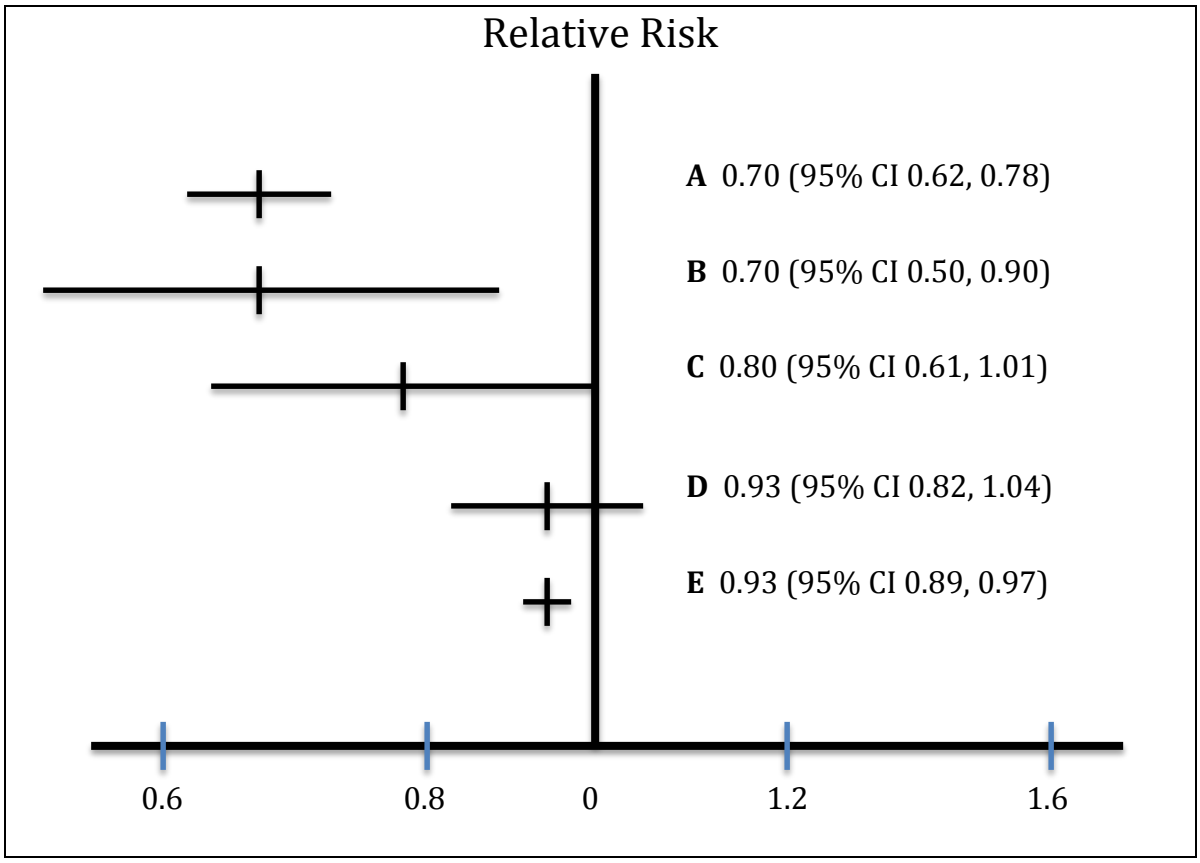
EVIDENCE APPRAISAL WORK BOOK

Numerology of EBM: The Basics and Beyond

Understanding numbers in EBM is essential. Without a solid understanding, it will be very easy to be led astray. Once you are comfortable with the numbers, it becomes simpler to review evidence and see when information is being misinterpreted or misrepresented.

1) **What do Confidence Intervals tell us?**



i) What is the difference between A & B (in words)?

_____

_____

ii) What is the difference between D & E (in words)?

_____

_____

iii) How would you explain C to a patient (be more simplistic)?

_____

_____

iv) How would you explain C to a resident (be more detailed)?

_____

_____

2) **Relative and Absolute Numbers**

When you hear relative and absolute numbers, understanding differences are essential. In prevention, the NNT (number needed to treat) to benefit one person is often much larger (worse) than when we are treating people with a problem. Therefore, it is most important to differentiate relative and absolute risk in prevention.

In the shingles prevention study, 38,546 adults 60 years of age or older, were followed a median of 3.1 years. Here is what they said in the abstract results section: "reduced the incidence of herpes zoster by 51.3 percent (P<0.001)." Here are the results.

Table 2. Effect of Zoster Vaccine on the Burden of Illness in Herpes Zoster in the Modified Intention-to-Treat Population.*

Group of Subjects	Vaccine Group			Placebo Group			VE_{BOI} (95% CI)§
	No. of Confirmed Cases/No. of Subjects	BOI Score†	Incidence per 1000 Person-Yr‡	No. of Confirmed Cases/No. of Subjects	BOI Score†	Incidence per 1000 Person-Yr‡	%
All subjects	315/19,254	2.21	5.42	642/19,247	5.68	11.12	61.1 (51.1–69.1)

i) What is the Event Rate with and without vaccine?

_____

ii) What is the relative risk and relative risk reduction?

_____

iii) What is the absolute risk reduction?

_____

iv) How many need to be vaccinated to prevent one shingles (over 3.1yrs)?

_____

v) How can you make this number better (what if 10yr risk was 10%)?

_____

vi) What assumptions are made when interpreting these results?

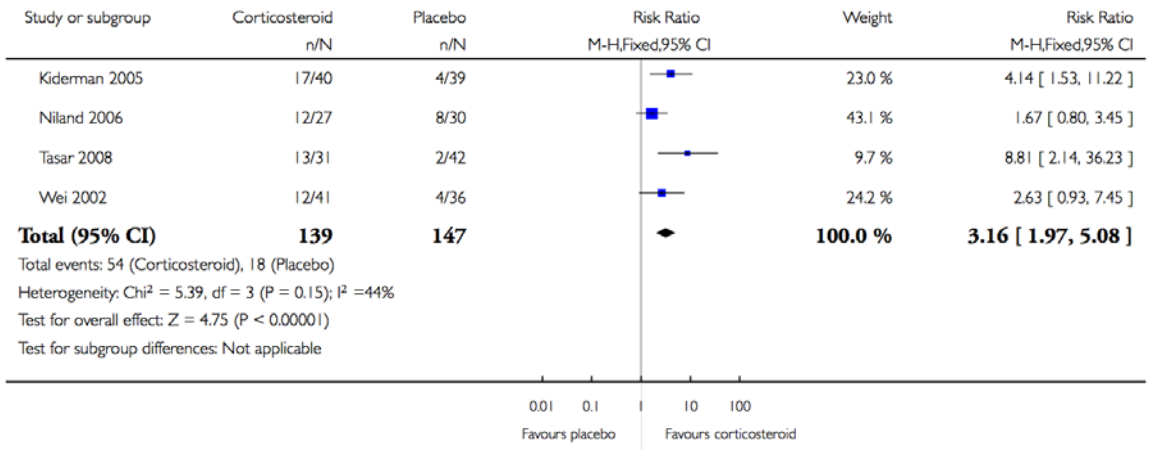
_____

Reference: Oxman MN, Levin MJ, Johnson GR, et al. A vaccine to prevent herpes zoster and postherpetic neuralgia in older adults. N Engl J Med. 2005;352:2271-84.

3) **How to get NNT from a Meta-graph**

In many meta-analyses, they don't give the NNT and it is often not readily apparent how many in each group had events. There are three ways to calculate NNT from meta-analyses. Let's examine corticosteroid use for sore throat, and the number of people with complete recovery at 24 hours.

Reference: Hayward G, Thompson MJ, Perera R, Glasziou PP, Del Mar CB, Heneghan CJ. Corticosteroids as standalone or add-on treatment for sore throat. Cochrane Database Syst Rev. 2012 Oct 17;10:CD008268.



Method 1: When available, use the numbers on the meta-graph.
Corticosteroid: Pain resolved in 24 hours for 54 out of 139.
Placebo: Pain resolved in 24 hours for 18 out of 147.

i) The event rates were: _____ Corticosteroid and _____ Placebo.
ii) The absolute difference was _____ and the NNT _____.

Note: This is considered a poorer method as it does not take into account weighting, etc.

Method 2: When summed numbers are not available, you can use the placebo event rate from the trial that best reflects your practice, your clinical question, or your specific patient. Here are the basics of the population.

	Country	Age	Severity
Kiderman	Israel	18-65 (mean 34)	57% GAS, 87% exudative
Niland	USA	4-21 (mean 7.7)	100% GAS, 57% exudative
Tasar	Turkey	18-65 (mean 31)	?
Wei	USA	>15 (mean 28)	27% GAS, 43% exudative

So, if you're seeing children, you might pick Niland. In Niland, the event rate was 8/30 or 26.7%. The relative risk (or risk ratio) is 3.16.

i) What would be the event rates with steroid? _____
ii) What would be the NNT? _____
iii) If you use the Wei study numbers (for adults in North America), what would be the event rate and NNT? _____
iv) Although this is the "preferred" EBM method, you can see some of the limits. Consider a few reasons? _____

_____

_____

Method 3: Instead of using the baseline (or placebo) event rates in the RCTs of the meta-analysis, use event or prevalence rates from your community or cohorts that match you community. A good example of this approach is considering the benefits of statins for the patient in front of you. Low-moderate potency statins reduce CVD by 25% (relative risk 0.75).

Your patient is a 55 year old male smoker non-diabetic with 130 systolic BP, 5 mmol/L total cholesterol and 1 mmol/L HDL: His risk of CVD over 10 years is 20%.

i) What would be his new risk if he took a statin? _____

ii) What would be the absolute reduction in his risk? _____

iii) What would be his NNT? _____

4) **Continuous Outcomes and Scales**

In many cases, the clinical outcomes are not dichotomous (yes or no answers) like death, hospitalization or stroke. Instead, clinical problems like pain or depression are often assessed on scales. We are now opening Pandora's Box.

Let's say you were doing a study of gabapentin for chronic neuropathic pain:

i) How would you measure your outcomes?

_____

_____

ii) Can you come up with at least three more outcomes?

_____

_____

_____

iii) Can you think of way to make the continuous outcome (pain) in to a dichotomous one?

_____

_____

Note: Only a dichotomous outcome can be turned in to a NNT.

5) **A Potpourri of Number Tricks**

Trick 1: **Report a Statistical difference that is not clinically meaningful.**

Let's consider the evidence that leads suggesting escitalopram is superior to citalopram. Meta-analysis shows that on MADRS scale, escitolopram is superior to citalopram: 1.13-point reduction (CI, 0.18 to 2.09).

Reference: Allan GM, Virani AS, Ivers N. Second-generation antidepressants. Can Fam Physician. 2011;57(10):1143.

i) What information about MADRS would be helpful in interpreting this?

_____

ii) What is the name of the "cut-off" that is helpful in interpreting scales?

_____

_____

iii) What else may cause bias in favour of one over another?

_____

_____

iv) Is it possible that a important difference still exists?

_____

_____

Trick 2: **Make multiple comparisons and find a few by chance**

Look at the results in the abstract of this study examining when women took HRT to see if that influenced events. Consider the small note in the methods.

i) What percent of results do you expect to be positive by chance?

_____

ii) Therefore, if study had 80 statistical comparisons, how many would be positive by change?

_____

iii) Is there a way to tell which results are positive by chance and which are real?

_____

_____

Results In the combined trials, there were 396 cases of CHD and 327 cases of stroke in the hormone therapy group vs 379 cases of CHD and 239 cases of stroke in the placebo group. For women with less than 10 years since menopause began, the hazard ratio (HR) for CHD was 0.76 (95% confidence interval [CI], 0.50-1.16); 10 to 19 years, 1.10 (95% CI, 0.84-1.45); and 20 or more years, 1.28 (95% CI, 1.03-1.58) (*P* for trend=.02). The estimated absolute excess risk for CHD for women within 10 years of menopause was –6 per 10 000 person-years; for women 10 to 19 years since menopause began, 4 per 10 000 person-years; and for women 20 or more years from menopause onset, 17 per 10 000 person-years. For the age group of 50 to 59 years, the HR for CHD was 0.93 (95% CI, 0.65-1.33) and the absolute excess risk was –2 per 10 000 person-years; 60 to 69 years, 0.98 (95% CI, 0.79-1.21) and –1 per 10 000 person-years; and 70 to 79 years, 1.26 (95% CI, 1.00-1.59) and 19 per 10 000 person-years (*P* for trend=.16). Hormone therapy increased the risk of stroke (HR, 1.32; 95% CI, 1.12-1.56). Risk did not vary significantly by age or time since menopause. There was a nonsignificant tendency for the effects of hormone therapy on total mortality to be more favorable in younger than older women (HR of 0.70 for 50-59 years; 1.05 for 60-69 years, and 1.14 for 70-79 years; *P* for trend=.06).

Statistical tests were undertaken at the .01 level to partially account for multiple testing issues and the post hoc nature of some of the tests. Forty-two tests for trend, 33 additional interaction tests, and 62 comparisons of HRs were performed (a total of 137 tests). Two *P* values were significant (1-2 were expected by chance). For consistency with previous WHI studies, HRs and 95% CIs were used. An HR of less than 1 favored hormone therapy and greater than 1 favored placebo. The 95% CIs were estimated in 182 subgroups. Of these 182, 19 did not include 1 (9 were expected by chance). Statistical analyses were performed using SAS version 9 (SAS Institute Inc, Cary, NC).

Rossouw JE, Prentice RL, Manson JE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. JAMA. 2007;297: 1465-77.

Trick 3: **It's hard to improve symptoms in people not really sick: False Negative**
In a well-designed RCT, 128 people in Queensland Australia were given 5000 IU/day of Vitamin D or placebo and depression was assessed with the Beck Inventory.

	Beck score at Start	Beck score at Finish	Statistical difference
Vitamin D	7.2	6.4	None
Placebo	5.7	5.4	

i) Please provide at least two reasons this study failed to show a difference?

_____

_____

ii) What information do you need to know about the Beck inventory tool?

_____

_____

iii) What other trial information would have been helpful?

_____

_____

Trick 4: **Including the wrong people**: two examples

A. When you include general public they might not mirror patients you would see. A Cochrane review of back pain, look at the results of exercise for two different groups of patients.

	Improvement (out 100) over placebo	
	Pain	Function
Recruited from community	7.3 (3.7-10.9)	2.5 (1.0-3.9)
Actual Patients seeking care	13.3 (5.5-21.1)	6.9 (2.2-11.7)

i) Which group is getting more benefit?

_____

_____

ii) What explains the difference in benefit between the groups?

_____

_____

Reference: Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev. 2005 Jul 20;(3):CD000335.

B. Studies of people in referral clinics can also generate results that may be different. Examine the results from two trials of orlistat for weight loss, one for primary care one in secondary, both using 120mg TID.

	Mean percent weight loss	Percent with 5% or more weight loss
Primary Care study	1.3% more	13%
Tertiary Care study	3% more	22%

i) Which group is getting more benefit?

_____

_____

ii) What explains the difference in benefit between the groups?

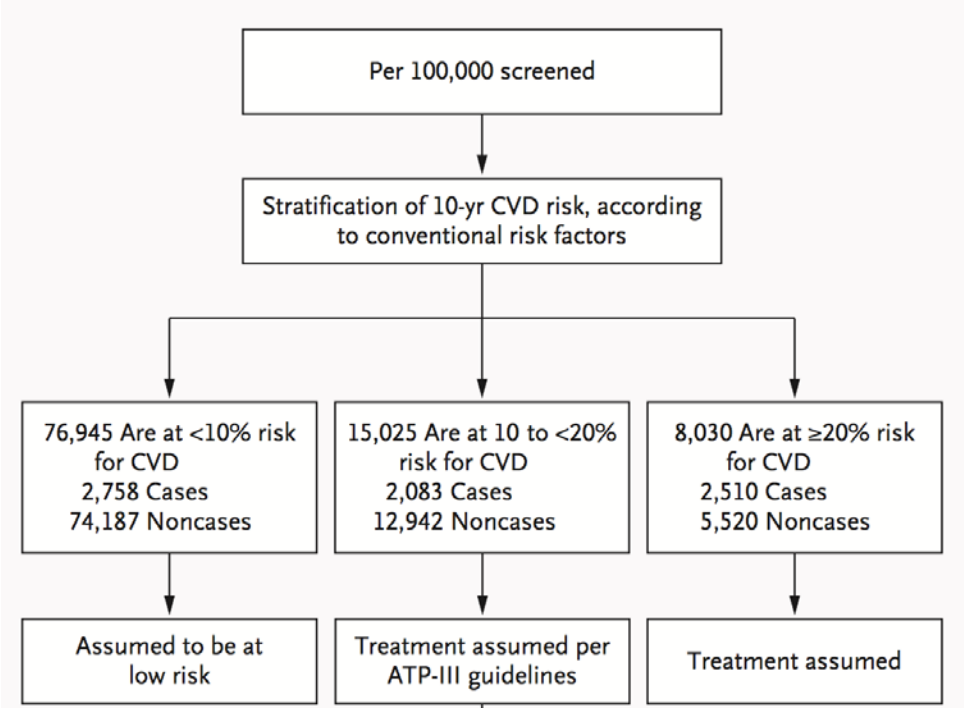
_____

_____

References:
Lindgärde F. The effect of orlistat on body weight and coronary heart disease risk profile in obese patients: the Swedish Multimorbidity Study. J Intern Med. 2000;248(3):245-54.
Davidson MH, Hauptman J, DiGirolamo M, et al. Weight control and risk factor reduction in obese subjects treated for 2 years with orlistat: a randomized controlled trial. JAMA. 1999;281(3):235-42.

Trick 5: **The numerator means little without a denominator.**
You may have heard that more people with normal cholesterol have heart attacks (than people with high cholesterol). It is also known that within primary care, more patients assessed as having low risk (<10% CVD over 10 years) will have a CVD event than those assessed as having high risk. So, how is this possible?

Look at the flowchart below:



Emerging Risk Factors Collaboration, et al. C-reactive protein, fibrinogen, and cardiovascular disease prediction. N Engl J Med. 2012;367:1310-20.

i) Which group is the largest?

_____

_____

ii) How is it that low risk has more events?

_____

_____

iii) What are the actual event rates per group?

_____

_____

Possible outcomes for a pain study

Here are some examples of ways to look at continuous data.

1) Use a 10 Point VAS scale:

Continuous Pain Outcome

- Mean Improvement in Scale vs placebo
- Mean Final Pain Scores vs placebo
- Mean Percent Improvement over baseline versus placebo
- Mean attained score (as ratio of baseline) versus placebo
- Time to 50% of population having resolution in pain (variable)
- Time to 100% of population having 50% improvement in pain (variable)

Dichotomous

- Number attaining a specific VAS score (e.g. 2/10) (variable)
- Number with a certain numeric improvement (e.g. better by 2 points) (variable)
- Number with certain percent improvement (50% reduction in pain) (variable)
- Number attaining a specific "minimal clinical important difference"

2) Use a Specific Pain Scale (like the NPS – Neuropathic Pain Scale)

- There may be a dozen types of scales (such as in depression)
- All of the above again

3) Examine any outcome at **Different Time Points**

- Multiple options: 1 week, 1 month, 3 months, etc
- All of the above again

4) Examine **Function**

- There may be a dozen types of scales
- All of the above again

5) Examine **Quality of Life**

- There may be multiple scales
- All of the above again

If your start to combine these, it is easy to see how 12 RCTS of gabapentin had 40 primary outcomes and 180 secondary outcomes. (N Engl J Med 2009;361:1963-71.)

Ideally, in any of these studies, you want two main outcomes

1) Mean improvements compared to placebo – Continuous outcome
2) Number in each group attaining the MCID – Dichotomous outcome.

MADRS is out of 60 and 2 is the MCID

BECK is out of 63, 0-9 is considered not depressed. Maybe even 14 before depression starts. The MCID is 4-5 (The Journal of Pain, Vol 9, No 2 (February), 2008: pp 105-121).

META-ANALYSIS: Pitfalls and Tricks

A common challenge in evidence is multiple systematic reviews and meta-analyses, sometimes with different results/findings.

There are many reasons why we get multiple systematic reviews. Here are some:

1) The quality of the research was poor.
2) Subtle changes in the enrolment / inclusion may change results.
3) There is dissatisfaction with the answer.
4) There is eagerness among some to tweak things and get an easy publication.
5) Other (sometimes multiple funding agencies (e.g. AHRQ) simultaneously will pay different groups to do a systematic review on a topic).

Meta-analyses: Task 1

Let's examine the idea of aggressive versus conventional management of blood glucose (A1c) in Type Diabetes. Here are 10 of the systematic reviews, spanning 3 years (2009-2011)

- Diabetes Res Clin Pract. 2009 Dec;86 Suppl 1:S57-62.
- Diab Vasc Dis Res. 2010 Apr;7(2):119-30.
- Lancet. 2009 May 23;373(9677):1765-72.
- BMJ. 2011 Nov 24;343:d6898. doi: 10.1136/bmj.d6898.
- BMJ. 2011 Jul 26;343:d4169. doi: 10.1136/bmj.d4169.
- Ann Med. 2010 May 6;42(4):305-15.
- Heart Lung Circ. 2011 Oct;20(10):647-54.
- Ann Intern Med. 2009 Sep 15;151(6):394-403.
- Nutr Metab Cardiovasc Dis. 2009 Nov;19(9):604-12.
- Diabetologia. 2009 Nov;52(11):2288-98.

Let's look at an early one and a later one: Just examining CVD mortality.

Examine the two meta-analyses and consider

i) What is the primary difference between the two?

_____

_____

ii) What do the new trials add to your understanding?

_____

_____

iii) What, if any, is the relevance of the differences in studies (heterogeneity)?

_____

_____

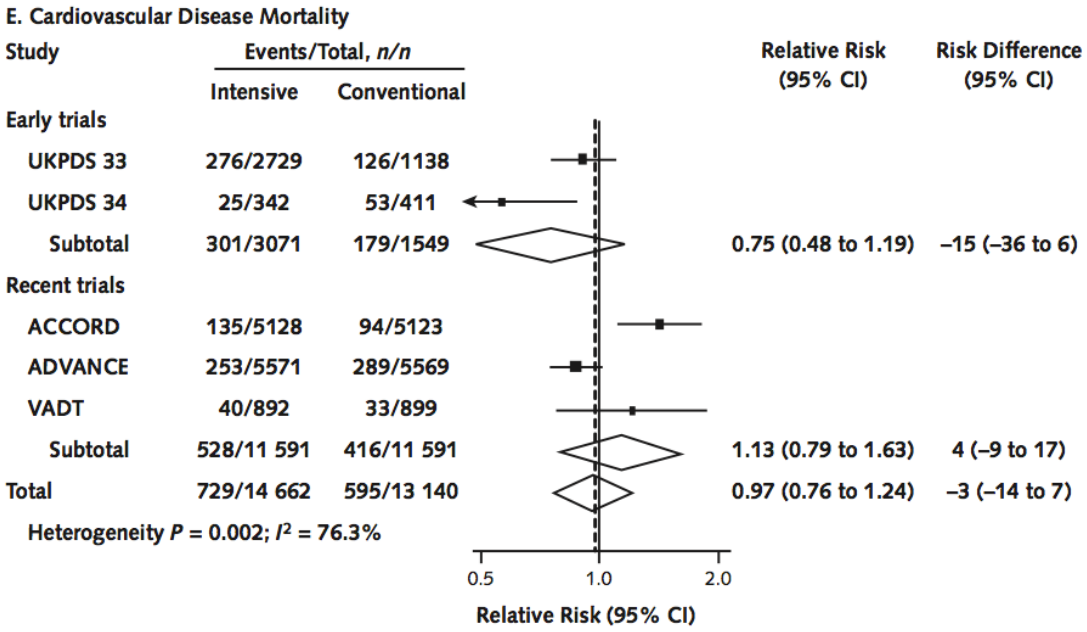
iv) What, if anything, do the individual studies tell us here?

_____

_____

META-ANALYSIS: Pitfalls and Tricks

1) Ann Intern Med. 2009 Sep 15;151(6):394-403.



2) BMJ. 2011 Nov 24;343:d6898. doi: 10.1136/bmj.d6898.

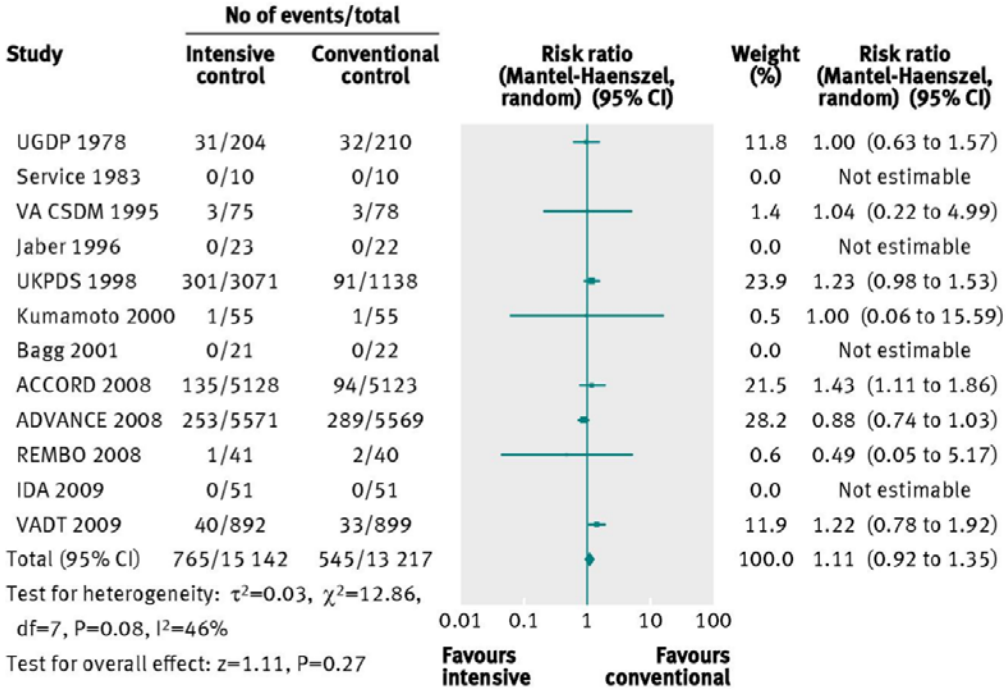


Fig 4 Forest plot for cardiovascular mortality

Meta-Analyses: Task 2

Next, we'll examine if Vitamin D can prevent Respiratory Tract Infections. Three systematic reviews with meta-analysis addressed these questions.

1. Charan J, Goyal JP, Saxena D, Yadav P. Vitamin D for prevention of respiratory tract infections: A systematic review and meta-analysis. J Pharmacol Pharmacother. 2012;3(4):300-3.
2. Bergman P, Lindh AU, Björkhem-Bergman L, Lindh JD. Vitamin D and Respiratory Tract Infections: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. PLoS One. 2013;8(6):e65835.
3. Mao S, Huang S. Vitamin D supplementation and risk of respiratory tract infections: a meta-analysis of randomized controlled trials. Scand J Infect Dis. 2013;45(9):696-702.

Examine the three meta-analyses and consider:

i) What are the primary differences between the three?

_____

_____

ii) What information is missing (from each)?

_____

_____

iii) Can you propose reasons why results varied?

_____

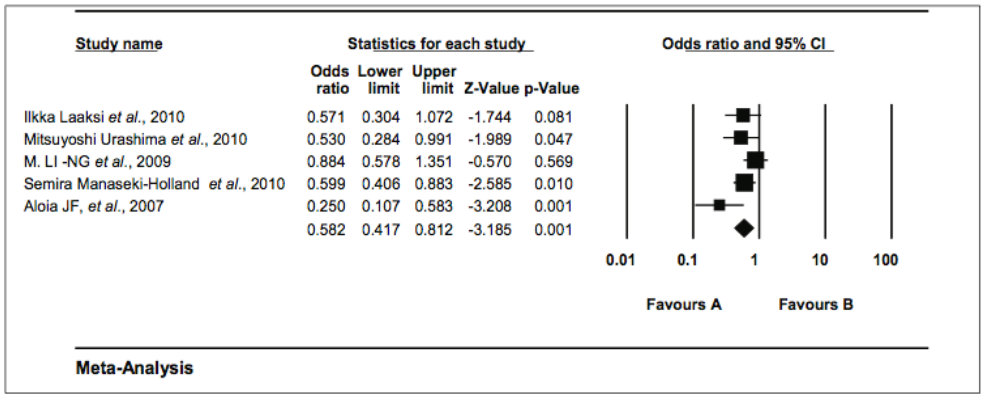
_____

iv) What are your thoughts on heterogeneity and weighting?

_____

_____

1) J Pharmacol Pharmacother. 2012;3(4):300-3.



2) PLoS One. 2013;8(6):e65835.

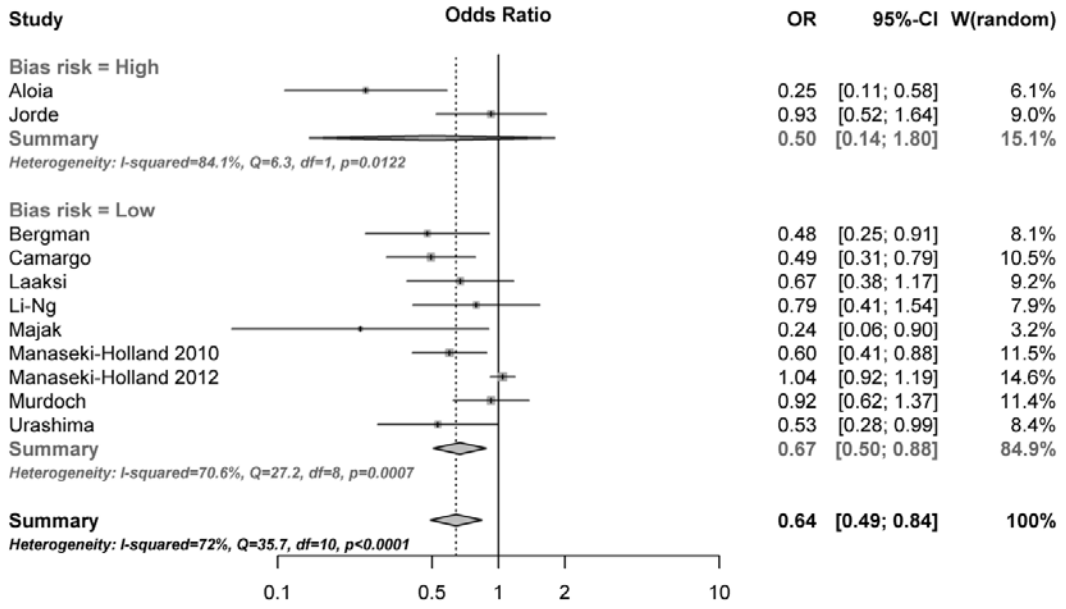
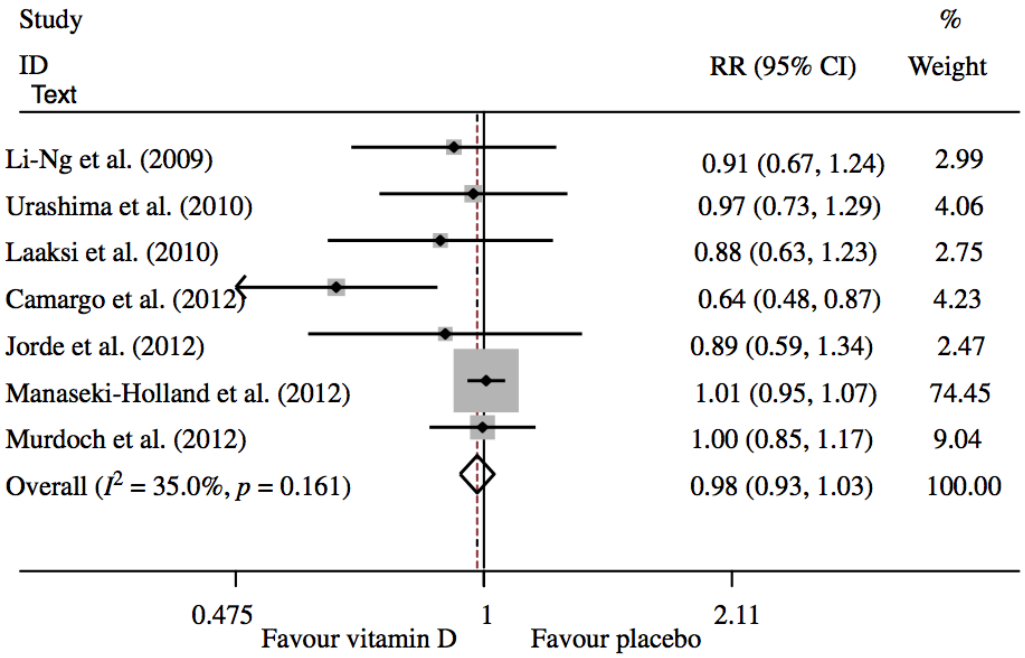


Figure 2. Efficacy of vitamin D for prevention of respiratory tract infections. Error bars indicate 95% confidence intervals.
doi:10.1371/journal.pone.0065835.g002

3) Scand J Infect Dis. 2013;45(9):696-702.



META-ANALYSIS: Pitfalls and Tricks

Meta-analyses: Task 3

Sometimes meta-analyses come out and contradict each other. This can happen for a number of reasons but we'll examine two.

A) Let's consider the question "Do statins reduce mortality in primary prevention?" Many meta-analyses have examined this but we are going to focus on two from large impact journals.

1) Ray KK, Seshasai SRK, Erqou S, et al. Statins and all-cause mortality in high-risk primary prevention: a meta-analysis of 11 randomized controlled trials involving 65,229 participants. Arch Intern Med 2010, 170:1024-1031.
2) Tonelli M, Lloyd A, Clement F, et al. Efficacy of statins for primary prevention in people at low cardiovascular risk: a meta-analysis. CMAJ. 2011;183:E1189-202.

Examine the two meta-analyses and consider:

i) What are the primary differences in results between the two?

_____

_____

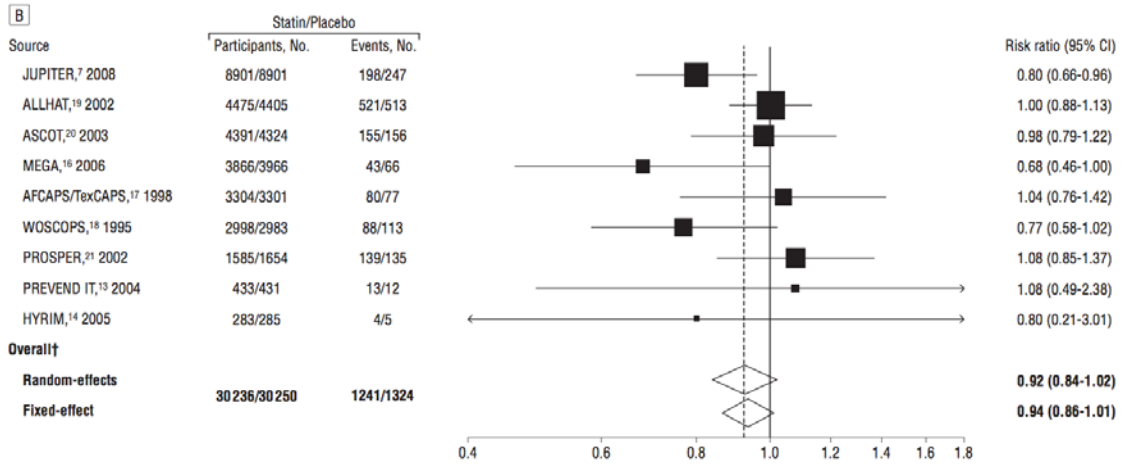
ii) To what degree, if any, did they find something different?

_____

_____

1) Arch Intern Med 2010, 170:1024-1031.

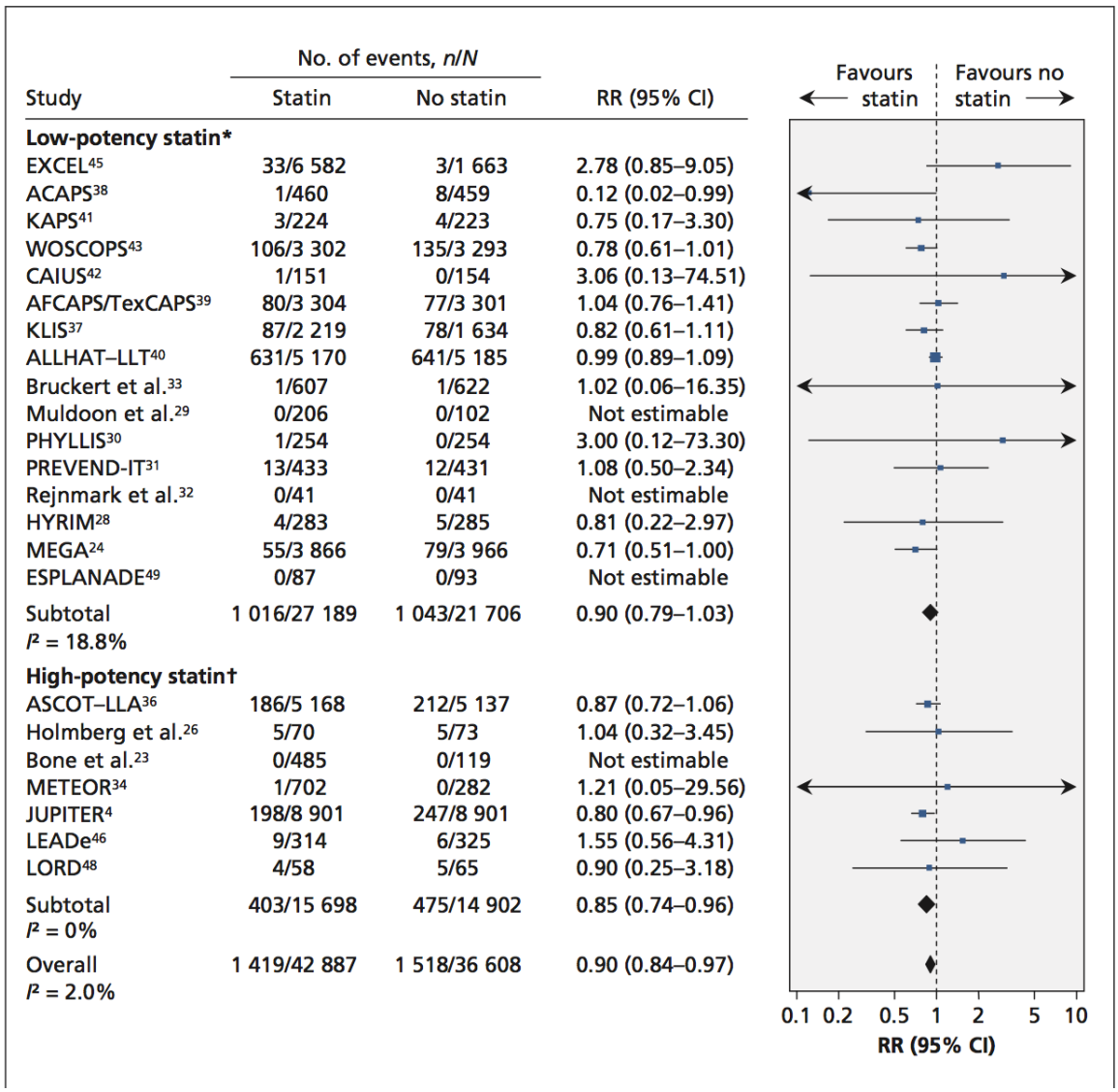
Conclusion: "This literature-based meta-analysis did not find evidence for the benefit of statin therapy on all-cause mortality in a high-risk primary prevention set-up."



Heterogeneity (I2=24%, P=.23).

2) CMAJ. 2011;183:E1189-202.

Conclusion: "Statins were found to be efficacious in preventing death and cardiovascular morbidity in people at low cardiovascular risk."



Study	No. of events, *n/N* Statin	No statin	RR (95% CI)
Low-potency statin*			
EXCEL[45]	33/6 582	3/1 663	2.78 (0.85–9.05)
ACAPS[38]	1/460	8/459	0.12 (0.02–0.99)
KAPS[41]	3/224	4/223	0.75 (0.17–3.30)
WOSCOPS[43]	106/3 302	135/3 293	0.78 (0.61–1.01)
CAIUS[42]	1/151	0/154	3.06 (0.13–74.51)
AFCAPS/TexCAPS[39]	80/3 304	77/3 301	1.04 (0.76–1.41)
KLIS[37]	87/2 219	78/1 634	0.82 (0.61–1.11)
ALLHAT–LLT[40]	631/5 170	641/5 185	0.99 (0.89–1.09)
Bruckert et al.[33]	1/607	1/622	1.02 (0.06–16.35)
Muldoon et al.[29]	0/206	0/102	Not estimable
PHYLLIS[30]	1/254	0/254	3.00 (0.12–73.30)
PREVEND-IT[31]	13/433	12/431	1.08 (0.50–2.34)
Rejnmark et al.[32]	0/41	0/41	Not estimable
HYRIM[28]	4/283	5/285	0.81 (0.22–2.97)
MEGA[24]	55/3 866	79/3 966	0.71 (0.51–1.00)
ESPLANADE[49]	0/87	0/93	Not estimable
Subtotal I^2 = 18.8%	1 016/27 189	1 043/21 706	0.90 (0.79–1.03)
High-potency statin†			
ASCOT–LLA[36]	186/5 168	212/5 137	0.87 (0.72–1.06)
Holmberg et al.[26]	5/70	5/73	1.04 (0.32–3.45)
Bone et al.[23]	0/485	0/119	Not estimable
METEOR[34]	1/702	0/282	1.21 (0.05–29.56)
JUPITER[4]	198/8 901	247/8 901	0.80 (0.67–0.96)
LEADe[46]	9/314	6/325	1.55 (0.56–4.31)
LORD[48]	4/58	5/65	0.90 (0.25–3.18)
Subtotal I^2 = 0%	403/15 698	475/14 902	0.85 (0.74–0.96)
Overall I^2 = 2.0%	1 419/42 887	1 518/36 608	0.90 (0.84–0.97)

B) Let's consider the question "Does regular self monitoring of blood glucose (SMBG) improve glycemic control." There are many meta-analyses (again) but we are going to look at two based on the different wording in the conclusion.

1) Farmer AJ, Perera R, Ward A, et al. Meta-analysis of individual patient data in randomised trials of self monitoring of blood glucose in people with non-insulin treated type 2 diabetes. BMJ. 2012 Feb 27;344:e486.
2) Allemann S, Houriet C, Diem P, Stettler C. Self-monitoring of blood glucose in non-insulin treated patients with type 2 diabetes: a systematic review and meta-analysis. Curr Med Res Opin. 2009;25(12):2903-13.

META-ANALYSIS: Pitfalls and Tricks

Examine the two meta-analyses and consider:

i) What are the primary differences in results between the two?

_____

_____

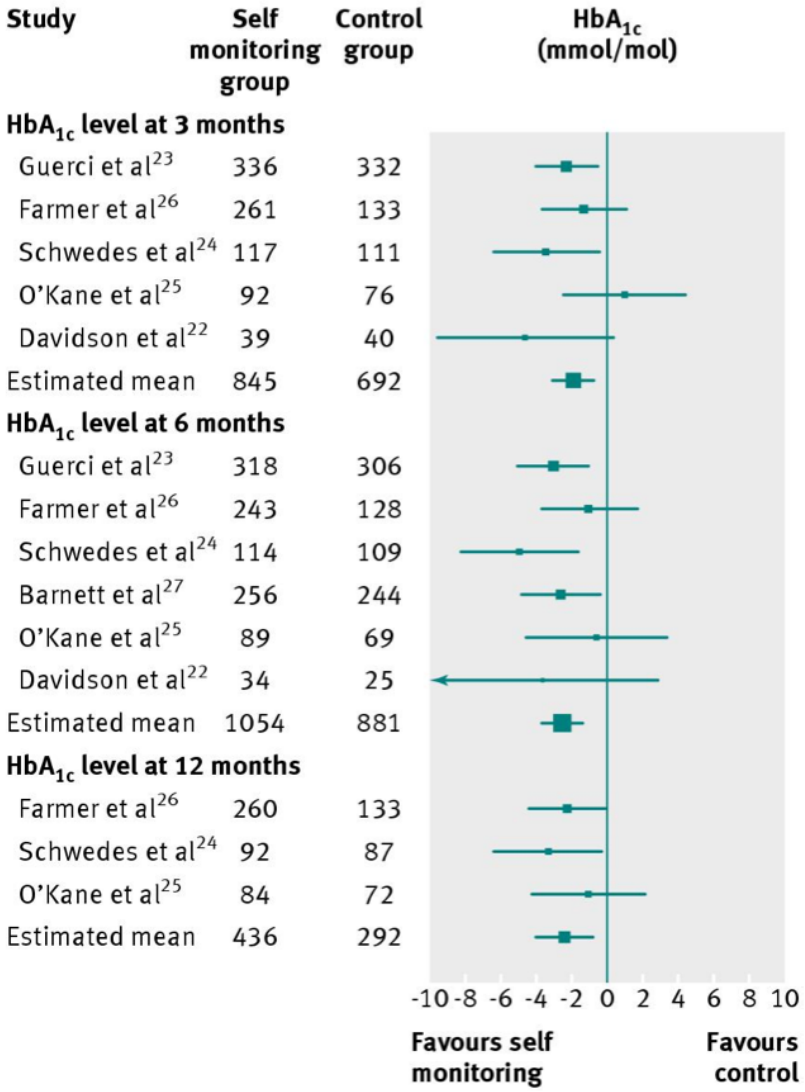
ii) To what degree, if any, did they find something different?

_____

_____

1) BMJ. 2012 Feb 27;344:e486.

Conclusions: "Evidence from this meta-analysis of individual patient data was not convincing for a clinically meaningful effect of clinical management of non-insulin treated type 2 diabetes by self monitoring of blood glucose levels compared with management without self monitoring,.."



2) Curr Med Res Opin. 2009;25(12):2903-13.

Conclusion: "SMBG compared with non-SMBG is associated with a significantly improved glycaemic control in non-insulin treated patients with type 2 diabetes."