Journal Club Critical Appraisal Worksheets

Dr David Walbridge

The Four Part Question

The purpose of a structured EBM question is:

- To be directly relevant to the clinical situation or problem
- To express the problem in a way which facilitates searching for a precise answer

The four parts are:

- The patient or problem being addressed
- The intervention, exposure or prevalence being considered
- The comparison intervention, exposure or prevalence as relevant
- The clinical outcomes of interest

Examples:

- Therapy: In adult patients with depression, does a daily dose of amitriptyline in the range 25-75mg yield the same remission rate as a daily dose in the range 125-150mg, remission being assessed by the severity of reported depressive symptoms?

- Diagnosis: When compared to a clinical history, how well does an EEG rule out epileptic seizure activity in a patient with unexplained collapses and a low pre-test probability of clinical epilepsy?

- Prognosis: What is the prevalence within the next year of death by suicide in depressed patients who have taken a recent non-fatal overdose of antidepressant medication, and what if any clinical features identify patients more or less likely to die by suicide, the comparison group being depressed patients who have not taken a recent overdose?

- Aetiology: Two possible ways of framing the question, retrospectively or prospectively:

    - Do adult female patients with anorexia nervosa have a higher prevalence of exposure to childhood sexual abuse than the general population?
    - Do females exposed to sexual abuse in childhood have a higher prevalence of anorexia nervosa in adult life than the general population?

- Guidelines: In considering the options for medication in a young man with first onset schizophrenia, are there any satisfactory guidelines on when to offer an atypical rather than a typical antipsychotic, bearing in mind the many and complex outcome measures to be weighed against each other in a resource-limited health service?

Additional notes:

The four part question does not necessarily have to be phrased as a fluent sentence, although the act of doing so may aid communication and discussion. Nor does the sentence inevitably have to open with the formulaic “In a patient with...”. The final question should however contain all four elements whatever the structure.

The type of question determines the format used for the critical appraisal of the evidence. A therapy question may lead to a paper describing either a single study or a systematic overview, and there are accordingly two formats available for this type of evidence.

Number Needed to Treat (NNT)

Definition of NNT:

NNT is a measure of the importance of outcome differences in a study comparing two treatments or exposures, and provides a standardised way of presenting the magnitude of such differences. It is applicable to Therapy, Prognosis, Systematic Review and Harm/Aetiology appraisals. Ideally, the outcome differences should already have been shown to be significantly different in statistical terms.

If the outcome measure is recovery for example, NNT=5 indicates that in a group of 5 patients treated with the experimental therapy, the number who recover will be one more than the number who recover in a group of 5 patients receiving the control therapy. The absolute number who recover in each group is not indicated, just the relative advantage in the experimental group.

There is no set cut-off by which an NNT is automatically considered important; it depends on context and subjective judgement. For example, a new vaccine with NNT=10,000 might be highly important if the outcome measure was prevention of death, there were no adverse side effects, and 1 million children were to be treated: 100 deaths would be prevented by using the new vaccine instead of the control treatment. By contrast, NNT=10 might not be very important if the outcome measure was weight gain with a new antipsychotic and there were many other factors to be considered also.

By convention, the term **Number Needed to Harm (NNH)** is used where the outcome measure is undesirable, for example a side effect of therapy or an adverse aetiological exposure. The statistical calculations are identical however. Both numbers are conventionally rounded to whole patients.

Other definitions:

- **Control Rate (CR)**. This is the rate of occurrence of an outcome measure in a group of patients who received the control, or reference, therapy or who were not exposed to a presumed aetiological factor. It is often reported as a percentage, although in further calculations is always expressed as a decimal, e.g. 70% is expressed as 0.7 and so on. It is vital that the paper reports this, as full NNT calculations require it.

- **Experimental Rate (ER)**. This is the rate of occurrence of an outcome measure in a group of patients who received the experimental therapy or who were exposed to a presumed aetiological factor. Depending on the study design, it may be greater or smaller than the CR. Ideally the paper will report this also (and it would surely be unusual not to), but it is not strictly necessary provided that at least one of the following statistics is reported instead, because it can always be calculated from CR and one other.

- **Absolute Risk Difference (ARD)**. Often called Absolute Risk Increase or Absolute Risk Reduction depending on the study design, i.e. whether ER is greater than CR or not. Note that Risk can refer to both desirable and undesirable outcomes and really just means Chance. It is often reported as a percentage, although in further calculations is always expressed as a decimal. It is always calculated to be a positive number:

  ARD=(ER-CR) or **ARD=(CR-ER)**

- **Relative Risk Difference (RRD).** Often called Relative Risk Increase or Relative Risk Reduction depending on the study design, i.e. whether ER is greater than CR or not. It is often reported as a percentage, although in further calculations is always expressed as a decimal. Its main use is as an intermediate step in calculating an adjusted NNT where your patient's PEER (see below) is likely to be different to the study CR. It is always calculated to be a positive number, and as can be seen, the relationship with ARD is RRD=ARD/CR:

  RRD=(ER-CR)/CR or **RRD=(CR-ER)/CR**

- **Relative Risk (RR)**. Often used in prospective randomised trials or cohort studies of Harm/Aetiology, this is the risk of the outcome in the exposed group divided by the risk of the outcome in the control group. It is an intuitively understandable expression of the relative chance of the outcome in the experimental group, e.g. RR=3.0 simply means that the chance of the outcome is three times greater than in the control group. It may be greater or smaller than 1.0 depending on the study design; there are two tables for deriving NNT from CR and RR accordingly.

RR=ER/CR and **ER=RRxCR**

- **Odds Ratio (OR)**. Also called Relative Odds. Often used in retrospective case-control studies of Harm/Aetiology and especially in Systematic Reviews, this is a mathematically powerful term but less intuitive. It may be greater or smaller than 1.0 depending on the study design; there are two sets of equations and two tables for deriving NNT from CR and OR accordingly.

OR=ER(1-CR)/CR(1-ER)

- **Patient Expected Event Rate (PEER)**. This is used to extrapolate the paper's NNT to your individual patient when considering Applicability in critical appraisal. PEER is your estimate of the outcome for your patient under control conditions. It is likely to be similar to the CR, and could well be identical if the patients in the paper are a close match to yours.

- **Confidence Intervals (CI)**. The paper should report these for each of the statistics described above that it uses. 95% CI that cross zero for ARD and RRD, or that cross 1.0 for RR and OR indicate that the outcome in the two groups is not significantly different at the 5% level. You can still calculate NNT if you wish, but don't attempt to calculate CI for the NNT as a mathematical anomaly will arise: the upper limit will be a negative number of patients.

Calculating NNT:

1. From ER and CR, or from ARD alone. Example: 73% patients recovered on Venlafaxine, 65% patients recovered on the control antidepressants (all SSRIs). The difference was significant, perhaps because the groups were very large, but was it important?

CR=0.65 ER=0.73 **Note**: ER>CR here

ARD=ER-CR=0.08

NNT=1/ARD

NNT=12 to achieve one more recovery.

2. From RR and CR. Example: 15% patients discontinue control antidepressants (all tricyclics), the Relative Risk for discontinuing Paroxetine is 0.73. Proceed by one of two methods:

- CR=0.15 ER=CRxRR=0.15x0.73=0.11 **Note**: ER<CR here

  ARD=CR-ER=0.04

  NNT=1/ARD

  NNT=25 to prevent one case of non-compliance.

- Use the following look-up tables for the nearest estimate.

CONTROL RATE (CR)	RELATIVE RISK (RR)												
	0.90	**0.85**	**0.80**	**0.75**	**0.70**	**0.65**	**0.60**	**0.55**	**0.50**	**0.45**	**0.40**	**0.35**	**0.25**
0.05	200	133	100	80	67	57	50	44	40	36	33	31	27
0.1	100	67	50	40	33	29	25	22	20	18	17	15	13
0.15	67	44	33	27	22	19	17	15	13	12	11	10	9
0.2	50	33	25	20	17	14	13	11	10	9	8	8	7
0.25	40	27	20	16	13	11	10	9	8	7	7	6	5
0.3	33	22	17	13	11	10	8	7	7	6	6	5	4
0.35	29	19	14	11	10	8	7	6	6	5	5	4	4
0.4	25	17	13	10	8	7	6	6	5	5	4	4	3
0.45	22	15	11	9	7	6	6	5	4	4	4	3	3
0.5	20	13	10	8	7	6	5	4	4	4	3	3	3
0.55	18	12	9	7	6	5	5	4	4	3	3	3	2
0.6	17	11	8	7	6	5	4	4	3	3	3	3	2
0.65	15	10	8	6	5	4	4	3	3	3	3	2	2
0.7	14	10	7	6	5	4	4	3	3	3	2	2	2
0.75	13	9	7	5	4	4	3	3	3	2	2	2	2
0.8	13	8	6	5	4	4	3	3	3	2	2	2	2
0.85	12	8	6	5	4	3	3	3	2	2	2	2	2
0.9	11	7	6	4	4	3	3	2	2	2	2	2	1

CONTROL RATE (CR)	RELATIVE RISK (RR)												
	1.1	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**2.0**	**2.5**	**3.0**	**3.5**	**4.0**
0.05	200	100	67	50	40	33	29	25	20	13	10	8	7
0.1	100	50	33	25	20	17	14	13	10	7	5	4	3
0.15	67	33	22	17	13	11	10	8	7	4	3	3	2
0.2	50	25	17	13	10	8	7	6	5	3	3	2	2
0.25	40	20	13	10	8	7	6	5	4	3	2	2	
0.3	33	17	11	8	7	6	5	4	3	2	2		
0.35	29	14	10	7	6	5	4	4	3	2			
0.4	25	13	8	6	5	4	4	3	3	2			
0.45	22	11	7	6	4	4	3	3	2				
0.5	20	10	7	5	4	3	3	3					
0.55	18	9	6	5	4	3	3						
0.6	17	8	6	4	3	3							
0.65	15	8	5	4	3								
0.7	14	7	5	4									
0.75	13	7	4										
0.8	13	6											
0.85	12												
0.9	11												

3. From OR and CR. Example: 15% patients discontinue control antidepressants (all tricyclics), the Odds Ratio for discontinuing Fluoxetine is 0.65. Proceed by one of three methods:

CR=0.15 OR=0.65

- Here, the OR is less than 1.0. Use the equation: **NNT= $\frac{\text{1-(CRx(1-OR))}}{\text{(1-CR)xCRx(1-OR)}}$**

  NNT=21 to prevent one case of non-compliance.

- When the OR is greater than 1.0, use the equation: **NNT= $\frac{1+(CR\times(OR-1))}{(1-CR)\times CR\times(OR-1)}$**

- Use the following look-up tables for the nearest estimate.

CONTROL RATE (CR)	ODDS RATIO (OR)												
	0.90	**0.85**	**0.80**	**0.75**	**0.70**	**0.65**	**0.60**	**0.55**	**0.50**	**0.45**	**0.40**	**0.35**	**0.30**
0.05	209	139	104	83	69	59	52	46	41	37	34	31	29
0.1	110	73	54	43	36	31	27	24	21	19	17	16	15
0.15	77	51	38	30	25	21	18	16	15	13	12	11	10
0.2	61	40	30	24	20	17	14	13	11	10	9	8	8
0.25	52	34	25	20	16	14	12	11	9	8	8	7	6
0.3	46	30	22	18	14	12	10	9	8	7	7	6	5
0.35	42	28	20	16	13	11	9	8	7	6	6	5	5
0.4	40	26	19	15	12	10	9	8	7	6	5	5	4
0.45	39	25	18	14	12	10	8	7	6	6	5	4	4
0.5	38	25	18	14	11	9	8	7	6	5	5	4	4
0.55	38	25	18	14	11	9	8	7	6	5	5	4	4
0.6	39	25	18	14	11	9	8	7	6	5	4	4	3
0.65	41	26	19	15	12	10	8	7	6	5	4	4	3
0.7	44	28	20	16	13	10	9	7	6	5	5	4	3
0.75	49	32	23	17	14	11	9	8	7	6	5	4	4
0.8	58	37	26	20	16	13	11	9	8	6	5	5	4
0.85	72	46	33	25	19	16	13	11	9	8	6	5	5
0.9	101	64	46	34	27	22	18	15	12	10	9	7	6

CONTROL RATE (CR)	ODDS RATIO (OR)												
	1.1	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**2.0**	**2.2**	**2.5**	**2.8**	**3.0**
0.05	211	105	70	53	42	35	30	26	21	18	14	12	11
0.1	111	56	37	28	22	19	16	14	11	9	8	6	6
0.15	78	39	26	20	16	13	11	10	8	7	5	5	4
0.2	63	31	21	16	13	11	9	8	6	5	4	4	4
0.25	53	27	18	13	11	9	8	7	6	5	4	3	3
0.3	48	24	16	12	10	8	7	6	5	4	4	3	3
0.35	44	22	15	11	9	8	7	6	5	4	3	3	3
0.4	42	21	14	11	9	7	6	6	5	4	3	3	3
0.45	40	20	14	10	8	7	6	5	4	4	3	3	3
0.5	40	20	13	10	8	7	6	5	5	4	3	3	3
0.55	40	20	14	10	8	7	6	5	5	4	4	3	3
0.6	42	21	14	11	9	7	6	6	5	4	4	3	3
0.65	44	22	15	11	9	8	7	6	5	4	4	4	3
0.7	48	24	16	12	10	8	7	7	5	5	4	4	4
0.75	53	27	18	14	11	9	8	7	6	5	5	4	4
0.8	63	31	21	16	13	11	9	8	7	6	5	5	5
0.85	79	39	26	20	16	14	12	10	9	8	7	6	6
0.9	111	56	37	28	23	19	17	15	12	10	9	8	7

Calculating NNT Confidence Intervals:

You require to know CR, ER and the number of patients in each of the two groups. Follow the steps:

1. NNT is the reciprocal of ARD, i.e. NNT=1/ARD and ARD=1/NNT. If you haven't already derived ARD in calculating NNT, do so now.

2. Calculate the 95% CI for the ARD using the formula:

$$\mathbf{ARD} \pm \mathbf{1.96} \times \sqrt{\frac{\mathbf{ER} \times \mathbf{(1-ER)}}{\mathbf{N\,of\ experimental\ \ pts}} + \frac{\mathbf{CR} \times \mathbf{(1-CR)}}{\mathbf{N\,of\ control\ \ pts}}}$$

3. You now have upper and lower 95% CI for the ARD. The lower and upper 95% CI for NNT are the reciprocals of these.

Example: 85% patients on Venlafaxine (N=130) complain of sexual dysfunction compared to 70% on Fluoxetine (N=280). Consider Fluoxetine to be the control group so that NNT applies to the consequences of a decision to prescribe Venlafaxine instead (and for a side effect it might be referred to as NNH).

CR=0.7 ER=0.85 ARD=0.15 NNT=6.7 (rounded to NNT=7).

$$95\%\,\mathrm{CI\,of\ ARD} = 0.15 \pm 1.96 \times \sqrt{\frac{0.85 \times 0.15}{130} + \frac{0.7 \times 0.3}{280}}$$

95% CI of ARD = 0.15 ± 0.0815 = 0.0685 to 0.2315

95% CI of NNT = 1/0.2315 to 1/0.0685 = 4.3 to 14.6 (rounded to NNT = 4 to 15)

The fact that the NNT confidence interval does not cross zero also indicates that the difference between treatments on this outcome measure is significant, even though the original data did not report this directly.

Adjusting NNT to apply to your patient:

In the paper under critical appraisal, the description of the patients' demographic and clinical details may lead you to doubt that the Control Rate of an outcome would exactly match the PEER for your individual patient if they received the control intervention or exposure. The NNT calculated for the paper's patients may not therefore be fully applicable to your patient.

Providing that you believe it is possible to estimate the PEER, you can use this to adjust the NNT. For example, a paper might report a 60% response to Amitriptyline as the CR, and then compare this with another treatment that you had been investigating as a possibility for your patient. The 60% response to standard treatment might not seem to apply to your patient however; perhaps they have a co-morbidity with their depression, or perhaps they have already failed to respond to several antidepressant drugs prescribed in clinic. It would then be reasonable to assume that the PEER is lower than the CR in the paper.

There are many ways to perform the calculations, but the simplest is to use Relative Risk Difference as follows:

1. Derive the ARD for the paper if you haven't already done so: ARD=1/NNT.

2. RRD=ARD/CR where CR is the figure reported in the paper.

3. Adjusted NNT=1/(PEERxRRD). Confidence Intervals are not applicable here.

Example: The paper reports that 60% depressed patients recover on SSRIs alone, 80% recover on SSRIs plus Cognitive Therapy. You have been considering Cognitive Therapy for your depressed patient, who also misuses alcohol, but could not find a paper reporting a trial of combining medication and Cognitive Therapy specifically in this situation.

The paper's CR is thus 60%, and the NNT is 5 for the combined treatment, but it is unlikely that your patient would fare this well. You know from other sources that depressed patients who misuse alcohol are about half as likely to recover on standard antidepressants as those who do not. You estimate your patient's PEER accordingly as 30%. Remember to express PEER as a decimal in calculations, i.e. as 0.3 in this example.

ARD=1/NNT=1/5=0.2

RRD=ARD/CR=0.2/0.6=0.3333

Adjusted NNT=1/(PEERxRRD)=1/(0.3x0.3333)=10.0

<u>NNT=10</u> as an estimate for your individual patient.

D.G. Walbridge 2002

CRITICAL APPRAISAL WORKSHEET - DIAGNOSIS EVIDENCE

Citation:

Are the results of this diagnostic study valid?

1. Was the new diagnostic test evaluated in a sample representative of the full range of patients?

2. Were all patients who entered the trial accounted for at its conclusion?

3. Was there an independent comparison with a reference “gold standard” diagnostic test?

4. Was each diagnostic test applied uniformly regardless of other results or findings?

5. Were patients and clinicians “blind” to the first test’s result when they performed the second?

6. Does the new diagnostic test show acceptable reliability?
   - **a)** Test-retest reliability.
   - **b)** Inter-rater reliability as appropriate.
   - **c)** Internal consistency as appropriate.

7. Does the paper report the authors’ Declaration of Interest?

Are the results of this diagnostic study important?

1. What are the main findings?

2. Are the main findings reported clearly, and analysed with appropriate statistical techniques?

3. Are the sensitivity and specificity of the new test reported?

4. Does the new test generate clinically useful Post-test Probabilities?

Post-test Probability:

Example: 100 patients with clinical dementia were investigated by SPECT scan for the signs of multi-infarct dementia (MID), and the "true" diagnosis was established eventually at post mortem:

	MID at post mortem (25)	Non-MID at post mortem (75)
SPECT positive	20	15
SPECT negative	5	60

Sensitivity: Proportion of "true" diagnosis detected by the new test, i.e. 20 out of 25 = **80%**
Specificity: Proportion of "true" non-diagnosis detected by the new test, i.e. 60 out of 75 = **80%**

Pre-test Probability: The chance that a patient has the diagnosis of interest. For a sample in a study, this is known exactly from the "gold standard" prevalence, i.e. 25 out of 100 = 0.25 or **25%**

Pre-test Odds = Pre-test Probability/(1 - Pre-test Probability), i.e. 0.25/0.75 = **0.33**
Likelihood Ratio (LR) for a positive result = Sensitivity/(1 - Specificity), i.e. 0.8/(1-0.8) = **4.0**
Likelihood Ratio (LR) for a negative result = (1 - Sensitivity)/Specificity, i.e. (1-0.8)/)0.8 = **0.25**
Post-test Odds = Pre-test Odds x LR, i.e. 0.33 x 4 = **1.33** if SPECT positive
0.33 x 0.25 = **0.083** if SPECT negative

Post-test Probability (PTP): The chance that a patient has the diagnosis of interest once the results of the new test are known. The magnitude of the difference between the Pre-test Probability and the Post-test Probabilities is an indication of the **importance** and hence clinical usefulness of the test.

PTP = Post-test Odds/(Post-test Odds + 1), i.e. 1.33/2.33=0.57 or **57%** if SPECT positive
0.083/1.083=0.08 or **8%** if SPECT negative

LRs are a property of the test itself and can therefore be used to calculate the PTP for any patient providing an estimate of their Pre-test Probability can be made. For example, if your patient had a history of myocardial infarction, you might estimate that their Pre-test Probability of a vascular aetiology to their dementia was twice that of the usual prevalence, i.e. 50%. Repeating the above calculations, or using the short-cut **PTP=(p x LR)/[(p x LR)+(1-p)]** where p is the Pre-test Probability, gives a PTP for your patient of 80% for a positive SPECT scan and 20% for a negative SPECT scan.

Can you apply this evidence about a diagnostic test in caring for your patient?

1. Was the diagnostic test described in sufficient detail to allow you to replicate it?

2. Could the test apply to your patient?
   - **a)** Can a good estimate of their Pre-test Probability be made?
   - **b)** Are the Post-test Probabilities likely to be clinically useful?
   - **c)** Could the results move you across a treatment threshold?

3. Would the test be feasible (available, accurate and affordable) in your clinical setting?

4. Would your patient accept this test, and would the overall consequences benefit them?

CRITICAL APPRAISAL WORKSHEET - HARM/AETIOLOGY EVIDENCE

Citation:

Are the results of this harm or aetiology trial valid?

1. Were there clearly defined groups of patients?

2. Were the groups similar in all important ways apart from exposure to a presumed cause?

3. Were exposures and clinical outcomes measured the same way in both groups?

4. Were those who rated outcomes kept “blind” as to which group the patient was in?

5. Were outcome measures objective (e.g. death), or shown to be reliable (if only by citation)?

6. Were all patients who entered the trial accounted for at its conclusion?

7. Was the follow-up long enough for the final outcome to be a valid clinical assessment?

8. And were all clinically important outcomes considered?

9. Does the paper report the authors’ Declaration of Interest?

10. Do the results satisfy the “diagnostic tests for causation”?

    d) Is it clear that the exposure preceded the onset of the outcome?

    e) Is there a dose-response gradient?

    f) Is there positive evidence from a “withdrawal-rechallenge” study?

    g) Is the association consistently found from study to study?

    h) Does the hypothesised causation make biological and clinical sense?

Are the results of this trial important (refer to NNT guidelines)?

1. What type of study is this: prospective RCT, prospective cohort, or retrospective case-control?

2. What are the main findings?

3. Are the main findings reported clearly, and analysed with appropriate statistical techniques?

4. What is the NNH for each relevant outcome measure?

5. How precise is each NNH in terms of its 95% confidence intervals?

Should these results change the treatment of your patient?

1. Is your patient so different from those in the trial that its results can't help you at all?

2. Could some of these results apply to your patient?

3. Calculate an adjusted NNH for each relevant finding if necessary.

4. For treatment, what alternatives are available and what are their consequences?

5. For exposure, what are the alternatives and their consequences if you try to prevent it?

6. What are your patient's preferences, concerns and expectations?

CRITICAL APPRAISAL WORKSHEET - THERAPY EVIDENCE (Single Study)

Citation:

Are the results of this single preventive or therapeutic trial valid?

1. Was the assignment of patients to treatments randomised?

2. And was the randomisation list concealed?

3. Were the groups similar at the start of the trial?

4. Were patients and clinicians kept “blind” to which treatment was being received?

5. Were all patients who entered the trial accounted for at its conclusion?

6. And were they analysed in the groups to which they were randomised?

7. Apart from the experimental treatment, were the groups treated equally?

8. Were the outcome measures used shown to be reliable (if only by citation)?

9. And were all clinically important outcomes considered?

10. Was the follow-up long enough for the final outcome to be a valid clinical response?

11. Does the paper report the authors’ Declaration of Interest?

Are the results of this trial important (refer to NNT guidelines)?

1. What are the main findings?

2. Are the main findings reported clearly, and analysed with appropriate statistical techniques?

3. What is the NNT or NNH for each relevant outcome measure?

4. How precise is each NNT or NNH in terms of its 95% confidence intervals?

Can you apply this evidence about a treatment in caring for your patient?

1. Is your patient so different from those in the trial that its results can't help you at all?

2. Could some of these results apply to your patient?

3. Calculate an adjusted NNT or NNH for each relevant finding if necessary.

4. How great would the potential benefit of therapy actually be for your individual patient?

Are your patient's needs and preferences met by this treatment and its consequences?

1. Do your patient and you have a clear assessment of their needs and preferences?

2. Are they met by this regimen and its consequences?

3. Are the likely treatment benefits worth the potential harms and costs?

CRITICAL APPRAISAL WORKSHEET - THERAPY EVIDENCE (Systematic Review)

Citation:

Are the results of this systematic review of therapy valid?

1. Does the review address a focussed clinical question?

2. Does the methods section describe finding and including all the relevant trials?

3. Does the review assess the individual trials' validity?

4. Are the issues of effect size and publication bias addressed?

5. Are the individual trials' results converted to standardised outcome measures?

6. Were all relevant outcomes considered?

7. How consistent were the results from trial to trial?

8. Were the assessments of the included studies shown to be reliable and reproducible?

9. Does the paper report the authors' Declaration of Interest?

10. Should you believe apparent differences in the efficacy of therapy in some patient subgroups?
    - **i)** Do the differences make biological and clinical sense?
    - **j)** Were they hypothesised at the outset, rather than arising from "dredging the data"?
    - **k)** How many subgroup analyses were carried out in searching for a significant result?
    - **l)** Are the subgroups delineated in an understandable and justifiable way?
    - **m)** Are differences confirmed in other independent studies or reviews?

Are the results of this trial important (refer to NNT guidelines)?

1. What are the main findings?

2. Are the main findings reported clearly, and analysed with appropriate statistical techniques?

3. What is the NNT or NNH for each relevant outcome measure?

4. How precise is each NNT or NNH in terms of its 95% confidence intervals?

5. Are any differences between subgroups also clinically important?

Can you apply this evidence about a treatment in caring for your patient?

1. Is your patient so different from those in the trial that its results can't help you at all?

2. Could some of these results apply to your patient?

3. Calculate an adjusted NNT or NNH for each relevant finding if necessary.

4. How great would the potential benefit of therapy actually be for your individual patient?

Are your patient's needs and preferences met by this treatment and its consequences?

1. Do your patient and you have a clear assessment of their needs and preferences?

2. Are they met by this regimen and its consequences?

3. Are the likely treatment benefits worth the potential harms and costs?

CRITICAL APPRAISAL WORKSHEET - PROGNOSIS EVIDENCE

Citation:

Are the results of this prognosis trial valid?

1. Was there a clearly defined group of patients?

2. Was the sample studied representative of the full range of clinical presentations?

3. Were the patients all at a common (preferably early) point in the course of their disease?

4. Were those who rated outcomes kept “blind” to any possible prognostic factors?

5. Were outcome measures objective (e.g. death), or shown to be reliable (if only by citation)?

6. Were all clinically important outcomes considered?

7. Were all patients who entered the trial accounted for at its conclusion?

8. Was the follow-up long enough for the final outcome to be a valid clinical assessment?

9. Were the findings validated against an independent set of patients (if only by citation)?

10. Does the paper report the authors’ Declaration of Interest?

11. If subgroups with different prognoses were identified:

    n) Do the differences make biological and clinical sense?

    o) Are differences confirmed in other independent studies or reviews?

    p) Was the overall group prognosis adjusted for the prevalence of important factors?

Are the results of this trial important?

5. Are the main findings reported clearly, and analysed with appropriate statistical techniques?

6. How likely are the outcomes over specified time periods?

7. How precise is each prognosis in terms of its 95% confidence intervals?

Confidence Intervals (CI) for a prognosis estimate:

Proportion (p) is the rate of a prognostic outcome, expressed as a decimal.
Number (N) is the number of patients in the group studied.

$$95\%\ \text{CI of } p = p \pm 1.96 \times \sqrt{\frac{p \times (1-p)}{N}}$$

Example: 70% of alcohol dependent patients were abstinent (or nearly so) at 1 year after attending an intensive Alcohol Treatment Programme, N=40.

p=0.7 N=40

95% CI of p = 0.7±1.96x√(0.7x0.3/40) = 0.7±0.14 = 56% to 84%

Can you apply this evidence about prognosis in caring for your patient?

5. Is your patient so different from those in the trial that its results can't help you at all?

6. Could some of these results apply to your patient, one of the subgroups perhaps?

7. Is it clinically possible that treatment could move your patient between prognostic subgroups?

8. If so, calculate the NNT using the subgroup prognoses as CR and ER (see NNT guidelines).

9. Will this evidence change your approach to what you offer, advise or tell your patient?