Critical Appraisal of Intervention Studies

Developed by Donna Ciliska, RN, PhD
Professor, McMaster University, and
Scientific Director, National Collaborating Centre for Methods and Tools

How do I use this learning module?

Estimated total time: about 5 hours

Objective: To be able to decide if an intervention study is of sufficient quality that it can be applied to your own situation. In order to do this, you will understand and be able to apply the criteria for critical appraisal of an intervention study.

Process: This module is built on a scenario that will allow you to understand and apply each criterion for critical appraisal. After having read the scenario, you will be able to follow sequentially through the questions that allow you to critique and make a decision about the use of the study. (Time estimates are in brackets.)

Links: Each time you see the word scenario, it is linked to the actual scenario and will take you there if you click on it. Similarly, the key terms are linked to a definition in a glossary.

Overview

1. Scenario (0.25hours)
2. What is critical appraisal? Why bother doing it? (0.5 hours)
3. Critical appraisal tools and criteria for intervention/prevention studies (0.5 hours)
4. Application of critical appraisal criteria
    a) Read article and complete answer sheet (1 hour)
    b) Are the results valid? (1 hour)
    c) What are the results? (1 hour)
    d) How can I apply the results? (0.5 hours)
    e) Resolution of scenario (0.25 hours)
5. Optional review practice
6. Useful references
7. Glossary

1. Scenario

(0.5 hours)

Last winter, you had the flu shot, but it felt like you almost always had a cold. You recovered and were free of symptoms for a week or two then the symptoms returned. You are aware of advertising of ginseng products for prevention and treatment of the common cold. In addition, your friends and family have asked if you think the ginseng products work. You decide to look in the health-related literature for an answer.

1. You clearly frame the PICO question:

P atient / **P** opulation:	healthy adults
I ntervention	oral ginseng preparation
C omparison	no ginseng
O utcome	number/duration of common cold in a season

2. You search on PubMed (http://www.ncbi.nlm.nih.gov/PubMed/). On the left side, you see "Clinical Queries"; when you click on that, a dialogue box appears that allows you to search a number of different types of studies. You choose "therapy", with "narrow and specific". (For more information on searching, please see the module: Evidence Informed Decision-Making.) When you type "ginseng and colds" one study appears (which happens to have free full-text). You read the abstract and decide to access the full article.

Your read:

Predy, G.N., Goel, V., Lovlin, R., Donner, A., Stitt, L., Basu, T.K. (2005). Efficacy of an extract of North American ginseng containing poly-furanosyl-pyranosyl-saccharides for preventing upper respiratory tract infections: a randomized controlled trial. *CMAJ, 173* (9), 1043-1048. http://www.cmaj.ca/cgi/content/full/173/9/1043

Questions:

Will you take a ginseng preparation this winter to prevent or treat common colds? What will you tell your family and friends about the effectiveness of ginseng?

2. What is critical appraisal? Why bother doing it?

(0.5 hours)

Evidence-informed decision-making is about applying the best *available* evidence to answer a specific question. You may be lucky and find a pre-appraised article where someone else had done the critical appraisal for you, such as the case with a synopsis from an evidence-based journal. If you cannot find that, you will have to assess for yourself, the methods of the study. This process is known as "critical appraisal". What you are judging is the quality of the study methods and if the study can be applicable to your own situation, whether your situation involves a population, an individual patient, a policy or yourself. You are trying to answer the question:
Were the methods used in this study good enough that I can be confident in the findings?
It is Step 3 in evidence-informed decision-making, where the process is:

1. *Ask*. How do I frame the question?
2. *Acquire*. How can I find the best evidence in 5 minutes or less?
3. *Appraise*. How can I decide if the particular study is good enough to apply?
4. *Integrate*. How do I decide which of multiple studies to use?
5. *Adapt*. How do I use the information from #5 in decision-making / a policy brief?
6. *Apply*. How do I develop the implementation the plan?
7. *Evaluate* How do I know if the plan worked?

(Note: for an overview of all steps above, see Module 1 in this series on Evidence-Informed Decision Making url)

How do you handle any one of the multiple situations that have arisen where one study found a drug or therapy to be helpful while another study found it to be ineffective or even harmful? Does Vitamin E prevent heart disease or increase cardiovascular risk? Does vitamin C prevent colds or just result in expensive urine? Do maggots contribute to venous ulcer debridement and subsequent wound healing or just increase anxiety and the "yuck" factor for patients and healthcare workers?

How do you know which study results to believe? The best we can do is an appraisal of the methods of each study in order to decide which studies the best methods to control for possible confounders or bias. Those studies would then constitute for you, the best available evidence, and you would base your practice or policy decisions using that evidence as part of the picture. We have grown to be wise consumers of advertising, critically analyzing claims that are made. We need to have the same ability to critically analyze results of health care studies.

There are key quality criteria for any types of studies that you find. In your search, you would always go to systematic reviews first, as they constitute a body of research on a topic that has already been critically appraised. (See Module 1 on Evidence-Informed Decision Making. url) However, in order to understand the critical appraisal of the systematic review (Module 3 in this series), you need to be able to understand the critical appraisal of single studies. This learning module will detail the critical appraisal process for single studies of intervention (therapy) or prevention.

⚠ A word of caution! Newbies to critical appraisal sometimes throw out relatively well-done studies from consideration because they are not perfect. There are **no** perfect studies. As you become more familiar with the process, you will see that there are some criteria that relate to larger concerns, and would therefore be 'fatal flaws' for which you would reject the study. However, some other criteria are not so critical and, even if the study has not fulfilled that particular criterion, you would still consider implementing the intervention.

3. Critical appraisal tools and criteria for intervention and prevention studies

(0.5 hours)

Most of the available critical appraisal tools for quantitative research are based on key criteria developed by the Evidence-Based Medicine Working Group. A series was first published as “Readers' Guides” in the *Canadian Medical Association Journal*, beginning in 1981, later revised and extended in *JAMA* as the "Users’ Guides", between 1993 and 2000 and, finally, collected in a book (Guyatt & Rennie, 2002). The Evidence-Based Medicine Working Group has produced some twenty-five tools for many different types of clinical questions and study designs (for example: treatment, systematic review, causation, diagnosis, economic analysis, clinical prediction guides, practice guidelines, health services research).

This learnng module is about critical appraisal of intervention (also called therapy or treatment) and prevention studies. Does low molecular-weight heparin prevent deep vein thrombosis? Does tight glycemic control in patients with diabetes prevent cardiovascular complications? Does aromatherapy increase relaxation? Does hormone replacement therapy reduce ‘hot flashes’ associated with menopause? Can an intensive educational program reduce rates of teen pregnancies? All these questions are considered intervention or prevention research questions.

The basic criteria for critical appraisal of intervention studies are:

Box 1. CRITICAL APPRAISAL FOR INTERVENTION AND PREVENTION STUDIES

Are the results valid?
- Were participants randomized?
- Was randomization concealed?
- Were participants analyzed in the groups to which they were randomized?
- Were participants in each group similar with regard to known prognostic variables?
- Were participants aware of group allocation?
- Were clinicians aware of group allocation?
- Were outcome assessors aware of group allocation?
- Was follow-up complete?

What are the results?
- How large was the treatment effect?
- How precise was the estimate of the treatment effect?

How can I apply the results?
- Were study participants similar to my own situation?
- Were all clinically-important outcomes (harms and benefits) considered?

Based on Guyatt & Rennie, 2002

You will note that the critical appraisal questions are asking about randomization and imply a randomized controlled trial (RCT). Where possible, an RCT is the most appropriate design to answer intervention questions, as random assignment allows for known and unknown determinants of outcome to be evenly distributed among the groups. Consequently, you can be more confident that, if there are differences in outcome, the differences are more likely to be due to the actual intervention as opposed to any underlying differences in the groups. In other words, randomized trials have the greatest ability to control for confounders or bias. However, it is not always possible to answer intervention or prevention studies with randomized trials. For example, it is not ethically possible (nor feasible) to randomize women to breastfeed or bottle-feed their newborns in order to assess the impact of exclusive breast milk on the prevention of asthma. A non-randomized two group before-after design would be used. In the case of non-randomized trials, the critical appraisal criteria are still useful, but you must realize that there is a greater possibility of underlying differences in the groups attributing to any differences in the results. (For more on that, see the module: Evidence-Informed Decision Making.) See Section 5 “Optional review practice” for an example of a question for which there are no RCTs, where the best evidence is a two group before/after design (non-randomized study).

Section #4 will use the scenario to illustrate and apply the criteria.

RECOMMENDED RESOURCE

For ease of access and understanding, you should consider using the critical appraisal criteria from *The Critical Appraisal Skills Program* (CASP) of the Public Health Resources Unit in the U.K. They have produced a series of tools based on the "Users' Guides" (Guyatt & Rennie, 2002). The advantages of their tools are that explanations of the criteria are built into the tool and they are freely accessible on-line for personal use. Although developed by the Public Health Unit, they are not specific to public health alone (http://www.phru.nhs.uk/Pages/PHD/resources.htm).

Reference:

Guyatt, G. & Rennie, D. (Eds) (2002). *Users' Guides to the Medical Literature: A manual for Evidence-Based Clinical Practice*. American Medical Association.

4. Application of critical appraisal criteria

a) Read article and complete answer sheet
(1 hour)

Go back to the **scenario** in section #1, about the effectiveness of ginseng for prevention or treatment of colds.

You find this article:

Predy, G.N., Goel, V., Lovlin, R., Donner, A., Stitt, L., Basu, T.K. (2005). Efficacy of an extract of North American ginseng containing poly-furanosyl-pyranosyl-saccharides for preventing upper respiratory tract infections: a randomized controlled trial. *CMAJ, 173* (9), 1043-1048. http://www.cmaj.ca/cgi/content/full/173/9/1043

Here is where you get to try your answers! You will use this article and answer each question sequentially on the *Critical Review Form For Intervention*.

NOTE: It will be helpful for working through the section if you print or view the pdf version of the article so the page references will be consistent with the discussion that follows.

1. Please read the entire article.
2. Answer the critical appraisal questions in this guide.

CRITICAL REVIEW FORM FOR INTERVENTION

Citation:

Guide	Comments
I) Are the Results Valid?	
Were participants randomized?	
Was randomization concealed?	
Were participants analyzed in the groups to which they were randomised?	
Were participants in treatment and control group similar with respect to known prognostic factors?	
Were participants aware of group allocation?	
Were clinicians aware of group allocation?	
Were outcome assessors aware of group allocation?	
Was follow-up complete?	
II. What are the Results?	
How large was the treatment effect?	
How precise was the treatment effect?	
III. How can I apply the results?	
Were the study participants similar to my own situation?	
Were all clinically important outcomes (harms and benefits) considered?	

4. b) Are the results valid?
(1 hour)

i) Were participants randomized?

The importance of using randomization is to ensure that groups are similar in all factors, other than the outcome, that might affect the outcomes (e.g., age, sex or socioeconomic status). This helps to reduce possible bias. A randomized trial is considered the highest level of evidence for a single study, with the caution that not all questions can be subjected to an RCT – either ethically or practically. Of course, a systematic review of a number of trials is a higher level of evidence. (See Module 1, Evidence Informed Decision-Making, in this series.)

There are some alternate ways to do the randomization and it is important to make sure that the study does true randomization rather than a 'quasi-randomization' such as days of the week or month of birth. True randomization is done with a table of random numbers or a computerized random number generator.

Q: *Were participants randomized?*

A: YES – On page 1044, 1st line under methods: "it was a randomized …trial" Pg 1045, 1st full para, authors tell us they used a computerized randomization scheme.

ii) Was randomization concealed?

Why should this matter? Other studies have told us that, if the person who recruits participants to the study knows the allocation sequence (what group assignment is coming up next), they may consciously or unconsciously make a choice (that is, substitute envelopes) if they think this person would be better served by being in the intervention. Strategies to ensure that randomization is concealed include use of sequentially numbered opaque envelopes or a call-in centre to give the allocation of the current participant recruit.

Q. *Was randomization concealed?*

A. YES
Pg 1045, 1st full para, authors tell us they used "numbered, opaque, sealed envelopes". That means the sequence could not be altered, and the person who was recruiting could not see through the envelope to determine the group allocation coming up next.

iii) Were participants analyzed in the groups to which they were randomized?

There are research horror stories of participants who dropped out of an intervention group and the researchers 'moved' them to the control group; or conversely, control group participants who somehow got the intervention outside of the study, who were then switched to the active intervention group. You can guess that excluding dropouts from the analysis of a smoking cessation or a weight loss study (where dropouts can be as high as 50-60%) might make the intervention look more effective than reality.

This criterion is ensuring that the participants will be kept in the analysis of their original group assignment regardless of whether they discontinue the treatment. Researchers call this "intention-to-treat" analysis. How do they include dropouts? They

do this by substituting either the baseline measurements or the 'last observation carried forward' of people who have dropped out, for the final outcome measurement.

Q. *Were participants analyzed in the groups to which they were randomized?*

A YES.
On page 1044, there is a flow diagram (Figure 1) that tells you that 149 people in the placebo group and 130 in the ginseng group started the intervention and all of those people were included in the analysis (last boxes in flow diagram Figure 1). As well, the researchers tell us that an "intention-to-treat analysis" was performed, using the last available observation carried forward in the analysis (pg 1045, 2nd column, 4th paragraph).

iv) *Were participants in each group similar with regard to known prognostic variables?*

Before the intervention begins, we want to know if there are differences between the groups that could potentially explain differences seen in outcomes at the end. Randomization should ensure that characteristics are relatively evenly distributed. Researchers check the adequacy of randomization by presenting the entry point characteristics thought to be possibly related to outcome. Some imbalances arise from a too-small sample size whereas others occur by chance. If the sample size is adequate, any remaining differences in the group can be accounted for in the analysis. In addition to presenting the 'unadjusted' results, researchers will often provide 'adjusted' results, where the adjustment takes into account baseline differences.

Q. *Were participants in each group similar with regard to known prognostic variables?*

A. YES.

Table 1 displays the age, sex, smoking status, number of colds per year and number of subjects with three or more colds per year. Also the researchers tell us there were no significant differences at baseline (pg 1046, 2nd column, 1st paragraph). Also, the *Results* section tells you there were no statistically significant differences in baseline characteristics (text on page 1046, 2nd column, 1st full paragraph) Are there any other variables that they should have considered? They excluded those with chronic or acute illness and those receiving medications, so it appears that this is a normal, healthy population.

v) Were participants aware of group allocation?

Blinding (or *masking*) is a term used to describe whether or not a variety of people know whether participants are in the active intervention group or the control. Research reports sometimes use the term 'single', 'double' or 'triple' blinded, but it is now considered important to specify *who* was blinded.

This criterion is related to *participants* being blinded. If participants know which group they are in, they may consciously or unconsciously have a greater awareness of favorable or unfavorable aspects of the intervention. In drug trials, placebos are usually difficult to discern from active treatment in that they look the same. In educational or

psychotherapies, it is much more difficult for participants to remain blinded; they know if they have been exercising or watching videos!

Q. *Were participants aware of group allocation?*

A. NO. On pg 1044, the study is described as a "double-blind". Participants were to take the unmarked preparations as instructed. Preparations (encapsulations) were identical (pg 1045, first column, end of first incomplete paragraph). In addition, after completion, participants were asked whether they thought they had taken the ginseng or the placebo (pg 1045, 2nd column, 1st paragraph); 69.8% of those taking ginseng and 77.3% of those taking the placebo thought they had been given the ginseng preparation.

vi) Were clinicians aware of group allocation?

This criterion is assessing whether the clinicians involved with participants knew which group their patients are in. They may, unconsciously, alter their treatment plan, provide additional care or heighten their vigilance for good or bad outcomes if they know the group allocation of their patients. If the clinician is the one delivering the intervention, it is impossible to blind that person to the intervention, but they may be kept blind to the research question, or at least to the comparison.

Q. *Were clinicians aware of group allocation*?**A**. NO. The study physician was blinded. For symptoms that suggested secondary complications, the study physician recommended family physician follow-up, but there was no mechanism for the family physician to know whether or not the participant was receiving ginseng or placebo (pg 1045, 1st column, last line).

vii) *Were outcome assessors aware of group allocation?*

As in v) above, this criterion is assessing whether the people who were conducting measurements of outcome knew group allocation of people they were assessing. Distortion of measurement may be more likely if an individual is required to do the measurement (e.g., blood pressure) while knowing the group allocation and having a belief about the likely effectiveness of the intervention. This potential bias is removed when tests are done by laboratory or computer equipment, such as blood samples for glycated hemoglobin.

Q. *Were outcome assessors aware of group allocation?*

A. Yes. In this case, the participants did most of the outcome assessment, using logs to keep track of symptoms (pg 1045, 1st column, last paragraph). Compliance with taking the medication was also checked by the weight of returned bottles. We see in this article that even the data analysts were blinded (pg 1045, 2nd column, 2nd full paragraph).

viii) Was follow-up complete?

There are two components to this question and the answers are dependent on the question of interest, as opposed to some global standard. First of all, were patients followed long enough to be able to see a result of treatment? For example, studying the

effect of taking vitamin C during adolescence on the rate of colon cancer would require upwards of 25 years of follow-up to determine effectiveness, as colon cancer is typically diagnosed in mid- to late adulthood. In contrast, testing the effectiveness of aloe vera extract on sunburn-related skin pain may take from a few hours to a few days to assess outcomes.

The second question related to follow-up is: How many participants dropped out of (or conversely, were retained in) the study before reaching the endpoint? What happened to the 'lost' participants and how might their outcomes be different than those who stayed in the study? For example, you might assume that dropouts from a weight loss or smoking cessation intervention are more likely to be 'treatment failures'.

Some people consider a 'gold standard' of less than 20% dropout is required for the study to be considered strong. Once again, this is dependent on the nature of the problem and participants of the study. For example, a dropout rate of 35% would be outstanding retention for a study involving street youth.

Q: *Was follow-up complete?*

A: YES

This study took place in Edmonton, Alberta, from the onset of the influenza season (November) for 4 months. This is adequate time to assess if the intervention (ginseng) is going to have an impact on the outcome (upper respiratory tract infections)

The flow diagram, Figure 1, pg 1044, gives you the information of flow of participants and reasons for dropout through this study period. In this study, there was a retention rate of 88% (149/170) in the placebo group and 85% (130/153) in the intervention group. This could also be expressed as 12% and 15% dropout rates, respectively. The authors indicated the reason for non-participation and discontinuation where it was known. This is a very acceptable rate of follow-up.

The authors go a step further to look at the potential differences in baseline data between those who did and did not start the intervention (Table 1, pg 1045). Further, on pg 1046, 1st full paragraph, the authors indicate there were no significant differences in baseline characteristics.

4. c) What are the results?
(1 hour)

In answering the questions in 4 b), you get a sense of the study methods and if the results are likely to be valid. If the answer is affirmative, you would go on to look at the actual results and to identify if the results of the study are important.

i) How large was the treatment effect?

The benefits (and harms) of any intervention may be measured by multiple outcomes. These outcomes may be dichotomous (either/or) outcomes such as dead versus alive, infection versus no infection, healed/not healed; or continuous such as # of sneezes per day, length of stay, respiratory rate, fasting glucose).

In reporting results of studies using dichotomous outcomes, comparisons can be measured by rates (49% healed ulcers in the intervention group versus 25% in the control

group). These rates may then be expressed in other ways such as absolute risk difference, relative benefit increase (or the converse which is relative risk reduction) or number needed to treat (or the number needed to harm). Results of studies using continuous outcomes (number of colds per season, sperm count) report differences in the means

You should consider if the difference between groups was statistically significant. The true effect of a treatment cannot be known; what we know is an *estimate* of effect. Confidence intervals are a statistical device to let us know the level of uncertainty around an estimate. The 95% confidence interval (CI) represents the range within which we are 95% certain that the true value of the effect lies. If the range for the 95% CI of an odds ratio or relative risk includes 1, there is no statistically significant difference between the treatment groups. Similarly, if the 95% CI for mean difference includes 0, there is no statistically significant difference.

Statistical significance can also be conveyed via the p value. By convention, we agree that if the p value is below 0.05, it is statistically significant. That is, we are willing to accept the probability is less than 1 in 20 that the result is occurring by chance alone.

Q. *How large was the treatment effect?*

A. The primary outcome was number of colds reported per subject (and verified with the Jackson criteria). The mean number of colds was 0.68 in the ginseng group and 0.93 in the placebo group (mean difference was 0.25 colds per person; 95% CI 0.04-0.45) (pg 1045, Table 2, and 2nd column, 3rd full paragraph). The CIs do not include 0, so you can tell that it is a statistically significant finding. In the text, the authors also give us the corresponding p value (p =0.017), which is less than 0.05 and is statistically significant.

The authors also examined the number of people who had "1 cold" or "2 or more colds" over the season. Let's look at the latter; on page 1045, Table 2, 10% of the people in the ginseng group reported 2 or more colds versus 22.8% in the placebo group. This is an absolute risk reduction of 12.8% (95% CI, 4.3 to 21.3). You can tell this is statistically significant because the CI does not include 1.

ii) *How precise was the estimate of the treatment effect?*

Precision can only be known by the Confidence Interval (see *4 c) i* above). If the CI is wide, the estimate of true effect lacks precision and we are unsure about the treatment effect. If the confidence interval is narrow, precision is high, and we can be more confident in the results. Larger sample sizes produce more precise results, so you must be wary of (i.e., not confident in) small sample sizes and large confidence intervals.

Finally, you need to decide if the statistically significant finding is clinically (or personally) meaningful. For example, a statistically significant weight loss of 4 kg is not clinically or personally meaningful for morbidly obese patients. Also, you can use the smallest possible effect size (the lower end of the confidence interval) to help you determine whether, if the effect were this small, it would still be worth doing.

Q. *How precise was the estimate of the treatment effect?*

A. As noted in the previous section, the result for number of colds was a mean difference of 0.25 colds per person (95% CI 0.04-0.45). The width of the confidence interval tells

you that, at the extremes, it may be an average of as little as 0.04 colds less per year or as much as approximately half a cold per year. This seems to have moderate precision. Even though statistically significant, is it clinically meaningful? If the difference is as great as a mean reduction of half a cold per year (or one cold every two years), that could have a clinical benefit in terms of quality of life, reduced days absent from work or school, or productivity.

When you consider the recurrence of colds, the results of a 12.8% reduction (95% CI, 4.3 to 21.3) again seems to be of moderate precision. Even at the lowest end, a 4.3% reduction in recurrent colds may be worth it (clinically meaningful).

Reference

Lipman, M.M. (2008). No safety in numbers. *Consumer Reports on Health*, June, 11. Presents a discussion geared to non-statisticians of 'absolute risk reduction', 'number needed to treat', and 'number needed to harm'.

4 d) How can I apply the results?
(0.5 hours)

i) Were study participants similar to my own situation?

You need to judge the generalizability from the study participants to your own patients, clients or situation. By agreement of the large medical journals, study participants are usually described in Table 1 of the study report. You need to consider if there were differences in age, gender mix, socioeconomic status, illness acuity or co-morbidities, for example, which would mean the outcomes would likely be different in your situation. It is rare that the participants will be exactly like your situation, so look instead for reasons why you should *not* apply the results of the study.

Consider, also, if the treatment is feasible in your situation. This includes comparing health care systems, estimated costs of treatment delivery, skills required to deliver the intervention, availability of special equipment and staff resources as well as likely acceptability to your patients.

Q. *Were study participants similar to my own situation?*

A. The participants in this study were otherwise healthy adults from the general population, with no chronic physical or mental health conditions and taking no medications. The mean age was 43, there slightly more females than males and there were slightly fewer smokers than the general public (page 1045, Table 1). They resided in Edmonton and surrounding areas, where you would expect more people would have colds due to the drying effect of the predominant weather and heating systems on mucous membranes. However, you can be sufficiently comfortable to use these results with the general population in Canada.

ii) Were all clinically important outcomes (harms and benefits) considered?

Researchers may use several different outcomes to test the effects of treatment. In addition, they should look for evidence of harm, although the sample size within a trial may not be large enough. However, it is important to know, for example, if blood lipids are improved with a study drug, yet there is a higher mortality rate in the intervention

group. Health care systems are also questioning expenses of such treatments and may call for an economic analysis such as cost-benefit.

Q. *Were all clinically important outcomes (harms and benefits) considered*

A. In addition to total number of colds and the rate of recurring colds, the researchers examined symptoms and side effects. There were statistically significantly lower cold symptom scores in the ginseng group compared to placebo (page 1046, Table 3). Also, adverse events were very similar with no statistically significant difference between groups (pg 1046, 2nd column, 4th full paragraph; and page 1047, Table 4).

4 d) Resolution of scenario

Back to the scenario, you were asked:

Will you take a ginseng preparation this winter to prevent or treat common colds? What will you tell your family and friends about the effectiveness of ginseng?

The article is quite strong, in terms of methods. You can be confident in the findings. The participants were truly randomized, remained blinded as to their group assignment, and follow-up was complete. While the average number of colds per person was not dramatically reduced (mean difference of 0.25 colds per person), there was also a reduction in recurrence of colds and cold symptoms. Participants are well adults in Canada, and there were no more side effects in the ginseng group than in the placebo group. All of that has convinced you to go to the pharmacy and to calculate the cost of taking 2 tablets per day for the season. You decide that if works out to less than $5 per week, it will be worth it to try your own experiment on yourself. You are going to tell your friends and relatives your conclusion as well.

ANSWERS: CRITICAL REVIEW FORM FOR INTERVENTION

Citation: Predy, G.N. et al. (2005). Efficacy of an extract of North American ginseng containing poly-furanosyl-pyranosyl-saccharides for preventing upper respiratory tract infections: a randomized controlled trial. *CMAJ, 173* (9): 1043-1048.

Guide	Comments
I) Are the Results Valid?	
Were participants randomized?	YES – on page 1044, "it was a randomized …trial" - pg 1045, authors tell us they used a computerized randomization scheme.
Was randomization concealed?	YES – pg 1045 "they used numbered opaque, sealed envelopes".
Were participants analyzed in the groups to which they were randomised?	YES – pg 1044, flow diagram - pg 1045 "intention to treat analysis was performed"
Were participants in treatment and control group similar with respect to known prognostic factors?	YES – Table 1 – no important differences in age, sex, smoking status, # of colds/subject or # of subjects with 3 or more colds/year
Were participants aware of group allocation?	NO – pg 1044, intervention and control preparations were identical
Were clinicians aware of group allocation?	NO – study physicians were blinded
Were outcome assessors aware of group allocation?	YES – participants kept logs, assessed their own outcomes
Was follow-up complete?	YES – 4 months over winter adequate to see if intervention will affect # of colds; -pg 1044, 15% dropout in intervention group and 12% in control group
II. What are the Results?	
How large was the treatment effect?	0.68 in the ginseng group and 0.93 in the placebo group (mean difference was 0.25 colds per person; (p =0.017) – this is statistically significant. 10% of the people in the ginseng group reported 2 or more colds versus 22.8% in the placebo group. This is an absolute risk reduction of 12.8% (95% CI, 4.3 to 21.3 (statistically significant because the CI does not include1)
How precise was the treatment effect?	0.68 in the ginseng group and 0.93 in the placebo group (mean difference was 0.25 colds per person; 95% CI 0.04-0.45; 10% of the people in the ginseng group reported 2 or

Guide	Comments
	more colds versus 22.8% in the placebo group. This is an absolute risk reduction of 12.8% (95% CI, 4.3 to 21.3 Both results are of medium precision. Reduction of 0.25 colds/person may not be clinically meaningful, but 12.8% risk reduction in having 2 or more colds is clinically meaningful.
III. How can I apply the results?	
Were the study participants similar to my own situation?	YES - Participants seem representative of the general healthy population of adults in Canada
Were all clinically-important outcomes (harms and benefits) considered?	YES - Researchers looked at symptoms and side effects: statistically significantly lower cold symptom scores in the ginseng group compared to placebo (pg 1046). Also, adverse events were very similar with no statistically significant difference between groups (pg 1046

5. Optional review practice

1. Scenario:
(0.5 hours)

You belong to a community multi-disciplinary group that is concerned about youth crime in your neighborhood. Public health, education, social services, police services and a variety of community groups are all represented. At one meeting, you brainstormed different possible solutions. One such solution was to offer primary school-based interventions with teacher training and, perhaps, parent training. You offered to search the literature to see if there were any evaluations of such interventions and to report back your findings at the next meeting.

You clearly frame the PICO question:

P opulation:	primary school children
I ntervention	curriculum enhancement, teacher training
C omparison	usual
O utcome	youth crime rates

You search on PubMed (http://www.ncbi.nlm.nih.gov/PubMed/). In the text box at the top, you type in "crime prevention" and "school curriculum". You get several hits, but one title looks to be particularly important, as it assessed outcomes in adulthood AND it has free full text on-line access.
You read:

Hawkins, J.D., Kosterman, R., Catalano, R.F., Hill, K.G., & Abbott, R.D. (2005). Promoting positive adult functioning through social development intervention in childhood: long-term effects from the Seattle Social Development Project. *Archives of Pediatric and Adolescent Medicine, 159*, 25-31.

Questions:
How will you summarize this article for your group?
Will you recommend a school curriculum intervention to reduce crime?

1. Read the entire article. If you want to download and/or print, please use the *pdf* format so that your page numbers will match the the answer sheet.
2. Answer the critical appraisal questions on the “Critical Review Form For Interventions”. For this exercise, only consider the full intervention versus the control. (There is another comparison with a ‘late’ intervention which you can disregard.)
3. Compare your answers with the completed answer sheet.

⚠ A word of caution! In REAL LIFE, you would conduct a thorough literature search. For this exercise, you are ‘pretending’ this is the ONLY study you found.

CRITICAL REVIEW FORM FOR INTERVENTIONS

Citation:

Guide	Comments
I) Are the Results Valid?	
Were participants randomized?	
Was randomization concealed?	
Were participants analyzed in the groups to which they were randomised?	
Were participants in treatment and control group similar with respect to known prognostic factors?	
Were participants aware of group allocation?	
Were clinicians aware of group allocation?	
Were outcome assessors aware of group allocation?	
Was follow-up complete?	
II. What are the Results?	
How large was the treatment effect?	
How precise was the treatment effect?	
III. How can I apply the results?	
Were the study participants similar to my own situation?	
Were all clinically important outcomes (harms and benefits) considered?	

ANSWERS: CRITICAL REVIEW FORM FOR INTERVENTIONS

Citation: Hawkins, J.D. et al (2005). Promoting positive adult functioning through social development intervention in childhood: long-term effects from the Seattle Social Development Project. *Archives of Pediatric and Adolescent Medicine, 159,* 25-31.

Guide	Comments
I) Are the Results Valid?	
Were participants randomized?	NO. In the abstract, it is stated that it is a non-randomized trial. On pg 25, it started out to be a randomized trial, but expanded during the study to include additional schools, which were "assigned non-randomly".
Was randomization concealed?	Not applicable, as it was not a randomized trial.
Were participants analyzed in the groups to which they were randomised?	YES – pg 28 "we conducted conservative intention-to-treat analysis". Participants were old enough to assess youth crime, your outcome of interest. Longer follow-up would be useful for assessing impact on adult crime rates.
Were participants in treatment and control group similar with respect to known prognostic factors?	YES Pg 26 – there is only a brief description of the sample at baseline, which indicates overall ethnicity, gender, and eligibility for free lunches. Authors do not compare across groups in this article. However, they cite an earlier report of this study (ref #19), where they reported that the groups did not differ on: - residential stability as measured by mean number of years living in Seattle by age 12 years and by the mean number of residences in which participants lived from age 5 to 14 years; - socioeconomic status, as measured by years of parental education or proportion eligible for the school lunch program; - proportion from single-parent families; - proportion of boys; - proportion of whites or non-whites. Also found: - roughly equivalent proportions of students in both the full intervention and control groups were living in disorganized neighborhoods at age 16 years as indicated by students' self-reports of rundown housing, crime, poor people, drug-selling, gangs, and disorderly and undesirable neighbors in their neighborhoods.
Were participants aware of group allocation?	NOT LIKELY Pg 26 Parents consented to the participation of students

Guide	Comments
	in the intervention and the participants consented to the follow-up interview. Since the intervention was given to the entire class, students would not necessarily be aware that their class was different than usual curriculum. However, teachers would know that they were delivering a different curriculum and had different training to do so, and this may have created a "Pygmalion" effect.
Were clinicians aware of group allocation?	Not relevant to this study
Were outcome assessors aware of group allocation?	NO – The information was collected by self-report. Pg 28 – Crime rates were collected from state and national records.
Was follow-up complete?	YES 94% were interviewed two years following baseline measurement (excellent follow-up).
II. What are the Results?	
How large was the treatment effect?	Table 2, pg 29 (last two lines): Crime rates from records were court charges in past year and court charges in lifetime. Mean difference between control and full intervention group was: 1. Court charges in past year: 3% lower in full intervention group than control group 2. Court charges in lifetime: 11% lower in full intervention group than control group.
How precise was the treatment effect?	1. Court charges in past year: 3% lower in full intervention group; (reported as -0.03, Confidence Interval is -0.10 to 0.04). This crosses the line of no difference (includes 0), so is not statistically significant; also indicated by p value- 0.40. 2. Court charges in lifetime: 11% lower in full intervention group (reported as -0.11, Confidence Interval -0.21 to -0.01). This CI does not include 0 so is statistically significant; also indicated by p value of 0.04. This CI indicates that the differences in court charges may be as high as 21% or as low as 1%. This is a fairly wide confidence interval, i.e., not very precise. Is the difference meaningful? Considering that the respondents are only 21 years old, it is an outcome worth considering.

Guide	Comments
III. How can I apply the results?	
Were the study participants similar to my own situation?	It would be necessary to know more about your actual neighbourhood, but this article does not give much information. Much more info re participants and their socioeconomic status is located in Ref #19.
Were all clinically important outcomes (harms and benefits) considered?	YES Many other outcomes are included, which were based on self-report. Statistically significant differences showed that the intervention group did better than the control group in school achievement, job performance, emotional regulation, suicide thoughts, variety of crime, and selling drugs in the past year.

Resolution of scenario

In to the scenario, you were asked:

1. How will you summarize this article for your group?
2. Will you recommend a school curriculum intervention to reduce crime?

1. The study could have been done as a randomized trial, which would have eliminated any concerns we had about some unknown bias in the groups before the intervention started. However, sometimes this is the best evidence you can get, particularly in community or population level interventions. This study did have a very impressive follow-up rate, considering they tracked the children to age 21. In addition, it was a strength that they supplemented self-report (which is usually heavily influenced by social desirability) with state and national crime records.

A difference in crime rates of 11% can be quite meaningful at a community and individual level. However, the confidence intervals are wide and, in the "worst case scenario", the actual difference may be as low as 1%.

2. Your recommendation: The local school board should consider how curriculum could fit in the current "overload" situation. They should also draft a budget for what such a curriculum would cost the school board over the next 10 years, to also include an evaluation component.

⚠ A word of caution! Remember, in REAL LIFE, you would look at all the studies (or ideally, a review) and not base your decision on this one study.

6. Useful References

OTHER RESOURCES

Duke University Medical Center. *Introduction to Evidence-Based Medicine. Evaluating the Evidence*. http://www.hsl.unc.edu/services/tutorials/ebm/Evidence.htm

GRADE Working Group (2004). Grading quality of evidence and strength of recommendations. *British Medical Journal*, 328,1490-7.

Guyatt, G. & Rennie, D. (Eds) (2002). *Users' Guides to the Medical Literature: A manual for Evidence-Based Clinical Practice*. American Medical Association

7. Glossary

Absolute risk difference: arithmetic difference in the event rates between intervention and control groups (obtained by subtracting one event rate from the other), usually reported as a %. If the risk in the intervention group is less than the control group, we call that an **Absolute risk reduction**.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell.

Bias: a systematic error or departure from the truth in results.

Blinding (masking): in an experimental study, refers to whether patients, clinicians providing an intervention, people assessing outcomes, and/or data analysts were aware or unaware of the group to which patients were assigned.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell.

Dichotomous data: data that can take one of two values (e.g., dead or alive, symptoms present or absent). Also known as binary data.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell.

Cohort study: a group of people with a common set of characteristics or set of characteristics that are followed up for a period of time to determine the incidence of an outcome; there is no comparison group.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction*. Oxford: Blackwell

Case control study: an observational study that begins by comparing patients who have the health problem (cases) and control participants who do not have the health problem, and then looking back in time to identify the existence of possible causal factors, for example, identifying patients with and without lung cancer and looking back in time to determine past smoking behavior (exposure to tobacco).

Dawson-Saunders, B., Trapp, R.G. (1994). *Basic and Clinical Biostatistics.* Norwalk: Appleton & Lange

Confounder: a variable that affects the observed relationship between two other variables. For example, alcohol consumption is related to lung cancer but does not cause the disease; instead, both alcohol and lung cancer are related to smoking (the confounder), which causes lung cancer.

Crombie, I.K. (1996). *The pocket guide to critical appraisal: A handbook for Healthcare Professionals.* London: BMJ Publishing Group.

Confidence interval (CI): quantifies the uncertainty in measurement; usually reported as 95% confidence interval, which is the range of values within which we can be 95% sure that the true value for the entire population lies.

Continuous data: data with a potentially infinite number of values along a continuum (weight, blood pressure).

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell.

Evidence-informed decision-making: the use of evidence that contributes to decision making about particular problems or issues about best use of resources within institutions and across the healthcare system.

Canadian Health Services Research Foundation (2006). *Weighing Up the Evidence. Making evidence-informed guidance accurate, achievable, and acceptable. A summary of the workshop held on September 29, 2005. http://www.chsrf.ca/other_documents/evidence_e.php#definition, last downloaded May 2008).*

Intention-to-treat analysis: all patients are analysed in the groups to which they were randomised, even if they failed to complete the intervention or received the wrong intervention.

Evidence-Based Nursing, Glossary. http://ebn.bmj.com/cgi/reprint/11/2/65?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&titleabstract=Glossary&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT

Number needed to treat (NNT): number of patients who need to be treated to prevent 1 additional negative event (or to promote 1 additional positive event). This is calculated as 1/absolute risk reduction (rounded to the next whole number), accompanied by the 95% confidence interval.

Evidence-Based Nursing, Glossary. http://ebn.bmj.com/cgi/reprint/11/2/65?maxtoshow=&HITS=10&hits=10&RESULTFORMAT=&titleabstract=Glossary&searchid=1&FIRSTINDEX=0&resourcetype=HWCIT

Odds Ratio: describes the odds of a patient in the experimental group having an event divided by the odds of a patient in the control group having the event, or the odds that a patient was exposed to a given risk factor divided by the odds that a control patient was exposed to the risk factor.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell

p value: a statistical value that relates the probability that the obtained results are due to chance alone; a p value of < 0.05 means that there is less than a 1 in 20 probability that the result is occurring by chance alone.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell

Randomized controlled trial (RCT): a study design in which individuals are randomly allocated to receive alternative preventive, therapeutic or diagnostic interventions and then followed up to determine the effect of the interventions (one of the alternatives might be no intervention).

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell.

Relative Risk (RR): proportion of patients experiencing an outcome in the treatment (exposed) group divided by the proportion experiencing the outcome in the control (unexposed) group.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction.* Oxford: Blackwell

Relative benefit increase (RBI): the proportional increase in the rates of good outcomes between experimental and control participants; it is reported as a percentage (%).
It is calculated by dividing the rate of the good outcome in the experimental group (EER), minus the rate of the good outcome in the control group (CER) by the rate of the good outcome in the control group: EER-CER/CER.

DiCenso, A., Guyatt, G., & Ciliska, D. (2005*). Evidence-Based Nursing: A guide to clinical practice*. St Louis: Mosby.

Statistical significance: indicates that results obtained in an analysis are unlikely to have occurred by chance and the null hypothesis is rejected (meaning that there is a difference in outcome between the groups). When statistically significant, the probability of finding the result by chance falls below a specified level of probability (most often $p<0.05$).

Systematic review: a research summary of all evidence that relates to a particular question; the question could be one of intervention effectiveness, causation, diagnosis or prognosis. The systematic review process follows a rigorous methodology for searching, retrieval, relevance and quality rating, data extraction, data synthesis and interpretation.

Cullum, N., Ciliska, D., Haynes, R.B., & Marks, S. (2008). *Evidence-Based Nursing. An Introduction*. Oxford: Blackwell