

Critical Appraisal of Drug Studies^{6,7}

A) Is the study valid?

- Were patients **randomized** to treatment (tx) groups & was **allocation concealed (AC)**. (Without concealment, 37% bias in favor of tx. Sealed, opaque envelopes or central registry used to attain AC)^{8,9}
- Was everyone (patients, physicians, investigators, assessors) **blinded** to tx? (Especially important for assessors of subjective outcomes.)
- Was the study **controlled**? (e.g. inclusion of placebo or active control group/arm; in an "N of 1" trial, patient is own control.)
- Were treatment & control **groups similar** in prognostic factors for outcome of interest at beginning of study? If not, were adjustments made?
- Were **all patients accounted for** at end? (Missing patients addressed?)
- Was data analyzed based on groups patients were initially randomized to? (Intent to treat or **ITT**: protects integrity of prognostic randomization; per protocol (PP) analysis may also be of interest (e.g. non-inferiority trials))
- Were patient **groups treated similarly** except for study intervention?
- How was the study **funded** (role of funder)? Was study stopped early?
- Was active comparator drug & dose a good choice?

B) What are the study results?

- What was the primary (1^o) endpoint? What were the secondary (2^o) endpoints? Were endpoints & subgroups pre-specified?¹⁰ *Avoid data mining!*
- What was the difference between treatments? (Harm vs Benefits)
- Were the differences **statistically significant**? **Clinically significant**? (What were the 95% confidence intervals (CIs) or p values? Does the CI cross line of no effect?)
- What are the **absolute** and **relative** risk reductions or increases?
- What is the number needed to treat (NNT) or harm (NNH)?

C) Does this study matter to my patients?

- How clinically relevant/important are the outcomes?
- Were the patients similar to those in my practice? (Consider **inclusion & exclusion criteria**: very sick, old, young, drug interactions & complicated/comorbidity patients often excluded.)
- Do treatment benefits outweigh the risks, costs & impact on life?

Types of Studies (from low to high level of evidence)¹¹

- Case-control study**: a retrospective observational study which selects patients with the outcome of interest (cases) & patients without that outcome (controls); attempts to find features linked to the outcome.
- Cohort study**: an observational study in which two groups (cohorts) are observed over time for an outcome of interest. One cohort has exposure to a condition or treatment that the other does not. (Observational studies: Association does not prove causation! Strength of association: RR: 1.01-1.5 *weak*; 1.51-3 *moderate*; >3 *strong*.¹²)
- Crossover study**: a design in which each patient receives both treatments in two phases separated by a washout period. Each patient serves as own control, thus less variability in outcomes, & smaller sample size required; period effects may limit findings.
- Randomized controlled trial (RCT)**: a prospective study in which patients are randomized to treatment or control groups (equal chance at being assigned to any group). Groups are followed for outcome of interest.
- Systematic Review (SR)**: a systematic collection, review & presentation of available studies addressing a clinical question using specific criteria & methods; may include meta-analysis. e.g. *Cochrane*¹³/*Campbell*¹⁴ *Reviews* (Meta-analysis: the combining of studies meeting prespecified criteria & addressing a clinical question. Results are calculated for data from each study. Data is then pooled. ↑ sample size & statistical power useful if individual trials underpowered or subgroup analysis.)

[Level of evidence: SR > RCT > observational study > expert opinion.¹⁵ Caution: lots of low quality RCTs may not be better than 1 good quality RCT!]

Terms: Related To Validity

- Bias**: design flaws leading to over/underestimation of treatment effect e.g. recall bias, selection bias, publication bias; confounding factors esp observational studies
- Blinding**: if investigators & patient unaware of who receives tx vs control, they are less likely to inappropriately report better results with tx.

Study Results: Size Of The Treatment Effect^{16,17,18,19}

- Event rate (ER)**: the number of people experiencing the event as a proportion of total number of people in the population or group
 - Experimental ER (**EER**): (# events in experimental group / total in exp. group)
 - Control group ER (**CER**): (# events in control group / total in control group)
- Relative risk (RR)** or risk ratio: {EER/CER}
- Relative risk reduction (RRR)**: the RR subtracted from 1 {RRR=1-RR} [Whereas ARR varies with type of population treated, RRR is often more constant.]
- Absolute risk reduction (ARR)**: the arithmetic difference between the 2 event rates {CER - EER} [If ↑ risk: **ARI**= absolute risk increase]
- Number needed to treat (NNT)**: the number of patients who would have to be treated with the studied intervention for the studied time period for 1 of them to benefit. {NNT= 100 / ARR%}
- Number needed to harm (NNH)**: number of patients who would have to be treated with the studied intervention for the studied time period for 1 of them to experience an adverse event. {NNH = 100 / ARI%}
- Odds ratio (OR)**: = experimental event odds / control event odds; especially used in case-control studies where baseline risk is not known; also used in metaanalysis. When events are rare, the OR is similar to the RR; however, OR rate exaggerated relative to RR when events more common. {Link www.cebm.net: tool for converting OR to NNT²⁰}
- Point estimate**: the trial result used as best estimate of the true effect
- Hazard ratio (HR)**: like RR but more accurate; accounts for the time each participant was in the study before having event or withdrawing.

Study Results: Precision of Treatment Effect²¹

- Confidence Interval (CI)**: a 95% CI provides the range of values we are 95% certain that overlaps the true value. CI's indicates the precision of the estimate; where CIs are wide, they indicate less precise estimates of effect (just an estimate of the worst & best case scenario of the outcome) (For ratios, a CI that includes 1 means possibility of no difference. For ARR, ARI, NNT, NNH, a CI that includes zero means possibility of no difference between treatments. Non-significant trends may provide clues for future research.)
- Type 1 (or α) error**: *the false positive*; to find a difference when there is none. **p-value**: reflects type 1 error. A p <0.05 suggests a <1 in 20 probability that any difference is due to chance (statistically significant by convention). The smaller the p-value, the less likely that the result is due to chance.
- Type 2 (or β) error**: *the false negative*; to conclude there is no difference when there really is a difference (e.g. if not enough patients enrolled)
- Heterogeneity**: when study results within a meta-analysis have more variation than expected; may indicate inappropriate to combine studies.²²

Calculations Example: 1 yr trial

- 200 patients in Control grp
- 200 patients in Treatment (Tx) group
- Deaths**: Control grp: 40. CER=40/200=0.2
Tx grp: 30. EER=30/200=0.15

RRR = (0.20 - 0.15)/0.20 X 100 = 25% (risk of event is reduced by 25%)	ARR 20% - 15% = 5% {absolute risk of event is reduced by 5%}
For every 20 patients treated for 1yr, there is 1 less <i>death</i> , & for every 3 patients treated there will be 1 extra <i>headache</i> .	

Do the study results matter to me & my patients?

- Clinical significance vs statistical significance**: some studies may detect extremely small statistically significant differences between groups; however magnitude of effect (e.g. NNT) may be too small to change practice. Evaluate both 1) the endpoint, & 2) the NNT or NNH. {e.g. small cognitive score improvement not noticeable to patient.^{23,24}}
- Composite endpoints**: combining endpoints can increase a study's power allowing for smaller or shorter trials. Outcomes should have **similar value**. Examination of individual outcomes can be important in interpretation as one endpoint may be the primary *driver*. {e.g. In *DREAM*, outcome of diabetes diagnosis *the driver* or death example of unequal endpoints.²⁵}
- Surrogate endpoints**: an endpoint meant to reflect / be correlated with another endpoint (e.g. BP/LDL/A1c for CV events; CD4 cell count for HIV mortality). **Clinical outcomes are more important** since surrogate endpoints assume correlation with an outcome which may or may not always be true. {e.g. lower A1C target ≤6% in ACCORD led to ↑ death.}
- Other considerations**: What uncertainties remain? Has drug been studied in enough patients to detect serious rare adverse events? What duration of intervention is studied & what are the potential benefits & risks over a longer term of exposure? Does **real-world** experience appear to be consistent with clinical trial data? Cost? How benefits & risks are described will also affect decisions.²⁶
- What patient specific and societal values need to be considered?**

Heads Up! Know what the numbers are telling you.

⇒ You double your chance of winning a lottery if you buy a 2nd ticket; however your chance of winning is more related to whether 2 tickets or 2 million tickets are sold!

Beware of the Relatives ☹

- Benefits are often given as relative** numbers, whereas **harms are often given as absolute** numbers. This tends to exaggerate benefits & minimize the harms. ⇒ **Look for NNTs & NNHs**. {e.g. *Vioxx* monograph 2004²⁷: reported ~ 50% ↓ in GI complications with *Vioxx* 50mg/day vs naproxen 500mg BID & a thrombotic event rate of 1.8% (*Vioxx*) vs 0.6% (naproxen). Actual GI complications reductions 0.59% vs 1.37% (ARR=0.78; NNT=129); whereas thrombotic risk was worse (NNH=83).} {e.g. Oral contraceptives: risk of DVT in a younger, non-smoking ♀ may be ↑300% but absolute risk is <1.5/10,000 /yr & lower than risk in pregnancy}
- Non-Equivalent Durations & Risk/Benefit Perception**
 - Benefits are often given for total duration of trial** which may be several years, whereas **harms are often given per year**. {e.g. UKPDS-33: aggressive glucose control benefit on microvascular endpoints given per 10 years; risks of hypoglycaemia were given per year.²⁷}
 - Analysis: Pooling Together or Dividing Out**
 - Discussing the multiple benefits of a composite endpoint while individually sorting out risks **may minimize risk perception**. {e.g. In WHI, risk of just breast ca with HRT was 8/10,000 pt-years; yet risk of any harm (DVT, CHD, stroke, PE & breast cancer) was 1/66 over 5.2yrs. ²⁸}

A few NNTs / NNHs of interest

↓ mortality with simvastatin 20-40mg/day over 5.4yrs vs placebo in patients with CHD ⁴⁵	NNT 30 / 5.4yrs
↓ mortality with metformin 2550mg/day over 10 years vs non-intensive in obese T2DM patients UKPDS-34	14 / 10 yrs
↓ CV death/MI/stroke: clopidogrel 75mg/day + ASA vs ASA alone in ACS pts (↑ bleeding; NNH=99) CURE	48 / 9mo
↓ pain by ≥50% with TCAs (e.g. amitriptyline 100mg/day) vs placebo in neuropathic pain (short term trials)	2

What makes for a good NNT? *It all depends!!!* NNTs will vary greatly with variations in baseline population risk, duration of tx, & type & number of endpoints included in composite. The value of the endpoint also varies from patient to patient.

