Systematic reviews- a clinical perspective

Prof Sanjay Patole, MD, DCH, FRACP, MSc, DrPH
Centre for Neonatal Research and Education
University of Western Australia
Perth, Western Australia
# Hierarchy of evidence table: (Oxford CEBM)

<table>
<thead>
<tr>
<th>Level</th>
<th>Evidence Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1A</td>
<td>Systematic reviews of RCTs</td>
</tr>
<tr>
<td>1B</td>
<td>Individual RCTs (with narrow CI)</td>
</tr>
<tr>
<td>1C</td>
<td>All other RCTs</td>
</tr>
<tr>
<td>2A</td>
<td>Systematic reviews of cohort studies</td>
</tr>
<tr>
<td>2B</td>
<td>Cohort study</td>
</tr>
<tr>
<td>2C</td>
<td>“Outcomes&quot; Research; Ecological studies</td>
</tr>
<tr>
<td>3A</td>
<td>Systematic review of case-control studies</td>
</tr>
<tr>
<td>3B</td>
<td>Case-control study</td>
</tr>
<tr>
<td>4</td>
<td>Case-series</td>
</tr>
<tr>
<td>5</td>
<td>Expert opinion <em>without explicit critical appraisal</em></td>
</tr>
</tbody>
</table>
Systematic review and meta analysis

- **Systematic review**: When literature is the subject of research
- **Meta analysis**: Results of several studies are combined mathematically to provide a summary estimate

- **SR with/without meta analysis**: Quantitative/Qualitative
- **SR could be for RCTs, non-RCTs, diagnostic studies etc.**

**Note**: Today’s focus is on systematic reviews of RCTs
Advantages of systematic reviews

- High volume of publications; most RCTs are small

- SRs increase power and precision of effect size, provide summary of evidence

- Help DMCs in deciding whether to continue an RCT

- Help individual units to decide whether it is ethical to continue recruiting patients into a trial

- Can challenge existing practice, identify research priorities

- SRs are prerequisites for future trial design

*Iain Chalmers. BMJ Books 2001*
Probiotics reduce the risk of NEC in preterm infants

Deshpande et al Pediatrics 2010

<table>
<thead>
<tr>
<th>Study or sub-category</th>
<th>Probiotic n/N</th>
<th>No probiotic n/N</th>
<th>RR (fixed)</th>
<th>Weight %</th>
<th>RR (fixed)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kibajma 1997</td>
<td>0/45</td>
<td>0/46</td>
<td>1.15</td>
<td>11.15</td>
<td>1.04 [0.49, 2.3]</td>
</tr>
<tr>
<td>Dam 2002</td>
<td>4/296</td>
<td>8/290</td>
<td>9.72</td>
<td>9.72</td>
<td>1.22 [0.61, 2.4]</td>
</tr>
<tr>
<td>Casteloe 2003</td>
<td>5/51</td>
<td>6/56</td>
<td>13.79</td>
<td>13.79</td>
<td>0.21 [0.05, 0.9]</td>
</tr>
<tr>
<td>Bin Nun 2005</td>
<td>1/72</td>
<td>10/73</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10 [0.01, 0.79]</td>
</tr>
<tr>
<td>Lin 2005</td>
<td>2/180</td>
<td>10/187</td>
<td>0.21</td>
<td>0.21</td>
<td>0.21 [0.05, 0.94]</td>
</tr>
<tr>
<td>Manzoni 2006</td>
<td>1/39</td>
<td>3/41</td>
<td>4.04</td>
<td>4.04</td>
<td>0.34 [0.04, 3.2]</td>
</tr>
<tr>
<td>Mohan 2006</td>
<td>2/21</td>
<td>1/17</td>
<td>0.83</td>
<td>0.83</td>
<td>0.83 [0.16, 4.37]</td>
</tr>
<tr>
<td>Shibata 2007</td>
<td>0/38</td>
<td>3/31</td>
<td>5.31</td>
<td>5.31</td>
<td>0.51 [0.01, 2.19]</td>
</tr>
<tr>
<td>Lin 2008</td>
<td>4/217</td>
<td>14/217</td>
<td>19.38</td>
<td>19.38</td>
<td>0.25 [0.10, 0.85]</td>
</tr>
<tr>
<td>Serrant 2008</td>
<td>5/91</td>
<td>15/96</td>
<td>0.29</td>
<td>0.29</td>
<td>0.29 [0.13, 0.92]</td>
</tr>
<tr>
<td>Rouge 2009</td>
<td>2/45</td>
<td>1/49</td>
<td>1.32</td>
<td>1.32</td>
<td>2.18 [0.22, 25.21]</td>
</tr>
</tbody>
</table>

Total events: 26 (Probiotic), 71 (no probiotic)
Test for heterogeneity: Chi² = 7.66, df = 9 (P = 0.57), I² = 0%
Test for overall effect: Z = 4.84 (P < 0.00001)

Note: Majority of Australian neonatal units now use probiotics
We decided to continue participation in the ICE trial considering the small sample size (n=449) in this systematic review.
AI-VP shunt catheters may decrease shunt infections

How to conduct a systematic review

Clinical question must be clearly defined and should include

- Population of interest (P)
- Intervention (I)
- Comparator (C)
- Outcome (O)
- Study design (S)
- Time (T)

Register title, write protocol, receive feedback, start work
Key areas covered in the protocol

• Why?
• Which studies? (Inclusion - Exclusion criteria)
• Search strategy (What, where, how, who etc.)
• Study selection
• Method of data extraction
• Assessment of risk of bias
• Statistical methods used to combine data
• How the results will be disseminated
Literature search

- PubMed: Available free on internet.
- **Medline and Embase**: OVID platform from library
- 70% of the citations in Embase are not on PubMed
- **CINAHL**: EbscoHost platform
- Cochrane register of controlled trials (**CENTRAL**)
- Grey literature and experts
Bias vs. Error

- **Bias**: *Systematic deviations* from the true underlying effect (False positive or negative results)

- **Reasons**: Poor study design--conduct--analysis--interpretation, or issues with publication and review

- **Risk of bias**: Classified as Low/High/Unclear

- **Error**: This is a *mistake* (i.e. wrong entry of numbers)
Risk of bias (ROB)

- It is not necessary to exclude studies with high ROB
- Cochrane collaboration allows for quasi-random studies
- ROB could be used for sensitivity analyses
- Studies with lowest ROB are analysed together
- The results compared to the analysis of all studies
Assessing ROB in RCTs

Generation of random sequence

Low risk: Using computer generated random numbers

High risk: Sequence generated by
  • Odd or even date of birth
  • Day of admission
  • Clinician or patient’s preference
  • Availability of intervention
Allocation concealment

Intervention to be allocated to a participant can not be known in advance
- **Low risk:** Central tel./computer-based randomisation
- **High risk:** Envelopes

Blinding

Carers and patients should not know what intervention they are receiving
- **Low risk:** Placebo **High risk:** No placebo
- Blinding may not be feasible in some RCTs
Blinding of outcome assessors
• Important for subjective outcome measures (e.g. pain)
• Less important for measures such as mortality

Incomplete outcome data
• Some patients drop out from RCTs
• Need to detail the number of drop outs and reasons

Selective reporting
• **High ROB:** Not all pre-specified outcomes reported
Data synthesis

**Qualitative:** Summaries and Tables

**Quantitative:** Meta analysis

**Meta analysis**
- Mathematical pooling of data (RevMan or other softwares)
- Gives an effect size estimate/meta estimate
- **Produces a “Forest plot”**
Cardiovascular events are less with statins: RR: 0.73 (0.67, 0.80)
Forest plot

- Studies listed in chronological order, alphabetically or by study weight.

- Each study’s estimated effect size is represented by a square, with the line representing the corresponding 95% confidence interval.

- Size of a study’s square indicates its weight toward overall summary effect

- Weight is determined by sample size, baseline risk etc.
Forest plot

- The summary estimate is represented by a diamond
- Centre of the diamond: Point estimate
- Tips of the diamond: 95% Confidence interval
Analytical models for meta analysis

Fixed effects model
- Assumes that intervention is equally effective across all studies. (*Confident* assumption) Ignores "Between study" variation
- **What is the best estimate of the effect?**

Random effects model
- Allows for ‘within’ as well as ‘between-study’ variability in effectiveness. (*Conservative* assumption)
- Being less confident, it usually has wider CIs and gives adequate emphasis on smaller studies.
- **What is the average effect?**

Note: Neonatal Cochrane group recommends FEM
Exploring heterogeneity

- Heterogeneity (differences in results) could be due to differences in study design, characteristics (PICO), and conduct

- If heterogeneity exists in a meta analysis, one must explore it.
Conceptual (clinical) heterogeneity

- Studies of clinically diverse treatments, populations, setting, design etc.

- Don’t pool data if significant clinical heterogeneity is present

- The results of studies should be combined only when the studies are homogenous (i.e. similar PICO and design)

Note: Don’t forget Apples vs. Oranges, different types of apples
Statistical heterogeneity

**Chi squared test (Q):** Is statistical heterogeneity present?

**I squared test:** Is the observed variability of effects greater than that expected by chance alone?

**I squared >50%:** Significant statistical heterogeneity, so results need to be interpreted cautiously
Long term antibiotics for prevention of recurrent symptomatic UTI
Williams and Craig, Cochrane review 2011

I squared statistic: 62%: Significant statistical heterogeneity was explored with sensitivity analysis
When only studies with low ROB were combined, there was no heterogeneity
Funnel plot: Assessing publication bias

- Scatter plot (X axis: Effect size, Y axis: Study precision)
- Study precision: Standard error (SE) of the effect size

- Effect sizes from smaller studies have larger SE, so will be located lower on the Y axis
- Effect estimates from smaller RCTs will scatter more widely at the bottom of the graph, with the spread narrowing among larger studies.

Note: In absence of bias and between study heterogeneity, the plot resembles a symmetrical inverted funnel.
Symmetrical funnel plot: The outer dashed lines indicate the triangular region within which 95% of studies are expected to lie in.

Sterne JAC et al. BMJ 2011
Funnel plot asymmetry

If there is a genuine asymmetry, the pooled effect estimate in a meta-analysis will overestimate the treatment effect. *Egger 1997*

**Statistical tests for funnel plot asymmetry**

- Do not use if less than 10 studies
- Power is too low to differentiate chance from real asymmetry
- Not routinely recommended

*Sterne et al, BMJ 2011*
Reporting a systematic review and meta analysis

Preferred Reporting Items for Systematic Reviews and Meta analyses (PRISMA statement)

Moher et al J Clin Epidemiol 2009
Pitfalls in systematic reviews

Pitfalls in conducting

- Single author
- Not searching all relevant databases
- Not including non-English studies
- Deviating from the protocol depending on the results
Influence of ROB on effect size estimates

- *Unpublished trials underestimate effect size by ~10%*
- Trials published in languages other than English will overestimate by 10%
- Trials not indexed in Medline will overestimate by 5%,
- Trials with inadequate or unclear concealment of allocation will overestimate by 30%
- Trials not double blinded will overestimate by 15%

*Egger et al Int J Epidemiol 2002*
Odds ratio vs. Risk ratio

• **Risk ratio:** 0.82, a 18% decrease in *risk* of infection.

• **Odds ratio:** 0.41, a 59% decrease in *odds* of infection.

• Clinicians can misinterpret OR as RR and overestimate the efficacy of protective intervention

Note: Neonatal Cochrane group recommends relative risk
### Effect of cooling on death or major disability among survivors

*Jacobs et al Cochrane 2013*

<table>
<thead>
<tr>
<th>RR: 0.75 vs. OR: 0.53</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>RR</th>
<th>OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.75</td>
<td>0.53</td>
</tr>
</tbody>
</table>

[Table and diagram content not transcribed due to OCR limitations]
Controversies

A well conducted systematic review with meta analysis can represent the pinnacle of evidence based evaluation.
Meta analysis vs. Mega RCT

ISIS-4: International study of infarct survival (N=58050)
  • No difference in mortality in MgSo4 vs. Control group
  • 2216/29011 (7.6%) vs. 2103/29039 (7.2%)

*Lancet 1995*

• These results overruled previous meta analysis that showed benefit (7 RCTs, N=1300, OR: 0.45, p<0.001).

Note: Clinicians have to treat patients using the best possible current evidence (systematic review) rather than waiting for a future RCT
Results from four concordant and four discordant pairs of meta-analysis and large scale RCT

Egger M et al. BMJ 1997
FEM and REM estimates: Effect of IV Magnesium on mortality after MI

FEM showed no difference, because it gave 90% weight to the ISIS-4 trial.
REM showed beneficial effect because smaller studies received adequate weight.

<table>
<thead>
<tr>
<th>Trial</th>
<th>Events</th>
<th>Relative risk (95% CI)</th>
<th>Weight (%)</th>
<th>Relative risk (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>Control</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Morton 1984</td>
<td>1/40</td>
<td>0.09</td>
<td>0.45</td>
<td>(0.04 to 4.76)</td>
</tr>
<tr>
<td>Rasmussen 1986</td>
<td>9/135</td>
<td>0.98</td>
<td>0.39</td>
<td>(0.19 to 0.81)</td>
</tr>
<tr>
<td>Smith 1986</td>
<td>2/200</td>
<td>0.30</td>
<td>0.29</td>
<td>(0.06 to 1.36)</td>
</tr>
<tr>
<td>Abraham 1987</td>
<td>1/48</td>
<td>0.04</td>
<td>0.96</td>
<td>(0.06 to 14.87)</td>
</tr>
<tr>
<td>Feldstedt 1988</td>
<td>10/150</td>
<td>0.34</td>
<td>1.23</td>
<td>(0.50 to 3.04)</td>
</tr>
<tr>
<td>Shechter 1989</td>
<td>1/59</td>
<td>0.39</td>
<td>0.11</td>
<td>(0.01 to 0.81)</td>
</tr>
<tr>
<td>Ceremuzynski 1989</td>
<td>1/25</td>
<td>0.13</td>
<td>0.31</td>
<td>(0.03 to 2.74)</td>
</tr>
<tr>
<td>Bentschat 1989</td>
<td>0/22</td>
<td>0.07</td>
<td>0.32</td>
<td>(0.01 to 7.42)</td>
</tr>
<tr>
<td>Singh 1990</td>
<td>6/76</td>
<td>0.47</td>
<td>0.54</td>
<td>(0.21 to 1.38)</td>
</tr>
<tr>
<td>Perelra 1990</td>
<td>1/27</td>
<td>0.30</td>
<td>0.14</td>
<td>(0.02 to 1.08)</td>
</tr>
<tr>
<td>Shechter 1 1991</td>
<td>2/89</td>
<td>0.54</td>
<td>0.15</td>
<td>(0.03 to 0.65)</td>
</tr>
<tr>
<td>Golf 1991</td>
<td>5/23</td>
<td>0.46</td>
<td>0.55</td>
<td>(0.23 to 1.33)</td>
</tr>
<tr>
<td>Thogersen 1991</td>
<td>4/130</td>
<td>0.35</td>
<td>0.47</td>
<td>(0.14 to 1.52)</td>
</tr>
<tr>
<td>LIMIT-2 1992</td>
<td>90/1159</td>
<td>5.04</td>
<td>0.76</td>
<td>(0.59 to 0.99)</td>
</tr>
<tr>
<td>Shechter 2 1995</td>
<td>4/107</td>
<td>0.72</td>
<td>0.24</td>
<td>(0.08 to 0.68)</td>
</tr>
<tr>
<td>ISIS-4 1995</td>
<td>2216/29011</td>
<td>89.76</td>
<td>1.05</td>
<td>(1.00 to 1.12)</td>
</tr>
</tbody>
</table>

Fixed-effect (M-H) estimate: $I^2=67\%$, $P=0.000$ 2353/31 301 2343/31 306
Random-effects (D+L) estimate

$100.0 \quad 1.01 \quad (0.95 to 1.06)$
$0.53 \quad 0.38 \quad (0.70 to 0.75)$
• It is better to compare the FEM and REM estimates of the treatment effect.

• If REM estimate appears more beneficial, treatment was more effective in smaller studies because weight given to each study by REM is less influenced by sample size.

• If there is no evidence of heterogeneity between studies, the FEM and REM estimates will be identical.
Checklist for systematic review

- **Methodology**: Robust, Comprehensive, Transparent, and Reproducible?

- **Type of studies** (RCTs, Non-RCTs)

- **Risk of bias** in included studies, Publication bias

- **Time span**
Checklist for forest plots: 10 points

- Number and type of studies, sample sizes, and total sample size
- Number of events and denominators in intervention vs control group
- Confidence intervals and their overlap
- Tests for heterogeneity: \( \chi^2 \) (Q statistics) and its P value, \( I^2: (%) \)
- Pooled effect (Z) size, P value, and statistical vs. clinical significance
- Risk vs. Odds, RR, AR, ARD
- Model/s used for analysis, Concordance/Discordance of results
- Weightage to different studies? Any study driving the results? Outliers?
- Type of outcome: Primary vs. Secondary
- Labelling of intervention and comparison groups and plotted results
Other clinically important issues

- Benefits vs. Risks (short and long-term)
- NNT, NNH
- Translational potential
Thank you