Rapid-fire critical appraisal for START assessments

Dr Ian Morris
Consultant Neonatologist
University Hospital of Wales
## Curriculum

<table>
<thead>
<tr>
<th>Curriculum Domain</th>
<th>Level 1</th>
<th>Level 2</th>
<th>Level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>11. Research</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GPC 9</td>
<td>Adopts an evidence-based approach to paediatric health practice and the critical appraisal of existing published research.</td>
<td>Implements an evidence-based approach to practice to inform decision-making and enhance patient care and patient outcomes.</td>
<td>Demonstrates independent development and revision of guidelines and procedures to improve service delivery, centred around current clinical research and evidence-based healthcare.</td>
</tr>
</tbody>
</table>
Learning objectives

• Understand how appraisal is relevant to practice

• Understand how and why it is included in START

• Develop a framework for rapid appraisal

• Understand basics of study design, bias and key terms
Relevance to practice

• Making decisions

• Justifying decisions
  – Parents
  – Colleagues
  – Nursing staff
  – Yourself!

• Delivering high quality, safe medicine
START

• ST7

• 12 x 8 minute stations (4 mins prep)
  – CBD type format

• 44 mins prep for prescribing and appraisal

• Standard: newly appointed consultant

• Form basis for PDP for ST8 year
Key points

- Dress smart

- Be confident and decisive

- Think about:
  - Leadership
  - Prioritisation
  - Delegation
  - Team working
  - Communication
  - Learning opportunities for trainees
  - Safeguarding
  - Quality and safety of service
This is a 8 minute scenario. Check the ‘Task’ below so you know how this scenario will be run. You will have up to 4 minutes before the start of the assessment to read this information and prepare yourself. You may make notes on the paper provided.

You should
• Enter the assessment room when the bell sounds
• Take this instruction sheet with you and any notes you have prepared
• Note that there will be a warning when you have approximately 2-minutes left
• Return the instruction sheet outside the room when you leave

<table>
<thead>
<tr>
<th>Context</th>
<th>You are a newly appointed consultant</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task</td>
<td>Read the scenario and prepare an appraisal of the provided paper (Archives of Disease in Childhood 2006; (5):414).</td>
</tr>
<tr>
<td>Information for trainee</td>
<td>A 3 year old with a history of a febrile convulsion at the age of 16 months arrives in the children’s emergency department. He is bright and alert but has an obvious upper respiratory tract infection. His temperature is 38.7°C. His parents and the nursing staff are keen to give paracetamol and ibuprofen together to reduce the temperature as quickly possible. Is this known to be helpful? Review the paper and decide to what extent this supports or doesn't support the concurrent use of these medications.</td>
</tr>
</tbody>
</table>
**Randomised controlled trial of combined paracetamol and ibuprofen for fever**

M D S Erlwyn-Lajunesse, K Coppens, LP Hunt, P J Chinnick, P Davies, I M Higginson, J R Benger


**RESULTS**

A total of 125 children were randomised between October 2004 and January 2005 (Table 1). Baseline data were similar in all three groups, except that more children were admitted to hospital in the combined group (13/36) compared to the ibuprofen (15/35) and paracetamol groups (11/34). A series of admissions showed that the children from the combined group took non-specifically well following the administration of study medication, during a quite different period from 19.5°C to 37°C in one hour. The drug was administered, monitored, and monitored, and was discharged three hours later. All other admissions were not related to the clinical events following the administration of study medication. One child in the paracetamol group received a dose of 2.8 mg/kg in error. The child did not suffer any adverse consequences from this event. There were no less other adverse events.

**Box 1: Exclusion criteria**

- Paracetamol or ibuprofen given in the previous six hours
- Severe or life-threatening infection
- Suspected chicken pox
- Gastrointestinal or other skin infection
- Known to be immunosuppressed
- Allergic to either ibuprofen or paracetamol
- Medication with warfarin, heparin, or antithrombotics
- Symptoms of acute gastrointestinal bleeding
- Known coagulopathy
- Acute jaundice
- Unlikely dehydration, defined as more than four episodes of diarrhea or vomiting in the previous 24 hours
- Asthma, defined as a need for regular "preventer" medication
- Chronic renal, liver, or cardiac failure

**DISCUSSION**

Combined paracetamol and ibuprofen were not different from each other, however, the effect is less than half a degree centigrade, and we do not believe that this is clinically important difference over this time period. We did not show a difference between combined therapy and ibuprofen. Our study was powered to detect a difference between combined therapy and paracetamol alone, and by doing so we have demonstrated proof of principle. The greater efficiency of ibuprofen might require a larger sample size to show a significant difference.

This was carried out in a paediatric emergency department, and therefore only examined the short term control of fever. A longer measurement period might produce different results, as the maximum decrease in temperature for both medicines is around three hours post-dose. In the emergency department, we were unable to keep children back for study drug exchange and in consequence, only examined the short term change.

**ACKNOWLEDGEMENTS**

The authors would like to thank Samantha Leroy-Turner and Jessica Perkins for their administrative help with this study. We would also like to thank the staff of the Child and the parents and children who participated.

**Table 1** Baseline data and mean temperatures (°C) at baseline and one hour for the three treatment groups

<table>
<thead>
<tr>
<th>Treatment Group</th>
<th>Paracetamol</th>
<th>Ibuprofen</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomisation</td>
<td>N = 43</td>
<td>N = 42</td>
<td>N = 42</td>
</tr>
<tr>
<td>Age (years)</td>
<td>9.8 (3.2)</td>
<td>9.7 (3.3)</td>
<td>9.7 (3.3)</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>21.0 (4.9)</td>
<td>20.6 (4.7)</td>
<td>20.7 (4.7)</td>
</tr>
<tr>
<td>Height (m)</td>
<td>1.16 (0.13)</td>
<td>1.16 (0.13)</td>
<td>1.16 (0.13)</td>
</tr>
<tr>
<td>Dose of paracetamol (mg/kg)</td>
<td>20 (2)</td>
<td>20 (2)</td>
<td>20 (2)</td>
</tr>
<tr>
<td>Dose of ibuprofen (mg/kg)</td>
<td>5 (3)</td>
<td>5 (3)</td>
<td>5 (3)</td>
</tr>
<tr>
<td>Time of treatment (h)</td>
<td>3.0 (0.5)</td>
<td>3.0 (0.5)</td>
<td>3.0 (0.5)</td>
</tr>
<tr>
<td>Time of measurement (h)</td>
<td>4.0 (0.5)</td>
<td>4.0 (0.5)</td>
<td>4.0 (0.5)</td>
</tr>
<tr>
<td>Time of measurement (min)</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Baseline temperature (°C)</td>
<td>37.9 (0.3)</td>
<td>37.9 (0.3)</td>
<td>37.9 (0.3)</td>
</tr>
<tr>
<td>Mean temperature (°C) at one hour</td>
<td>36.5 (0.3)</td>
<td>36.5 (0.3)</td>
<td>36.5 (0.3)</td>
</tr>
</tbody>
</table>

**REFERENCES**


**Authors’ affiliations**

- M D S Erlwyn-Lajunesse, K Coppens, LP Hunt, P J Chinnick, P Davies, I M Higginson, J R Benger, Children’s Emergency Department, Royal Hospital for Children, Brisbane, UK
- A. C. C. M., PA. M. N., and A. G. C. R. M., School of Public Health, University of Queensland, Brisbane, Australia
- A. C. C. M., PA. M. N., and A. G. C. R. M., School of Public Health, University of Queensland, Brisbane, Australia
Approach to appraisal

• Many different ways.....

• Many different checklists....

• 3 key areas:
  – Relevance
  – Internal validity
  – External validity
Relevance:
- Is this article important to me, my patients and my practice
- Title and abstract

Internal validity
- Soundness of the study
- Methods

External validity
- Applicability to my practice / generalisability
- Methods, results
<table>
<thead>
<tr>
<th>BASIS FOR COMPARISON</th>
<th>INTERNAL VALIDITY</th>
<th>EXTERNAL VALIDITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning</td>
<td>Internal validity is the extent to which the experiment is free from errors and any difference in measurement is due to independent variable and nothing else.</td>
<td>External validity is the extent to which the research results can be inferred to world at large.</td>
</tr>
<tr>
<td>Concerned with</td>
<td>Control</td>
<td>Naturalness</td>
</tr>
<tr>
<td>What is it?</td>
<td>It is a measure of accuracy of the experiment.</td>
<td>It checks whether the casual relationship discovered in the experiment can be generalized or not.</td>
</tr>
<tr>
<td>Identifies</td>
<td>How strong the research methods are?</td>
<td>Can the outcome of the research be applied to the real world?</td>
</tr>
<tr>
<td>Describes</td>
<td>Degree to which the conclusion is warranted.</td>
<td>Degree to which the study is warranted to generalize the result to other context.</td>
</tr>
<tr>
<td>Used to</td>
<td>Address or eliminate alternative explanation for the result.</td>
<td>Generalize the outcome.</td>
</tr>
<tr>
<td>BASIS FOR COMPARISON</td>
<td>INTERNAL VALIDITY</td>
<td>EXTERNAL VALIDITY</td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>Meaning</td>
<td>Internal validity is the extent to which the experiment is free from errors and any difference in measurement is due to independent variable and nothing else.</td>
<td>External validity is the extent to which the research results can be inferred to world at large.</td>
</tr>
<tr>
<td>Concerned with</td>
<td>Control</td>
<td>Naturalness</td>
</tr>
<tr>
<td>What is it?</td>
<td>It is a measure of accuracy of the experiment.</td>
<td>It checks whether the casual relationship discovered in the experiment can be generalized or not.</td>
</tr>
<tr>
<td>Identifies</td>
<td>How strong the research methods are?</td>
<td>Can the outcome of the research be applied to the real world?</td>
</tr>
<tr>
<td>Describes</td>
<td>Degree to which the conclusion is warranted.</td>
<td>Degree to which the study is warranted to generalize the result to other context.</td>
</tr>
<tr>
<td>Used to</td>
<td>Address or eliminate alternative explanation for the result.</td>
<td>Generalize the outcome.</td>
</tr>
</tbody>
</table>
Paper analysis

1) Describe the study
   – What type of study (method)
   – What type of question (diagnostic, therapeutic, economic)
   – Where was it done (including single/multi-centre)
   – Key feature (e.g. first of its kind)

Example: “This was a multicentre, randomised, controlled trial of therapy, and the latest of a number of studies assessing the use of probiotics to prevent necrotising enterocolitis in very preterm infants.”
2) Describe the research question
   – Population (who)
   – Intervention (treatment, test)
   – Comparison (control)
   – Outcome

Example:
   P – in infants born between 23 and 30 weeks’ gestation
   I – do daily probiotics (Bifodobacterium breve BBG-001)
   C – versus placebo
   O – reduce necrotising enterocolitis, late-onset sepsis, and death before discharge
3) State the relevance of this question
   – This is usually described in the introduction
   – Correlate it to your clinical practice / clinical scenario

4) Describe the methods
   – Use PICO format but expand

Example:
   
   P – 24 hospitals, SE England, 23+0 – 30+6
   - excluded potentially lethal or GI malformations identified within 48 hours
   - excluded those no realistic chance of survival
   IC – Randomised within 48 hours of birth (web-based, NPEU)
   - Masking of parents, clinicians, outcome assessors
   - OD B breve BBG-001 or placebo (identical packaging and preparation procedures)
   - started asap after randomisation, irrespective of enteral feeding
   - continued up to 36 weeks PMA / discharge
   - **Stool cultures + 16s rRNA for B Breve at 2 weeks and 36 wk PMA**
   - clinical data recorded
• O – any episode NEC bell stage 2 or 3
  - sepsis: any LOS (<46 PMA) with positive blood culture that is not skin commensal
  - death before discharge from hospital

secondary outcomes
  - number of cultures, cultures per infant, positive culture of B. Breve from sterile site, antibiotic use, antifungal use, time to full enteral feeds, LOS, BPD, ROP...........

Estimated 1300 infants needed
  - 5% sig level, 90% power to detect 40% risk reduction in primary outcomes (15% to 9.1%)
  - ITT analysis
5) Comment on internal validity
  – Various checklists e.g. CONSORT, STROBE, PRISMA
  – Is there a risk of **bias**?

  – Were experiment and control groups similar at baseline? (randomisation, concealment, ITT)
  – Did groups remain similar after study started? (blinding, follow-up)
  – Attempt to minimise type 1 and type 2 errors?
A **bias** is a systematic error, or deviation from the truth, in results or inferences.

- Multiple replications of the same study would reach the wrong answer on average.

  *cf*

- Random error or imprecision
- Quality of study

Sources of Bias - 1

• Selection bias (randomisation)
  – systematic differences between baseline characteristics of the groups that are compared
  – sequence generation; allocation sequence concealment

• Performance bias (blinding of operators)
  – systematic differences between groups in the care that is provided, or in exposure to factors other than the interventions of interest

• Detection bias (blinding of assessors)
  – systematic differences between groups in how outcomes are determined
Sources of Bias - 2

• Attrition bias (incomplete outcome data)
  – systematic differences between groups in withdrawals from a study (\textit{cf} exclusions)
  – “intention to treat” analysis

• Reporting bias (selective reporting)
  – systematic differences between reported and unreported findings (analyses with statistically significant differences between intervention groups are more likely to be reported than non-significant differences – “within-study publication bias”)

Chance

• Null hypothesis: Prediction that there will be NO significant difference in an outcome

• Type I vs Type II error

<table>
<thead>
<tr>
<th></th>
<th>Null hypothesis ($H_0$) is true</th>
<th>Null hypothesis ($H_0$) is false</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject null hypothesis</td>
<td>Type I error False positive</td>
<td>Correct outcome True positive</td>
</tr>
<tr>
<td>Fail to reject null hypothesis</td>
<td>Correct outcome True negative</td>
<td>Type II error False negative</td>
</tr>
</tbody>
</table>

In colloquial usage type I error can be thought of as "convicting an innocent person" and type II error "letting a guilty person go free".
6) Summarise the primary results

7) Summarise key secondary results

8) Describe the generalisability of these findings (external validity)
   - Context of your local population
   - Were inclusion / exclusion criteria reflective
   - Limitations that may impact
   - In conjunction with other studies

9) Conclude
   - Overall, does this paper help you answer your clinical question
Worked example

Study

This was a single-centre, open label, randomised trial of therapy, assessing the use of antipyretics in reducing fever in children.
Research question

Example:

P – in children between 6 months and 10 years with a fever

I – does combined ibuprofen and paracetamol

C – versus paracetamol or ibuprofen alone

O – reduce temperature at 1 hour
Relevance

Fever is a common presenting feature and cause for concern for parents and healthcare professionals alike. The use of multiple antipyretics is common, but unsupported by evidence. This paper thus seeks to answer an important and relevant question and is worth exploring.
Methodology

P – Children admitted to a UK Children’s ED with a fever > 38°C were included. A number of exclusions are listed including those receiving antipyretics in the previous 6 hours and those with severe infection.

IC – Patients were randomised to receive either suspensions of paracetamol 15mg/kg, ibuprofen 5mg/kg or both in an open-label design.

O - The primary outcome was temperature at 1 hour as measured by a tympanometric thermometer.
Internal validity

A sample size calculation suggested 40 participants per group were needed for an 80% chance of detecting a $1^\circ$C difference at the 95% confidence level.

The groups were similar at baseline and were analysed on an intention to treat basis.

A high and equal number of children were analysed at 1 hour in each group.

Participants and assessors were not masked to intervention.

There is thus low risk of selection and attrition but high risk of performance and detection bias.

Overall, the validity of this study is reasonable.
Key results

In total, 123 children were randomised and 108 were included in primary analysis with an equal number across the 3 groups.

There was a statistically significant difference in temperature between the combined group and those receiving paracetamol alone. This was of $0.35^{\circ}C$ mean reduction after adjusting for baseline variables.

No other differences in primary or secondary outcomes were seen.
Generalisability

Patients included in this study were similar to the population we would expect locally, with broad inclusion criteria. The observation time was in line with what would be relevant to our department. No comment was made on other clinical observations nor on important history points such as a history of febrile convulsions or source of fever.

Conclusion

This well designed, small study failed to show a significant clinical benefit of combined antipyretics for reducing fever at 1 hour in febrile children. It provides further evidence for a more conservative use of antipyretics, but I would consider results of similar studies before adjusting my practice.
Concepts to understand

• Study design
  – RCT, Observational studies, systematic reviews
  – Randomisation, blinding, intention to treat
• Chance
  – Type 1 and type 2 errors
  – Sample size calculations
• Bias
• Statistical Vs Clinical significance
• Basic statistics
  – Key types of test
  – P value
  – Means, confidence intervals
  – RR, OR, NNT, NNH, PPV, NPV
Sample size

• Importance: waste time, waste resources, ethics

• 4 components:
  – Type I error ($\alpha$) Usually 5%
  – Type II error ($\beta$) Usually 10 or 20%
  – Variability ($\sigma^2$) e.g. standard deviation
  – Effect size ($d$) Difference that would be clinically relevant

• Standard formulas, e.g. $N = 16\sigma^2 / d^2$

• Sample size increases if variance increases or effect size decreases
Common tests

- **Parametric:**
  assumes a form (usually normal) of distribution of data

- **Non-parametric:**
  no assumptions
  looks at rank order of values, ignores absolute differences
  Makes showing significance harder

<table>
<thead>
<tr>
<th>Parametric test</th>
<th>Equivalent non-parametric test</th>
<th>Purpose of test</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Two sample (unpaired) t-test</td>
<td>Mann-Whitney U test</td>
<td>Compares two independent samples drawn from the same population</td>
<td>To compare girls' heights with boys' heights</td>
</tr>
<tr>
<td>One sample (paired) t-test</td>
<td>Wilcoxon matched pairs test</td>
<td>Compares two sets of observations on a single sample</td>
<td>To compare weight of infants before and after a feed</td>
</tr>
<tr>
<td>One way analysis of variance (Friedman)</td>
<td>Kruskal-Wallis analysis of variance by ranks</td>
<td>Effectively, a generalisation of the paired t-test or Wilcoxon matched pairs test where three or more sets of observations are made on a single sample</td>
<td>To determine whether plasma glucose level is higher one hour, two hours, or three hours after a meal</td>
</tr>
<tr>
<td>Two way analysis of variance</td>
<td>Two way analysis of variance by ranks</td>
<td>As above, but tests the influence (and interaction) of two different covariates</td>
<td>In the above example, to determine if the results differ in male and female subjects</td>
</tr>
<tr>
<td>x² test</td>
<td>Fisher's exact test</td>
<td>Tests the null hypothesis that the distribution of a discontinuous variable is the same in two (or more) independent samples</td>
<td>To assess whether acceptance into medical school is more likely if the applicant was born in Britain</td>
</tr>
<tr>
<td>Product moment correlation coefficient (Pearson's r)</td>
<td>Spearman's rank correlation coefficient (r₁₉₃, d)</td>
<td>Assesses the strength of the straight line association between two continuous variables.</td>
<td>To assess whether and to what extent plasma HbA1c concentration is related to plasma triglyceride concentration in diabetic patients</td>
</tr>
<tr>
<td>Regression by least squares method</td>
<td>Non-parametric regression (various tests)</td>
<td>Describes the numerical relationship between two quantitative variables, allowing one value to be predicted from the other</td>
<td>To see how peak expiratory flow rate varies with height</td>
</tr>
<tr>
<td>Multiple regression by least squares method</td>
<td>Non-parametric regression (various tests)</td>
<td>Describes the numerical relationship between a dependent variable and several predictor variables (covariates)</td>
<td>To determine whether and to what extent a person's age, body fat, and sodium intake determine their blood pressure</td>
</tr>
</tbody>
</table>
Resources

• How to read a paper: Statistics for the non-statistician. I: Different types of data need different statistical tests
  How to read a paper: Statistics for the non-statistician. II: “Significant” relations and their pitfalls  BMJ 1997; 315 Greenhalgh

• https://www.bmj.com/about-bmj/resources-readers/publications/statistics-square-one

• https://lifeinthefastlane.com

• http://training.cochrane.org

• E-learning for health, https://portal.e-lfh.org.uk/
Welcome Ian | Log Out

My e-Learning > NICE Guidance (NICE)

- My e-Learning
  - Adolescent Health Programme (AHP)
  - Healthy Child Programme (HCP)
  - Hospital at Night (HaN)
  - Leadership for Clinicians: Clinical Leadership (CLE)
  - Leadership for Clinicians: Medical Leadership (MLE)
  - NICE Guidance (NICE)
  - Research, Audit and Quality Improvement (R&A)
  - Safe Prescribing (SPB)
  - Safeguarding Children and Young People (SGC)
  - Shared Decision Making (SDM)

- NICE 01 - How to put guidance into practice
- NICE 02 - How to make evidence based decisions
- NICE 03 - How to bring about change
- NICE 04 - How to use audit to improve patient care

e-LfH is a Health Education England Programme in partnership with the NHS and Professional Bodies
**Summary**

- Ability to appraise and apply EBM crucial

- Develop a simple, logical approach
  - Relevance
  - Internal validity
  - External validity

- Don’t get bogged down with statistics

- Build skills through free courses and work place practice
Questions?

Thank you

ian.morris3@wales.nhs.uk