The Keep it Simple Guide to Evidence-Based Medicine

Dr Simon Curtis
EVIDENCE-BASED MEDICINE

Many GPs have a fear of statistics and EBM. But this is an area where with some simple tools and knowledge can increase your appreciation of the literature and help you help your patients.

Exam candidates also commonly worry (it comprises 10% of the AKT component of the MRCP exam) but you can pick up easy marks here. What follows is a simple guide, explaining the terms and giving you some simple tools for appraisal and easy statistical calculations of the type you need to do for the AKT.

So what is evidence based medicine (EBM)?

Definitions vary, but possibly the best is: ‘EBM is defined as the integration of the best available evidence with our clinical expertise and our patients’ unique values and circumstances’ (Strauss and Haynes, EBM200712:1). So whilst the maths and statistics are important, just as important are your clinical judgement and the patient beliefs. The evidence may say that your elderly patient would benefit from anticoagulation, but if your clinical judgement is that it would be unsafe or she does not want to take it, you are still practising EBM by not prescribing.

In its pure form, EBM involves the 5 step approach of:

- Asking an answerable clinical question e.g. using PICO, see below
- Searching for the evidence
- Critically appraising the evidence
- Making a decision by integrating this evidence with your clinical opinion and the patient’s values and circumstances
- Evaluating your performance

PICO is the commonly used acronym for forming clinical questions when practising EBM.
P = patient or problem
I = intervention
C = comparison intervention
O = outcome
e.g. in a 50 year old man with diabetes and no heart disease (patient), does a statin (intervention) compared to placebo (comparison) reduce outcomes (CV events or mortality)?

We then have to critically appraise what we find by asking ourselves if the information we find is:

- Valid?
  - Is it the correct study type to answer the question, is the methodology of the study sound, are the results statistically significant?
- Relevant?
  - Particularly important in primary care is asking, but is the study population relevant to my particular patient, and is the result clinically meaningful?

However, searching for and then critically appraising information from scratch is time-consuming and difficult. GPs are being increasingly encouraged to use pre-appraised, secondary sources of information whenever possible such as Cochrane, Clinical Evidence,
Evidence Based Medicine

Evidence-Based guidelines etc. Personally, we think this is the only practical way to practise EBM, but it is important to be able to understand the different study types, the basics of critical appraisal and simple statistics.


The following pages take you through the A-Z of EBM, medical statistics and research terminology. Yes, it's sexy!!!!!

**Absolute risk and absolute risk reduction (ARR)**
The absolute risk is the actual, arithmetic risk of an event happening. The ARR (sometimes also called the Risk Difference) is the difference between 2 event rates e.g. AR of a MI with placebo over 5 years is 5% and with drug A is 3%, the ARR is simply 2%. This is the difference between the CER (control event rate) and the EER (experimental event rate). Knowing the absolute risk is essential when deciding how clinically relevant a study is. Absolute risk increase (ARI) similarly calculates an absolute difference in bad events happening in a trial e.g. when the experimental treatment harms more patients than the control. e.g. Drug B reduces the chance of a stroke from 20% (CER) to 17% (EER). What is the ARR? Answer 3%.

**Allocation Concealment**
This is considered by many to be the most important thing to check when appraising a RCT. To minimise bias in a blinded RCT it is essential that patients are allocated completely at random between the two groups. Researchers must conceal which group the patients will be allocated to from their clinicians, otherwise the clinicians will inevitably distort the randomisation process and it is no longer truly ‘blind’. For example: if patients in our clinic were being entered into a RCT of antibiotic vs. placebo for acute cough, and we could tell (by means fair or foul) which group our patient was going to be allocated to we would naturally want our sicker patients to go into the antibiotic group and the not-so-sick ones into the placebo group. So, researchers have to conceal allocation from the clinicians to avoid bias.

**Case-control study**
A retrospective, observational study type that helps answer questions about aetiology and potential harms of interventions. Cases and a ‘matched’ control group are identified, and the retrospective exposure of interest between the two groups is identified. Example: is mobile phone use associated with brain tumours? Patients with brain tumours and a control group would have their retrospective mobile phone use studied to see if there were an association. It allows odds ratios to be calculated but not absolute risks. It provides a higher level of evidence than a case-series, which does not have a control group, but less than a prospective cohort study. Confounding factors are difficult to control for but is useful for looking for very rare events. Case-control studies may also be ‘nested’ within the same group. See ‘nested case-control’.

**Cluster randomisation**
You often see ‘cluster randomised controlled trials’ in primary care research – what are they? Simply, it means a natural group, or cluster, which is randomized to the same intervention and then compared to a control group. A cluster may be a family, a school, an area or (commonly) a group of GPs or practices which are then compared to another group.
**Cohort study**
A prospective observational study that helps answer questions about aetiology and prognosis. It identifies a cohort of patients who receive the exposure of interest (e.g. they live near high voltage power cables) and a matched control group who do not, and they follow them for a long period of time to see if there is a difference in the incidence of the outcome of interest (e.g. development of leukaemia). The most famous cohort study of all established the link between the higher incidence of lung cancer in a cohort of doctors who smoke versus one that did not.

They can also look at therapeutics e.g. a recent interesting example, is the observation from a large cohort study that patients with colorectal cancer who take aspirin have reduced mortality compared to those who do not (JAMA2009;302:649). This creates an interesting hypothesis, but it will need a prospective RCT to prove if it is true...There have been many examples of cohort studies that have suggested benefits that RCTs have subsequently found not to be true (e.g. that HRT is cardio-protective, or that lowering homocysteine with folate reduces heart disease).

Cohort studies give a higher level of evidence than a case-control study, but are more expensive, difficult and time-consuming. They are lower down the evidence hierarchy than a RCT, but are cheaper and simpler and for many research questions a RCT is not practical.

**Confidence Intervals (CI)**
Very useful. You can tell a lot about a study from the CI. **A CI is a range of numbers within which there is a 95% chance that the true result lies.** So if the result is a NNT 30 (95% CI 12-56) it tells us that the result of this trial is a NNT of 30 but the true result could lie anywhere between 12 and 56 i.e. if we repeated this trial 100 times, 95 times the result would lie between 12 and 56.

With CI you can easily tell if statistical significance has been reached, without doing any maths! **If the CI includes the value that reflects ‘no-effect’ the result is statistically non-significant.** This value of no-effect is obviously 1 for results that are expressed as ratios (e.g. Relative Risk, Odds Ratio) and 0 for measurements (e.g. NNT, percentages or ARR).

So CI give two very useful bits of information: **firstly they tell us if a result has reached statistical significance, like a p value does, but they also give us an idea of the precision of the result.** A wide CI means a less precise result and a narrow CI a more precise result. Generally speaking the larger the numbers in the study, the smaller the CI and the more sure you are that the result is true. This is why meta-analyses are considered at the top of the evidence hierarchy because they have tighter CI.

**Confounding variable**
Important to consider when appraising a study, confounding variables are factors that you are not actually interested in but will affect the result. They should be spread equally amongst the control and experimental groups but are very difficult to control for, especially in case-control studies.

**Control Event Rate (CER)**
The rate at which events occur in the control group e.g. in a RCT of aspirin v placebo to prevent MI, a CER of 10% means that 10% of the placebo group had a MI. It is sometimes represented as a proportion (10% = 10/100= 0.1). See Experimental Event Rate.

**Cross sectional study or survey**
An observational study that observes a defined population at a single point in time or time interval and measures the prevalence of risk factors. Cheap and simple but subject to bias.
Descriptive studies
What they say on the tin. These include case reports, case series, qualitative studies and surveys. At the bottom of the evidence hierarchy, but can produce useful learning points and can help answer some of the questions that RCTs or systematic reviews cannot.

Diagnostic validation study
Answers questions about the efficacy of a diagnostic test, symptom or sign. Expresses results using LR, sensitivity and specificity. See sensitivity below.

Diagnostic odds-ratio (DOR)
Used when evaluating diagnostic tests. The DOR is the odds of a positive test in patients with the disease, compared to the odds of a positive test in those without the disease e.g. if the DOR is 6, the odds of having a positive test are 6 times higher in someone with the disease than in someone without.

End-point
Another important point to appraise when looking at a RCT, both for the validity and relevance of the study, is: what was the pre-defined end-point that the study was looking at? Was it a clinically relevant end-point (e.g. a bone fracture) or a surrogate end-point (e.g. reduced bone mineral density)? Is the trial reporting the result of the principal, pre-defined end-point of the study or is it reporting sub-group analyses or secondary end-points? If it is a composite end-point, what does that include (it may lump together a number of outcomes of varying clinical significance, so the overall endpoint may be reached but would the result have been positive if only the most clinically significant event were included?)

Experimental Event Rate (EER)
The rate at which events occur in the experimental group e.g. in the CER example above, an EER of 9% (or 0.09) means that 9% of the aspirin group had a MI.

Exclusion criteria
Essential information when appraising a RCT for relevance to your patient. This is very pertinent in primary care where often we see patients with co-morbidity, and these tend to be excluded from RCTs. For example an early statin study showed an increase in breast cancer cases (CARE study), and since then all patients with a previous history of cancer have been excluded from statin trials. Is a recent statin study applicable then to the patient in front of you with a recent history of cancer?

Follow-up
Very important when appraising a RCT, is looking for drop-out rates. If the drop-out rates are high, how confident can you be in the final results? What if all the drop-outs had a bad outcome? If less than 80% are followed up it is generally recommended that the result is ignored.

Forest Plot
The pictorial representation, also known as a blobbogram, of the results of a meta-analysis. You know what these look like and they are very, very useful as they give you rapidly accessible information. For example, see below from a recent study looking at steroids in acute sore throat (BMJ2009:339:b2976). Just by ‘eyeballing’ the plots below you can tell in about 10 seconds that although individual studies have shown conflicting results, and half of
them have not reached statistical significance, pooling the results together gives a statistically significant result in favour of steroids with a narrow CI. If unsure how, read on...

**Fig 2 Effect of corticosteroids on number of patients experiencing complete pain relief at 24 and 48 hours. See web appendix for references**

<table>
<thead>
<tr>
<th>Study</th>
<th>Steroid Events Total</th>
<th>Placebo Events Total</th>
<th>Relative risk (Floyd) (95% CI)</th>
<th>Weight (%)</th>
<th>Relative risk (Floyd) (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>24 hours</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kiderman</td>
<td>17</td>
<td>40</td>
<td>39</td>
<td>23.0</td>
<td>4.14 (1.53 to 11.22)</td>
</tr>
<tr>
<td>Niland</td>
<td>12</td>
<td>27</td>
<td>8</td>
<td>43.1</td>
<td>1.67 (0.80 to 3.44)</td>
</tr>
<tr>
<td>Tanser</td>
<td>13</td>
<td>51</td>
<td>2</td>
<td>7.7</td>
<td>8.81 (2.14 to 36.23)</td>
</tr>
<tr>
<td>Weir</td>
<td>12</td>
<td>41</td>
<td>4</td>
<td>24.2</td>
<td>2.63 (0.93 to 7.68)</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>139</td>
<td>18</td>
<td>100.0</td>
<td>3.16 (0.97 to 5.08)</td>
</tr>
<tr>
<td>Heterogeneity: $I^2=44$%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 48 hours  |                      |                      |                               |            |                               |
| Kiderman  | 13                   | 40                   | 39                            | 27.5       | 1.73 (1.03 to 3.98)           |
| Niland    | 22                   | 27                   | 17                            | 33.6       | 1.44 (1.00 to 2.06)           |
| Tanser    | 29                   | 31                   | 22                            | 39.0       | 1.79 (1.32 to 2.32)           |
| Total     | 76                   | 98                   | 111                           | 100.0      | 1.65 (1.32 to 2.06)           |
| Heterogeneity: $I^2=0$% |

Forest plots are easy and very useful to interpret. Each individual study should be annotated, and it has a square, or ‘blob’, with a horizontal line representing the 95% CI. The blob in the middle is the reported result of the study, and the relative size of it represents the weighting that this individual study has in the overall analysis. The vertical line down the middle is the ‘line of no effect’, which for a ratio is 1. Studies are statistically significant if the CI line does not cross the value of no-effect i.e. the vertical line. If the horizontal line of the study crosses the vertical ‘line of no effect’ it is non-significant. The overall result pooled by meta-analysis is represented by a diamond, the length of which represents the CI. If the diamond does not cross the line of no effect, this is a positive result. If it does, it means the overall result is non-significant. See also meta-analysis.

**Funnel Plot**
A method of graphing results in a meta-analysis to see if the results have been affected by publication bias.

**Hazard ratio**
A way of expressing the relative risk of an adverse event i.e. if an adverse event was twice as likely to happen with a particular intervention, it would have a HR of 2.

**Heterogeneity**
This is important when appraising systematic reviews. Consider how heterogeneous the results are both clinically and statistically. Clinically you have to use your judgement to see how heterogeneous the studies are to warrant combining them (e.g. is a systematic review of sinusitis that includes groups of patients with just facial pain and those with a CT scan confirmed diagnosis valid?). Statistically consider if the individual results contradict each
Evidence Based Medicine

other. This can be assessed by ‘eye balling’ the Forest plot. Anoraks use formal statistical tests such as the Cochran Q (chi-squared) test to do this.

**Incidence**
The percent of a population who will develop a disease during a specified interval e.g. a study that found an incidence of chlamydia amongst new college students was 2% per annum, means that 2% of the students contracted chlamydia during the year.

**Intention to treat**
This is a method of analysing results of a RCT which is intended to minimize bias. It is very counter-intuitive! If a patient is allocated drug B at the start of the trial, at the end he must be included in the drug B group for the final analysis EVEN IF he never ended up taking drug B or even switched to drug A or C. It happens quite often that patients have to drop-out or swap drugs during a trial but they have to be analysed according to how they were intended to be treated at the beginning. Otherwise researchers could pick and choose who makes it or not to the final analysis which would bias results. Intention to treat ensures that the original randomized groups remain comparable. The alternative (‘per protocol analysis’), is known to cause bias.

**Kaplan-Meier Plot**
A simple graph very commonly used in clinical trials to represent the time to an event (e.g. MI), or survival rates of participants as they move through time. Time on horizontal access, percentage reaching primary end point on vertical axis, curves in different colours represent experimental and control groups.

**Lead-time bias** This is important in prognosis studies. *Detecting disease early does not necessarily improve mortality*. If patients are not all recruited at a similar stage of the disease differences in outcome over time may just reflect earlier diagnosis e.g. a study that shows apparent improved survival in breast cancer: have the patients really lived longer, or actually died at the same time but just been aware for longer of the diagnosis? (e.g. Imagine two women develop breast cancer aged 50. Woman A is diagnosed by screening at age 52. Woman B is not screened, and the cancer manifests clinically at age 56. Both women die aged 60. Screening appears to improve survival with breast cancer from 4 years to 8 years, but in fact has just diagnosed it earlier)

**Levels (or hierarchy) of evidence**
Frequently used in guidelines to grade the strength of recommendations. Variable, but a useful general guide (www.cebm.net) is:

<table>
<thead>
<tr>
<th>Grade recommendation</th>
<th>of Level evidence</th>
<th>of Level evidence</th>
<th>Which means…</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1a to 1c</td>
<td>SR of RCTs or RCT with narrow CI</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2a to 3b</td>
<td>SR of cohort or case control studies, or good quality cohort or CC studies</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>Case series, poor quality cohort and case control studies</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>5</td>
<td>Expert opinion, or based on physiology or ‘first principles’</td>
<td></td>
</tr>
</tbody>
</table>
**Likelihood ratio**
A measure used in studies looking at diagnosis. It tells us how useful a test, symptom or sign is for establishing a diagnosis and is considered the most useful overall measure of its efficacy. Example: a study is looking at a new near-patient test for streptococcal throat, the LR is the ratio of the probability of a positive result amongst patients who really do have strep throat to the probability of a positive result in patients who do not have strep throat.

**Meta-analysis**
A statistical technique used to integrate the quantitative results of pooled studies, for example in a systematic review. For mathematical reasons, they tend to express the results in terms of odds ratios. See systematic review.

**Nested case-control study.**
This is a case control study which is ‘nested’ within a defined cohort. Cases of a disease which occur within this cohort are identified, and compared to matched controls within the same cohort which do not develop the disease. Although a less robust level of evidence than a cohort study, they are cheaper and easier to do than a full cohort study and can help answer useful questions about factors which contribute to the development of a condition. For example, an excellent recent nested-case control study looked at death in patients with epilepsy (BJGP2011;61;341). The cohort was patients diagnosed with epilepsy, cases were identified who died and for each of these two controls with epilepsy were selected who were still alive. Cases and controls are then matched for age and sex and the researchers then identify factors which are associated with death in epilepsy.

**‘Non-inferiority trials’ and ‘Equivalence’ trials**
These are trials specifically designed to see if the new drug is ‘no worse than’ or at least ‘as good as’ the standard treatment. Non-inferiority trials require smaller sample sizes, are cheaper and quicker to do and less likely to produce disappointing results for new drugs so are increasingly used in Pharma sponsored research! The hope is that by demonstrating non-inferiority the new drug will then become a ‘me too’ option, a second-line therapy or stimulate further research.

**Number needed to treat or harm (NNT and NNH)**
A clinically useful measure of the absolute benefit or harm of an intervention expressed in terms of the number of patients who have to be treated for one of them to benefit or be harmed. Calculated as 1/ARR. Example: the ARR of a stroke with warfarin is 2% (=2/100 = 0.02), the NNT is 1/0.02 = 50. e.g. Drug A reduces risk of a MI from 10% to 5%, what is the NNT? The ARR is 5% (0.05), so the NNT is 1/0.05 = 20.

**Odds ratios**
Is another way of expressing a relative risk and is simply the odds of an event (either beneficial or harmful) happening versus it not happening. For statistical reasons they are favoured in meta-analysis and case-control studies. They are calculated from the CER and EER e.g. consider a trial of a new drug. 10 out of 200 patients died in the control arm. 5 out of 200 died in the treatment arm.

- What are the odds of death in the treatment arm? The EER is 5/200 i.e. 2.5% which is 0.025
- What are the odds of death with control? The CER is 10/200 i.e. 5% which is 0.05
- What is the OR of the study? This is the ratio of the odds, i.e. 0.025/0.05 = 0.5 i.e. the treatment halves the risk of death compared to control
Evidence Based Medicine

Odds ratios can be converted into an absolute measure such as NNT, provided all the baseline data are known, but this is complicated and you will not be asked to do it in the AKT! You can convert OR to NNT using the ‘visual treatment calculator’ on-line at www.nntonline.com or at www.cebm.net

**P value**
A measure that an event happened by chance alone e.g. p = 0.05 means that there is a 5% chance that the result occurred by chance. For entirely arbitrary reasons p<0.05 means statistically significant and p<0.01 means highly significant. The lower the value, the more likely the effect is real. It is always sobering to remember that if the makers of a new bottled water called HEADPAINGONE did a RCT comparing their product with tap water to cure headaches, and they repeated this RCT 20 times they would eventually get a statistically significant result.

**Pragmatic Clinical Trial**
What it says on the tin. This is a RCT which is based in patients usual care settings, without rigorous exclusion criteria and often using non-blinded treatments which can be flexibly given or used according to clinical need. The aim is to produce a research setting which more closely mimics ‘real life’ care and ‘real’ patient populations. Such a trial is also likely to be easier and cheaper to do. The downside of course is that because it is less controlled it will be more subject to bias.

**Prevalence**
The probability of a disease in a population at any one point in time. Example, the prevalence of diabetes in the population is 2% simply means that 2% of the population at the time of the study have diabetes. See incidence.

**Publication bias**
An important source of bias. Negative trials are just as valid as positive ones, but are less likely to be published. Systematic reviews should search for all data, including unpublished, to try and eliminate this. Also occurs in a SR if the search for studies is incomplete.

**Qualitative study**
These studies use observation and interview to shed light on beliefs, thoughts and motivations. They can help ‘fill the gaps’ in knowledge that numbers cannot answer. For example, why do parents worry so much about fever in children? Why do men want a PSA test? Although by definition they are qualitative rather than quantitative, and tend to be small, they should still be designed in a rigorous and systematic way.

**Randomized control trial (RCT)**
Obviously the gold standard study type for answering questions about treatment, and the best way of minimizing bias and confounding variables. An ‘experimental’ study that can be used to test hypotheses that may have been generated by ‘observational’ studies, such as cohort studies. But for RCTs very large numbers are needed to find a difference in rare events, so they prone to **type 1 and 2 errors**. Also very expensive, so often funded by manufacturers with strong vested interests in the outcome.

**Relative risk, Relative risk reduction (RRR) and Risk Ratio (RR)**
The relative risk, or **risk ratio**, is the ratio of the risk of an event in experimental group compared to the control group i.e. \( RR = \frac{EER}{CER} \). The RRR is the proportional reduction seen in an event rate between the experimental and control groups. For example if a drug
Evidence Based Medicine

Evidence-based medicine reduces your risk of an MI from 6% to 3%, it halves your risk of a MI i.e. the RRR is 50%. But note that the ARR is only 3%.

Relative risks and odds ratios are used in meta-analyses as they are more stable across trials of different duration and in individuals with different baseline risks. They remain constant across a range of absolute risks. Crucially, if not understood, they can create an illusion of a much more dramatic effect. Saying that this drug reduces your risk of a MI by 50% sounds great; but if your absolute risk was only 6%, this is the same as reducing it 3%. So, saying this drug reduces your risk by 50% or 3% are both true statements but sound very different, so guess which one that drug companies tend to prefer!

However relative risks are still useful to know as you can then apply them yourself to an individual’s absolute risk. So, if you know an intervention reduces the relative risk by a third, you can easily calculate and inform your two patients with a 30% and a 9% risk of heart disease that the intervention will reduce their individual risk by about a third i.e. to 20% and 6% respectively.

Relative risk increase (RRI) similarly describes increases in rates of bad events in a trial or the development of a disease in a cohort study.

Risk difference. See Absolute Risk

Sensitivity and specificity

In diagnostic studies, sensitivity is the probability of a positive test amongst patients with the disease. A very sensitive test will have few false negatives and be good at picking up disease. Specificity is the probability of a negative test among patients without the disease. A very specific test will have few false positives and be good at ruling a disease out. SnNOUT means if a test is very Sensitive (Sn) a Negative test rules the diagnosis out. SpPIN means if a test is highly Specific (Sp) a Positive result rules the diagnosis in.

Imagine a new test for whooping cough. The results are presented as:

<table>
<thead>
<tr>
<th>Test result</th>
<th>Present</th>
<th>Absent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Negative</td>
<td>c</td>
<td>d</td>
</tr>
</tbody>
</table>

The sensitivity of the test is a/a+c; the specificity of the test is d/b+d

The positive predictive value (PPV) of the test is a/a+b; The negative predictive value (NPV) of the test is d/c+d

The PPV is the percentage of patients who test positive for Bordetella who really do have it, and the NPV is the percentage who test negative who really do not have it. Importantly these are depend on the background prevalence of the disorder in the population. If a disease is rare, the PPV will be lower (but sensitivity and specificity remain constant). So often with tests the PPV is higher in a secondary care or sieved population than it is in primary care. The likelihood ratio takes this into account and gives the most accurate information on test accuracy.
Systematic review (SR)
A paper in which the authors have systematically searched for, appraised and summarized all the known literature (including non-published studies) on a topic. Quantitative reviews then use the statistical method of meta-analysis to synthesize the pooled results. This creates statistically powerful studies with narrow CI. Compared to individual studies this improves the precision of the result, increases confidence that the result has not occurred by chance and reduces the chance of type 1 and 2 error (see below). They are critically dependant on the quality of the studies that are included (rubbish in=rubbish out) and can sometimes be blunt tools for looking at some of the very specific problems seen in primary care.

Type 1 and 2 error
A Type 1 error occurs if a result is statistically significant, but this is a chance finding and in fact there is no real difference. A Type 2 error occurs if the study finds no-significant difference when in fact there is a real treatment difference. Small studies with wide CI are prone to these errors.

If you see an unexpected positive result (e.g. a small trial shows willow bark extract is effective for back pain) think: could this be a type 1 error? After all, every RCT has at least a 1 in 20 chance of a positive result and a lot of RCTs are published…
If a trial shows a non-significant result, when perhaps you might not have expected it, think could this be a type 2 error? Is the study under-powered to show a positive result? Systematic reviews, which increase study power and reduce CI, are therefore very useful at reducing Type 1 and 2 error

The frequent outcome when reading/learning about statistics!