difference you are seeing is unlikely to be a coincidence.

- If the ARR for MI is 1% and the P value is 0.05
  - There is a 5% chance of observing a difference as large as we did even if treatment A is the same as placebo
    (i.e. 5% chance the difference you observed is a coincidence)
  - There is a 95% chance that if we compared MI rates in groups treated identically (i.e. both groups received same active treatment) we would show a risk difference that was smaller

One Tail vs. Two-Tail p Values

- P values presented in the medical literature are almost always <u>two-tailed</u>
- Start with the **null** hypothesis:
  - the rate of MI for patients receiving Treatment A is the same for pts receiving Control
- A two-tailed test allows us to reject the null hypothesis if we demonstrate that:
  - Treatment A is better than Control
  - Control is better than Treatment A

Normal Data Distribution



What does the Confidence Interval Mean?

- The mean result you calculate from a sample of patients is not exactly equal to the actual population mean
  - Small, variable samples of patients provide mean results that may be far from the actual population mean
  - Larger samples give a mean result that is likely closer to the population mean
  - Statistical methods combining sample size and sample variability (standard deviation) generate confidence intervals.
  - The 95% confidence interval gives you a range of results within which the population mean falls.
    - you can be 95% sure that the actual population mean falls somewhere within confidence interval

Wrapping this discussion up

- So if the ARR is 1% and the P-value is 0.05 and the 95% confidence interval is 0.05% - 4.8%
  - You are 1% less likely to have an MI if you take treatment A compared with placebo
  - There is a 5% chance the risk difference observed is purely coincidental and that treatment A is the same as placebo and…
  - You can be 95% sure that the real ARR for treatment A compared with placebo falls somewhere between 0.05% and 4.8%
  - If the p value is <0.05, you can be sure that the 95% confidence interval won't cross zero.

What is Power?

		Reality	
		Treatment Works	Treatment Doesn't Work
Study Result	Treatment Works	A	B
	Treatment Doesn't Work	C	D

A =	Study shows treatment works and it really does. A true positive
B =	Study shows treatment works and in reality it doesn't. False positive. Type I or alpha error, → the P value by convention set at 0.05

What is Power?

		Reality	
		Treatment Works	Treatment Doesn't Work
Study Result	Treatment Works	A	B
	Treatment Doesn't Work	C	D

C = Study shows treatment doesn't work and it really does. A false negative. **Type II or Beta error**

D = Study shows treatment doesn't work and in reality it doesn't. True negative.

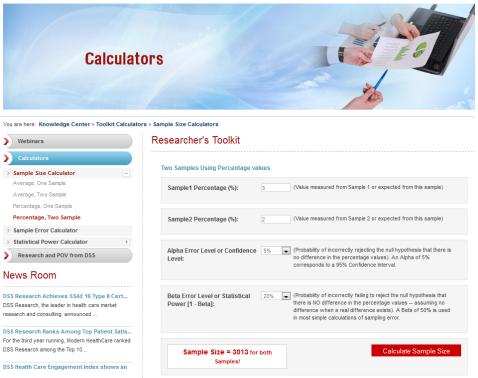
Power

Situation C: Study shows treatment doesn't work and it really does. Type II or Beta error

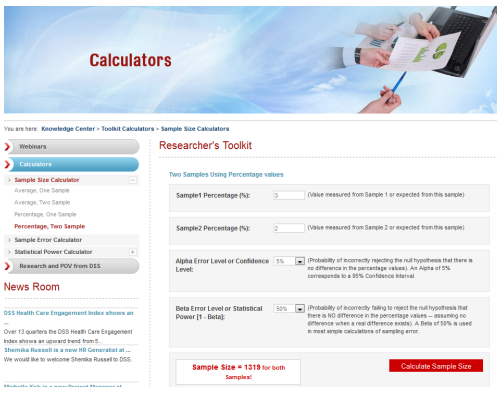
- Power: the ability of a study to show a difference between treatment group if one exists
- Beta error is often set at 80%:
  - 80% power to detect a difference of a pre-determined size assuming there is one
- Greater statistical power requires larger sample sizes
- The smaller the difference you want to be able to detect, the larger the sample size needed

Lets design a study

- Pretend you are designing a study comparing Treatment A (a statin) to Placebo
- You want to achieve type I (alpha) error level of 0.05 and type II (beta) error of 80%
- You believe the rate of events in your placebo group will be 3% (baseline event rate)
- Previous research suggests Treatment A will decrease event rate to 2% (ARR 1%)
- How many people do you need in each randomization group?



http://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/samplesizecalculators.aspx



EvidenceUpdates - Windows Internet Explorer

http://plus.mcmaster.ca/evidenceupdates/Default.aspx

File Edit View Favorites Tools Help

EvidenceUpdates

Search all BMJ Products Search

BMJ BMJ Journals BMJ Careers BMJ Learning BMJ Evidence Centre BMJ Gr

EvidenceUPDATES

FROM THE BMJ EVIDENCE CENTRE

Home | My Profile | My Alerts | Search | Tools | Help | Log Out

Home | About This Site | About BMJ Group

BMJ Group and McMaster University's Health Information Research Unit are collaborating to provide you with access to current best evidence from research, tailored to your own health care interests, to support evidence-based clinical decisions.

This service is unique: all citations (from over 120 premier clinical journals) are pre-rated for quality by research staff, then rated for clinical relevance and interest by at least 3 members of a worldwide panel of practicing physicians. Here's what we offer:

- A searchable database of the best evidence from the medical literature
- An email alerting system
- Links to selected evidence-based resources

Hit Parade: The most often read articles in all disciplines, in the past 30 days

1. Duloxetine for the management of pain in older adults with knee osteoarthritis: randomised placebo-controlled trial.
   Age Ageing *(Original)*
2. Basal Insulin and Cardiovascular and Other Outcomes in Dysglycemia.
   N Engl J Med *(Original)*
3. Colloid solutions for fluid resuscitation.
   Cochrane Database Syst Rev *(Review)*

My Hit Parade:

The most often read articles **in your discipline(s)**, in the past 30 days.

- Duloxetine for the management of pain in older adults with knee osteoarthritis: randomised placebo-controlled trial.
  Age Ageing. 2012 Jun 27. (Original)
- Basal Insulin and Cardiovascular and Other Outcomes in Dysglycemia.
  N Engl J Med. 2012 Jun 11. (Original)
- Colloid solutions for fluid resuscitation.
  Cochrane Database Syst Rev. 2012 Jun 13;6:CD001319. (Review)
- Screening for cognitive impairment: comparing the performance of four instruments in primary care.
  J Am Geriatr Soc. 2012 Jun;60(6):1027-36. doi: 10.1111/j.1532-5415.2012.03967.x. Epub 2012 May 30. (Original)
- Association of aspirin use with major bleeding in patients with and without diabetes.
  JAMA. 2012 Jun 6;307(21):2286-94. (Original)

update details

Signed in as...
Elizabeth Allen

Log out

Play video

BMJ Masterclas

Quick Search:

Use my clinical discip

Search

Advanced Search