Rapid Critical Appraisal of Scientific Literature

A/Prof Sue Skull
Head of CAHS Research Education Program
Deputy Director of Department of Child Health Research, CAHS

20 September 2019

Critical Appraisal of Scientific Literature

CONTENTS:

Government of Western Australia
Child and Adolescent Health Service

CAHS Research Education Program

Community Health | Mental Health | Perth Children's Hospital

Compassion Collaboration Equity Respect Excellence Accountability

1

Government of Western Australia
Child and Adolescent Health Service

RESEARCH SKILLS SEMINAR SERIES 2019
CAHS Research Education Program

Rapid Critical Appraisal of Scientific Literature

Assoc. Prof Sue Skull
Head of CAHS Research Education Program
Deputy Director of Department of Child Health Research, CAHS

Research Skills Seminar Series | CAHS Research Education Program
Department of Child Health Research | Child and Adolescent Health Service

ResearchEducationProgram.org

Overview

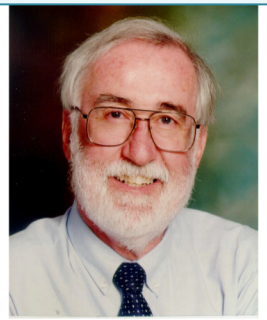
- What is Evidence-Based Clinical Practice
- Critical Appraisal and why we need it
- General critical appraisal strategies
- Examining validity
- Interpreting results
- Application of results

3

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



4

What is Critical Appraisal

5

"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."

Glasziou, Burls and Gilbert. BMJ 2008:337:704-705

6

Why do we need it?

7

Professional obligation

- Maintain knowledge and skills →
- Apply up to date knowledge →
- Provide good clinical care

8

So much to know…

National Library of Medicine MetaThesaurus

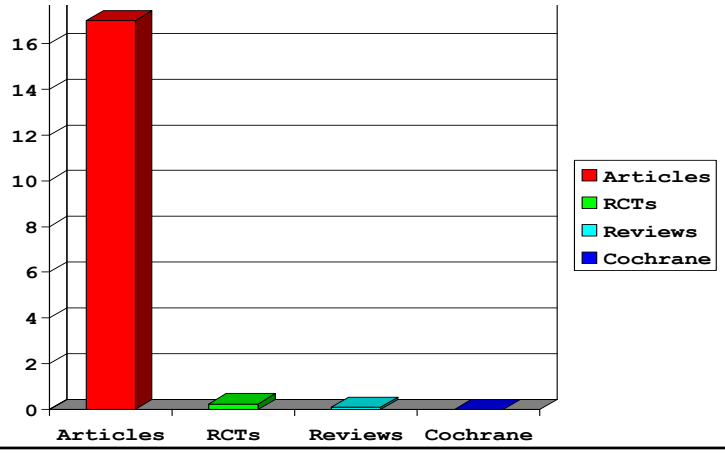
- 1 million biomedical concepts
- 5 million concept names

DiagnosisPro

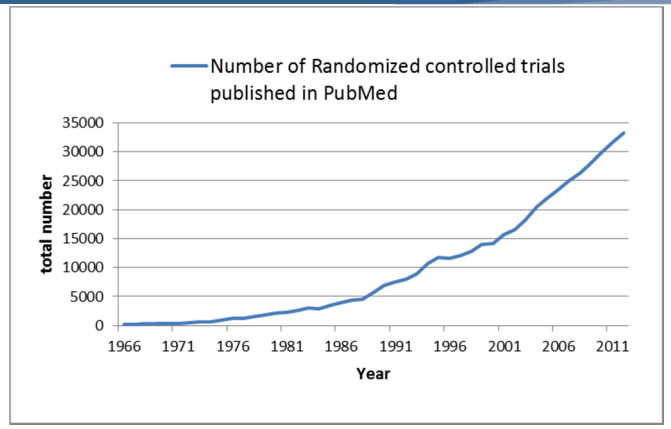
- 11,000 diseases - 1 new disease per day for 30 years
- 30,000 abnormalities (symptoms, signs, lab, X-ray,)
- 3,200 drugs (and FDA's 18,283 products, ≈50 new drugs/yr)

"To cover the vast field of medicine in four years is an impossible task."
William Olser

How much research?



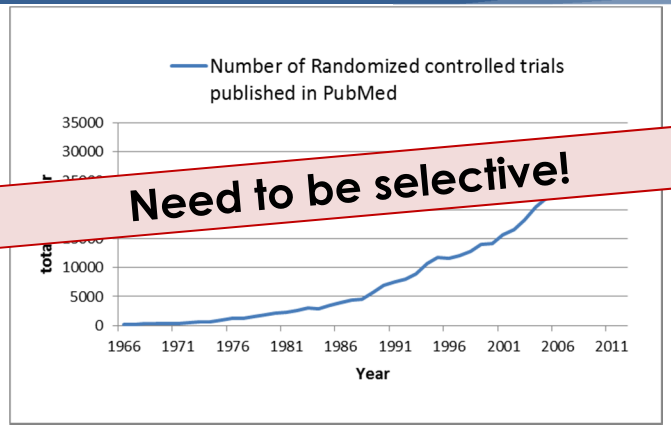
How much research?



Source: Carl Heneghan

11

How much research?



Source: Carl Heneghan

12

"Just in case" Learning

Traditionally "Just in case" or "Push" learning

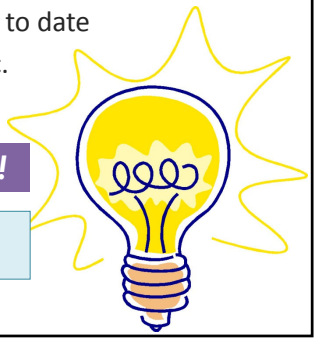
- Alerts to new information e.g. Journal articles, books etc.
- Too much to possibly keep up AND 50% out of date in 5y
- "Meat" in clinical journals?

Journal	% of high quality clinically relevant articles
N Eng J Med	12.6
JAMA	7.2
Lancet	6.2
Ann Intern Med	7.6
BMJ	4.4
Arch Intern Med	2.4
Circulation	1.7
Am J Med	1.1
J Intern Med	1.2
Others	<0.1

13

"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.



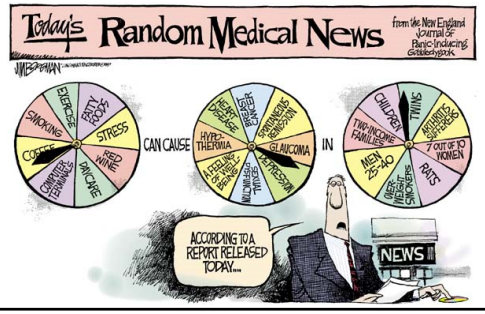
ONLY read articles to answer questions!

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

14

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"

What to read depends on your type of question!

Question	Question type	Study design
How common is the problem?	Prevalence	Survey
Is early detection worthwhile?	Screening	Cross-sectional
Is the diagnostic test accurate?	Diagnosis	Cross-sectional
What caused this problem?	Risk factors	Case-control
What will happen if we do nothing?	Prognosis	Cohort
Does this intervention help?	Treatment	RCT
What are the harms of an intervention?	Treatment	RCT

Adapted from CEBM University of Oxford, Prof Carl Heneghan

16

Types of Questions and Research

Has the best design been used for the question?

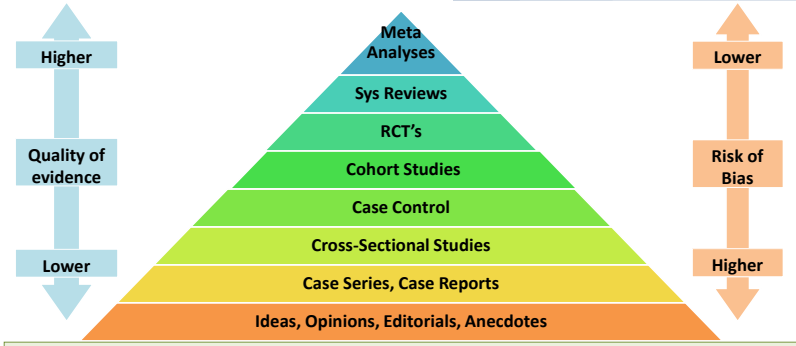
So what should we be reading?

- Must focus on the highest level of evidence where possible
- BUT - pick the right type of design for the type of Q you have

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

Choosing and Finding the "Best Evidence"



"The basis of a good research study is an appropriate study design, one that will best answer the questions you have set with the resources you have available"
East of England Research Development & Support Unit, Norfolk & Suffolk

18

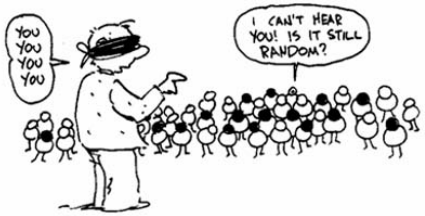
Can we tell "Good" from "Bad" research?

- Study of 607 BMJ reviewers
  - 14 deliberate errors inserted (9 major, 5 minor)

Actually…… No!

- 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.
J R Soc Med 2008; 101(10): 507–514



19

Not all research is useful!

- Ioannidis JPA (2016) **Why Most Clinical Research Is Not Useful**. PLoS Med 13(6): e1002049. 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

20

General Critical Appraisal Strategies

21

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?

If no……

22

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

23

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 and analysis features can improve**

25

Minimising Confounding: Design

- **Randomisation** *(intervention studies only)*
  - with allocation concealment check "baseline" table
  - If not possible, collect information on known confounders
    ↓
- **Restriction**
  - Eliminates variation in the confounder (e.g. males)
  - **+ve:** easy, cheap
  - **-ve:** limits eligible subjects, generalisability, others
- **Matching**
  - Comparison group forced to resemble index group
  - E.g. case-control studies
  - **-ve:** selection of controls, bias!, over matching, analyses

26

Minimising Confounding: Analysis

- **Must have collected information on known confounders**
  - **-ve:** can't fix everything! Unknown confounders
- **Stratification**
  - produce groups within which the confounder does not vary – works best if only one or two strata
  - E.g. analysis within "male" vs "female", or "old" vs "young"
  - If not possible ↓
- **Multivariable analysis (statistical packages required)**
  - **+ve:** can handle multiple potential confounders
  - **-ve:** 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

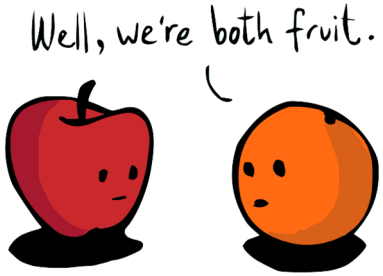
Controlled Trials: Look for Dissimilar Groups

- 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 differences (adjusted vs not)
  - Key reason for failure of randomisation is poor allocation concealment

29

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



30

Loss to Follow-up

- Small is ok, large is not (5 and 20 rule of thumb)
  - Do sensitivity analyses
- Equal is "more OK", <u>selective is not,</u> e.g.:
  - Rapid cure rate in intervention group
  - Unpalatable study intervention
  - Increased mortality in one group

31

Blinding

Figure 1: The authors: double blinded versus single blinded

Figure 2: The authors blinded and masked

Read Schulz and Grimes. Lancet, 2002

32

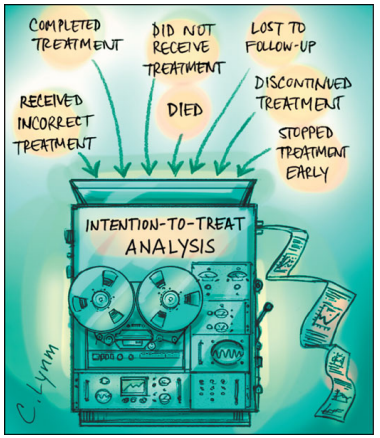
Blinding

- Prevents opinions about efficacy of an intervention affecting:
  - **Treatment** (monitoring, co-interventions)
  - **Measurement** (encouragement, interpretation)
  - **Reporting** (interpretation)

  } →**BIAS**
- Blinding improves compliance and 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 determine:
  - Who: participants, investigators, assessors, analysts
  - How
  - Successful or 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???



34

Intention to Treat Analysis

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

35

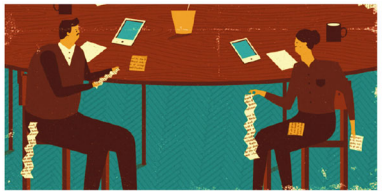
Interpreting the Results

36

Interpreting Results…

In 2013 Sutherland et al published a list (in Nature) to :

"help non-scientists interrogate advisers and grasp the limitations of evidence."



Twenty tips for interpreting scientific claims

This list will help non-scientists to interrogate advisers and to grasp the limitations of evidence, say **William J. Sutherland, David Spiegelhalter** and **Mark A. Burgman.**

intelligently interrogate experts and advisers, and to understand the quality, limitations and biases of evidence. We term these interpretive scientific skills. These skills are more accessible than those required to understand the fundamental science itself, and can form part of the broad skill set of most politicians.

To this end, we suggest 20 concepts that should be part of the education of civil servants, politicians, policy advisers and journalists — and anyone else who may have to interact with science or scientists. Politicians with a healthy scepticism of scientific advocates might simply prefer to arm themselves with this critical set of knowledge.

We are not so naive as to believe that improved policy decisions will automatically follow. We are fully aware that scientific judgement itself is value-laden, and that bias and context are integral to how data are collected and interpreted. What we offer is a simple list of ideas that could help decision-makers to parse how evidence can contribute to a decision, and potentially to avoid undue influence by those with vested interests. The harder part — the social acceptability of different policies — remains in the hands of politicians and the broader political process.

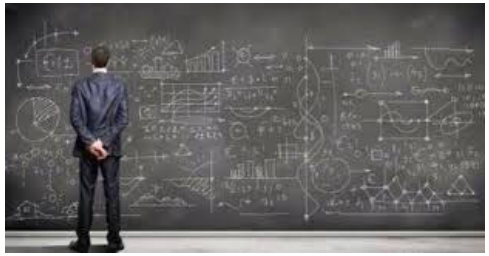
Of course, others will have slightly different lists. Our point is that a wider

21 NOVEMBER 2013 | VOL 503 | NATURE | 335

Interpreting the Results

And the big ticket items are…

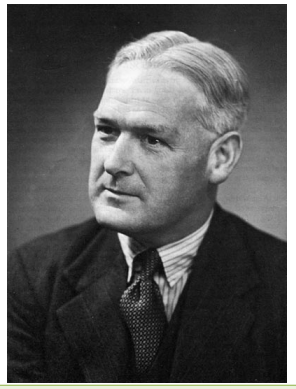
- Common sense
- Size of the effect
- Precision
- Applicability



Common Sense

Common Sense: Bradford Hill Criteria

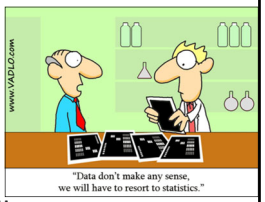
- Strength of association
- Consistency
- Specificity
- Temporal relationship
- Biological gradient
- Biological plausibility
- Coherence
- (*Relevance)



An "aid to thought": *Austin Bradford Hill, Proc R Soc Med 1965*

Don't be distracted by statistics!

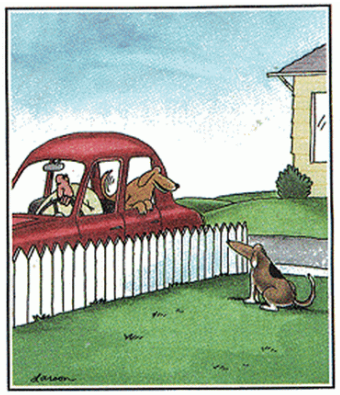
- Cannot remove all bias even with fancy adjustments
- Every test or formula has it own assumptions
- Statistics describe the impact of chance
  - They cannot assess **<u>importance</u>** or **<u>exclude bias</u>**
- **Judgement is required**
  - Check for: 1o and 2o outcomes, subgroups
    - Adverse outcomes/side effects
  - Make sure CIs are used
  - Extract the numbers eg exclusions, loss to follow up
  - Basic checks: sample size, normal distribution, multiple tests



And check author interpretation

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

Beware of dodgy or selective reporting!


"Ha ha ha, Biff. Guess what? After we go to the drugstore and the post office, I'm going to the vet's to get tutored."

42

Size of the Effect

43

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*

44

Why such a "Big" difference?

- Always potential for confounders
- Measures of effect (e.g. 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*

45

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)

46

Measure Effect

Prevalence	No. of events / no. people who could have it *no time component, not a rate (e.g. 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

47

Is it really a "Big Effect"? Beware!

Risk	**Accompanying explanation**	**% Accepting Therapy**
RD	2.5% had a heart attack vs 3.9% - a difference of 1.4%	42%
RRR	A cholesterol lowering drug resulted in a 34% reduction in heart attacks	88%
NNT	Would need to treat 71 patients for 5 years to prevent 1 heart attack	31%

Hux & Naylor. Med Decis Making 1995;15;152-7

48

Is it really a "Big Effect"? Beware!

Risk	Accompanying explanation	% Accepting Therapy
RD	2.5% had a heart attack vs 3.9% - a difference of 1.4%	42%
NNT	Would need to treat 71 patients for 5 years to prevent 1 heart attack	31%
RRR	A cholesterol lowering drug resulted in a 34% reduction in heart attacks	88%

Risk	Accompanying explanation
RD	0.25% had a heart attack vs 0.39% - a difference of 0.14%
NNT	Would need to treat 710 patients for 5 years to prevent 1 heart attack
RRR	A cholesterol lowering drug resulted in **a 34% reduction in heart attacks**

Hux & Naylor. Med Decis Making 1995;15;152-7

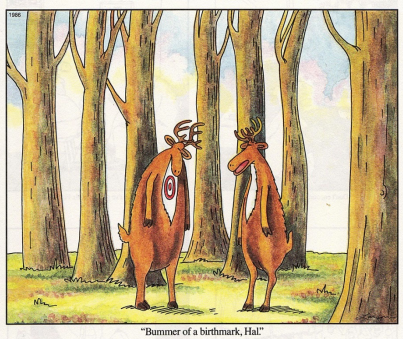
49

Precision

50

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)



51

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 results "true"
- So **don't** use / accept a p-value on its own

*P-values/CIs covered in **Introductory Biostatistics**
Available at ResearchEducationProgram.org

52

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 between groups

Formulae for different CIs: means, proportions, differences in means/proportions etc.

Can I apply the results in my setting? Application of the Results

54

Generalisability ("external validity")

- 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)

55

Applicability

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

56

Important Questions to Ask

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

57

Checklists

- **CASP** (Critical Appraisal Skills Program)
  - Systematic reviews, RCT, qualitative research, economic review, case-control studies, cohort studies
- **JAMA** (Journal American Medical Association)
  - User's guides/teaching papers
  - Set of excellent teaching papers
- **AGREE site** (Appraisal of Guidelines Research and Evaluation)
  - Clinical practice guideline checklist

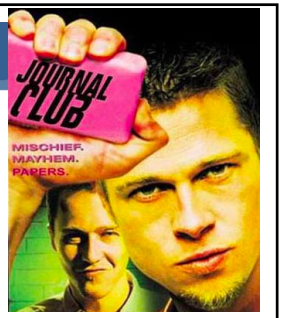
58

Checklists

- **CONSORT**
  - Randomised controlled trials
- **TREND**
  - Non-randomised controlled trials
- **STROBE**
  - Observational studies in epidemiology
- **PRISMA**
  - Systematic reviews and meta-analyses
- And many others…

59

Evidence-Based Journal Clubs



- Focus on **<u>current</u>** 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)

60

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!

61

2 CRITICAL APPRAISAL – ADDITIONAL NOTES AND RESOURCES

2.1 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

2.2 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).

2.3 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)

2.4 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.

2.5 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.

2.6 WHAT ARE THE RESULTS? (STUDY EFFECT SIZE AND INTEPRETATION 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.
Read: Hux & Naylor Med Decis Making 1995;15;152-7.
http://www.ncbi.nlm.nih.gov/pubmed/7783576

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)

2.7 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 CRITICAL 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)

2.8 KEY RESOURCES

Evidence-Based Clinical Practice

- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS: Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71-2. http://www.bmj.com/content/312/7023/71
- Centre for Evidence-based Medicine. University of Oxford (including Critical Appraisal Sheets) http://www.cebm.net/
- Polythenia gravis: the downside of evidence based medicine. *BMJ* 1995;311:1666 (for fun) http://www.bmj.com/content/311/7021/1666

Critical Appraisal

- What is critical appraisal. Amanda Burls. University of Oxford http://www.bandolier.org.uk/painres/download/whatis/What_is_critical_appraisal.pdf
- The Environment and Disease: Association or Causation? Sir Austin Bradford Hill, Professor Emeritus of Medical Statistics, University of London) http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1898525/pdf/procrsmed00196-0010.pdf
- Schroter S, Black N, Evans S, Godlee F, Osorio L, Smith R. What errors do peer reviewers detect, and does training improve their ability to detect them? *J R Soc Med* 2008; 101(10): 507–514 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2586872/
- Critical Appraisal Skills Programme, UK http://www.casp-uk.net/

Reviewing the Evidence

- How to review the evidence: systematic identification and review of the scientific literature. NHMRC. http://www.nhmrc.gov.au/publications/synopses/cp65syn.htm
- Diagnostic Study Appraisal CEBM http://www.cebm.net/wp-content/uploads/2014/06/Diagnostic-studies_Annette-Pluddemann.pdf
  And "Appraising Diagnostic Studies" Jeremy Howick http://www.cebm.net/resources/cebm-presentations/
- Critical Appraisal Tools: International Centre for Allied Health Evidence, University of SA http://www.unisa.edu.au/Research/Sansom-Institute-for-Health-Research/Research/Allied-Health-Evidence/Resources/CAT/
- Jackson et. al., The GATE frame: critical appraisal with pictures, EBM 2006 http://ebm.bmj.com/content/11/2/35.extract

Useful Checklists for Critical Appraisal

- Critical Appraisal Tools (Check lists). CEBM Oxford. http://www.cebm.net/critical-appraisal/
- BMJ Best Practice. Critical Appraisal Checklists https://bestpractice.bmj.com/info/toolkit/ebm-toolbox/critical-appraisal-checklists/
- CASP Checklists https://casp-uk.net/casp-tools-checklists/

- Critical Appraisal Tools: International Centre for Allied Health Evidence, University of SA http://www.unisa.edu.au/Research/Sansom-Institute-for-Health-Research/Research/Allied-Health-Evidence/Resources/CAT/
- 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 https://www.cebm.net/wp-content/uploads/2014/04/cebm-prognosis-worksheet.pdf
- CONSORT randomised controlled trials http://www.consort-statement.org/media/default/downloads/consort%202010%20checklist.pdf
- TREND non-randomised controlled trials https://www.cdc.gov/trendstatement/pdf/trendstatement_TREND_Checklist.pdf
- STROBE observational studies in epidemiology e.g. for case-control studies: https://www.strobe-statement.org/fileadmin/Strobe/uploads/checklists/STROBE_checklist_v4_case-control.pdf
- PRISMA systematic reviews and meta-analyses http://prisma-statement.org/prismastatement/Checklist.aspx

Communicating Results

- Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making* 1995;15;152-7 http://www.ncbi.nlm.nih.gov/pubmed/7783576
- SD Carley et al, Moving towards evidence based emergency medicine: use of a structured critical appraisal journal club, *Lancet* (1997) **349:**301-5 http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1343126/pdf/jaccidem00025-0010.pdf

Journal Clubs

- Phillips RS, Glasziou P. What makes evidence-based journal clubs succeed? *Evid Based Med* 2004;9:36-37 http://ebm.bmj.com/content/9/2/36.full

ResearchEducationProgram.org
ReserachEducationProgram@health.wa.gov.au