RACGP | Foundation

Evidence-based practice
'Train the trainer'
national workshop

Navigating the maze: Better searching, sifting and critical analysis of research

Workbook

Workshop presented by
Professor Paul Glasziou, Professor Chris Del Mar and Dr Scott Preston

racgp.org.au/foundation

Evidence-Based Practice Workbook

Contents

1.1 What practice changes have you made, and why?

Evidence-based practice is concerned with keeping up to date with clinically important and valid research that will help you manage your patients' problems. Before we look at how that might work, let's consider your current ways of keeping up to date.

What changes you have made in your clinical practice in the past year? These could be new treatments or new indications for old treatments, new tests, modifications to who or how you did practices, or something you stopped doing.

Exercise: List as many changes as you can in the table below. Then note the source – that is, who or where you got the information from, and what evidence was the change based on?

The Clinical Practice Change	Source: Who/where?	What Evidence was it based on?

Q1. What did you observe about the changes you have made?

_____

_____

Q2. What problems, if any, do you notice from the table? How might you improve that?

_____

_____

1.2 Exercise

An elderly patient comes in for regular review. At the end of the consultation she asks about her husband, who you have been treating for intractable depression.
"I heard a news item about ketamine for depression" she says. *'have you heard of that? New research… I am at my wits end, Doctor"*
Yes, you remember a recent news item in '6-Minutes':

www.6minutes.com.au/news

How would you respond?

Would you consider offering this?

(Respond before looking at the summary overleaf)

George D, Gálvez V, Martin D, Kumar D, Leyden J, Hadzi-Pavlovic D, et al. Pilot Randomized-Controlled Trial of Titrated Subcutaneous Ketamine in Older Patients with Treatment Resistant Depression. The Am J Geriatric Psych. 2017. [ePub ahead of print]

Abstract extract:

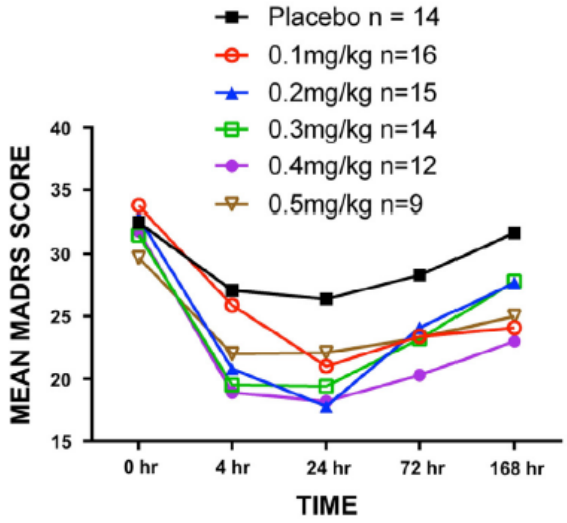
Methods

In this double-blind, controlled, multiple-crossover study with a 6-month follow-up (randomized controlled trial [RCT] phase), 16 participants (≥60 years) with treatment-resistant depression who relapsed after remission or did not remit in the RCT were administered an open-label phase. Up to five subcutaneous doses of ketamine (0.1, 0.2, 0.3, 0.4, and 0.5 mg/kg) were administered in separate sessions (≥1 week apart), with one active control (midazolam) randomly inserted (RCT phase). Twelve ketamine treatments were given in the open-label phase. Mood, hemodynamic, and psychotomimetic outcomes were assessed by blinded raters. Remitters in each phase were followed for 6 months.

Results

Seven of 14 RCT-phase completers remitted with ketamine treatment. Five remitted at doses below 0.5 mg/kg. Doses ≥ 0.2 mg/kg were significantly more effective than midazolam. Ketamine was well tolerated. Repeated treatments resulted in higher likelihood of remission or longer time to relapse.

Fig 2 MADRS scores for midazolam and all ketamine dose levels in the RCT phase.



MADRS is a depression rating scale

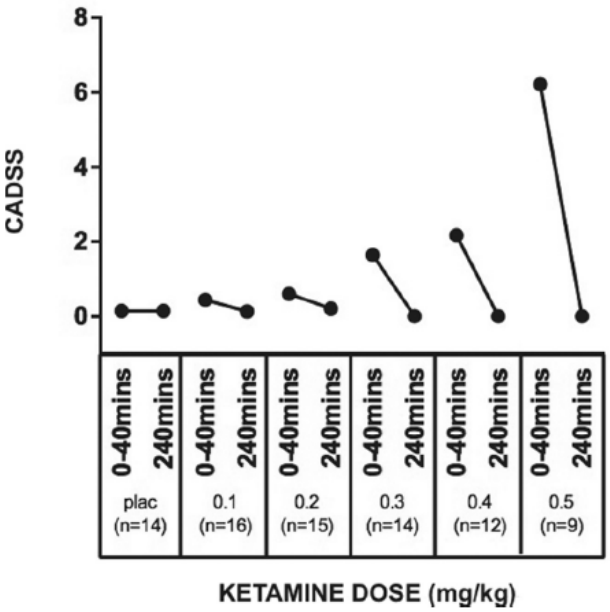
placebo

|

| ketamine

|

Fig 3 Psychotomimetic and dissociative side effects (measured by Clinician Administered Dissociative Symptoms Scale, CADSS)



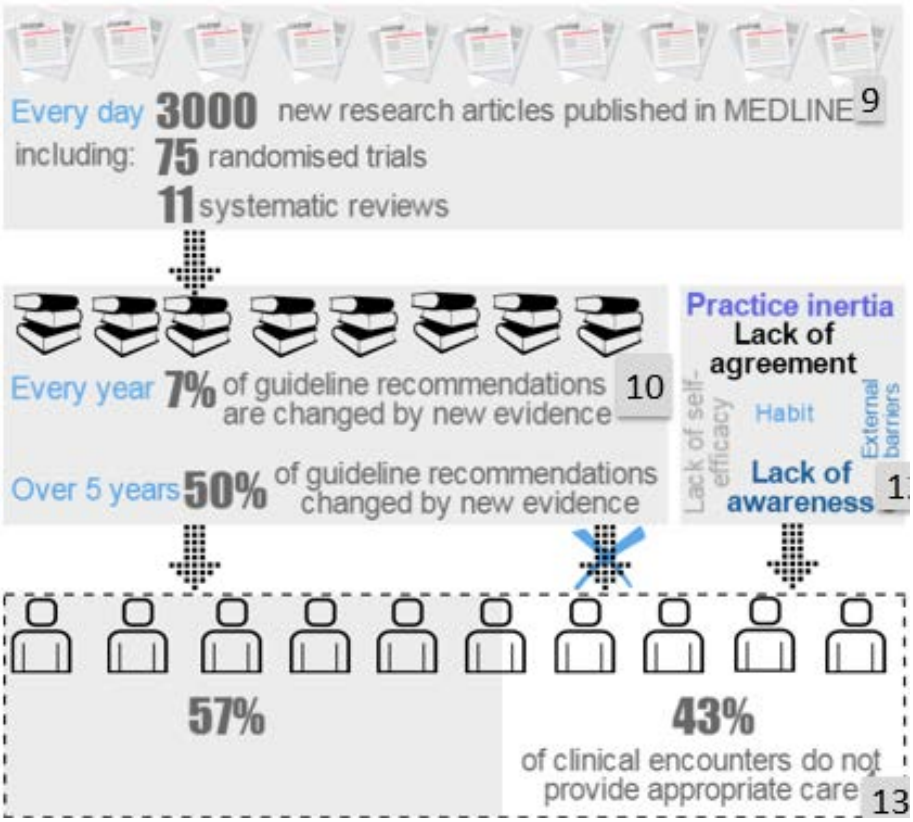
Do these Methods or Results change your inclination to offer this.

1.3 Why do we need Evidence Based Practice? (EBP) Module – Discussion and reading

There are 3 main reasons:

1. We don't make many changes in clinical practice (as we have just seen!)

2. New research is being generated faster than you might guess have a look at this figure,



How many of these new research articles do you think you read every year,

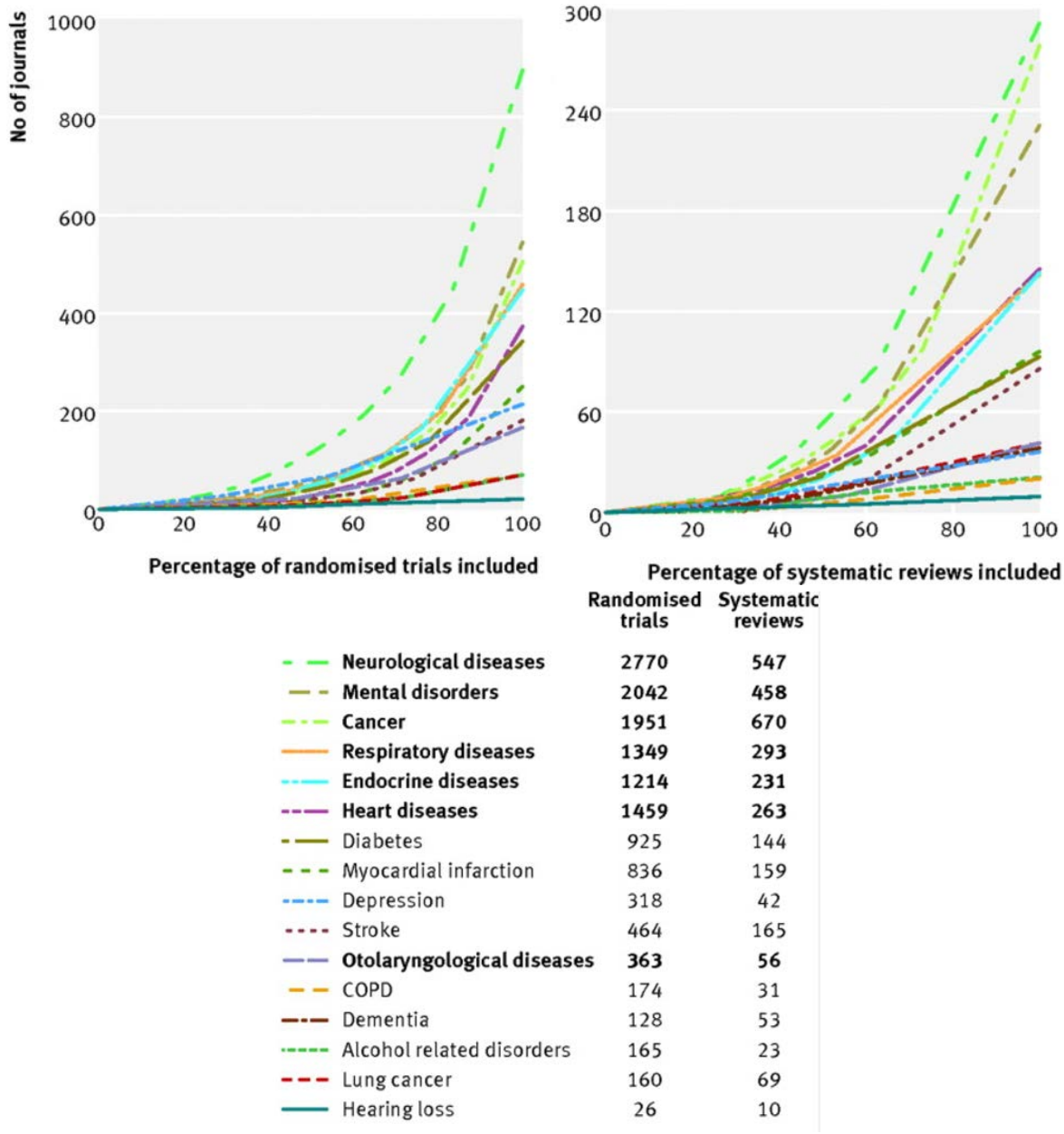
a) directly? _____

b) as a report (eg via the media)? _____

Reflect on any dissonance with Exercise 1.1

How easy is it to be notified about new research?

1.4 The Scatter of Research Information



	Randomised trials	Systematic reviews
Neurological diseases	**2770**	**547**
Mental disorders	**2042**	**458**
Cancer	**1951**	**670**
Respiratory diseases	**1349**	**293**
Endocrine diseases	**1214**	**231**
Heart diseases	**1459**	**263**
Diabetes	925	144
Myocardial infarction	836	159
Depression	318	42
Stroke	464	165
Otolaryngological diseases	**363**	**56**
COPD	174	31
Dementia	128	53
Alcohol related disorders	165	23
Lung cancer	160	69
Hearing loss	26	10

Hoffmann TC, Erueti C, Thorning S, Glasziou PP BMJ 2012;344:e3223

a) How many **Journals** would you have to read to find 50% of the **trials** in neurological disease?

_____

b) How many **Journals** would you have to read to find 50% of the **Systematic Reviews** in neurological disease?

_____

2.1 Basic Study Designs

Research uses many study designs, with different types of questions having different ideal designs. The best evidence comes from studies where the methods used maximise the chance of eliminating bias. The major basic designs are shown in the Table below:

DESIGN*	DESCRIPTION	STRUCTURE
Randomised controlled trial	Subjects are randomly allocated to treatment or control groups and outcomes assessed. **Ideal study for: treatments.**	
Cohort study	Outcomes are compared for matched groups with and without exposure or risk factor (prospective study). **Ideal study for: prognosis**	
Case-control study	Subjects with and without outcome of interest are compared for previous exposure or risk factor (retrospective study). **Good study for: rare events**	
Cross-sectional study	Preferably an independent, blind, comparison with 'gold standard' test. **Ideal for prevalence, diagnostic accuracy**	

Exercise: What is the STUDY DESIGN?

For each of the descriptions below, decide which study design above was used.

Question (from abstract)	Design?
1. "We investigated 1-year outcome and the pattern of recovery in 141 cardiac arrest survivors."	
2. "59 participants presenting mild-to-moderate symptoms of depression were randomized either to a behavioural intervention (n = 20), a physical activity intervention (n = 19) or a wait-list control (n = 20)."	
3. "The accuracy of the min-Cog was assessed in 400 patients with suspected dementia and compared to a geriatrician's diagnosis."	
4. "We compared lifetime migraine prevalence in patients with unruptured intracranial aneurysm to those with no aneurysm history."	
5. "Peoples' interest in having genetic testing for heart disease susceptibility was surveyed among a nationally representative sample of 1,960 British adults."	

2.2 The 'PICO' structure of research questions

Dissecting clinical questions into their component parts is an essential first step in EBP. Most questions can be divided into four components:

Patients	The relevant people or patients for the clinical problem.
Intervention (Indicator or Index text)*	The treatment (intervention), or other factor of interest, eg: • A treatment such as drug, surgery or diet (**intervention**) • Exposure (**indicator**) to an environmental chemical, a physical feature (eg overweight), or behaviour (smoking) • a test (**index test**), such as pathology, sign, or scan.
Comparator	The alternative or control exposure or test for comparison
Outcome(s)	The consequences you or your patient are concerned about.

Below is a study abstract showing the PICO elements as you might mark them up.

Pilot Randomized Controlled Trial of Titrated Subcutaneous Ketamine in Older Patients with Treatment-Resistant Depression

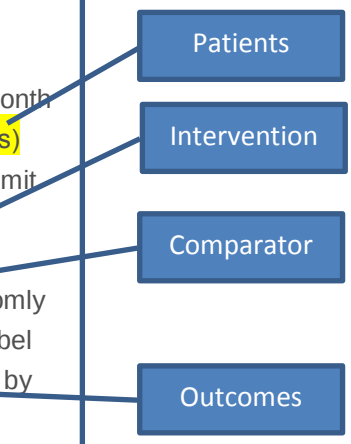
DuncanGeorge et al. The American Journal of Geriatric Psychiatry, June 2017.

Objective: To assess the efficacy and safety of subcutaneous ketamine for geriatric treatment-resistant depression. Secondary aims were to examine if repeated treatments were safe and more effective in inducing or prolonging remission than a single treatment.

Methods: In this double-blind, controlled, multiple-crossover study with a 6-month follow-up (randomized controlled trial [RCT] phase), 16 participants (≥60 years) with treatment-resistant depression who relapsed after remission or did not remit in the RCT were administered an open-label phase. Up to five subcutaneous doses of ketamine (0.1, 0.2, 0.3, 0.4, and 0.5 mg/kg) were administered in separate sessions (≥1 week apart), with one active control (midazolam) randomly inserted (RCT phase). Twelve ketamine treatments were given in the open-label phase. Mood, hemodynamic, and psychotomimetic outcomes were assessed by blinded raters. Remitters in each phase were followed for 6 months.

Results: Seven of 14 RCT-phase completers remitted with ketamine treatment. Five remitted at doses below 0.5 mg/kg. Doses ≥ 0.2 mg/kg were significantly more effective than midazolam. Ketamine was well tolerated. Repeated treatments resulted in higher likelihood of remission or longer time to relapse.

Conclusion: Results provide preliminary evidence for the efficacy and safety of ketamine in treating elderly depressed. Dose titration is recommended for optimizing antidepressant and safety outcomes on an individual basis. Subcutaneous injection is a practical method for giving ketamine. Repeated treatments may improve remission rates (clinicaltrials.gov; NCT01441505)



In the next exercises, we use the PICO structure to break down each type of clinical question.

Exercise: Study designs

Read the following abstracts from published studies and see if you can work out the researchable clinical question – the "PICO" or "PO".

Then for each, decide whether the question type is one of:
Treatment (so best would be a **randomized trial**)
Prognosis (so best would be an **inception cohort study**)
Diagnosis (so best would be a **cross-sectional study**)
Prevalence (s best would be a **survey / cross sectional study**)

Then work out what design they used.

Designation of levels of evidence according to type of research question

	Level	Treatment	Diagnosis[2]	Prognosis[1]	Aetiology[1,3]	Frequency (Point prevalence)
Least biased	I	Systematic review of level II studies	Systematic review of level II studies	Systematic review of level II studies	Systematic review of level II studies	Systematic review of level II studies
	II	Randomised controlled trial	Cross-sectional study among consecutive presenting patients	Inception cohort study	Prospective cohort study	Cross-sectional study of a representative sample
	III	One of the following: • non-randomised experimental study (eg controlled pre- and post-test intervention study) • comparative (observational) study with a concurrent control group (eg cohort study, case-control study)	One of the following: • cross-sectional study among non-consecutive patients • diagnostic case control study	One of the following: • untreated control patients in a randomised controlled trial • retrospectively assembled cohort study	One of the following: • retrospective cohort study • case-control study (Note: these are the most common study types for aetiology, but see level III for intervention studies for other options)	
Most biased	IV	Case series	Case series	Case series, or a cohort study of patients at different stages of disease	A cross-sectional study	

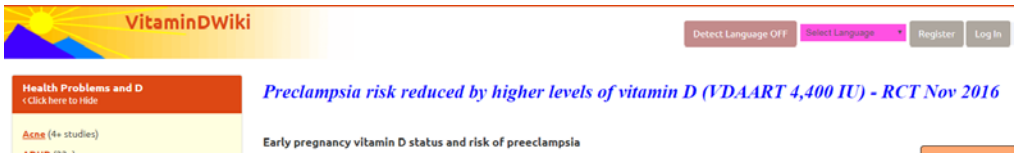
Abstract 1

Haugen, M. et al. Vitamin D supplementation and reduced risk of preeclampsia in nulliparous women. *Epidemiology. 2009 Sep;20(5):720-6.*

BACKGROUND: A recent study showed that nulliparous women who develop preeclampsia had low concentrations of vitamin D in serum sampled in midpregnancy. The aim of the present study was to estimate the association between intake of vitamin D during pregnancy and the risk of preeclampsia in 23,423 nulliparous pregnant women taking part in the Norwegian Mother and Child Cohort Study.
METHODS: Participating women answered questionnaires at gestational week 15 (general health questionnaire), at week 22 (food frequency questionnaire), and at week 30 (general health questionnaire). Pregnancy outcomes were obtained from the Medical Birth Registry. Nutrient intake was calculated from foods and dietary supplements. We estimated relative risks as odds ratios, and controlled for confounding with multiple logistic regression.
RESULTS: The odds ratio of preeclampsia for women with a total vitamin D intake of 15-20 microg/d compared with less than 5 microg/d was 0.76 (95% confidence interval = 0.60-0.95). Considering only the intake of vitamin D from supplements, we found a 27% reduction in risk of preeclampsia (OR = 0.73 [0.58-0.92]) for women taking 10-15 microg/d as compared with no supplements. No association was found between intake of vitamin D from the diet alone and the occurrence of preeclampsia.
CONCLUSIONS: These findings are consistent with other reports of a protective effect of vitamin D on preeclampsia development. However, because vitamin D intake is highly correlated with the intake of long chain n-3 fatty acids in the Norwegian diet, further research is needed to disentangle the separate effects of these nutrients.

Question	Answer
1. What is the question (PICO) of the study?	P
	I
	C
	O
2. What type of question does this study address?	
3. What is the study design?	
4. Is this the highest level of evidence for this question?	
5. If not, why do you think they used this study design?	

Abstract 2



Mirzakhani H, et al. Early pregnancy vitamin D status and risk of preeclampsia. J Clin Invest. 2016 Dec 1;126(12):4702-4715.

BACKGROUND: Low vitamin D status in pregnancy was proposed as a risk factor of preeclampsia.

METHODS: We assessed the effect of vitamin D supplementation (4,400 vs. 400 IU/day), initiated early in pregnancy (10-18 weeks), on the development of preeclampsia. The effects of serum vitamin D (25-hydroxyvitamin D [25OHD]) levels on preeclampsia incidence at trial entry and in the third trimester (32-38 weeks) were studied. We also conducted a nested case-control study of 157 women to investigate peripheral blood vitamin D-associated gene expression profiles at 10 to 18 weeks in 47 participants who developed preeclampsia.

RESULTS: Of 881 women randomized, outcome data were available for 816, with 67 (8.2%) developing preeclampsia. There was no significant difference between treatment (N = 408) or control (N = 408) groups in the incidence of preeclampsia (8.08% vs. 8.33%, respectively; relative risk: 0.97; 95% CI, 0.61-1.53). However, in a cohort analysis and after adjustmentfor confounders, a significant effect of sufficient vitamin D status (25OHD ≥30 ng/ml) was observed in both early and late pregnancy compared with insufficient levels (25OHD <30 ng/ml) (adjusted odds ratio, 0.28; 95% CI, 0.10-0.96). Differential expression of 348 vitamin D-associated genes (158 upregulated) was found in peripheral blood of women who developed preeclampsia (FDR <0.05 in the Vitamin D Antenatal Asthma Reduction Trial [VDAART]; P < 0.05 in a replication cohort). Functional enrichment and network analyses of this vitamin D-associated gene set suggests several highly functional modules related to systematic inflammatory and immune responses, including some nodes with a high degree of connectivity.

CONCLUSIONS: Vitamin D supplementation initiated in weeks 10-18 of pregnancy did not reduce preeclampsia incidence in the intention-to-treat paradigm. However, vitamin D levels of 30 ng/ml or higher at trial entry and in late pregnancy were associated with a lower risk of preeclampsia. Differentially expressed vitamin D-associated transcriptomes implicated the emergence of an early pregnancy, distinctive immune response in women who went on to develop preeclampsia.

Question	**Answer**
1. What is the question (PICO) of the study?	P
	I
	C
	O
2. What type of question does this study address?	
3. What is the study design?	
4. Is this the highest level of evidence for this question?	
5. If not, why do you think they used this study design?	

Abstract 3

Caglayan Sozmen S, Povesi Dascola C, Gioia E, Mastrorilli C, Rizzuti L, Caffarelli C. Diagnostic accuracy of patch test in children with food allergy. ***Pediatr Allergy Immunol. 2015 Mar 23. doi: 10.1111/pai.12377.***

BACKGROUND: The gold standard test for confirming whether a child has clinical hypersensitivity reactions to foods is the oral food challenge. Therefore, there is increasing interest in simpler diagnostic markers of food allergy, especially in children, to avoid oral food challenge. The goal of this study was to assess the diagnostic accuracy of atopy patch test in comparison with oral food food challenge.
METHODS: We investigated 243 children (mean age, 51 months) referred for evaluation of suspected egg or cow's milk allergy. Skin prick test and atopy patch test were carried out, and after a 2 weeks elimination diet, oral food challenge was performed.
RESULTS: Two hundred and forty-three children underwent OFC to the suspected food. We found clinically relevant food allergies in 40 (65%) children to egg and in 22 (35%) to cow's milk. The sensitivity of skin prick test for both milk and egg was 92%, specificity 91%, positive predictive value 35%, and negative predictive value of 93%. Sensitivity, specificity, positive predictive value, and negative predictive value of atopy patch test for both milk and egg were 21%, 73%, 20%, and 74%, respectively.
CONCLUSION: Our study suggests that there is insufficient evidence for the routine use of atopy patch test for the evaluation of egg and cow's milk allergy. OFC remains gold standard for the diagnosis of egg and milk allergy even in the presence of high costs in terms of both time and risks during application.

Question	**Answer**
1. What is the question (PICO) of the study?	P
	I
	C
	O
2. What type of question does this study address?	
3. What is the study design?	
4. Is this the highest level of evidence for this question?	
5. If not, why do you think they used this study design?	

Abstract 4

Brna, P., Dooley, J., Gordon, K., Dewan, T. (2005). The prognosis of childhood headache: a 20-year follow-up. *Archives of Pediatric and Adolescent Medicine* 159:1157-60.

BACKGROUND: Headaches affect most children and rank third among illness-related causes of school absenteeism. Although the short-term outcome for most children appears favorable, few studies have reported long-term outcome.
OBJECTIVE: To evaluate the long-term prognosis of childhood headaches 20 years after initial diagnosis in a cohort of Atlantic Canadian children who had headaches diagnosed in 1983.
METHODS: Ninety-five patients with headaches who consulted 1 of the authors in 1983 were previously studied in 1993. The 77 patients contacted in 1993 were followed up in 2003. A standardized interview protocol was used.
RESULTS: Sixty (78%) of 77 patients responded (60 of the 95 of the original cohort). At 20-year follow-up, 16 (27%) were headache free, 20 (33%) had tension-type headaches, 10 (17%) had migraine, and 14 (23%) had migraine and tension-type headaches. Having more than 1 headache type was more prevalent than at diagnosis or initial follow-up (*P*<.001), and headache type varied across time.
CONCLUSIONS: Twenty years after diagnosis of paediatric headache, most patients continue to have headache, although the headache classification often changes across time.

Question	**Answer**
1. What is the question (PICO) of the study?	P
	I
	C
	O
2. What type of question does this study address?	
3. What is the study design?	
4. Is this the highest level of evidence for this question?	
5. If not, why do you think they used this study design?	

Abstract 5

Fitzpatrick, M.F., Martin, K., Fossey, E., Shapiro, C.M. et al. (1993). Snoring, asthma and sleep disturbance in Britain: a community-based survey. *European Respiratory Journal* 6(4):531–535.

A questionnaire was sent to a random sample of adults in eight locations throughout Britain, to investigate the prevalence of snoring, asthma and sleep complaints in community-based British adults. Of the 1478 respondents (831 females, 647 males; mean +/- SD age 45 +/- 18 years), 37% reported snoring at least occasionally, and 11% reported snoring on at least four nights per week (frequent snorers). Frequent snorers reported spending less time asleep at night, falling asleep accidentally during the day more often, taking planned daytime naps, and falling asleep whilst driving or operating machinery more often than the other respondents.

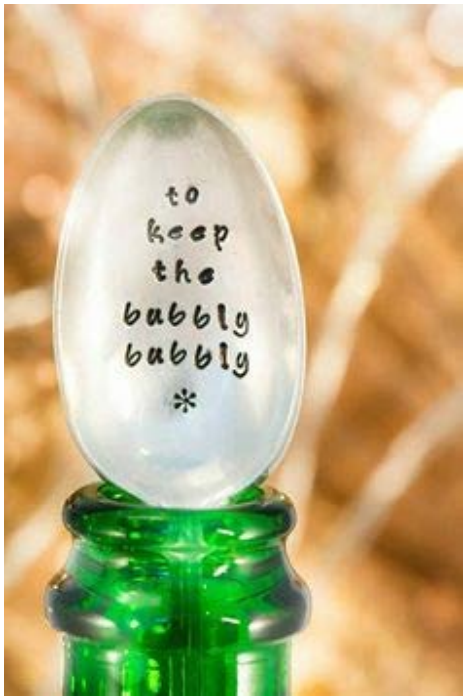
Using ordinal logistic regression analysis to allow for the age and sex of the respondents, both accidental daytime sleep and planned daytime naps were commoner in frequent snorers than other respondents. Six per cent of all respondents and 6% of those aged under 40 years reported that they had asthma (asthmatics). Seven per cent of respondents aged less than 40 years reported wheezing on three or more occasions per year, and had been prescribed oral or inhaled bronchodilators (young wheezers).

Question	**Answer**
1. What is the question (PICO) of the study?	P
	I
	C
	O
2. What type of question does this study address?	
3. What is the study design?	
4. Is this the highest level of evidence for this question?	
5. If not, why do you think they used this study design?	

3.1 Critical Appraisal

So we know that a randomized controlled trial is “the best” study for assessing the effects, positive and negative, or new treatments. But there are good trials and poor trials, and everything in between. This section looks at the critical features to look for to tell whether it is a good trial or a poor trial. Before we start on a full trial appraisal though, let’s think about champagne.

Exercise: How would you test if this worked?



Do you know the trick with the spoon and the champagne bottle?

A spoon put into an opened champagne bottle is supposed to keep the champagne fresh and bubbly for longer.

But is there any truth to this? **And how would you devise a study to test the claim?**

(Do this for yourself first, then discuss with your group)

When you have devised your own “test”, appraise the one over the page.

3.2 Is this a reasonable test?

The editors of a scientific journal tested this theory in a simple experiment many years ago. To do this, they put not just one, but two, half-empty open bottles in a refrigerator overnight: one with a spoon, one without.

Over the course of the next few days, they had volunteers sample both bottles several times, without telling them which one contained the spoon. This way, if the volunteers believed in the spoon's effect, this wouldn't influence their judgment. The first finding was that the champagne stayed drinkable for a surprisingly long time, taking more than four days to go flat. The second finding was that the participants could not tell the difference between champagne from a bottle with a spoon and from one without. Both bottles went flat equally fast.

For this study, what do you consider to be:

The Strengths?

_____

_____

_____

The Weaknesses?

_____

_____

_____

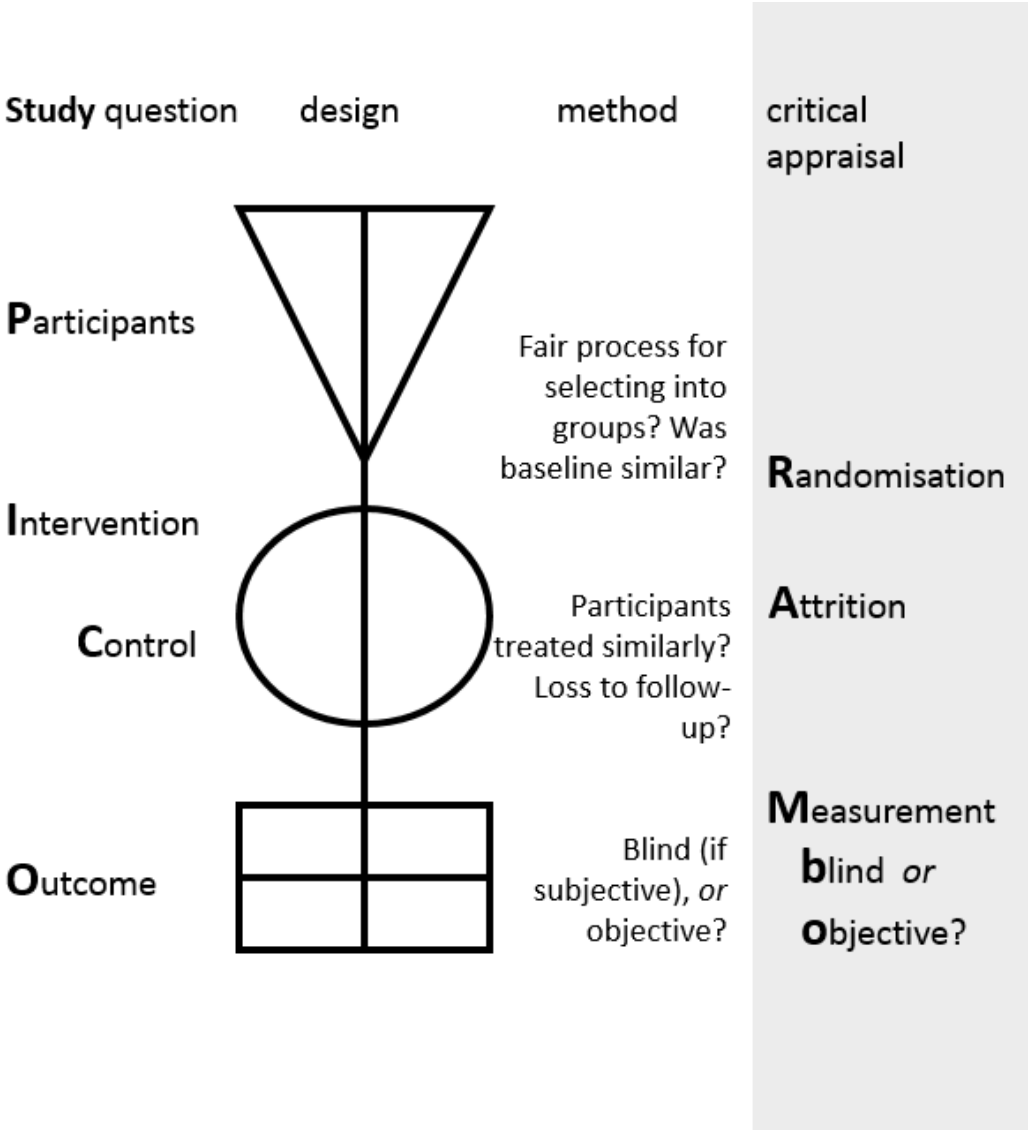
Does the suggested study have:

- ☐ Comparable groups?
- ☐ Low Attrition?
- ☐ Outcome measurements that are blinded OR objective?
- ☐ Sufficient numbers to have a reliable conclusion?

Footnote: For description of a more detailed experiment see: Richard Zare, Champagne bubble myth burst: Forget the silver spoon http://news.stanford.edu/pr/94/941221Arc4008.html

3.3 Critical appraisal in 2 mnemonics

Here are the two mnemonics for remembering the critical appraisal steps: PICO & RAMbo



The left hand side is the research question – the PICO – of the study, which can be seen as a flow from (1) recruiting the participants, (2), to allocating them intervention or control (or measuring exposed vs unexposed) and (3) following up to measuring the outcomes.
The right hand side – the RAMbo – shows the 3 key methodological issues we should check over this research flow: the randomization (or random selection for some questions), the attrition, and if the outcomes were objective or, if subjective, measured blind to treatment.

Let's now use this to appraise at a more complex study of a medical treatment.

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?	
What is best?	Where do I find the information?
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.	The *Methods* should tell you how patients were allocated to groups and whether or not randomisation was concealed.
This paper: Yes ☐ No ☐ Unclear ☐ Comment:	
1b. R- Were the groups similar at the start of the trial?	
What is best?	Where do I find the information?
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).	The *Results* should have a table of "Baseline Characteristics" comparing the randomized groups on a number of variables that could affect the outcome (ie. age, risk factors etc). If not, there may be a description of group similarity in the first paragraphs of the *Results* section.
This paper: Yes ☐ No ☐ Unclear ☐ Comment:	
2a. A – Aside from the allocated treatment, were groups treated equally?	
What is best?	Where do I find the information?
Apart from the intervention the patients in the different groups should be treated the same, eg., additional treatments or tests.	Look in the *Methods* section for the follow-up schedule, and permitted additional treatments, etc and in *Results* for actual use.
This paper: Yes ☐ No ☐ Unclear ☐ Comment:	
2b. A – Were all patients who entered the trial accounted for? – and were they analysed in the groups to which they were randomised?	
What is best?	Where do I find the information?
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*'.	The *Results* section should say how many patients were 20andomised (eg., Baseline Characteristics table) and how many patients were actually included in the analysis. You will need to read the results section to clarify the number and reason for losses to follow-up.
This paper: Yes ☐ No ☐ Unclear ☐ Comment:	
3. M - Were measures objective or were the patients and clinicians kept "blind" to which treatment was being received?	
What is best?	Where do I find the information?
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.	First, look in the *Methods* section to see if there is some mention of masking of treatments, eg., placebos with the same appearance or sham therapy. Second, the *Methods* section should describe how the outcome was assessed and whether the assessor/s were aware of the patients' treatment.
This paper: Yes ☐ No ☐ Unclear ☐ Comment:	

What were the results?

1. How large was the treatment effect?

Most often results are presented as dichotomous outcomes (yes or not 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.

What is the measure?	What does it mean?
Risk Ratio (RR) = risk of the outcome in the treatment group / risk of the outcome in the control group. In our example, the RR = 0.10/0.15 = 0.67	The relative risk tells us how many times more likely 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 no effect. An RR < 1 means that the treatment decreases the risk of the outcome. An RR > 1 means that the treatment increased the risk of the outcome. Since the RR < 1, the treatment decreases the risk of death.
Absolute Risk Reduction (ARR) = risk of the outcome in the control group - risk of the outcome in the treatment group. This is also known as the absolute risk difference. In our example, the ARR = 0.15 - 0.10 = 0.05 or 5%	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 no effect. The absolute benefit of treatment is a 5% reduction in the death rate.
Relative Risk Reduction (RRR) = 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) In our example, the RRR = 0.05/0.15 = 0.33 or 33% Or RRR = 1 - 0.67 = 0.33 or 33%	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. The treatment reduced the risk of death by 33% relative to that occurring in the control group.
Number Needed to Treat (NNT) = inverse of the ARR and is calculated as 1 / ARR. In our example, the NNT = 1/ 0.05 = 20	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. We would need to treat 20 people for 2 years in order to prevent 1 death.

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? (ExternalValidity/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?

Does a single cup of caffeinated drink significantly increase blood pressure in young adults? A randomised controlled trial

Cheong Lieng Teng, Wee Yang Lim, Chen Zhi Chua, Richard Soon Kiat Teo, Kenny Tze Hoe Lin, Jie Cong Yeoh

This paper was accepted as part of AFP's commitment to, on occasion, publish research that will 'make people laugh, then think' as highlighted in the annual IgNobel Prizes.

Background

Previous studies have shown that the blood pressure elevating effect of acute caffeine consumption was variable because of the heterogeneity of study participants, dosage of caffeine and study designs.

Objective

This research aimed to examine the effect of a single cup of coffee on the blood pressure of young adults.

Methods

Normotensive adults were randomised to receive either a cup of caffeinated drink (intervention group) or a cup of decaffeinated drink (control group). The main outcome measure was mean change in systolic blood pressure (SBP) and diastolic blood pressure (DBP) between intervention and control groups.

Results

Enrolled participants (n = 104) were randomly assigned to the intervention group (n = 53) or the control group (n = 51). The mean differences in SBP and DBP of the two groups were +2.77 mmHg (P = 0.05) and +2.11 mmHg (P = 0.64), respectively. Therefore, the rise in both SBP and DBP after caffeine consumption was not statistically significant.

Discussion

Our study confirmed that drinking a single cup of coffee (containing 80 mg of caffeine) does not have a significant impact on the blood pressure of healthy normotensive young adults one hour after the drink.

The potentially deleterious effect of coffee intake on cardiovascular health has been a subject of research interest.[1,2] Systematic reviews of randomised controlled trials on the chronic consumption of coffee/caffeine confirmed a small but statistically significant increase in systolic blood pressure (SBP; 2.0–2.4 mmHg) and diastolic blood pressure (DBP; 0.7–1.2 mmHg).[3,4] Interestingly, the reviews identified a somewhat larger blood pressure elevating effect of caffeine versus coffee[3] and a greater effect among younger study participants.[4] Reviews of the hypertensive effect of acute consumption of coffee/caffeine highlighted the heterogeneity of study participants, dosage of caffeine and study designs.[5–7]

In a review by Nurminen et al,[5] a single dose of caffeine (200–250 mg, equivalent to two to three cups of coffee) increased SBP by 3–14 mmHg and DBP by 4–13 mmHg in normotensive study participants. On the other hand, a review by James estimated that dietary caffeine will raise mean population blood pressure by approximately 4/2 mmHg.[6] The Australia New Zealand Food Standards states that there is 80 mg of caffeine in a 250 ml cup of caffeinated beverage, whereas percolated coffee has 60–120 mg of caffeine per cup.[8] However, in the US, a standard 8-ounce (240 ml) cup of coffee contains around 100 mg of caffeine, whereas espresso has a much higher content of caffeine (240–720 mg per 8 ounces).[9] Thus, we think it is more useful to ascertain if a typical cup of coffee (with 80 mg of caffeine) can produce a significant hypertensive effect.

Methods

Our study design was a randomised, double-blind, placebo-controlled clinical trial. The study participants were medical

students from the International Medical University (IMU) Clinical School in Seremban, Malaysia. The intervention and placebo were standard coffee drinks (caffeinated and decaffeinated, respectively) in Styrofoam cups (250 ml). The drinks were made with two teaspoons of instant coffee (2.6 g), 250 ml of warm water and one packet each of non-dairy creamer and sugar. The instant coffee chosen consisted of the caffeinated and decaffeinated varieties of a popular brand available commercially. The amount of caffeine in the intervention and placebo drinks was verified by a private laboratory (name available from the investigators) as 82.2 mg and undetectable, respectively. The amount of caffeine received by the participants in the intervention group was approximately 1.4 mg/kg of body weight. The drinks (labelled as 'A' and 'B' respectively) were freshly prepared by an investigator who was not involved in the subsequent part of the study.

A quiet room in the IMU library was used for this study. Medical students studying in the library were invited to participate. Eligible study participants were healthy male and female students aged >18 years and were free of chronic diseases, including hypertension. After obtaining informed consent, study participants were randomly assigned to either the intervention or control group. The randomisation list was created using the online Research Randomizer software,[10] and 110 codes (labelled as 'A' and 'B') were placed in opaque envelops, which were opened at enrolment. The investigators who recorded the study participants' blood pressure did not know which drink was caffeinated.

After randomisation, study participants completed a brief questionnaire about their sociodemographic data and personal history of hypertension. Then, their blood pressure was measured twice using the Omron SEM-1 digital blood pressure monitor. Measurement of blood pressure followed the recommendations of a guideline.[11] Study participants who had elevated blood pressure (≥140/90 mmHg) were excluded from the study. Sixty minutes after consumption of the coffee, the same study participants were invited for repeat blood pressure measurement (measured twice). The timing of the second set of blood pressure readings was based on the fact that peak plasma concentration of caffeine after a typical drink is reached within 40–60 minutes.[12]

Sample size estimation

We assumed that SBP elevation by 5 mmHg was clinically significant on the basis of data from Nurminen et al.[5,6] Using StudySize 2.0.4 software,[13] and assuming SD = 10 mmHg, at one-sided test of significance, we estimated the sample size to be 100 study participants (50 for each group). We increased the sample size by 10% (110, 55 per group) to account for possible dropouts.

Statistical analysis

We analysed the data using SPSS version 19. The main outcome measure was mean change in blood pressure after consuming the intervention or placebo drinks (mean of two blood pressure readings one hour after drink minus mean of two blood pressure readings just before drink). The mean difference in blood pressure and pulse rate between caffeinated and decaffeinated groups was compared using a paired t-test.

In view of the slight imbalance of sociodemographic variables in the two treatment groups, we used multiple linear regression to check if these variables independently influenced the change in SBP, and also to verify the value and direction of the standardised coefficient of treatment group as an independent variable.

The proportion of participants with ≥5 mmHg of blood pressure change in the intervention and control groups was compared using a chi-square test. Test of agreement using kappa statistic was used to check whether the participants could guess correctly the type of drink given. Statistical significance was set at $P < 0.05$.

The study was carried out from 9 to 20 December 2013 in the IMU library. The research proposal received approval from the International Medical University Joint Research & Ethics Committee (Grant Number: CSc/Sem6[23]2013) and was registered with the National Medical Research Registry (NMRR ID: NMRR-14-868-20259).

Results

Study participants

One hundred and ten participants were enrolled in this study. Of these, six participants were excluded from analysis as their initial blood pressure was ≥140/90 mmHg. All enrolled participants completed the randomised controlled trial. Their mean age was 22.41 years (SD = 1.02; range 19–26 years; Table 1).

Baseline and outcome data

A comparison of the study participants' baseline characteristics revealed similarities in the intervention and placebo groups (Table 1). As shown in Table 2, in the intervention group, both SBP and DBP showed a small increase (0.65 and 0.62 mmHg, respectively) one hour after coffee intake, but pulse rate showed a small reduction (–3.68 beats per minute). In the placebo group, SBP, DBP and pulse rate showed small decreases (–2.12 mmHg, –1.49 mmHg and –3.05 beats per minute, respectively) one hour after coffee intake. Comparing the intervention and placebo groups, there was no statistically significant change in SBP ($P = 0.05$). We explored the change in SBP using multiple linear regression (dependent variable: change in systolic blood pressure; independent variables: treatment groups, age, gender, ethnicity and daily coffee consumption) but did not find a statistically significant change in SBP by any of the five independent variables. The standardised coefficient of the treatment group was +0.172 (95% CI: 0.099, –0.468), but the change in SBP was not statistically significant). The change in DBP and pulse rate between the two groups did not reach statistical significance.

Other data

All study participants were asked to guess whether they were given the caffeinated drink. Forty-five participants (43.7%) guessed their drink correctly (one participant in the placebo group did not give an answer; kappa = –0.13; $P = 0.17$). We noted that 20 participants in the intervention group had an SBP elevation of at least 5 mmHg. The proportion of participants showing a change in SBP of ≥5 mmHg did not reach statistical significance ($\chi^2 = 3.591$, $P = 0.058$). However, when we stratified the above data by their coffee consumption, we found that participants who did not drink coffee daily were statistically more likely to have a change in SBP of ≥5 mmHg (Table 3).

Discussion

This randomised controlled trial showed that a single cup of coffee containing about 80 mg of caffeine produced a small change in SBP and DBP (3 and 2 mmHg, respectively). This change in blood pressure was not statistically significant and did not reach the a priori 5-mmHg elevation that we had set. We note that most of the difference in SBP between the intervention and control group was due to the lowering of blood pressure by around 2 mg in the decaffeinated group. However, we found that participants who did not consume coffee daily were more likely to show a change in SBP of ≥5 mmHg. This is in keeping with other observations that caffeine tolerance diminishes the acute effect of caffeine on blood pressure.[6]

Nurminen et al[5] identified 20 controlled studies that produced data on the hypertensive effect of caffeine on normotensive participants, with SBP elevation of 2–12 mmHg. Only two small controlled studies used a caffeine dose that was close to that in our study: Passmore[14] (n = 8, caffeine dose = 90 mg, SBP change = 5 mmHg) and Astrup[15] (n = 6, caffeine dose = 100 mg, SBP change = 2 mmHg). In comparison with the aforementioned studies, our randomised controlled trial has the strength of a larger sample size evaluating a specific research question.

The European guideline on home blood pressure measurement recommends avoiding caffeinated drink for at least 30 minutes prior to blood pressure measurement.[16] This guideline does not provide a reference for their recommendation, while Mort and Kruse

Table 1. Baseline characteristics of study participants

Characteristics	Intervention (n = 53)	Placebo (n = 51)
Age, years*	22.55 (1.12)	22.27 (0.90)
Male gender	23 (43.4%)	31 (60.8%)
Chinese ethnicity	42 (79.2%)	34 (66.7%)
Overweight/obese	12 (22.6%)	15 (29.4%)
Daily coffee drinker	12 (22.6%)	10 (19.6%)
Number of cups of coffee consume per week†	1 (0–15)	1 (0–21)
Blood pressure, mmHg*	112/65 (13/9)	116/69 (11/7)
Pulse rate, beats per minute*	83 (15)	81 (11)

**Data are means (SD) or numbers (%)*
†Data are median (range)

Table 2. Effect of caffeine on blood pressure and pulse rate

Outcome	Intervention (n = 53)	Placebo (n = 51)	Mean difference (95% CI)
Mean SBP before (mmHg)	112.11 (12.93)	115.88 (11.48)	
Mean SBP after (mmHg)	112.76 (12.12)	113.77 (12.08)	
SBP change (mmHg)	+0.65 (7.81)	–2.12 (6.28)	+2.77mmHg (0.00, +5.53)*
Mean DBP before (mmHg)	65.27(8.54)	66.87 (7.42)	
Mean DBP after (mmHg)	65.90 (7.77)	65.38 (8.07)	
DBP change (mmHg)	+0.62 (6.46)	-1.49 (4.91)	+2.11mmHg (–0.40, +4.07)†
Mean PR before (beats per minute)	82.46 (14.83)	81.34 (10.58)	
Mean PR after (beats per minute)	78.78 (14.46)	78.29 (9.48)	
PR change (beats per minute)	–3.68 (6.91)	–3.05 (5.05)	–0.63 beats per minute (–2.99, +1.73)‡

CI, confidence interval; DBP, diastolic blood pressure; PR, pulse rate; SBP, systolic blood pressure
**t =1.986, df = 102, P = 0.05*
†t =1.873, df = 102, P = 0.64
‡t = –0.529, df =102, P = 0.60

Table 3. Systolic blood pressure change of at least 5 mmHg among study participants stratified by coffee consumption

Coffee consumption	Treatment group	SBP change ≥5 mmHg	SBP change <5 mmHg	Total
Daily*	Intervention	2 (20.0%)	8 (80.0%)	10
	Placebo	3 (25.0%)	9 (75.0%)	12
		5 (22.7%)	17 (77.3%)	22
Not daily†	Intervention	11 (26.8%)	30 (73.2%)	41
	Placebo	4 (9.8%)	37 (90.2%)	41
		15 (18.3%)	67 (81.7%)	82

*x^2=0.078, P = 0.781
†x^2=3.998, P = 0.046

mentioned that blood pressure changes due to caffeine occur within 30 minutes and peak in 1–2 hours.[7] Our study confirms that drinking a single cup of coffee (containing 80 mg of caffeine) does not have a major impact on blood pressure in healthy, normotensive young adults one hour after the drink, except in those who are not habitual coffee drinkers. We cannot exclude the possibility of a small but statistically significant blood pressure elevation (<5 mmHg) due to caffeine, which may be shown by a larger sample study. The findings of our study cannot be extrapolated to older adults or those with pre-existing hypertension, who may have a greater pressor response to caffeine.[17] In view of the common usage of home blood pressure monitors, this information will be of some help in giving clear advice to young adults prior to their home blood pressure measurement.

Acknowledgement

We wish to thank Zuhanariah Mohd Nordin, Chief Librarian of IMU Library, for allowing the study to be conducted in the library.

Authors

Cheong Lieng Teng MFamMed, FRACGP, Professor, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia. tengcl@gmail.com

Wee Yang Lim, Medical Student, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia

Chen Zhi Chua, Medical Student, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia

Richard Soon Kiat Teo, Medical Student, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia

Kenny Tze Hoe Lin, Medical Student, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia

Jie Cong Yeoh, Medical Student, Department of Family Medicine, International Medical University, Clinical School, Jalan Rasah, Negeri Sembilan, Malaysia

Competing interests: None.

Provenance and peer review: Not commissioned, externally peer reviewed.

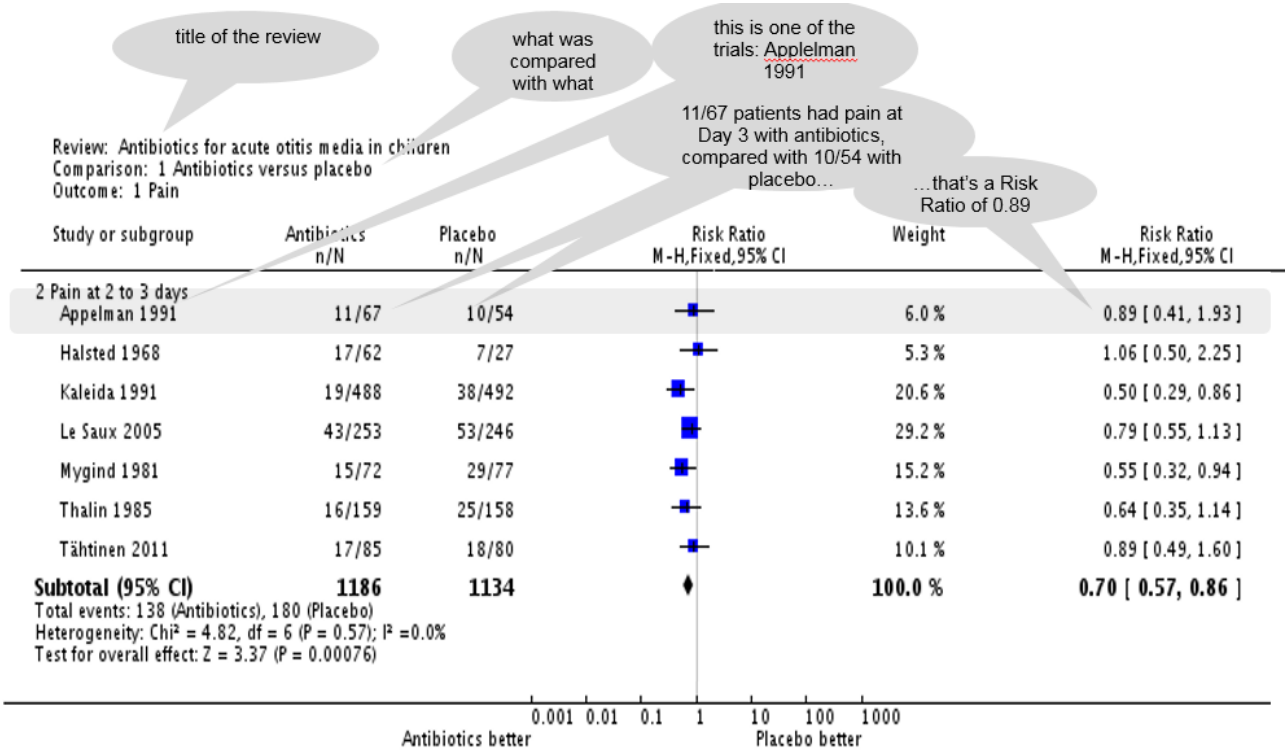
References

1. Kleemola P, Jousilahti P, Pietinen P, Vartiainen E, Tuomilehto J. Coffee consumption and the risk of coronary heart disease and death. Arch Intern Med 2000;160:3393–400.
2. Higdon JV, Frei B. Coffee and health: A review of recent human research. Crit Rev Food Sci Nutr 2006;46:101–23.
3. Noordzij M, Uiterwaal CSPM, Arends LR, Kok FJ, Grobbee DE, Geleijnse JM. Blood pressure response to chronic intake of coffee and caffeine: A meta-analysis of randomized controlled trials. J Hypertens 2005;23:921–28.
4. Jee SH, He J, Whelton PK, Suh I, Klag MJ. The effect of chronic coffee drinking on blood pressure: A meta-analysis of controlled clinical trials. Hypertension 1999;33:647–52.
5. Nurminen ML, Niittynen L, Korpela R, Vapaatalo H. Coffee, caffeine and blood pressure: A critical review. Eur J Clin Nutr 1999;53:831–39.
6. James JE. Critical review of dietary caffeine and blood pressure: A relationship that should be taken more seriously. Psychosom Med 2004;66:63–71.
7. Mort JR, Kruse HR. Timing of blood pressure measurement related to caffeine consumption. Ann Pharmacother 2008;42:105–10.
8. Australia New Zealand Food Standards. Caffeine. Canberra, ACT: ANZFS, 2014. Available at www.foodstandards.gov.au/consumer/generalissues/Pages/Caffeine.aspx [Accessed 14 June 2015].
9. Heckman MA, Weil J, Gonzalez de Mejia E. Caffeine (1, 3, 7-trimethylxanthine) in foods: A comprehensive review on consumption, functionality, safety, and regulatory matters. J Food Sci 2010;75:R77–87.
10. Urbaniak GC, Plous S. Research Randomizer. Social Psychology Network. Available at www.randomizer.org [Accessed 7 July 2015].
11. Pickering TG, Hall JE, Appel LJ, et al. Recommendations for blood pressure measurement in humans and experimental animals: Part 1: Blood pressure measurement in humans: A statement for professionals from the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Hypertension 2005;45:142–61.
12. Undem BJ. Chapter 27: Pharmacotherapy of asthma. In: Brunton LL, Lazo JS, Parker KL, editors. Goodman and Gilman's The pharmacological basis of therapeutics. 11th edn. New York: McGraw-Hill, 2006;728.
13. Orloffson B. StudySize 2.0.4. Sweden: CreoStat HB, 2008.
14. Passmore AP, Kondowe GB, Johnston GD. Renal and cardiovascular effects of caffeine: A dose-response study. Clin Sci (Lond) 1987;72:749–56.
15. Astrup A, Toubro S, Cannon S, et al. Caffeine: A double-blind, placebo-controlled study of its thermogenic, metabolic, and cardiovascular effects in healthy volunteers. Am J Clin Nutr 1990;51:759–67.
16. Parati G, Stergiou GS, Asmar R, et al. European Society of Hypertension Practice Guidelines for home blood pressure monitoring. J Hum Hypertension 2010;24:779–85.
17. Izzo JL Jr, Ghosal A, Kwong T, Freeman RB, Jaenike JR. Age and prior caffeine use alter the cardiovascular and adrenomedullary responses to oral caffeine. Am J Cardiol 1983;52:769–73.

correspondence **afp@racgp.org.au**

4. Systematic reviews. These are pooled results of studies. Finding your way round a forest plot

Can you find the **Title**, and the **PICO** elements? **(Patients; Intervention; Comparator; Outcome)**



A. Now let us look in detail at the third trial (Kaleida 1991).

Q1. How many patients had pain at Day 3 with antibiotics? _____

Q2. How many had pain at Day 3 with 10/54 with placebo? _____

Q3. What was the risk of having pain *with* antibiotics? _____

_____

Q4. What was the risk of having pain *without* antibiotics? _____

_____

_____

Q5. What was the ratio of these risks? (This will give you the chance of a child who took antibiotics of having pain compared to a child taking placebo, at Day 2-3)

_____

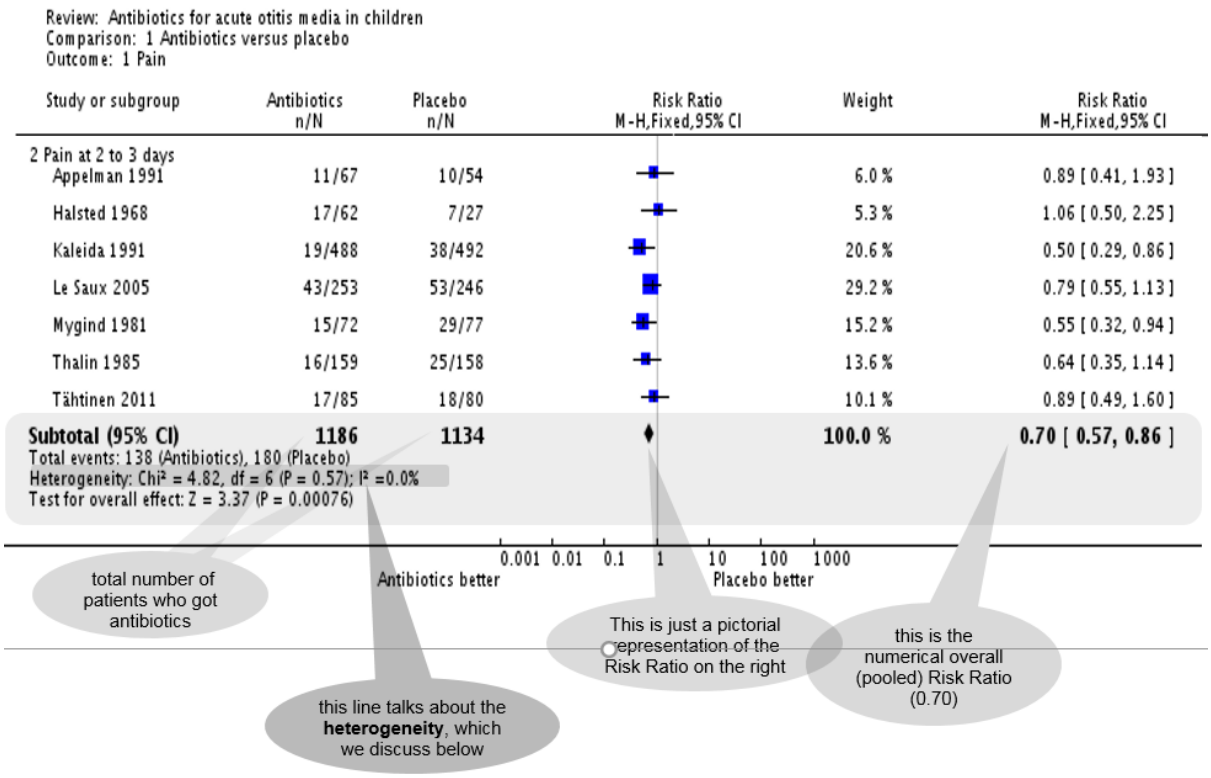
_____

Q6. What does the "95% confidence interval (95% CI)" mean?

_____

_____

B. Now let's look at the bottom row, called the Subtotal. This is where the results of these 7 trials are pooled ('meta-analysed').



Q1. How many trials were combined for this analysis? _____

Q2. What was the total number of participants in the trials combined? _____

Q3. What was the meta-analysed Risk Ratio? _____

_____

Q4. How many trials showed a significant reduction of pain for antibiotics? _____

Q5. Was this **statistically significant?** (Hint: look at the upper and lower limit of the 95% CI)

_____

Q6. What does this mean: “the risk ratio was 0.70”?

_____

_____

Q7. What is **Heterogeneity**, and why is it important?

_____

_____

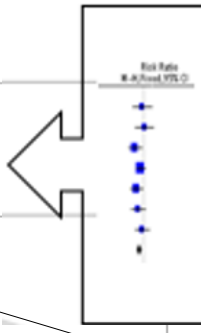
Summary of Findings Table

These are very useful for gaining a rapid overview of a systematic review. You can gain a picture of evidence, the size of any effect, and the quality of the evidence, all in one Table!

Antibiotics versus placebo for acute otitis media in children

Patient or population: children with acute otitis media
Settings: primary care and secondary care
Intervention: antibiotics versus placebo

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	No of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Antibiotics versus placebo				
Pain - pain at 24 hours	Study population		RR 0.89 (0.78 to 1.01)	1394 (5 studies)[1]	⊕⊕⊕⊕ high	
	426 per 1000	379 per 1000 (332 to 431)				
Pain - pain at 2 to 3 days	Study population		RR 0.70 (0.57 to 0.86)	2320 (7 studies)	⊕⊕⊕⊕ high	
	159 per 1000	111 per 1000 (90 to 137)				
Pain - pain at 4 to 7 days	Study population		RR 0.76 (0.63 to 0.91)	1347 (7 studies)[1]	⊕⊕⊕⊕ high	
	241 per 1000	183 per 1000 (152 to 220)				
Pain - pain at 10 to 12 days	Study population		RR 0.33 (0.17 to 0.66)	278 (1 study)	⊕⊕⊕⊕ moderate[2]	
	216 per 1000	71 per 1000 (37 to 142)				
Abnormal tympanometry - 2 to 4 weeks	Study population		RR 0.82 (0.74 to 0.90)	2138 (7 studies)	⊕⊕⊕⊕ high	
	481 per 1000	395 per 1000 (356 to 433)				
Abnormal tympanometry - 3 months	Study population		RR 0.97 (0.76 to 1.24)	809 (3 studies)	⊕⊕⊕⊕ high	
	241 per 1000	234 per 1000 (183 to 299)				
Vomiting, diarrhoea or rash	Study population		RR 1.38 (1.19 to 1.59)	2107 (8 studies)	⊕⊕⊕⊕ high	
	196 per 1000	270 per 1000 (233 to 311)				



Each row of the SoF Table represents a different forest plot – for example, a different outcome

1. this is the outcome discussed above

2. the absolute difference is calculated from …

3. … the risk ratio

4. this describes how strong the evidence is: GRADE

C. More questions!

Q1. How many outcomes are displayed in this Summary of Findings (SoF) table? _____

Q2. How many outcomes describe 'adverse events' of the intervention? _____

5. Guidelines: is it evidence-based?

Not all guidelines are well developed and evidence-based. Below is the "Mini-AGREE" checklist which the RACGP Quality Committee uses for assessing guidelines. The 8 questions are a subset of the full AGREE checklist. If we were to suggest only 2, then we would look at:

A. The guideline authors are an appropriate mix of disciplines with minimal or managed conflicts of interest (#4)
B. The guidelines are evidence based – with a systematic search and evidence criteria (# 5)

Methodological Guideline Quality – Mini-Checklist

1. The guideline has been written in a generally comprehensible manner and its key recommendations are easy to identify.

□ YES □ TO SOME EXTENT □ NO

2. The guideline's target audiences and scope of application were specified.

□ YES □ TO SOME EXTENT □ NO

3. The background, the objectives of the guideline, and the patients for whom the guideline is relevant were clearly described.

□ YES □ TO SOME EXTENT □ NO

4. The persons that developed the guideline are named, and their financial independence and any conflicts of interest are clearly documented.

□ YES □ TO SOME EXTENT □ NO

5. The search for evidence was systematic and the criteria used to select evidence were described.

□ YES □ TO SOME EXTENT □ NO

6. The guideline recommendations are unambiguous and the evidence they are based on is clearly presented.

□ YES □ TO SOME EXTENT □ NO

7. Different treatment options are presented that take account of potential benefits, side effects and risks.

□ YES □ TO SOME EXTENT □ NO

8. Clear information is provided on how up-to-date the guideline is and for how long this is expected to be the case.

□ YES □ TO SOME EXTENT □ NO

Exercise: Read the first page below and indicated your YES/NO/SOME for each question.

CMAJ GUIDELINES

Recommendations for preventing fracture in long-term care

Alexandra Papaioannou MD MSc, Nancy Santesso RD PhD, Suzanne N. Morin MD MSc, Sidney Feldman MD, Jonathan D. Adachi MD, Richard Crilly BSc MD, Lora M. Giangregorio PhD, Susan Jaglal PhD, Robert G. Josse MBBS, Sharon Kaasalainen PhD, Paul Katz MD, Andrea Moser MD MSc, Laura Pickard MA, Hope Weiler RD PhD, Susan Whiting PhD, Carly J. Skidmore MSc, Angela M. Cheung MD PhD; for the Scientific Advisory Council of Osteoporosis Canada

CMAJ Podcasts: author interview at https://soundcloud.com/cmajpodcasts/141331-guide

The 2010 clinical practice guideline for the diagnosis and management of osteoporosis in Canada[1] focused on the care of adults living in the community. However, the fracture rate for adults living in long-term care (residents) is two to four times that of adults of similar age living in the community, and one-third of older adults who experience hip fracture are residents in long-term care.[2] Hip fracture is one of the most serious consequences of osteoporosis and also one of the leading causes of admission to hospital.[3] When residents return to long-term care after a hospital stay, they need additional hours of specialized care.[4,5] In addition, fracture pain and delirium frequently associated with analgesia are distressing for residents and their families. Vertebral fractures are also a concern for residents, and the reported prevalence is up to 30% (for at least one moderate to severe fracture).[6] Multiple vertebral fractures can be a substantial cause of pain, anxiety, depression, reduced pulmonary function[7] and agitation.

Frail older adults at high risk of fracture in long-term care face other challenges. More than 40% have dementia,[8] a similar percentage experience swallowing difficulties,[9,10] and over 20% may have renal insufficiency.[11,12]

It may be difficult to identify residents at high risk of fracture, as the current fracture risk assessment tools (the Canadian Association of Radiologists and Osteoporosis Canada tool[13] [CAROC; www.osteoporosis.ca/multimedia/pdf/CAROC.pdf] and the Canadian WHO Fracture Risk Assessment Tool [FRAX; www.shef.ac.uk/FRAX/]) provide 10-year fracture risk and have not been validated in long-term care, where over 20% of residents may die within one year of admission.[14,15] Most research regarding risk assessment and pharmacologic therapies has not included those wi

Scope

This document provides guidance regarding strategies for the prevention of fractures directed toward interprofessional teams caring for frail older adults in long-term care.

Methods

This guideline, which has been endorsed by Osteoporosis Canada, was developed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach[18,19] (www.gradeworkinggroup.org), in a process led by a GRADE methodologist (N.S.). The guideline panel comprised the authors, other multidisciplinary health care providers and researchers, and representatives from resident and family councils (see Appendix 1, available at www.cmaj.ca/lookup/suppl/doi:10.1503/cmaj.141331/-/DC1). The panel was first surveyed to prioritize questions and important outcomes. In addition to fractures (hip, vertebral and nonvertebral), the group as a whole identified pain, quality of life, loss of activities of daily living and mobility, death and adverse events requiring medical attention as important outcomes. Family members of residents ranked prevention of pain and maintenance of mobility as most important.

We conducted systematic searches of the literature for published network meta-analyses, system-

Competing interests: See end of article.

This article h been peer reviewed.

Corres ce to:

KEY POINTS

- In older adults living in long-term care (residents agitation, immobility and transfers to hospital.
- Residents identified as being at high risk of fractu prior fracture of the hip or spine, those with more and those with one prior fracture and recent use of glucocorticoids.
- Recommendations for preventing fracture in long-term care were developed using the GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach, with consideration of the quality of the available evidence, the balance between benefits and harms, the preferences of residents and their care providers, and the resources required to implement the recommendations.
- Strategies to prevent fractures, including vitamin D and calcium supplementation, use of hip protectors, exercise, multifactorial interventions to prevent falls and pharmacologic therapies, should be tailored to each resid s level of fracture risk, mobility, life expectancy, renal function and swallow.

4. developers named and COIs documented

4. developers named and COIs documented

3. objectives of the guideline

5. search for evidence was systematic; criteria used to select evidence

5. search for evidence was systematic; criteria used to select evidence

2. audiences and scope
3. relevant patients

7. different treatment options are presented that take account of potential benefits, side effects and risks

6. Searching skills

The basic steps are:

- ❑ define your clinical question [PICO]
- ❑ choose key search terms for each of the PICO elements

Some additional discretionary steps:

- ❑ synonyms, wildcards, quotes
- ❑ combined with “Boolean” operators

The PICO (elements of the Q) forms the basis of the search

Synonyms, wildcards, quotes

Synonyms: Interchangeable medical terms

Examples and exercises

Example	**Other terms**
myocardial infarction	AMI, coronary thrombosis heart attack
Exercise	**Your other terms**
acute otitis media	
grommets	

Spelling variants

Examples and exercises

Example	**Other Terms**
pediatric	Paediatric
randomized	randomised
Exercise	**Your variants**
haemoglobin	

Some special tips

Wildcards

The * symbol searching for child* will yield 'child', 'children', 'childhood' …

Note: the wildcard symbol is different in different databases. Check first!

Exercise:

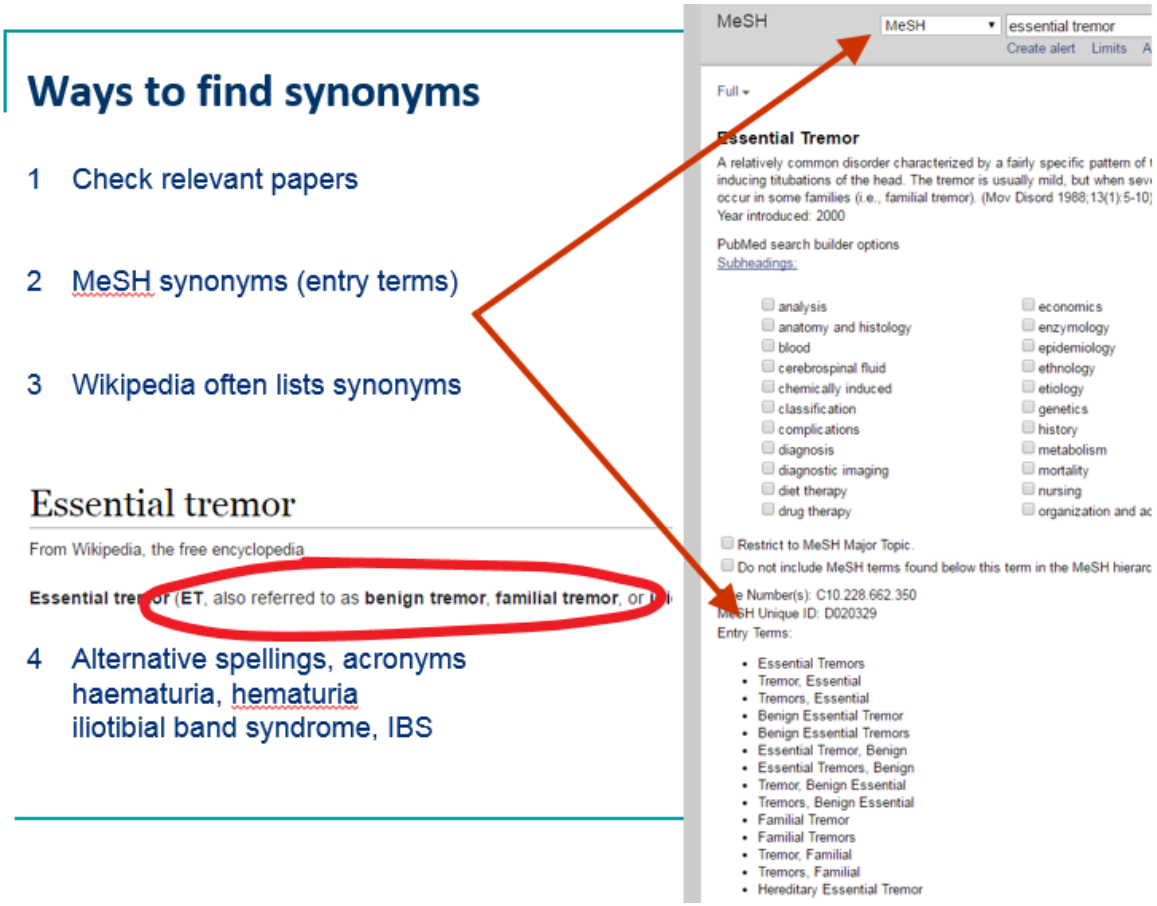
How would you use wildcards for:

1. Obesity, obese, etc _____
2. Anaemic, anaemia, etc _____

Quotes

These force the search to only yield your search terms if

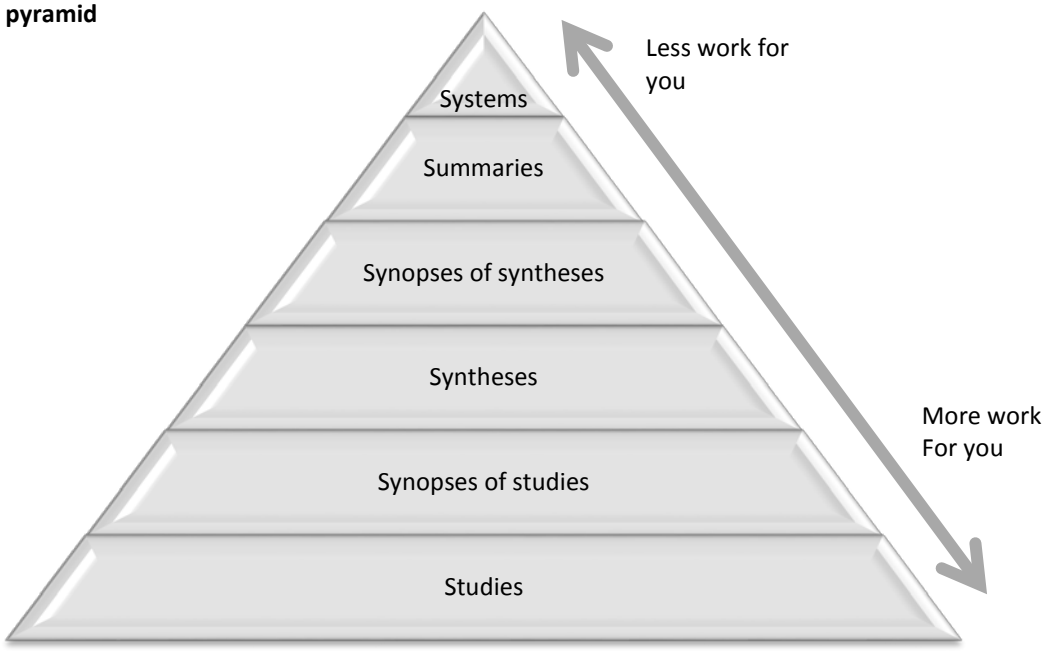
- they are found together; *and* in the same order:
  eg. "myocardial infarct" "acute otitis media"



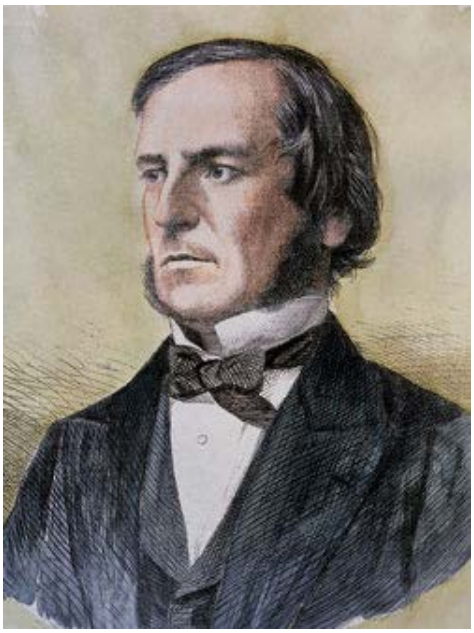
What's the *type* of evidence?

Evidence comes in several levels, and it is important to know which sort you are chasing.

The '6 S' pyramid



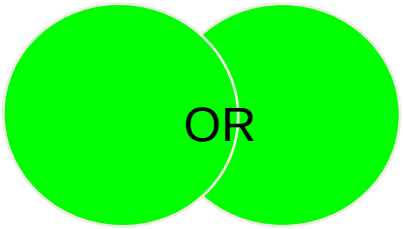
Combining words with Boolean operators



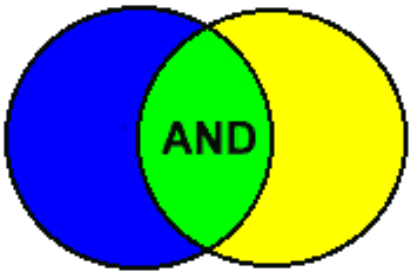
http://schools.keldysh.ru/sch444/museum/1_17-19.htm

George Boole was a mathematician-philosopher who anticipated computers in the 1800s, and developed a logic language for automating some tasks.

There are 2 terms that are worth learning (the others are optional)



OR → combines terms



AND → only returns overlaps

(Other Boolean terms include **NOT** and **NEAR**, but they are not necessary)

Then you can combine synonyms together with the elements of the evidence question (the PICO)

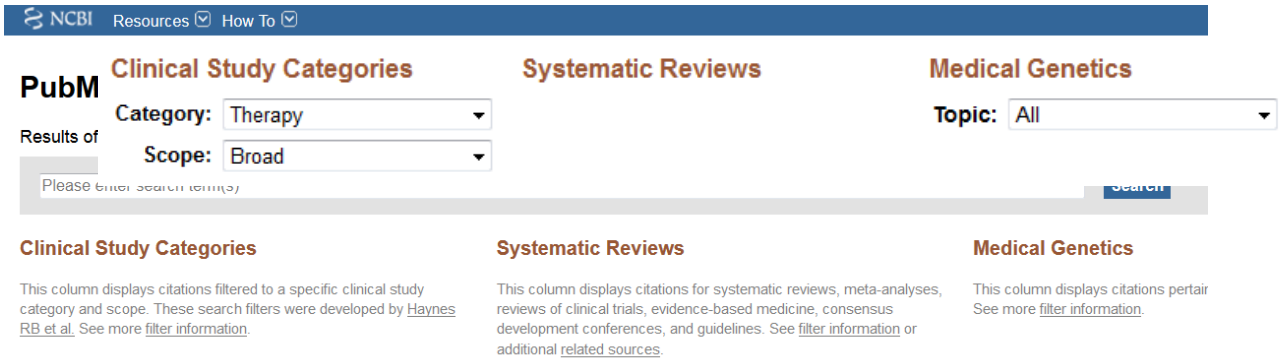
For example, you are wondering about which of two treatments, systemic antivirals or corticosteroids for Bell's palsy are most effective:

P		I		C		O
Bell's palsy	**AND**	antivirals	**AND**	corticosteroids	**AND**	paralysis
OR		**OR**		**OR**		**OR**
facial paralysis		acyclovir		dexamethasone		
OR		**OR**		**OR**		**OR**
fifth cranial nerve palsy				prednisone		
OR				**OR**		
				prednisolone		

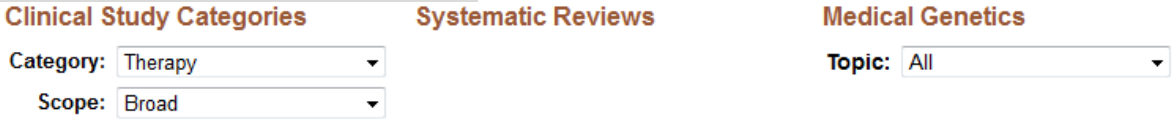
Searching PubMed Clinical Queries

This is really useful for doing rapid searches.

Go to http://www.ncbi.nlm.nih.gov/pubmed/clinical, (or Google 'pubmed clinical queries')



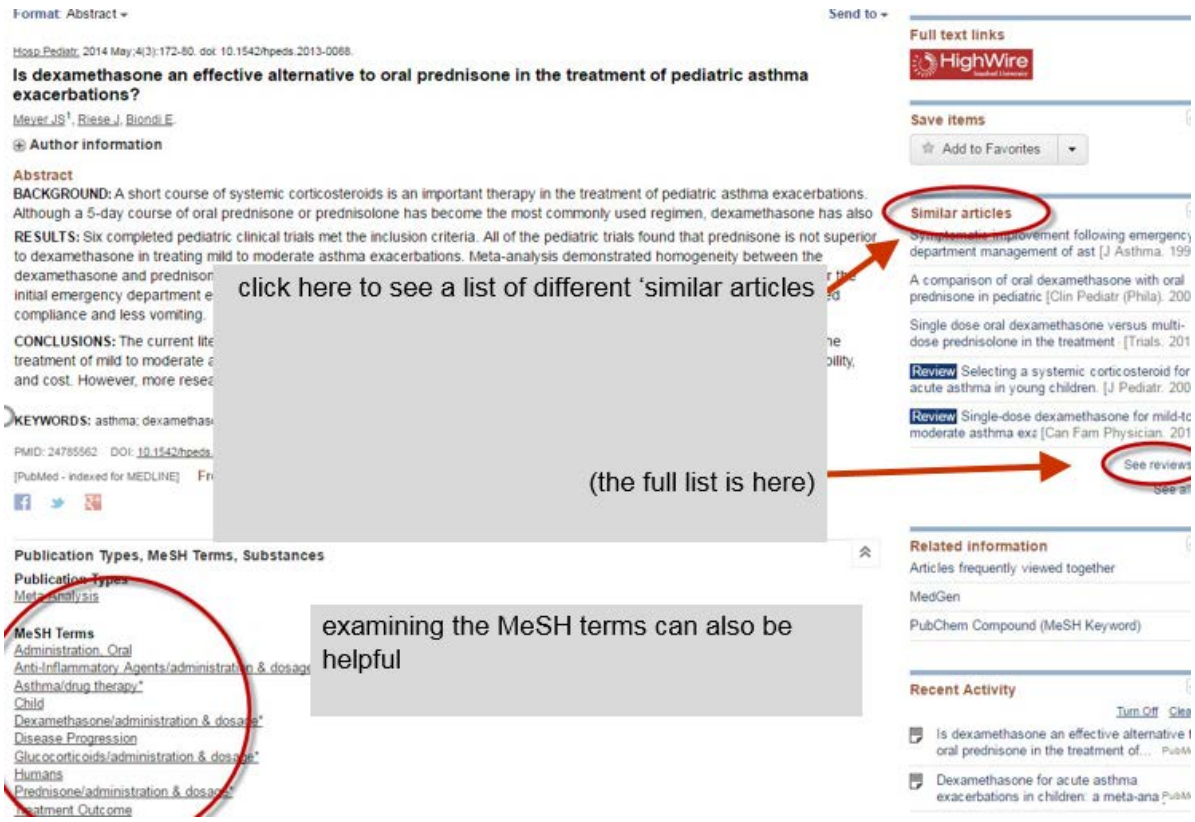
Then choose 'broad' or 'narrow'



Clinical Queries deals with the 'study-type' question for you!

‘Similar articles’

This is a feature of PubMed that is useful if you have found one good article, and you want to check if there are similar ones.



Exercise

Finding synthesised evidence

1. You've seen articles in the medical newspapers about fish oil curing psychosis, you're sceptical. Use PubMed Clinical Queries to find the most synthesised evidence you can.

Clinical guideline (paste the title here):

Systematic review (paste the title here):

2. A colleague mentions that he used an age-adjusted D-dimer test for an elderly patient with suspected deep vein thrombosis. You had heard of age-adjusted D-dimers, but want to brush up your knowledge. Using any search platform you know about find the most synthesised evidence you can. You can copy and paste this search:

   ("deep vein thrombosis" OR "venous thrombosis" OR DVT) AND (age* OR elderly) AND d-dimer

   Paste the title of two sources you find here:

7. Putting it all together

Consider your "changes" that you listed in the first exercise. How can you improve the quantity and quality of the changes you make in practice? Three activities you might consider are:

1. Evidence Alert services

There are several good services which will alert you to new important and valid articles.

The McMaster Plus clinical alerts service provides a weekly email which you can personalize: https://plus.mcmaster.ca/ACCESSSS

Richard Lehmann, a UK GP, provides a weekly summary of the key general medical journals: http://blogs.bmj.com/bmj/category/richard-lehmans-weekly-review-of-medical-journals/

Do you know of others? Discuss which of these you might regularly monitor.

2. Clinical Questions log

During your clinical practice, keep a "logbook" (paper or electronic) of important clinical questions which arise. Chose 1 (or more) of these each week to try out the 4 step process of EBM (see next page).

3. Journal Club

Based on both the alert service and clinical questions you have answered, try running a regular journal club in your practice. Read the attached article on the processes, and discuss the "*Factors that engender success in EBM journal clubs*".

There are several quite distinct processes, which can be remembered as 'the 4 As', summarised:

Ask… an 'answerable' question	This is often the most difficult first step. … • being humble enough to admit we don't know things; • converting everyday clinical questions into a format that can be answered. **P** (patient; population) **I** (intervention; index) **C** (comparator; control) **O** (outcome – what it is that measure) to help you remember the essential components of the question. • remember: some questions aren't answerable by EBP.
Access…… the literature to find suitable studies	Usually this means **searching** electronic databases of published research. This is not hard to learn, but it takes practice. Further on, we devote some tips and shortcuts to make it easier.
Appraise…… the best studies to decide whether the results are valid.	This is called 'critical appraisal'. We have mnemonics to help you to check the important bits of the study. Most commonly used are:- For randomised trials: **R**andomisation **A**ttrition **M**easurement – was this either **b** 'blinded', *or* **o** 'objective'? For systematic reviews of randomised trials: **F** 'find' (where all the trials found?) **A** 'apprasied' (was the quality appraised?) **I** 'included' (were all trials included – whatever their appraisal?) **T** 'totalled' (were the trials combined?) **H** 'heterogeneity' (was this considered and adjusted for?)
Apply… …the evidence to the patient (or clinical setting which prompted the question)	This step ideally uses a process called shared decision making (SDM). SDM incorporates the evidence with communication skills useful to patients and their clinicians when making difficult clinical decisions. It has these steps: • Options: list them (not forgetting to include 'do nothing') • Benefits of each option • Harms of each option • Preferences: what's the patient's preferences for each? Include postponing the decision until later. • Shared Decision: come to an agreed choice

Evidence Informed Practice: handbook of technical terms

[more can be found here: http://getitglossary.org/]

Results are often presented as *risks*, *odds*, *hazards*

term	
risk	**the chance of an event divided by the total.** examples: 1 Out of 1000 women delivering babies, 498 males were born. The risk of having a boy is therefore 498/1000 = ~0.5, or ~"50%" 2 Out of 67 children, 11 in one arm of the trial had pain at Day 3 in a RCT. The risk of having pain is therefore 11/67 = 0.16 3 Out 100,000 in the population, 50 will develop melanoma annually. The risk of developing melanoma is 50/100,000 = 0.0005
odds	**the chance of an event divided by the chance of *not* having the event.** examples: 4 Out of 1000 women delivering, 498 males were born. The odds of having a boy is therefore 498/502 = ~1, or "1:1" 5 Out of 67 children, 11 in one arm of the trial had pain at Day 3 in a RCT. The odds of having pain is therefore 11/(67-11) = 11/56 = 0.196 6 Out 100,000 in the population, 50 will develop melanoma annually. The odds of developing melanoma is 50/99,950 = 0.0005 **Notes** Clinicians prefer risks. But epidemiologists (and horserace gamblers) often generate results as odds, for technical reasons. Risks and odds are nearly the same if the denominator is large (see example 6 above).
hazard	the same as a risk (see above), of a bad outcome (death, pain, etc)

Results of different outcomes can be presented as *ratios*

term	the ratio of…	example
risk ratio (RR) ("relative risk")	…two risks	For example, in an RCT, the risk of having pain at Day 3 *with* antibiotics was 11/67 (or 0.16), and the risk *without* antibiotics (ie placebo) was 10/54 (or 0.19). The risk ratio is 0.16/0.19, = 0.89 This means that a child with acute otitis media who took antibiotics has 84% of the chance of having pain than a child taking placebo.
odds ratio	…two odds	
hazard ratio	…two hazards	

Relative risk reduction, (RRR)

This is the proportion of risk (say, of a bad event) that is reduced by some factor (say, a treatment).

> Example [Applemann 1991 http://www.bmj.com/content/303/6815/1450.short full paper free]
> In an RCT, the risk of having pain at Day 3 *with* antibiotics was 11/67 (or 0.16), and the risk *without* antibiotics (ie placebo) was 10/54 (or 0.19).
> The relative risk reduction (RRR) is the relative difference of these risks,
> or 0.19-0.16/0.16 = 0.03.

If you like formulae,

$$\text{RRR} = \frac{\text{(event rate in the intervention group) minus (event rate in the control group)}}{\text{(event rate in the control group)}}$$

Results presented as absolute risk reduction (ARR)

Thinking about results in absolute differences is necessary to account for the prevalence of the outcome.

> Doubling the risk may sound serious,
> but if it is an excessively rare event, it may not be important.
>
> Reducing the risk by only 5% may sound trivial,
> but if it is a very common event, it may be important.

This makes it hard to make a judgement about *relative* differences without prevalence information. Let's work out how to do that.

The ARR is the *absolute* reduction in risk (meaning the 'difference in numbers')

> Example [Applemann 1991]
> In an RCT, the risk of having pain at Day 3 *with* antibiotics was 11/67 (or 0.16), and the risk *without* antibiotics (ie placebo) was 10/54 (or 0.19).
> The absolute risk reduction (RRR) is the numerical difference of these risks, or 0.19-0.16 = 0.03, or 3% less chance of having pain at Day 3.

This is easier to understand. (Do you think it is an important enough difference? There are differences of opinion on this – there's no right answer!)

If you like formulae,

> ARR = (event rate in the intervention group) minus (event rate in the control group)

Notice this has to be estimated from the ratios, (see above).

Number Needed to Treat (NNT) also known as Number Needed to Treat for Benefit (NNTB), or for Harm (NNTH)

NNT is a version of the ARR which is easier for clinicians (but not patients)

It is 'the number of patients who must be treated for one to benefit (or be harmed)'
If you like formulae,

NNT = 1/ARR

Example [Applemann 1991 http://www.bmj.com/content/303/6815/1450.short]
The absolute risk reduction (RRR) is 0.03
Therefore, the NNT is 1/0.03 = 33
This means a GP has to treat 33 children with AOM for one to be free of pain at Day 3.
The other 32 children gain no benefit for this outcome.

Many clinicians find this even easier to understand. (Do you think it is an important enough difference? Again, there are differences of opinion on this – there's no right answer!
But first we have to make sure the difference was not to chance)

The play of chance

Differences between groups being compared may be the consequence of *chance* rather than a biological process.

For example, could the difference between the two proportions of children pain-free at Day 3, [11/67 (or 0.16) in the antibiotic group, and 10/54 (or 0.19) in the control group (a RR of 0.84)] be attributed to chance? [Appelman 1991]

There are two ways of investigating this.

P-value [P stand for 'probability']

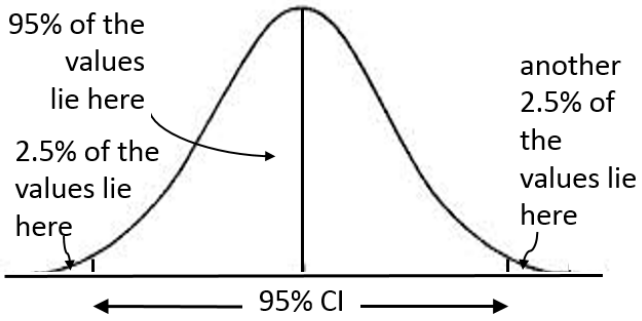
We estimate whether the difference would have occurred by chance.
If that chance is less than an arbitrary 5%, (or 0.05), then we declare that 'this difference is unlikely to be attributable to chance', and that the result is statistically significant.

In Appelman 1991, the p-value for the difference above was 0.381 (from Table 3 of their paper). This is above the arbitrary 0.05. This means that this values between control and intervention are not statistically significantly different from each other.

Confidence Intervals [usually abbreviated to '95%CIs']

Another way of approaching this entirely is to estimate the uncertainty (variance) around the values. This can be calculated statistically.

Figure: the range of values in a 95% confidence interval



For Appelman 1991, the 95% Confidence Intervals (the 95% means set at 'the range of values containing 95% of the value, if the sample was repeated endlessly').

11/67 antibiotics group and (ie) was 10/54 placebo group, with a risk ratio of 0.89 (95%CI 0.41, 1.93)

Since this value includes the null (which is always 1 for a ratio), then, again, we say the values between control and intervention are not statistically significantly different from each other.

Jenny Doust
BMBS, PhD, FRACGP, is Associate Professor and PHCRED facilitator, Discipline of General Practice, University of Queensland. j.doust@uq.edu.au

Chris B Del Mar
FRACGP, MD, is Dean, Faculty of Health Sciences and Medicine, Bond University, Queensland.

Brett David Montgomery
MBBS, DCH, FRACGP, is a general practitioner, Hospital General Practice, Fremantle, Western Australia.

Clare Heal
MBChB, FRACGP, DRACOG, is Senior Lecturer, Department of General Practice and Rural Medicine, James Cook University, Townsville, Queensland.

EBM journal clubs in general practice

The context of general practice makes the translation of evidence into clinical practice difficult. General practitioners interested in implementing evidence met at The Royal Australian College of General Practitioners' 2006 Annual Scientific Convention. We discussed evidence based medicine (EBM) journal clubs as a solution to this problem, including keys to success and barriers to overcome. The aim of this article is to provide suggestions for those wishing to set up their own EBM journal club. This will be supported as a Category 1 CPD activity by The Royal Australian College of General Practitioners in the 2008–2010 triennium.

Evidence based medicine (EBM) changes the way medicine is practised by integrating individual clinical expertise with the best available external clinical evidence and patient values.[1] In hospitals, clinicians can incorporate evidence into clinical decision making during ward rounds and team consultations.[2] In general practice however, this can be more challenging; consultations are shorter (often with patients backed up waiting to be seen), clinical decision making is more solitary, and there is less access to library services. Searching for and using evidence during the consultation process is only rarely feasible. One solution is the establishment of EBM journal clubs.[3]

Why are EBM journal clubs better?

In a traditional journal club, participants choose a recent journal article to discuss. In an EBM journal club, a topic is chosen because of a question from clinical practice. This fundamental difference leads to information being accessed in a 'pull' rather than a 'push' fashion,[4] thereby increasing the relevance of the information. Two other key differences are that the group critically appraises the research to identify potential biases and that the group considers how the results might be applied in their own practice.

The learning process is active. This process has educational advantages over other methods, such as didactic lectures, and also has the advantage that participants can detect and disregard irrelevant or poor quality information.

Elements for a successful journal club

Elements for running a successful EBM journal club include:

- Critical appraisal skills
  - courses, registrars or medical students are likely to have these skills and will benefit from practising them
  - go through examples previously worked up by another group
- Time
  - remember: you should be able to replace some conventional

54 Reprinted from AUSTRALIAN FAMILY PHYSICIAN Vol. 37, No. 1/2, January/February 2008

Rachel Bidgood
MBBS, BSc, DRCOG, MRCGP, FRACGP, is a general practitioner, Mackay, Queensland.

David Jeacocke
MBBS, MMedSci, PhD, FRACGP, is Senior Medical Advisor, Medicare Australia.

Gary Bourke
MBBS, FRACGP, is Senior Lecturer, Rural General Practice, University of Melbourne, Victoria.

Geoffrey Spurling
MBBS, DTM&H, FRACGP, is Senior Lecturer, Discipline for General Practice, University of Queensland.

CPD with EBM journal clubs
- missing the odd session is manageable if there is good communication

- Skills in asking questions
- first try some worked up examples
- Ability to retrieve the literature
- PubMed and the Cochrane Library are free
- The Royal Australian College of General Practitioners (RACGP) John Murtagh Library for members and RACGP Fellows provides an excellent service by fax or mail of full text articles from many journals within days of request

- A venue
- practices, division offices, local hospitals and local university facilities are all possibilities.

Some critical appraisal skills are necessary, although this is less of a barrier with current undergraduate and postgraduate training programs. Being part of a journal club gives participants a valuable opportunity to experience and practise these skills.

The lack of recognition of these activities for continuing professional development (CPD) points has been a considerable barrier for some participants in the past. The RACGP QA&CPD Committee recently agreed to grant Category 1 CPD points for EBM journal club activities for the 2008–2010 triennium (see www.racgp.org.au).

Trying to find a time when a group can meet is often difficult. However, accepting that not everyone will always be able to attend is important and does not devalue the process. Circulating a summary of the discussion for those who cannot attend is useful. Some divisions of general practice have facilitated organising a venue and food for meetings.

People will only be interested in an EBM journal club if the information flows through to changes in clinical practice. Many questions in general practice have not been well studied, and it can be discouraging if you ask too many questions that have no evidence to answer them. On the other hand, one group used a question from their journal club that had no high quality evidence to initiate their own research and perform their own randomised controlled trial.[5] Two other examples of questions are shown in *Table 1*.

Factors that engender success in EBM journal clubs include:

- At least 2–3 GPs committed to meeting regularly. Divisions of general practice may be able to facilitate 'hub' meetings
- Volunteers for minor tasks such as keeping a log of the clinical questions and keeping summaries of the meetings
- Skills in critical appraisal
- Group facilitation skills. Like any meeting, sessions can run off track and lose relevance without good facilitation
- A method for selecting the clinical questions. It is important that no-one dominates the selection and that the questions are of interest to all member of the group

Table 1. Examples of questions and answers arising from EBM journal clubs, and how they may change practice

Does salmeterol cause harm in asthma?

A trial report showed that salmeterol improves symptoms but causes approximately one death in 600 patients each 6 month period. These potentially harmful effects may be offset by the use of concomitant steroid use. Based on this evidence, we ran a search of all patients who were currently prescribed a long acting beta agonist. We recalled all such patients and commenced them on a combination product[6]

Is atenolol effective for hypertension?

A systematic review of placebo controlled trials containing 6825 patients followed up for a mean of 4.6 years showed no effect on all cause mortality or myocardial infarction.[7] This finding generated other questions such as is metoprolol effective for hypertension? We are still trying to decide what to do with patients currently taking atenolol for hypertension

Table 2. A possible structure for EBM journal clubs

- Discuss possible questions; GPs might find it easier to remember questions if they keep a log of questions on their desk
- Search the literature. We find it easier to do this before the meeting so that relevant papers can be copied and distributed
- The person who initiated the question outlines the reason why the question is being asked
- Discuss the results of the search – who are the participants, what is the intervention, what is it being compared to and what outcomes are being addressed?
- Critically appraise the paper. Don't jump to the discussion of the paper. There are several checklists available (www.cebm.net/critical_appraisal.asp)
- Discuss the results of the paper. Do the results answer the question being asked? Are their potential harms or unexpected consequences? How might the results change your own clinical practice? Should the practice be implementing a change as a whole?

Reprinted from AUSTRALIAN FAMILY PHYSICIAN Vol. 37, No. 1/2, January/February 2008 55

- An internet connection is helpful. Often the initial question generates further questions and immediate access to the internet can help provide answers
- Several journal clubs report that providing food with the meeting encourages attendance.

A possible structure for EMB journal clubs is listed in *Table 2*.

What next?

We are keen to link with other GPs running similar activities or to hear from others who have had less successful attempts so that we can try and identify barriers to a successful journal club and ways such barriers can be overcome.

We are also keen to encourage and support others who would like to start up their own EBM journal club, and welcome anyone who would like to do so, to contact the authors.

Conflict of interest: none declared.

References

1. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996;312:71–2.
2. Sackett DL, Straus SE. Finding and applying evidence during clinical rounds: the 'evidence cart'. JAMA 1998;280:1336–8.
3. Phillips RS, Glasziou P. What makes evidence based journal clubs succeed? ACP J Club 2004;140:A11–2.
4. Glasziou PP, Del Mar CB, Salisbury J. Evidence based medicine workbook. London: BMJ Books, 2002.
5. Heal C, Buettner P, Raasch B, Browning S, et al. Can sutures get wet? Prospective randomised controlled trial of wound management in general practice. BMJ 2006;332:1053–6.
6. Spurling G, Doust J. Evidence based answers: is salmeterol safe in asthma? Aust Fam Physician 2006;35:625–6.
7. Lindholm LH, Carlberg B, Samuelsson O. Should beta blockers remain first choice in the treatment of primary hypertension? A meta-analysis. Lancet 2005;366:1545–53.

AFP CORRESPONDENCE afp@racgp.org.au

RACGP

Royal Australian College *of* General Practitioners

Healthy Profession.
Healthy Australia.