Workshop 8 — Appraising a study on therapy

Diagnosing and treating diseases are the most common and important daily activities of a clinician. After reaching at a high enough level of certainty on the diagnosis, the clinician would have to make a decision on whether treatment (medical intervention) should be given, and if so what treatment(s) is/are to be given. As distinct from complementary and alternative medicine, in modern medical practice (including public health interventions) such decisions should be based on evidence.

Randomised controlled trials (RCTs) are often regarded as the 'gold standard' for evidence in therapy although there are other study designs that also provide information on the effects of certain interventions.[1] Despite the many efforts at summarising evidence from the medical literature to guide medical practice (systematic reviews and meta-analyses of RCTs, 'Guidelines', etc.), the ability to critically appraise an original RCT remains fundamental, both for readers and authors. The RCT is an experimental design that aims to make the comparison groups as similar as possible in background characteristics, so that whatever effects observed can be attributed to the intervention or treatment and not other differences between the groups (confounding). However, one must still be aware that RCTs are not bias-proof.

In this Workshop, the four major questions to be answered when appraising a study on a RCT are discussed.

(1) How large were the reported effects? Do the results support benefits (vs harm) of the treatment or intervention over no intervention or other treatment alternatives?

Depending on the outcome measures used in the RCT, the effect (benefit) can either be classified dichotomously (with or without the adverse event) or otherwise (grading of response or changes of measurements on a continuous scale). In the former case, the relative risk (RR) and/or relative risk reduction (RRR) is usually reported to reflect the benefit of the treatment/intervention. The smaller the RR (below 1) or the greater the RRR (1 – RR), the greater the benefit. To assist clinical decision making, the absolute risk reduction (ARR) and its reciprocal, the number needed to treat (NNT) are also frequently reported. One must be aware that the ARR and NNT depend very much on the probability of a certain outcome in the control group [see discussions on (4) below]. Outcomes reporting changes in grading or measurements on a continuous scale involve less concern on application in another setting, but the magnitude of change with the intervention should be of clinical importance and not just of statistical significance. Given enough resources (money and sample size), a very small incremental benefit (eg reduction of blood pressure by 1 mm Hg, lowering LDL-cholesterol level by 1%, or reduction of pain score by 1 on a scale of 0 – 100) associated with the intervention can be detected and shown to be highly statistically significant (eg $P<0.001$). However, such a difference might not be of importance to an individual patient or of interest to the clinician.

One must also be aware that any harmful side-effects are adequately reported. "All substances are poisons: there is none which is not a poison …" Paracelsus (1493-1541),[2] and all medical interventions can have the potential to harm. A good RCT paper should report the side-effects of the intervention in addition to its efficacy, so as to allow clinicians and patients to make informed decisions. Regrettably, the reporting of side-effects is frequently limited to acute and/or mostly minor, effects. The serious, and often chronic adverse effects surface only after years or decades. In the case of a pharmaceutical product, this may ensue a long time after it has been marketed.

(2) Are the results about the benefits over harm basically valid?

Despite the plethora of guides/rules promulgated for appraising RCTs, the validity of results have seldom been assessed systematically by examining the sources of bias. The specific questions to be answered for ascertaining the validity of study results in a study of therapy are structured under the three major sources of bias[3] (see Box on p.420).

(3) Are the results reasonably reliable or precise?

The precision of the intervention effect (eg 95% confidence interval) should be reported to enable clinicians to judge the clinical importance of the result, by examining both the upper and lower bounds of the estimated benefit.

(4) Can the results be applied to a specific patient or in another setting?

A specific patient similar to those included in the study (eg age, gender, nature and stage of disease, co-morbidities, etc) is more likely to benefit from applying the valid results than another one with a very dissimilar background. Of course, one needs to balance the benefits with the potential harms associated with the treatment. Moreover, incremental benefits over other treatment alternatives should be compared to incremental costs.

BOX. Validity of study results

Validity — selection bias

- Was the source of study subjects described, as to the inclusion and exclusion criteria? Was a representative sample selected from all eligible subjects?
- Was the response or participation rate for the sampled subjects reported and reasonably high?
- Was the randomisation or allocation concealed or masked? Examine whether any participating subjects were removed (by care provider or self) from the trial or not accounted for after random allocation but before the intervention was administered.
- Was follow-up complete or loss to follow-up reported? Were the destination/outcomes of all study subjects known?

Validity – measurement / information bias (including misclassification)

- Were patients blinded to the group allocation and not aware of that throughout the observation period? The use of placebo in drug trials may help, but blinding of patients is not always possible in RCTs, eg surgical procedures.
- Were care providers blinded to the group allocation and not aware of that throughout the observation period? Very dissimilar co-interventions between the comparison groups administered by the care providers may provide a hint on inadequate/ineffective blinding.
- Were outcome assessors blinded to the group allocation and not aware of that throughout the observation period? The use of more objective health outcomes (eg blood chemistry, death) may help to reduce information bias resulting from inadequate/ineffective blinding.
- Was the follow-up duration sufficient for observing the relevant health outcomes (benefits and harm)?
- Was compliance assessed and reasonably good?
- Was contamination between groups considered and avoided (especially for educational interventions)?

Validity – confounding

- Was randomisation carried out effectively, so that the comparison groups were similar with respect to all known prognostic factors apart from the intervention? Be aware of substantial differences between groups that may not be statistically significant for small trials.
- Were statistical adjustments carried out if the comparison groups were not similar enough?
- Were co-interventions or changes in important prognostic factors during the study period reported and similar enough between the groups?
- Were study subjects classified in their originally assigned groups using an intention-to-treat analysis?

In applying the treatment in another setting requires certain further considerations. Results produced from RCTs conducted in tertiary hospitals under strictly controlled settings, including standardised diets, abstinence from certain medications, etc (efficacy studies), may not be applicable to free living patients in the community. The latter group may be more varied in terms of co-morbidities, dietary habits, and other co-interventions. Moreover, drug-drug and drug-diet interactions could be much more complex than in a controlled experimental environment. Any RCT conducted in a community setting may generate more realistic results on benefits (effectiveness studies) for patients in primary care. Similar to applying a diagnostic test to predict disease status (positive and negative predictive values),[4] the probability of a certain outcome in the patient or patient group would affect the utility of results from a RCT. As mentioned earlier, ARR and NNT could change in different settings with different probabilities of the health outcome, although the RR and RRR may remain fairly stable across settings. Certain health events are more likely to occur in patients under tertiary care than those in the community. Likewise, the probability of having a certain health outcome may be much higher in the country where the RCT was conducted than locally. One needs to factor in the differences in health event rates (probability) when calculating the NNT (which could therefore be much larger or smaller). Thus, applicability must be considered in the local setting, instead of injudicious application of the NNT reported in given study.

Ignatius TS Yu
Shelly LA Tse
Clinical Epidemiology Group
Hong Kong Medical Journal

References

1. Yu IT, Tse SL. Clinical Epidemiology Workshop—Introduction. Hong Kong Med J 2011;17:315-6.
2. Society of Toxicology website: http://www.toxicology.org/AI/EO/principl.asp. Accessed 4 Sep 2012.
3. Yu IT, Tse SL. Clinical Epidemiology Workshop 2—General approach to critical appraisal of a medical journal paper. Hong Kong Med J 2011;17:405-6.
4. Yu IT, Tse SL. Clinical Epidemiology Workshop 7—Appraising a study on diagnosis. Hong Kong Med J 2012;18:325-6.