Rapid Critical Appraisal of Scientific Literature

Presenter: Assoc. Professor Sue Skull
Head, Research Education Program
Deputy Director, Department of Child Health Research

4 May 2018

RESEARCH SKILLS SEMINAR SERIES 2018
Part of the Research Education Program
Department of Child Health Research
Princess Margaret Hospital for Children

ResearchEducationProgram.org
ResearchEducationProgram@health.wa.gov.au
Critical Appraisal of Scientific Literature

CONTENTS:

1 PRESENTATION .................................................................................................................................. 1
  1.1 WHAT IS EVIDENCE-BASED CLINICAL PRACTICE? ................................................................. 16
  1.2 SO WHAT IS CRITICAL APPRAISAL .......................................................................................... 16
  1.3 WHY DO WE NEED CRITICAL APPRAISAL? ............................................................................ 16
  1.4 KEY CONCEPTS UNDERPINNING CRITICAL APPRAISAL .................................................... 16
  1.5 ARE THE RESULTS VALID? (STUDY QUALITY) ......................................................................... 17
  1.6 WHAT ARE THE RESULTS? (STUDY EFFECT SIZE AND INTERPRETATION INCLUDING PRECISION) .... 18
  1.7 CAN WE APPLY THE RESULTS IN OUR SETTING? ................................................................. 19
  1.8 KEY RESOURCES .......................................................................................................................... 20
  1.9 EXAMPLE: DIAGNOSTIC ACCURACY STUDY: ARE THE RESULTS OF THE STUDY VALID? .......... 21
  1.10 EXAMPLE: THERAPY STUDY: ARE THE RESULTS OF THE TRIAL VALID? (INTERNAL VALIDITY) ....... 23
  1.11 EXAMPLE: SYSTEMATIC REVIEW: ARE THE RESULTS OF THE REVIEW VALID? ....................... 26
  1.12 EXAMPLE: QUESTIONS TO ASK OF YOUR PAPER (with thanks to T KEARNS) ......................... 29
  1.13 EXAMPLE: A CHECKLIST FOR APPRAISING RANDOMISED CONTROLLED TRIALS .................. 32

2 CONSUMER AND COMMUNITY INVOLVEMENT – 18 MAY ................................................................. 33

3 SURVEY DESIGN & TECHNIQUES – 25 MAY .................................................................................... 34

4 RESEARCH SKILLS SEMINAR SCHEDULE 2018 ............................................................................. 35

5 SEMINAR FEEDBACK FORM .............................................................................................................. 37
If today’s practitioners are to retain their professionalism, clinicians’ information and research appraisal skills need to be improved urgently. Otherwise they risk being rapidly overtaken by administrators and patients who may not be able to use a stethoscope but are comfortable using Google, Wikipedia, and the internet.


Overview

- What is Evidence-Based Clinical Practice
- Critical Appraisal and why we need it
- General critical appraisal strategies
- Examining validity
- Interpreting results
- Application of results

Evidence-based Clinical Practice

"The integration of best research evidence with clinical expertise and patient values" (David Sackett, 1996)

- Formulate an answerable question (PICOT)
- Find the best evidence
- Critically appraise that evidence
- Integrate with clinical expertise and patient values
- Evaluate our effectiveness and efficiency
What is Critical Appraisal

“Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, its value and relevance in a particular context.”

(Burls, 2009)

“A 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system.”


Why do we need it?

So much to know…

- National Library of Medicine MetaThesaurus
  - 875,255 concepts
  - 2.14 million concept names
- DiagnosisPro
  - 11,000 diseases - 1 new disease per day for 30 years
  - 30,000 abnormalities (symptoms, signs, lab, X-ray)
  - 3,200 drugs (and FDA’s 18,283 products)

“To cover the vast field of medicine in four years is an impossible task.”

William Osler

How Much Research?
How Much Research?

Source: Carl Heneghan

"Just in case" learning

Traditionally "Just in case" / "Push" learning

- Alerts to new information e.g. Journal articles, books etc.
- Too much to possibly keep up AND 50% out of date in 5y
- Meat in clinical journals?

Not all research is useful!

https://doi.org/10.1371/journal.pmed.1002049

- Many studies fail to be useful because of their design
- Consider problem, context, information gain, pragmatism, patient centeredness, value for money, feasibility, and transparency
“Just in case” → “Just in time” learning

- Need a shift in focus: find current best answer for current Q →
- "Just in time" or “Pull” learning – use whenever Qs arise
- Access information when needed: relevant, up to date
- Critically Appraised Topics (CATs), Cochrane etc.

**ONLY read articles to answer questions!**

Might improve on the 50% of valid evidence that is never implemented!

---

**What Skills are needed?**

- To find evidence more efficiently
- To appraise the quality of the evidence more effectively
- To use good quality evidence more systematically

---

**Choosing and Finding the Best Evidence**

```
Different designs fit different questions
- Diagnosis (e.g. cross-sectional studies)
- Prognosis (e.g. cohort studies)
- Treatment (generally most important e.g. RCTs)
- Frequency, Aetiology (cohort, case-control studies)
```

Different designs have different potential impact
- Descriptive: the results are the results (e.g. mortality rate)
- Hypothesis-generating: results are interesting & should be tested e.g. observational studies, qualitative studies
- Hypothesis-testing: results should change practice e.g. well-conducted RCTs or other studies with large effects

---

**Types of Questions & Research**

```
Random Medical News
```

Has the best design been used for the question?
Can we tell “Good” from “Bad” Research?

- Study of 607 BMJ reviewers
  - 14 deliberate errors inserted (9 major, 5 minor)
- Detection rates
  - On average <3 of 9 major errors detected
  - Poor randomisation (by name or day) - 47%
  - No intention-to-treat analysis - 22%
  - Poor response rate - 41%

Schroter S et al.  

So we need Critical Appraisal … how do we approach it exactly?

General Critical Appraisal Strategies

Critical Appraisal

Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, its value and relevance in a particular context (Burls, 2009).

- Does the study have a clearly focused question?
- Did the study use valid methods to address the question?
- Are the valid results of this study important?
- Are these valid, important results applicable in my setting?

Key Concepts of Critical Appraisal

1. No study is perfect → Examine validity
2. There is more than one possible explanation for the result reported in a study
   - Truth
   - Bias and confounding
   - Chance
3. Not all results can be applied in your setting
4. ***Don’t change practice based on one paper
Examining Validity

"Your accomplishments speak for themselves. Unfortunately for you I’m completely fluent in exaggeration."

Validity

- Different checklists for internal validity by study type
  - http://www.cebm.net/ (Oxford)
- Depends on how the study was
  - Designed
  - Conducted
  - Analysed and interpreted
- Key concepts
  - Bias: systematic errors (bad): recruitment, recall, measurement, etc
  - Noise/ “random” errors (not so bad)
  - Confounding
- Study design, conduct & analysis features can minimise

Minimising Confounding: Design

- **Randomisation:** (intervention studies only) with blinding
  - Check "baseline" table.
  - If not possible, collect information on known confounders →
- **Restriction:** eliminates variation in the confounder (e.g. males)
  - Adv: easy, cheap
  - Disadv: limits eligible subjects, generalisability, others
- **Matching:** comparison group forced to resemble index group
  - E.g. case-control studies
  - Disadv – selection of controls, over-matching, bias! analyses

Minimising Confounding: Analysis

- **Must collect information on known confounders**
- **Disadvantage:** can’t fix everything! Unknown confounders
- **Stratification:** produce groups within which the confounder does not vary – works best if only one or two strata, eg: analysis within “male” v “female” or “old” v “young” if not possible →
- **Multivariable analysis (statistical packages required)**
  - Adv: can handle multiple potential confounders
  - Disadv: residual confounding, misclassification of confounders, inadequate adjustment procedures
- **Other:** graphical, propensity scores, marginal structural models etc.
Common Sources of Bias

- Selection bias
- Failed/improper randomisation (allocation concealment? groups similar at start?)
- Compliance (did participants receive allocated intervention?)
- Contamination
- Co-interventions
- Measurement bias (blinding? training/monitoring?)
- Misclassification
- Drop outs? Loss to follow-up? Ineligible?
- Placebo effect (controls?)
- Intention to treat analysis fails
- Selective reporting

Controlled Trials: Look for Dissimilar Groups

- Reassuring if intervention/control groups similar at start
- For known prognostic factors (will never know unknowns)
- “Baseline” “Table 1” is the place to look
- Randomisation should balance groups
- Small studies may have failure of randomisation
- If dissimilar, can still adjust in analyses → look for reporting of differences (adj vs not)
- Key reason for failure of randomisation is poor allocation concealment

Were Groups Treated Equally?

- Should be similar in all respects except for intervention
- Same follow-up?
- Same non-study therapies (co-interventions)?
- *Should be described

Loss to Follow-up

- Small is OK, large is not (5 and 20 rule of thumb)
  - Do sensitivity analyses
- Equal is “more OK”, selective is not. E.g.:
  - Rapid cure rate in intervention group
  - Unpalatable study intervention
  - Increased mortality in one group
Blinding

- Prevents opinions about efficacy of an intervention affecting:
  - Treatment (monitoring, co-interventions)
  - Measurement (encouragement, interpretation)
  - Reporting (interpretation)
- All of which cause BIAS
- Blinding can improve compliance & retention of trial participants
- Traditionally via use of an indistinguishable placebo (blinds participants, treating staff)
- Blind assessors if can’t blind the intervention
- Need to work out:
  - Who: participants, investigators, assessors, analysts
  - How
  - Successful/not

Intention to Treat Analysis

- Analyse participants in the groups to which they were randomised
- Even if they:
  - Discontinue,
  - Never get the intervention (forget, refuse, too sick, die), or
  - Cross over to another group
- Why???

Intention to Treat Analysis

- Preserves randomisation
- Therefore any effect we see is due to the assigned treatment
Interpreting the Results

And the big ticket items are......

- Common sense
- Size of the effect
- Precision
- Applicability

In 2013 Sutherland et al published a list (in Nature) to:
“help non-scientists interrogate advisers and grasp the limitations of evidence.”

Common Sense
Common Sense II: Bradford Hill Criteria

- Strength of association
- Consistency
- Specificity
- Temporal relationship
- Biological gradient
- Biological plausibility
- Coherence
- (*Relevance)


Don’t be distracted by statistics!

Cannot remove all bias even with fancy adjustments
Every test or formula has its own assumptions
Statistics describe the impact of chance
- They cannot assess importance or exclude bias
Judgement is required
- Check for
  - 10 and 20 outcomes, subgroups
  - Adverse outcomes/side effects.
- Make sure CIs are used
- Extract the numbers
  - e.g. exclusions, loss to follow-up, drop outs etc.
- Basic checks: sample size, normal distribution, multiple testing etc.

And check author interpretation

- Overstatement of the results
- Selective statement of the results
- Misinterpretation

Beware dodgy or selective reporting!

Size of the Effect

- Overstatement of the results
- Selective statement of the results
- Misinterpretation
**Big Effects are Hard to Ignore!**

"No single epidemiological study is persuasive by itself unless the lower limit of its 95 percent confidence level falls above a threefold (200 percent) increased risk."

Sir Richard Doll

"As a general rule of thumb, we are looking for a relative risk of 3 or more (> 200 percent increased risk) [before accepting a paper for publication]."

Marcia Angell, fmr Editor, NEJM

"My basic rule is if the relative risk isn’t at least 3 or 4 [a 200 percent or 300 percent increased risk], forget it."

Robert Temple, CDER, FDA

**Why Such a “Big” Difference?**

- Always potential for confounders
- Measures of effect (eg RR) are only estimates
- Bias may be inherent in the method of measurement used
- Statistical manipulations and assumptions
- Publication bias

*We are talking about asking questions to potentially change practice*

**Summary Measures of Effect**

- **Categorical variables**
  - Prevalence
  - Incidence
  - Relative risk, RR (or Odds ratio, OR)
  - Risk difference (RD) or absolute risk reduction (ARR)
  - Relative risk reduction (RRR)
  - Number needed to treat/harm (NNT/NNH)

- **Continuous variables**
  - Mean difference (difference in means)
  - Standardised mean difference (/pooled SD)

**Measure of Effect**

- **Prevalence**: no. of events / no. people who could have it
  - no time component, not a rate (eg survey)

- **Incidence**: no. new events / population at risk in a time period
  - Not those who already have the event at the start

- **RR**: prevalence in group 1 / prevalence in group 2

- **RD**: prevalence in group 1 - prevalence in group 2

- **RRR**: (1-RR) x 100%

- **NNT**: 1/RD
Is it really a “Big Effect”? Beware!

<table>
<thead>
<tr>
<th>Risk</th>
<th>Accompanying explanation</th>
<th>% Accepting Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>2.5% had a heart attack vs 3.9%</td>
<td>42%</td>
</tr>
<tr>
<td>RRR</td>
<td>A cholesterol lowering drug resulted in a 34% reduction in heart attacks</td>
<td>88%</td>
</tr>
<tr>
<td>NNT</td>
<td>Would need to treat 71 patients for 5 years to prevent 1 heart attack</td>
<td>31%</td>
</tr>
</tbody>
</table>


Is it really a Big Effect? Beware!

<table>
<thead>
<tr>
<th>Risk</th>
<th>Accompanying explanation</th>
<th>% Accepting Therapy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RD</td>
<td>2.5% had a heart attack vs 3.9%</td>
<td>42%</td>
</tr>
<tr>
<td>RRR</td>
<td>A cholesterol lowering drug resulted in a 34% reduction in heart attacks</td>
<td>88%</td>
</tr>
<tr>
<td>NNT</td>
<td>Would need to treat 71 patients for 5 years to prevent 1 heart attack</td>
<td>31%</td>
</tr>
</tbody>
</table>


Precision – What to look at

- Size of the P-value (never by itself)
- Size of the confidence intervals around the outcome measure
- Have the authors considered all the important variables? Made adjustments? Were the findings robust?
- How was the effect of subjects refusing to participate evaluated? Sensitivity analyses (see validity)
P-values

- Hypothesis-testing
- Tell us NOTHING on their own about precision or sample size
- Arbitrarily set value to examine the possibility that the result could have occurred by chance
- If p<0.05 this doesn’t make the result “true”
- So don’t use/accept a p-value on its own

* p values/CIs covered in Introductory Biostatistics seminar
   Available on REP website

Confidence Intervals: Measure of Precision

- A range of values around the estimated effect that has a high probability of containing the “true” value
- Also arbitrarily set (95%)
- If the range includes 1 (e.g. RR, OR) then potentially NO difference between groups
- Bigger sample size → smaller confidence intervals
  = greater precision around the estimate
- *Overlapping CIs when making comparisons b/w groups

Formulae for different CIs – means, proportions, differences in means/proportions etc

Can I apply the results in my setting?

Application of the Results

- Inclusion/exclusion criteria
- Socioeconomic, cultural, demographic differences
- Typical patient in the study, range ("Table 1")
- Was the study population selection process appropriate?
- Was the study population similar to the source population?
  - Participation, drop-outs, loss to follow-up?
  - (sensitivity analyses)
Applicability

- Were all relevant harms AND benefits considered?
- Do the benefits outweigh the harms?
- Individual preference?
- Is it feasible? Affordable?

Important Questions to Ask

- What if I do nothing?
- What other options do I have?
- What are the benefits and harms of all the options?
- Do I have enough information to make a decision?
  - Other studies?

Check Lists

- CASP (Critical Appraisal Skills Program)
  - systematic reviews, RCT, qualitative research, economic review, case-control studies, cohort studies
- JAMA – User’s guides/teaching papers
- Journal American Medical Association
  - set of excellent teaching papers
- AGREE site – clinical practice guideline checklist
  - Appraisal of Guidelines Research & Evaluation

Check Lists

- CONSORT randomised controlled trials
- TREND non-randomised controlled trials
- STROBE observational studies in epidemiology
- PRISMA systematic reviews and meta-analyses
- And many others...
Evidence-Based Journal Clubs

- Focus on current real patient problems
- Bring questions, sense of humour, good food
- Distribute topics and roles.
- Bring enough copies
- Keep handy multiple copies of quick appraisal tools.
- Keep a log of questions asked and answered.
- Finish with the group’s bottom line, and any follow up actions (e.g., tools, flowchart, audits, and further searches).

Critical Appraisal: Take Home Messages

- Only read to answer important, current questions
- Choose the “best” sources to answer your question
- Don’t believe everything you read!
- Choose and use a checklist for efficient appraisal
- Assess validity
- Interpret results – size, precision +
- Use common sense – Bradford Hill
- Consider relevance in your setting
- Don’t forget practical issues – cost, feasibility
- Don’t change your practice based on one paper
- Consider a journal club!

Questions?

Upcoming Research Skills Seminars:

18 May 12:30-13:30 Consumer and Community Involvement
Anne McKenzie AM

25 May 12:30-13:30 Survey Design & Techniques
A/Prof Sue Skull

*Full 2018 RSSS calendar in back of handouts

Please give us feedback!
A survey is included in the back of your handout or complete it online via:
www.surveymonkey.com/r/criticalappraisal18

ResearchEducationProgram.org
ResearchEducationProgram@health.wa.gov.au
1. CRITICAL APPRAISAL – ADDITIONAL NOTES AND RESOURCES

1.1 WHAT IS EVIDENCE-BASED CLINICAL PRACTICE?
Evidence-based medicine is the integration of best research evidence with clinical expertise and patient values” (David Sackett). Evidence-based clinical practise involves critically interpreting available evidence, and applying it appropriately to clinical situations. The 5 key steps of EBCP are:

A) Formulate an answerable question – PICOT (to ensure all key elements of a question are included)
   - P patient or population
   - I intervention or exposure or test
   - C comparison
   - O outcome
   - T time

B) Track down the best evidence

C) Critically appraise the evidence for:
   - Validity
   - Impact (size of the benefit)
   - Applicability/usefulness

D) Integrate with clinical expertise and patient values

E) Evaluate our effectiveness and efficiency - keep a record; improve the process

1.2 SO WHAT IS CRITICAL APPRAISAL
Critical appraisal is the process of carefully and systematically examining research to judge its trustworthiness, its value and relevance in a particular context (Burls, 2009).

1.3 WHY DO WE NEED CRITICAL APPRAISAL?
   a) Mass of rapidly expanding scientific literature – need to have a relevant, efficient approach.
   b) Need to shift focus to current clinical issues/problems (“just in time” education), which is relevant to our practice, up to date and memorable (rather than “just in case” reading where we try to read everything that crosses our desk in case we might need it one day). RCTs, systematic reviews and meta-analyses provide the highest level of evidence and should be the focus wherever possible. It’s impossible to read everything relevant to your discipline.
   c) We are (currently) poorly equipped to tell good from bad research (read Schroter et al for further explanation – see below)

1.4 KEY CONCEPTS UNDERPINNING CRITICAL APPRAISAL
1. Key concept 1: No study is perfect. (→ why we must assess Study Validity)
   All research is flawed. We need to determine whether there are enough flaws to discard it, or interpret/use it. To meaningfully interpret results, as a minimum, a paper must have:
      a. Sufficient detail to assess the key elements making up study validity
      b. The right study design to be able to answer the study question. Different types of questions will require different kinds of evidence: Is the study design chosen able to answer the study question? There are different things to look for according to each study design/question type – and different check lists available (below) – CONSORT etc.
2. **Key concept 2: There is more than one possible explanation for a reported “effect” found in a study. The published result(s) may reflect:**

   - Truth: a real effect – what we hope from a “good” study
   - Chance: according to a p value pre-determined by the researcher (eg p=0.05). Any statistical assessment is one of probability – ie the result occurred by chance
   - Error: an erroneous result due to problems with study design/ implementation/analysis/ interpretation

3. **Key concept 3: Not all results can be applied in the setting in which you work.**
   Can we apply the results in our setting? – more detail below

4. **Key concept 4: Don’t rely on one study (unless there really is only one!) to change practice.**
   Studies on the same topic will always have different estimates of effect and often different conclusions. This is why well-conducted systematic reviews/meta-analyses provide stronger evidence than a single trial.

1.5 **ARE THE RESULTS VALID? (STUDY QUALITY)**

**Internal validity.** Is the study design, conduct, and analysis such that the study results are likely to reflect a true answer to the study question? To examine this, we must rule out the influences of bias and confounding that might be contributing to observed differences in outcomes between the treatment/exposure groups, or the measure of effect attributable to the study.

There are different check lists for internal validity according to the type of study: look at CEBM website (https://www.cebm.net/2014/06/critical-appraisal/) and at the end of these notes.

Internal validity will depend on how the study was:

- designed
- conducted
- analysed
- interpreted and reported

Important concepts:

- Bias or “systematic error” Occurs when measurements deviate systematically from the true state of the attribute (eg sick people more likely to remember an exposure than well people)
- Noise or “Random error” Occurs when repeated measurements of the same attribute do not agree, but there is no systematic deviation from the true state of the attribute (eg measuring head circumference three times)
- Confounding: a confounder is a baseline variable or intervention that is extraneous to the study question, but potentially related to the outcome and is differentially applied to the intervention and control groups. Or: Were there alternative factors which differed between the compared groups that could have accounted for the outcome?
  
  http://www.teachepi.org/documents/courses/fundamentals/Pai_Lecture8_Confounding_Part2.pdf
- Study design features can minimise bias and confounding. But bias cannot be “fixed” once it has occurred; only described as a study weakness to allow appropriate interpretation of results.
1.6 WHAT ARE THE RESULTS?
(Study effect size and interpretation including precision)

3 big things to consider:
1) Common sense, including clinical relevance (eg 1mmHg difference in BP) – Bradford Hill criteria, interpretation
2) Size of the effect (RR, OR, NNT, RD, ARR, etc) – big is relevant
3) Precision (was it measured appropriately, was the effect precise)

Summary measures of effect (measures of “occurrence”) – (using CIs wherever possible – see below)
• Categorical variables:
  • Prevalence
  • Incidence/absolute risk
  • Relative risk, Odds ratio
  • Absolute risk reduction or risk difference
  • Relative risk reduction,
  • Number needed to treat/harm
• Continuous variables:
  • Mean difference in final outcome measure (eg difference in means) (eg on average participants receiving an intervention scored 10 points lower than the control group)
  • Standardised mean difference (difference in means/pooled standard deviation of both groups – need SD and size of each group).

Prevalence = counts of events at one point in time / total number of people who could have had the event (the population at risk). It is not a rate as there is no time component.
Incidence = count of new events / population at risk over a given time period (a true rate described per unit of time). Those who already have the event at the start of the time period are excluded because they are not part of the population at risk.
Relative risk (RR) = prevalence in group 1/prevalence in group 2
Risk difference (RD) = prevalence in group 1 - prevalence in group 2 = same thing as absolute risk reduction (absolute difference between two rates)
Relative risk reduction (RRR) = (1-RR) x 100% (proportional difference between two rates)
Number needed to treat (NNT) = 1/RD

Patient and clinican acceptance DOES vary according to which results are presented.

Because measures of relative risk are relative to the comparator group, they can seem big or important. Very important to remember that the risk benefit ratio depends on the ABSOLUTE (baseline) risk.

Note that for diagnostic-type Questions, other measures of effect are generally used:
• Sensitivity, specificity
• Positive predictive value
• Negative predictive value
• Likelihood ratio positive
• Likelihood ratio negative

Please see the CEBM website for formulae etc – Specific diagnostic study materials are listed below.
Precision
How precise is the estimate of risk? Consider:
• Size of the P-value (never by itself)
• Size of the confidence intervals
• Have all the important variables been considered? Adjustments made? Were results robust?
• Was the effect of subjects refusing to participate evaluated? Sensitivity analyses? (see validity)

1.7 CAN WE APPLY THE RESULTS IN OUR SETTING?
• Consider external validity (generalizability), applicability, and individual preference (when considering patients or clients)

IMPORTANT QUESTIONS TO ASK
• What if I do nothing?
• What other options do I have?
• What are the benefits and harms of all the options?
• Do I have enough information to make a decision?

SUMMARY OF CONSIDERATIONS IN CRITICIAL APPRAISAL
• Overall validity and quality
• Consistency with other studies/evidence, Bradford-Hill criteria
• Interpretation of results
• Relevance to your patient
• Practical issues (e.g. local costs, feasibility)

EXAMPLE CHECK LISTS TO HELP WITH APPRAISING EVIDENCE (SEE SOME EXAMPLES BELOW)
Useful sites for Appraisal checklists:
• https://www.cebm.net/2014/06/critical-appraisal/

• Critical Appraisal Tools: International Centre for Allied Health Evidence, University of SA
• CASP (critical appraisal skills program) - UK site
  – systematic reviews, RCT, qualitative research, economic review, case-control studies, cohort studies
• JAMA – User’s guides/teaching papers
  – Journal American Medical Association- set of excellent teaching papers- also new book
• AGREE site – clinical practice guideline checklist: Appraisal of Guidelines Research & Evaluation
• Prognostic studies
• CONSORT randomised controlled trials
http://www.consort-statement.org/media/default/downloads/consort%202010%20checklist.pdf
• TREND non-randomised controlled trials
• STROBE observational studies in epidemiology e.g. for case-control studies:
• PRISMA systematic reviews and meta-analyses
1.8 KEY RESOURCES

- Centre for Evidence-based Medicine. University of Oxford (including Critical Appraisal Sheets) [http://www.cebm.net/]
- Critical Appraisal Tools (Check lists). CEBM Oxford. [http://www.cebm.net/critical-appraisal/]
- Critical Appraisal Skills Programme, UK [http://www.casp-uk.net/]
- Jackson et. al., The GATE frame: critical appraisal with pictures, EBM 2006 [http://ebm.bmj.com/content/11/2/35.extract]
### 1.9 EXAMPLE: DIAGNOSTIC ACCURACY STUDY: ARE THE RESULTS OF THE STUDY VALID?

#### R - Was the diagnostic test evaluated in a Representative spectrum of patients (like those in whom it would be used in practice)?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>It is ideal if the diagnostic test is applied to the full spectrum of patients - those with mild, severe, early and late cases of the target disorder. It is also best if the patients are randomly selected or consecutive admissions so that selection bias is minimized.</td>
<td>The <strong>Methods</strong> section should tell you how patients were enrolled and whether they were randomly selected or consecutive admissions. It should also tell you where patients came from and whether they are likely to be representative of the patients in whom the test is to be used.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐

Comment:

#### A – Was the reference standard ascertained regardless of the index test result?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ideally both the index test and the reference standard should be carried out on all patients in the study. In some situations where the reference standard is invasive or expensive there may be reservations about subjecting patients with a negative index test result (and thus a low probability of disease) to the reference standard. An alternative reference standard is to follow-up people for an appropriate period of time (dependent on disease in question) to see if they are truly negative.</td>
<td>The <strong>Methods</strong> section should indicate whether or not the reference standard was applied to all patients or if an alternative reference standard (e.g., follow-up) was applied to those who tested negative on the index test.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐

Comment:

#### Mbo - Was there an independent, blind comparison between the index test and an appropriate reference ('gold') standard of diagnosis?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>There are two issues here. First the reference standard should be appropriate - as close to the 'truth' as possible. Sometimes there may not be a single reference test that is suitable and a combination of tests may be used to indicate the presence of disease. Second, the reference standard and the index test being assessed should be applied to each patient independently and blindly. Those who interpreted the results of one test should not be aware of the results of the other test.</td>
<td>The <strong>Methods</strong> section should have a description of the reference standard used and if you are unsure of whether or not this is an appropriate reference standard you may need to do some background searching in the area. The <strong>Methods</strong> section should also describe who conducted the two tests and whether each was conducted independently and blinded to the results of the other.</td>
</tr>
</tbody>
</table>
This paper: Yes ☐  No ☐  Unclear ☐

Comment:

What were the results?

Are test characteristics presented?

There are two types of results commonly reported in diagnostic test studies. One concerns the accuracy of the test and is reflected in the sensitivity and specificity. The other concerns how the test performs in the population being tested and is reflected in predictive values (also called post-test probabilities). To explore the meaning of these terms, consider a study in which 1000 elderly people with suspected dementia undergo an index test and a reference standard. The prevalence of dementia in this group is 25%. 240 people tested positive on both the index test and the reference standard and 600 people tested negative on both tests. The first step is to draw a 2 x 2 table as shown below. We are told that the prevalence of dementia is 25% therefore we can fill in the last row of totals - 25% of 1000 people is 250 - so 250 people will have dementia and 750 will be free of dementia. We also know the number of people testing positive and negative on both tests and so we can fill in two more cells of the table.

<table>
<thead>
<tr>
<th>Reference Standard</th>
<th>+ve</th>
<th>-ve</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ve</td>
<td>240</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-ve</td>
<td>600</td>
<td></td>
<td>1000</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
<td>1000</td>
</tr>
</tbody>
</table>

By subtraction we can easily complete the table:

<table>
<thead>
<tr>
<th>Reference Standard</th>
<th>+ve</th>
<th>-ve</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ve</td>
<td>240</td>
<td>150</td>
<td>390</td>
</tr>
<tr>
<td>-ve</td>
<td>10</td>
<td>600</td>
<td>610</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>750</td>
<td>1000</td>
</tr>
</tbody>
</table>

Now we are ready to calculate the various measures.

<table>
<thead>
<tr>
<th>What is the measure?</th>
<th>What does it mean?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sensitivity (Sn)</strong> = the proportion of people with the condition who have a positive test result.</td>
<td>The sensitivity tells us how well the test identifies people with the condition. A highly sensitive test will not miss many people.</td>
</tr>
<tr>
<td>In our example, the Sn = 240/250 = 0.96</td>
<td>10 people (4%) with dementia were falsely identified as not having it. This means the test is fairly good at identifying people with the condition.</td>
</tr>
<tr>
<td><strong>Specificity (Sp)</strong> = the proportion of people without the condition who have a negative test result.</td>
<td>The specificity tells us how well the test identifies people without the condition. A highly specific test will not falsely identify many people as having the condition.</td>
</tr>
<tr>
<td>In our example, the Sp = 600/750 = 0.80</td>
<td>150 people (20%) without dementia were falsely identified as having it. This means the test is only moderately good at identifying people without the condition.</td>
</tr>
</tbody>
</table>
Positive Predictive Value (PPV) = the proportion of people with a positive test who have the condition.

In our example, the PPV = 240/390 = 0.62

This measure tells us how well the test performs in this population. It is dependent on the accuracy of the test (primarily specificity) and the prevalence of the condition.

Of the 390 people who had a positive test result, 62% will actually have dementia.

Negative Predictive Value (NPV) = the proportion of people with a negative test who do not have the condition.

In our example, the NPV = 600/610 = 0.98

This measure tells us how well the test performs in this population. It is dependent on the accuracy of the test and the prevalence of the condition.

Of the 610 people with a -ve test, 98% will not have dementia.

Application

1. Were the methods for performing the test described in sufficient detail to permit replication?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The article should have sufficient description of the test to allow its replication and also interpretation of the results.</td>
<td>The Methods section should describe the test in detail.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐

Comment:

1.10 EXAMPLE: THERAPY STUDY: ARE THE RESULTS OF THE TRIAL VALID? (INTERNAL VALIDITY)

What question did the study ask?

Patients –
Intervention -
Comparison -
Outcome(s) –

1a. R- Was the assignment of patients to treatments randomised?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Centralised computer randomisation is ideal and often used in multi-centred trials. Smaller trials may use an independent person (e.g., the hospital pharmacy) to “police” the randomization.</td>
<td>The Methods should tell you how patients were allocated to groups and whether or not randomisation was concealed.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐

Comment:
<table>
<thead>
<tr>
<th>1b. R - Were the groups similar at the start of the trial?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>If the randomisation process worked (that is, achieved comparable groups) the groups should be similar. The more similar the groups the better it is. There should be some indication of whether differences between groups are statistically significant (ie. p values).</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐  
Comment: 

<table>
<thead>
<tr>
<th>2a. A - Aside from the allocated treatment, were groups treated equally?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>Apart from the intervention the patients in the different groups should be treated the same, eg., additional treatments or tests.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐  
Comment: 

<table>
<thead>
<tr>
<th>2b. A - Were all patients who entered the trial accounted for? – and were they analysed in the groups to which they were randomised?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>Losses to follow-up should be minimal – preferably less than 20%. However, if few patients have the outcome of interest, then even small losses to follow-up can bias the results. Patients should also be analysed in the groups to which they were randomised – ‘intention-to-treat analysis’.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐  
Comment: 

<table>
<thead>
<tr>
<th>3. M - Were measures objective or were the patients and clinicians kept “blind” to which treatment was being received?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>It is ideal if the study is ‘double-blinded’ – that is, both patients and investigators are unaware of treatment allocation. If the outcome is objective (eg., death) then blinding is less critical. If the outcome is subjective (eg., symptoms or function) then blinding of the outcome assessor is critical.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐  
Comment: 

---

**Note:**

- The comments and questions are based on the structure and content of the provided table. The table is designed to help assess the comparability of groups at the start of the trial, adherence to the allocated treatment, and the objective nature of the outcome measures. Each section is followed by a question on the best approach and where to find the information in the results section.
What were the results?

### 1. How large was the treatment effect?

Most often results are presented as dichotomous outcomes (yes or no outcomes that happen or don’t happen) and can include such outcomes as cancer recurrence, myocardial infarction and death. Consider a study in which 15% (0.15) of the control group died and 10% (0.10) of the treatment group died after 2 years of treatment. The results can be expressed in many ways as shown below.

<table>
<thead>
<tr>
<th>What is the measure?</th>
<th>What does it mean?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Relative Risk (RR)</strong> = risk of the outcome in the treatment group / risk of the outcome in the control group.</td>
<td>The relative risk tells us <strong>how many times more likely</strong> it is that an event will occur in the treatment group relative to the control group. An RR of 1 means that there is no difference between the two groups thus, the treatment had <strong>no effect</strong>. An RR &lt; 1 means that the treatment decreases the risk of the outcome. An RR &gt; 1 means that the treatment increased the risk of the outcome.</td>
</tr>
<tr>
<td>In our example, the RR = 0.10/0.15 = 0.67</td>
<td>Since the RR &lt; 1, the treatment decreases the risk of death.</td>
</tr>
<tr>
<td><strong>Absolute Risk Reduction (ARR)</strong> = risk of the outcome in the control group - risk of the outcome in the treatment group. This is also known as the <strong>absolute risk difference</strong>.</td>
<td>The absolute risk reduction tells us the absolute difference in the rates of events between the two groups and gives an indication of the baseline risk and treatment effect. An ARR of 0 means that there is no difference between the two groups thus, the treatment had <strong>no effect</strong>.</td>
</tr>
<tr>
<td>In our example, the ARR = 0.15 - 0.10 = 0.05 or 5%</td>
<td>The absolute benefit of treatment is a 5% reduction in the death rate.</td>
</tr>
<tr>
<td><strong>Relative Risk Reduction (RRR)</strong> = absolute risk reduction / risk of the outcome in the control group. An alternative way to calculate the RRR is to subtract the RR from 1 (eg. RRR = 1 - RR)</td>
<td>The relative risk reduction is the complement of the RR and is probably the most commonly reported measure of treatment effects. It tells us the reduction in the rate of the outcome in the treatment group relative to that in the control group.</td>
</tr>
<tr>
<td>In our example, the RRR = 0.05/0.15 = 0.33 or 33% Or RRR = 1 - 0.67 = 0.33 or 33%</td>
<td>The treatment reduced the risk of death by 33% relative to that occurring in the control group.</td>
</tr>
<tr>
<td><strong>Number Needed to Treat (NNT)</strong> = inverse of the ARR and is calculated as 1 / ARR.</td>
<td>The number needed to treat represents the number of patients we need to treat with the experimental therapy in order to prevent 1 bad outcome and incorporates the duration of treatment. Clinical significance can be determined to some extent by looking at the NNTs, but also by weighing the NNTs against any harms or adverse effects (NNHs) of therapy.</td>
</tr>
<tr>
<td>In our example, the NNT = 1/ 0.05 = 20</td>
<td>We would need to treat 20 people for 2 years in order to prevent 1 death.</td>
</tr>
</tbody>
</table>
2. How precise was the estimate of the treatment effect?

The true risk of the outcome in the population is not known and the best we can do is estimate the true risk based on the sample of patients in the trial. This estimate is called the point estimate. We can gauge how close this estimate is to the true value by looking at the confidence intervals (CI) for each estimate. If the confidence interval is fairly narrow then we can be confident that our point estimate is a precise reflection of the population value. The confidence interval also provides us with information about the statistical significance of the result. If the value corresponding to no effect falls outside the 95% confidence interval then the result is statistically significant at the 0.05 level. If the confidence interval includes the value corresponding to no effect then the results are not statistically significant.

Will the results help me in caring for my patient? (External Validity/Applicability)

The questions that you should ask before you decide to apply the results of the study to your patient are:

- Is my patient so different to those in the study that the results cannot apply?
- Is the treatment feasible in my setting?
- Will the potential benefits of treatment outweigh the potential harms of treatment for my patient?

1.11 Example: Systematic Review: Are the Results of the Review Valid?

<table>
<thead>
<tr>
<th>What question (PICO) did the systematic review address?</th>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The main question being addressed should be clearly stated. The exposure, such as a therapy or diagnostic test, and the outcome(s) of interest will often be expressed in terms of a simple relationship.</td>
<td>The Title, Abstract or final paragraph of the Introduction should clearly state the question. If you still cannot ascertain what the focused question is after reading these sections, search for another paper!</td>
<td></td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐
Comment:

Is it unlikely that important, relevant studies were missed?

<table>
<thead>
<tr>
<th>What is best?</th>
<th>Where do I find the information?</th>
</tr>
</thead>
<tbody>
<tr>
<td>The starting point for comprehensive search for all relevant studies is the major bibliographic databases (e.g., Medline, Cochrane, EMBASE, etc) but should also include a search of reference lists from relevant studies, and contact with experts, particularly to inquire about unpublished studies. The search should not be limited to English language only. The search strategy should include both MESH terms and text words.</td>
<td>The Methods section should describe the search strategy, including the terms used, in some detail. The Results section will outline the number of titles and abstracts reviewed, the number of full-text studies retrieved, and the number of studies excluded together with the reasons for exclusion. This information may be presented in a figure or flow chart.</td>
</tr>
</tbody>
</table>

This paper: Yes ☐ No ☐ Unclear ☐
Comment:
<table>
<thead>
<tr>
<th>Were the criteria used to select articles for inclusion appropriate?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>The inclusion or exclusion of studies in a systematic review should be clearly defined a priori. The eligibility criteria used should specify the patients, interventions or exposures and outcomes of interest. In many cases the type of study design will also be a key component of the eligibility criteria.</td>
</tr>
<tr>
<td><strong>This paper:</strong> Yes ☐ No ☐ Unclear ☐</td>
</tr>
<tr>
<td><strong>Were the included studies sufficiently valid for the type of question asked?</strong></td>
</tr>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>The article should describe how the quality of each study was assessed using predetermined quality criteria appropriate to the type of clinical question (e.g., randomization, blinding and completeness of follow-up)</td>
</tr>
<tr>
<td><strong>This paper:</strong> Yes ☐ No ☐ Unclear ☐</td>
</tr>
<tr>
<td><strong>Were the results similar from study to study?</strong></td>
</tr>
<tr>
<td><strong>What is best?</strong></td>
</tr>
<tr>
<td>Ideally, the results of the different studies should be similar or homogeneous. If heterogeneity exists the authors may estimate whether the differences are significant (chi-square test). Possible reasons for the heterogeneity should be explored.</td>
</tr>
<tr>
<td><strong>This paper:</strong> Yes ☐ No ☐ Unclear ☐</td>
</tr>
</tbody>
</table>
What were the results?

How are the results presented?

A systematic review provides a summary of the data from the results of a number of individual studies. If the results of the individual studies are similar, a statistical method (called meta-analysis) is used to combine the results from the individual studies and an overall summary estimate is calculated. The meta-analysis gives weighted values to each of the individual studies according to their size. The individual results of the studies need to be expressed in a standard way, such as relative risk, odds ratio or mean difference between the groups. Results are traditionally displayed in a figure, like the one below, called a forest plot.

Comparison: B3 Treatment versus Placebo
Outcome: 01 Effect of treatment on mortality

<table>
<thead>
<tr>
<th>Study</th>
<th>Treatment n/M</th>
<th>Control n/M</th>
<th>OR (95% CI Fixed)</th>
<th>Weight %</th>
<th>OR (95% CI Fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown 1998</td>
<td>24 / 472</td>
<td>35 / 499</td>
<td></td>
<td>9.6</td>
<td>0.71 (0.42, 1.21)</td>
</tr>
<tr>
<td>Geoffrey 1997</td>
<td>120 / 2050</td>
<td>102 / 2250</td>
<td><strong>88</strong></td>
<td>1.0</td>
<td>0.54 (0.29, 0.91)</td>
</tr>
<tr>
<td>Mason 1996</td>
<td>86 / 2061</td>
<td>84 / 2290</td>
<td></td>
<td>24.4</td>
<td>0.63 (0.45, 0.82)</td>
</tr>
<tr>
<td>Febers 2000</td>
<td>5 / 691</td>
<td>4 / 728</td>
<td></td>
<td>1.1</td>
<td>1.22 (0.94, 1.58)</td>
</tr>
<tr>
<td>Scott 1999</td>
<td>31 / 798</td>
<td>45 / 732</td>
<td></td>
<td>15.1</td>
<td>0.60 (0.42, 0.85)</td>
</tr>
<tr>
<td>Total (95% CI)</td>
<td>200 / 6242</td>
<td>351 / 6227</td>
<td></td>
<td>100.0</td>
<td>0.68 (0.56, 0.83)</td>
</tr>
</tbody>
</table>

Test for heterogeneity chi-square = 0.92, df = 4, p = 0.92
Test for overall effect: z = 4.02, p < 0.0001

The forest plot depicted above represents a meta-analysis of 5 trials that assessed the effects of a hypothetical treatment on mortality. Individual studies are represented by a black square and a horizontal line, which corresponds to the point estimate and 95% confidence interval of the odds ratio. The size of the black square reflects the weight of the study in the meta-analysis. The solid vertical line corresponds to 'no effect' of treatment - an odds ratio of 1.0. When the confidence interval includes 1 it indicates that the result is not significant at conventional levels (P > 0.05).

The diamond at the bottom represents the combined or pooled odds ratio of all 5 trials with its 95% confidence interval. In this case, it shows that the treatment reduces mortality by 34% (OR 0.66 95% CI 0.56 to 0.78). Notice that the diamond does not overlap the 'no effect' line (the confidence interval doesn’t include 1) so we can be assured that the pooled OR is statistically significant. The test for overall effect also indicates statistical significance (p < 0.0001).

Exploring heterogeneity

Heterogeneity can be assessed using the “eyeball” test or more formally with statistical tests, such as the Cochran Q test. With the “eyeball” test one looks for overlap of the confidence intervals of the trials with the summary estimate. In the example above note that the dotted line running vertically through the combined odds ratio crosses the horizontal lines of all the individual studies indicating that the studies are homogenous. Heterogeneity can also be assessed using the Cochran chi-square (Cochran Q). If Cochran Q is statistically significant there is definite heterogeneity. If Cochran Q is not statistically significant but the ratio of Cochran Q and the degrees of freedom (Q/df) is > 1 there is possible heterogeneity. If Cochran Q is not statistically significant and Q/df is < 1 then heterogeneity is very unlikely. In the example above Q/df is < 1 (0.92/4 = 0.23) and the p-value is not significant (0.92) indicating no heterogeneity.

Note: The level of significance for Cochran Q is often set at 0.1 due to the low power of the test to detect heterogeneity.
1.12 EXAMPLE: QUESTIONS TO ASK OF YOUR PAPER (with thanks to T KEARNS)

1. What is the research question?
   - Is it concerned with evaluating an intervention, determining causality or determining the magnitude of a health problem?
   - Is there a clearly stated research question or hypothesis?

2. What is the study factor (intervention or exposure)?
   - Is there consistency in the intervention or is there potential for measurement error of the exposure

3. What is the comparator or control?
   - Are the controls appropriate for the exposed group?
   - Is the control intervention standard care?

4. What are the outcome factors (primary and secondary outcomes of interest)?
   - Are all relevant outcomes assessed?
   - how important are omitted outcomes
   - Are they well defined, including timing?
   - Is there measurement error?
   - could measurement error be an important source of bias?

5. What is the study type?
   - Is this the best study type for the question being investigated?
   - If not, why not?
   - If not, which study type would provide the best evidence to answer the question?
   - If not, how useful are the results produced by this type of study?

6. What are the main methodological issues that need to be considered in this type of study?
   - For experimental studies, how were participants assigned to groups?
   - How many reached final follow up?
   - For case-control studies, are the controls appropriate?
   - Have the authors considered any methodological issues in sufficient detail?
   - If not, what effects may this have on the results and conclusions drawn?
   - Analysis
     ▪ Were the analyses undertaken appropriate for the data?
     ▪ Unadjusted
     ▪ Adjusted
   - Is an estimate of effect size given?
     ▪ Are absolute effects given?
     ▪ Is an exact p-value presented
     ▪ Are confidence intervals given?
   - Is clinical significance discussed?

7. Is the sample size adequate to rule out a “chance” effect (or lack of effect)?
   - Is there evidence of a sample size calculation?
   - Is the sample size big enough to detect a clinically significant result?
8. Is there any potential for bias (consider selection bias and measurement bias)?
   - If yes, what are these?
   - If yes, have the authors dealt with them adequately?
   - If there is bias, how could that affect the result and conclusions, and the external validity of the study?

9. Is there any potential for confounding?
   - If yes, detail what those confounders might be
   - If yes, have the authors adequately dealt with these (i.e. examined and controlled for)

10. What is/are the main finding(s) of the research?
    - Do you believe the authors’ conclusions? Are they valid?
    - If you don’t believe them, why not?
    - If you don’t believe them, what should you do next to find “the truth”?
    - If you do believe them, are the results generalisable to the wider population (that is, outside the people who participated in this study?)
    - Do they answer the original research question?

Are the results of this aetiology (risk factor) study valid?

Did the study address a clearly focused issue?
A question can be focused in terms of:
   - the population studied
   - the risk factors studied
   - whether the study tried to detect a beneficial or harmful effect?

Did the authors use an appropriate method to answer their question?
Consider:
   - is the study method an appropriate way of answering the question under the circumstances? (Is the outcome rare or harmful?)
   - did it address the study question?

Were the cases recruited in an acceptable way?
HINT: We are looking for selection bias which might compromise the validity of the findings:
   - Are the cases defined precisely?
   - Were the cases representative of a defined population (geographically and/or temporally)?
   - Was there an established reliable system for selecting all the cases?
   - Are they incident or prevalent?
   - Is there something special about the cases?
   - Is the time frame of the study relevant to the disease/exposure?
   - Was there a sufficient number of cases selected?
   - Was there a power calculation?
Were the controls selected in an acceptable way?
HINT: We are looking for selection bias which might compromise the generalisability of the findings:
- Were the controls representative of a defined population (geographically and/or temporally)?
- Was there something special about the controls?
- Was the non-response high? Could non-respondents be different in any way?
- Are they matched, population based or randomly selected?
- Was there a sufficient number of controls selected?

Was the exposure accurately measured to minimise bias?
HINT: We are looking for measurement, recall or classification bias:
- Was the exposure clearly defined and accurately measured?
- Did the authors use subjective or objective measurements?
- Do the measures truly reflect what they are supposed to measure? (have they been validated?)
- Were the measurement methods similar in cases and controls?
- Did the study incorporate blinding where feasible?
- Is the temporal relation correct? (does the exposure of interest precede the outcome?)

What confounding factors have the authors accounted for?
Hint: List the other ones you think might be important, that the authors missed (genetic, environmental and socio-economic)

Have the authors taken account of the potential confounding factors in the design and/or in their analysis?
HINT: Look for restriction in design, and techniques, e.g. modelling, stratified-, regression-, or sensitivity analysis to correct, control or adjust for confounding factors.
1.13 EXAMPLE: A CHECKLIST FOR APPRAISING RANDOMISED CONTROLLED TRIALS

1. Was the objective of the trial sufficiently described?

2. Was a satisfactory statement given of the diagnostic criteria for entry to the trial?

3. Were concurrent controls used (as opposed to historical controls)?

4. Were the treatments well defined?

5. Was random allocation to treatments used?

6. Was the potential degree of blindness used?

7. Was there a satisfactory statement of criteria for outcome measures? Was a primary outcome measure identified?

8. Were the outcome measures appropriate?

9. Was a pre-study calculation of required sample size reported?

10. Was the duration of post-treatment follow-up stated?

11. Were the treatment and control groups comparable in relevant measures?

12. Were a high proportion of the subjects followed up?

13. Were the drop-outs described by treatment and control groups?

14. Were the side-effects of treatment reported?

15. How were the ethical issues dealt with?

16. Was there a statement adequately describing or referencing all statistical procedures used?

17. What tests were used to compare the outcome in test and control patients?

18. Were 95% confidence intervals given for the main results?

19. Were any additional analyses done to see whether baseline characteristics (prognostic factors) influenced the outcomes observed?

20. Were the conclusions drawn from the statistical analyses justified?

With thanks to Dr Peter Morris, MSHR Darwin
Every researcher should be actively involving consumer or community members to improve quality and increase impact of their research.

Community involvement is increasingly a requirement for funding agencies. This seminar provides a practical introduction and will cover basic principles of community and consumer involvement, the benefits and barriers, and what to put in place to get started.

Anne McKenzie AM
McKenzie AM is the Head of the Consumer and Community Health Research Network which is an enabling platform of the WA Health Translation Network. Anne is a senior consumer representative for state and national health committees and former Chair of the Health Consumers Council WA. In 2004 Anne established the inaugural role of the Parent Advocate at PMH and has a long history of consumer advocacy with WA health services.

Register Online
ResearchEducationProgram.eventbrite.com
Further information: ResearchEducationProgram@health.wa.gov.au
ResearchEducationProgram.org

*Hosted VC Sites Include:*
- Armadale Hospital
- Bunbury Hospital
- Curtin University
- Fiona Stanley Hospital
- Joondalup Health Campus
- King Edward Memorial Hospital
- Midland Community Health Centre
- Royal Perth Hospital
- Sir Charles Gairdner Hospital

*Online VC via Scopia App*

The Research Skills Seminar Series is part of the Research Education Program, Dept. of Child Health Research, Princess Margaret Hospital, WA Department of Health. Seminars are hosted by WA Department of Health.
Surveys, including clinical audits, are one of the most commonly conducted clinical research projects. There is a lot more to doing these well than meets the eye.

This seminar provides practical help for planning and conducting surveys. It includes good survey design, approval pathways, sampling and administration methods, writing high quality questionnaires & data collection instruments, maximising response rates and reducing data errors.

Associate Professor Sue Skull

A/Prof Sue Skull is a paediatrician, epidemiologist, public health physician and researcher. She is an experienced teacher of research methods and enjoys helping others improve and enjoy their research experience.
### Research Skills Seminar Series 2018

**All sessions held Fridays 12:30 – 1:30 PM**  
**Macdonald Lecture Theatre, Princess Margaret Hospital for Children**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic (abbreviated)</th>
<th>Presenter</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Feb 9</td>
<td>Research Fundamentals – Question and Protocol Development</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>2 Feb 23</td>
<td>Introductory Biostatistics – Understanding and reporting research results including P-values and Confidence intervals</td>
<td>Julie Marsh</td>
</tr>
<tr>
<td>3 Mar 9</td>
<td>Introduction into Good Clinical Practice</td>
<td>Natalie Barber</td>
</tr>
<tr>
<td>4 Mar 23</td>
<td>Scientific Writing</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>5 Apr 6</td>
<td>Ethics for Research - Practical Ethical Considerations: Obstacle or Powerful Tool?</td>
<td>Nik Zeps</td>
</tr>
<tr>
<td>6 May 4</td>
<td>Rapid Critical Appraisal of Scientific Literature</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>7 May 18</td>
<td>Consumer and Community Involvement in Research</td>
<td>Anne McKenzie</td>
</tr>
<tr>
<td>8 May 25</td>
<td>Survey Design &amp; Techniques</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>9 Jun 22</td>
<td>Getting the most out of Research Supervision – Tips for Students and Supervisors</td>
<td>Jonathan Carapetis</td>
</tr>
<tr>
<td>10 Jul 20</td>
<td>Knowledge Translation</td>
<td>Fenella Gill/Anthony Kicic</td>
</tr>
<tr>
<td>11 Aug 3</td>
<td>Ethics Processes within Health Research</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>12 Aug 10</td>
<td>Conducting Systematic Reviews</td>
<td>Sonya Girdler</td>
</tr>
<tr>
<td>13 Aug 24</td>
<td>Media and Communication in Research – How to refine and pitch your message for maximum impact</td>
<td>Liz Chester</td>
</tr>
<tr>
<td>14 Sep 7</td>
<td>Sample Size Calculations</td>
<td>Julie Marsh</td>
</tr>
<tr>
<td>15 Sep 14</td>
<td>Data Collection and Management</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>16 Oct 19</td>
<td>Conducting Research with the Indigenous Community</td>
<td>Michael Wright/Sue Skull</td>
</tr>
<tr>
<td>17 Oct 26</td>
<td>Statistical Tips for Interpreting Scientific Claims</td>
<td>Julie Marsh</td>
</tr>
<tr>
<td>18 Nov 2</td>
<td>Oral Presentation of Research Results</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>19 Nov 16</td>
<td>Research Governance</td>
<td>TBC</td>
</tr>
<tr>
<td>20 Nov 30</td>
<td>Grant Applications and Finding Funding</td>
<td>Sue Skull</td>
</tr>
<tr>
<td>21 Dec 7</td>
<td>Qualitative Research Methods <em>NEW</em></td>
<td>Dianne Wynaden</td>
</tr>
</tbody>
</table>

*Topics may move or change; email notice will be provided
*All seminars and corresponding handouts are regularly revised and updated
*Attendance Certificates are available upon request
*Location may change upon move to Perth Children’s Hospital
*For hosted videoconference sites, go to [ResearchEducationProgram.org](http://ResearchEducationProgram.org)
Contact: ResearchEducationProgram@health.wa.gov.au | ResearchEducationProgram.org
**SEMINAR FEEDBACK FORM**

**SEMINAR:** Rapid Critical Appraisal of Scientific Literature  
**DATE:** 4 May 2018  
**LECTURER/S:** Assoc. Prof Sue Skull

**INSTRUCTIONS:** - Please rate the following statements  
- Use a pen to mark circles  
- One response per item  
- Please return to coordinator

1. How did you attend the seminar?  
- Live Seminar at Princess Margaret Hospital  
- Hosted Site Video-Conference  
- Online via Scopia

2. The aim and learning objectives were clear.  
(I felt well informed about scope and content)

3. The learning materials were helpful.  
(handouts, slides, lecture notes supported learning)

4. The session content was well structured.

5. The material was presented in a way which maintained my interest.

6. The material challenged me to think more critically about the subject.  
(covered at the right level)

7. The lecturer/s communicated clearly.  
(lecture presentation, clear instructions)

8. The material presented extended my knowledge in this area.

9. What were the best aspects?

10. What changes would you suggest?

11. How did you hear about the seminar?  
You may select multiple responses

- Email Invitation from Research Skills Seminar Series  
- “Health Happenings” E-News  
- “What’s Happening at CAHS” E-News  
- Healthpoint Intranet  
- Displayed Poster/Flyer  
- Other (please specify):  
  e.g. Curtin email, CACH, colleague, supervisor etc.

THANK YOU - your comments will be provided to the presenter and inform other education activities.