How to CRITICALLY APPRAISE *an RCT in 10 minutes*

James McCormack
BSc(Pharm), Pharm D
Professor
Faculty of Pharmaceutical Sciences, UBC
Vancouver, BC, Canada
medicationmythbusters.com

CRITICAL APPRAISAL AND YOU

Critical appraisal skills have become as important as the use of a stethoscope or the ability to write a legible prescription. If the thought of reviewing a clinical study seems like an insurmountable task, READ ON, it's not that hard.

"Individual practitioners therefore need to be able to find and use evidence themselves—a 21st century clinician who cannot critically read a study is as unprepared as one who cannot take a blood pressure or examine the cardiovascular system"

IT'S EASIER THAN YOU THINK

With literally 100's of new randomized controlled trials coming out every week and with limited time, does the thought of reviewing even one of these give you a feeling of an absolutely insurmountable task or maybe that you can't even figure out where you should start and so you kind of just stand there and look longingly and just wondering what if? Well, I know exactly how you feel.

The bottom line with critical appraisal is it's really just a way for you to figure out if you can TRUST the results of a study you are reading. If you can trust the results, does the OUTCOME and MAGNITUDE of those results justify you incorporating them into your practice.

While one could, and many people do, spend literally hours critically appraising a study, what if I could show you how to do a pretty solid job of critically appraising an article in 10 minutes - would that be useful? If not, then I'm not sure why you got this book. However, if this is of interest to you then let's keep going by starting you off with the critical appraisal SET-UP .

THE
CRITICAL APPRAISAL SET-UP

Critical appraisal starts with a question - in a Patient with "X" condition, does an Intervention, compared to a Comparator change Outcomes - this is commonly referred to as a PICO question.

Virtually every study result will show numerical differences in outcomes of some magnitude between the intervention and the comparator groups. The reason you need to know how to do critical appraisal is to figure out what caused these numerical differences. In other words, to be somewhat glib, do we think the differences are due to the intervention or did the researcher make a "mistake". This is primarily a process of elimination.

1) was the difference seen due to CHANCE? - if we think it was a NON-CHANCE OCCURRENCE then

2) was it due to BIAS? - if NOT then, by process of elimination

3) we must assume the difference was DUE TO THE INTERVENTION

Statistics are used solely to give us an idea as to how often a particular result or one more extreme will occur by chance

alone if there really was no difference between the groups. That's it. Statistics can't do more than that.

Now if based on the statistics, we believe the difference seen was a non-chance event we need to figure out what caused that difference. That's why we go in search of bias. Were the groups different enough in their baseline characteristics that this could have caused the difference seen? Did the groups get assigned to the interventions in a biased way? Other than the interventions being studied, did the groups get treated the same? Did the subjects get followed for the whole study?

If we think none of the above contributed substantially to the differences seen then we must assume it was the intervention.

Then, YOU ALONG WITH YOUR PATIENT need to decide if the outcome(s) are important and if the MAGNITUDE OF THE DIFFERENCE IS CLINICALLY IMPORTANT.

While a more detailed critical appraisal can be very useful, the following approach is designed to show how you as a busy clinician can quickly figure all this out in around TEN MINUTES.

CRITICAL APPRAISAL AND YOU

Let's use an article from the New England Journal of Medicine (June 12, 2008) entitled "Effects of intensive glucose lowering in type 2 diabetes" as an example. It's 15 pages, 4500 words, 4 tables, 3 figures - YIKES - let's try to simplify - see below.

Simplicity is the Ultimate Sophistication

TURNING
15 pages, 4500 words, 4 tables, 3 figures
INTO
1 page, 10-30 words and 1 table

6 SIMPLE STEPS TO SUCCESSFUL CRITICAL APPRAISAL

(yes I know there are more than six steps in the picture)

1) First you have to figure out your NEEDS

4) Fourth, find out WHO was studied

5) Fifth, find out WHAT HAPPENED to them

2) Second, you have be able to SEARCH for BIAS

3) Third, you have to IGNORE much of the text

6) Finally, determine if you should CARE about the results

FINDING ALL YOUR NEEDS IN THE ABSTRACT

Effects of Intensive Glucose Lowering in Type 2 Diabetes

ABSTRACT

1) TITLE

3) OUTCOME

2) COMPARISON

4) DURATION

5) CONCLUSION

◀ ● 1 2 3 4 5 ▶

YOUR NEEDS

1) Read the TITLE – if the title doesn't have anything to do with your practice or doesn't interest you just STOP – don't go any further. There's no point. However if it does tickle your fancy you need to look further by looking for key statements in the abstract.

2) What was the COMPARISON – if what was compared is of interest you just keep going. This typically forms the basis of the clinical question - the PICO.

3) What were the OUTCOMES – are they important ones? If they are just surrogate markers like cholesterol, blood pressure, glucose you should likely stop right there as the results will unlikely be game-changers.

4) Look for the DURATION of the study

5) Now read the CONCLUSION – you don't read it because you want the answer - in my experience the conclusions can not infrequently bear little resemblance to the results - rather the conclusion is the "best that it can be". So, if the conclusion was true would it change your practice - if not - STOP - if yes, KEEP GOING. You now need to SEARCH FOR BIAS.

SEARCH FOR 5 WORDS
Random, Blind, Allocation, Intent, Follow
THEN LOOK FOR
Conflict

SEARCH FOR BIAS

SEARCH for the following words in the search bar of the PDF

RANDOM - you want to see if subjects were randomized to the different treatment groups – this is a key quality of a study as it helps to get subject factors equally distributed – in other words, it increases the chance the groups were similar at the start of the study which reduces the chance of bias.

Now type in BLIND – a study that is performed in a blinded fashion helps lessen the influence of the prejudices of either the subject or the researcher knowing what they are receiving – this is particularly important in studies that have more subjective endpoints like pain or other symptoms - single-blinded typically means the subject doesn't know what they are receiving and double-blinded means neither the subject nor the investigator knows. In addition, the people assessing the outcomes could also be blinded - sometimes called triple-blinded. The more the study is blinded the better, but unblinded studies can be fine as long as the endpoints are fairly objective.

Now type ALLOCATION – you are looking for allocation concealment – you want to know if the researcher could have figured out what treatment the next subject would receive as this may impact whether or not the researcher decides to enter the

subject into the study or what treatment the subject receives – that can affect the whole process of randomization. The success of randomisation in the study depends a lot on allocation concealment so if we know there was proper allocation concealment it minimizes the potential for selection bias. Sometimes you may not find the word allocation and then you have to look in the methods section to see how the randomization process was carried out. Central randomization by a third party is likely the best method to avoid bias.

Now type INTENT - you are looking for intention-to-treat – you want to know if all subjects who were randomized to one treatment or another got analyzed at the end of the study according to the treatment to which they were originally assigned – intention-to-treat is the most conservative way to look at the results - intention to treat analyses are done to avoid the effects of crossovers and drop-outs, which may break to some degree, the whole reason we randomized the study in the first place. However, a per-protocol analysis, where only those subjects who completed the entire study are evaluated, is considered more desirable for non-inferiority studies to reduce the chance of falsely concluding non-inferiority.

Next type FOLLOW – you want to know if all subjects were followed-up for the duration of the study or how many of them dropped out - you want as little as possible and preferably less than 20%. If during the study only a small % of subjects had an outcome, then even a smaller % loss-to-follow-up is required to decrease the chance of bias. If a greater number of sicker people dropped out in one group that can create a reason for the difference seen.

So how many of these things do you have to have for the results to be considered valid? There is no objective answer to that question it's really a very subjective gestalt type thing. The more the better.

Looking for Conflict

WHO PAID FOR THE STUDY

WHO GOT PAID
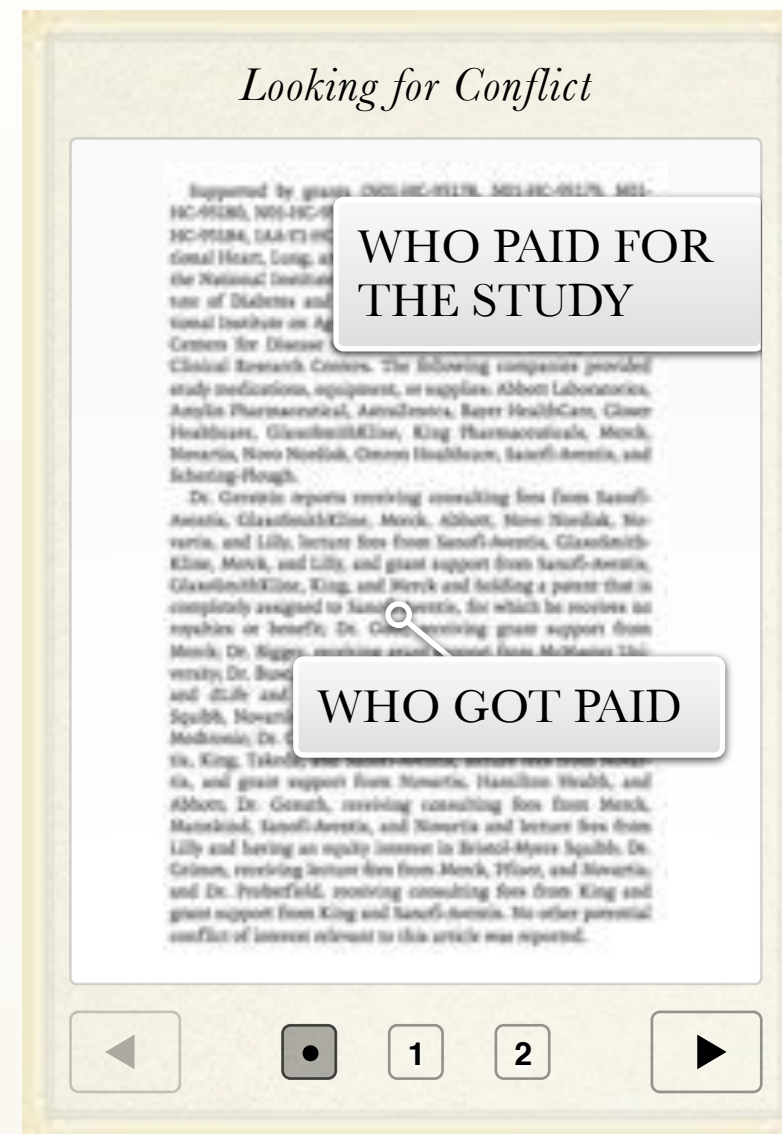
Lastly, you want to look for conflicts of interest - typically you can find this by jumping to the end of the paper – conflicts really have nothing to do with the "quality" of the study but you need to know and then you decide if you have any concerns – conflict of interest is a very individual assessment thing - the bottom line is disclosure needs to occur.

Now let's figure out what parts of the paper to basically IGNORE.

WHAT TO IGNORE
and
WHAT TO KEEP

"IGNORE" THE REST

The main trick to doing a 10-minute critical appraisal is to basically IGNORE much of the text in the article.

In general, as long as you can get the PICO from the abstract don't bother reading the INTRODUCTION - you already know the questions and you've convinced yourself this topic is important.

You can typically, unless you are doing a very detailed critical appraisal, skip the specific details of the METHODOLOGY and ignore the type of STATISTICAL TESTING done – it's usually done reasonably well – and if not you would need a degree in statistics to figure out if there were problems anyway. As long as you understand odds ratios, absolute numbers, relative numbers and confidence intervals - we'll get to that in a minute - you're good to go.

While the DISCUSSION will give you an idea of how the author plans to contextualize their findings it is just someone else's opinion of the importance of the results. The bottom line is it's YOUR opinion that counts.

To help inform your opinion you now need to now find out WHO was studied and then what happened to them.

LOOK FOR DIFFERENCES
and
WHO WAS IN THE STUDY
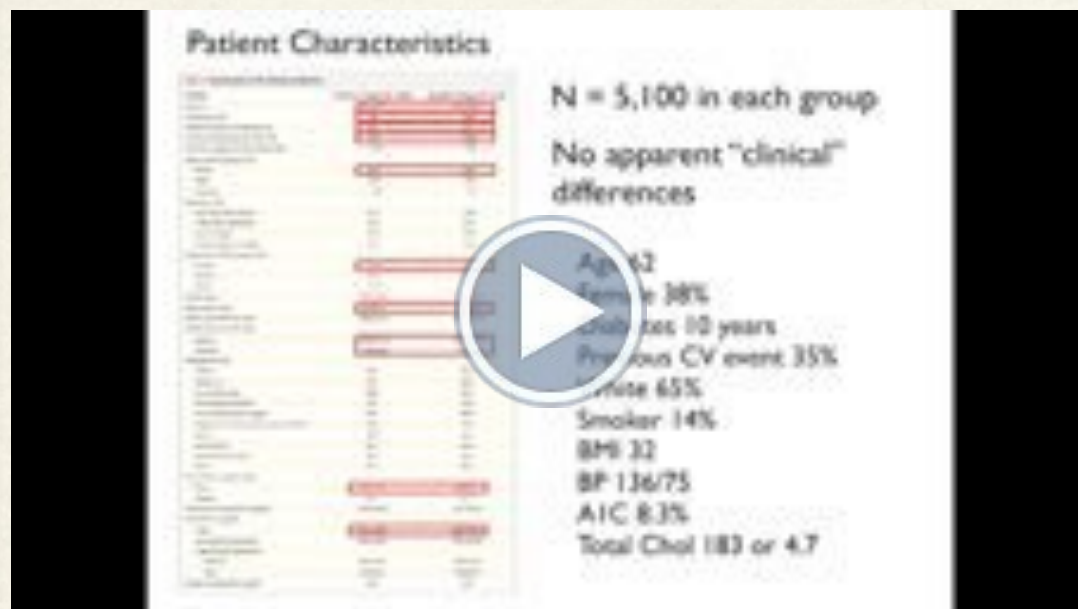
WHO WAS STUDIED

Typically baseline characteristics are found in Table 1 of a study. Here you'll find how many subjects were in each group. Now even if a study is randomized the groups will have numerically different baseline characteristics – so what you do is look at the specific characteristics and see if the GROUP DIFFERENCES are large enough to have led to the results seen. Statistics on baseline characteristics are unnecessary and really don't make sense – what you are looking for are clinically important differences between the groups and are they big enough to have played a role in the results of the study. For instance, let's say a particular study showed that a new form of birth control reduced the chance of a subject becoming pregnant. However if for some strange reason all the subjects in the new treatment group were male I'm sure you can figure out a potential reason for why a difference was seen.

After that you also use the table of baseline characteristics to give you an idea of WHO WAS STUDIED. Typically you look at average age, gender, ethnicity, previous medical conditions, lab values etc to describe the type of subjects enrolled. Now let's see WHAT happened to these subjects.

A SIMPLE LOOK AT THE OUTCOMES

A SIMPLE LOOK AT THE ADVERSE EVENTS

WHAT HAPPENED

To find out what happened to the subjects in each group you need to look for the table of results that lists the primary and secondary outcomes – this is often Table 2 but not always. Look for the OUTCOMES that are important to you or would likely be important to a patient. Typically these are outcomes like overall mortality, death, fatal or non-fatal MIs, fatal or nonfatal strokes, death from CV causes, heart failure, serious adverse events. Other things that effect quality of life like pain or other symptoms are also really important. Sometimes a number of these outcomes will be combined into what is called a PRIMARY OUTCOME. This doesn't necessarily mean this is the most important outcome but studies are typically designed around one outcome and the sample size and duration are based on the investigators best estimate of how frequently they think these events will occur. Endpoints are combined because studies typically will not be large enough to look at each endpoint separately and be able to find a difference. While these tables can often look complex they are fairly easy to understand. What you mainly need to know are the absolute numbers. In other words how many subjects in each group had an event.

This is calculated by taking the number of subjects who had the event and dividing it by the number of subjects in the group. If there were 1000 subjects in each group and 80 subjects in one of the groups had a heart attack that would

EXAMPLE	GROUP A	GROUP B	
Total # of subjects	1000	1000	2% absolute difference
Heart attacks	80	60	
Absolute %	8	6	25% relative difference

mean 8% had a heart attack. If in the other group 60 of the 1000 had a heart attack that would mean 6% in this group had a heart attack. The ABSOLUTE DIFFERENCE is what you really want to know as this tells you the magnitude of the difference between the groups which is why you did the study in the first place. In the example the absolute difference is 8% minus 6% or 2%. So there was a 2% absolute difference between the groups. This is often stated as a 2% ABSOLUTE RISK REDUCTION or ARR.

Now results are also very often expressed as relative numbers. These are the numbers you will often hear about on TV, or in advertisements. While relative numbers are correct they can be misleading if they aren't put into context. If you were to see something in a store that you wanted and it was 25% off, would you buy it? The very first questio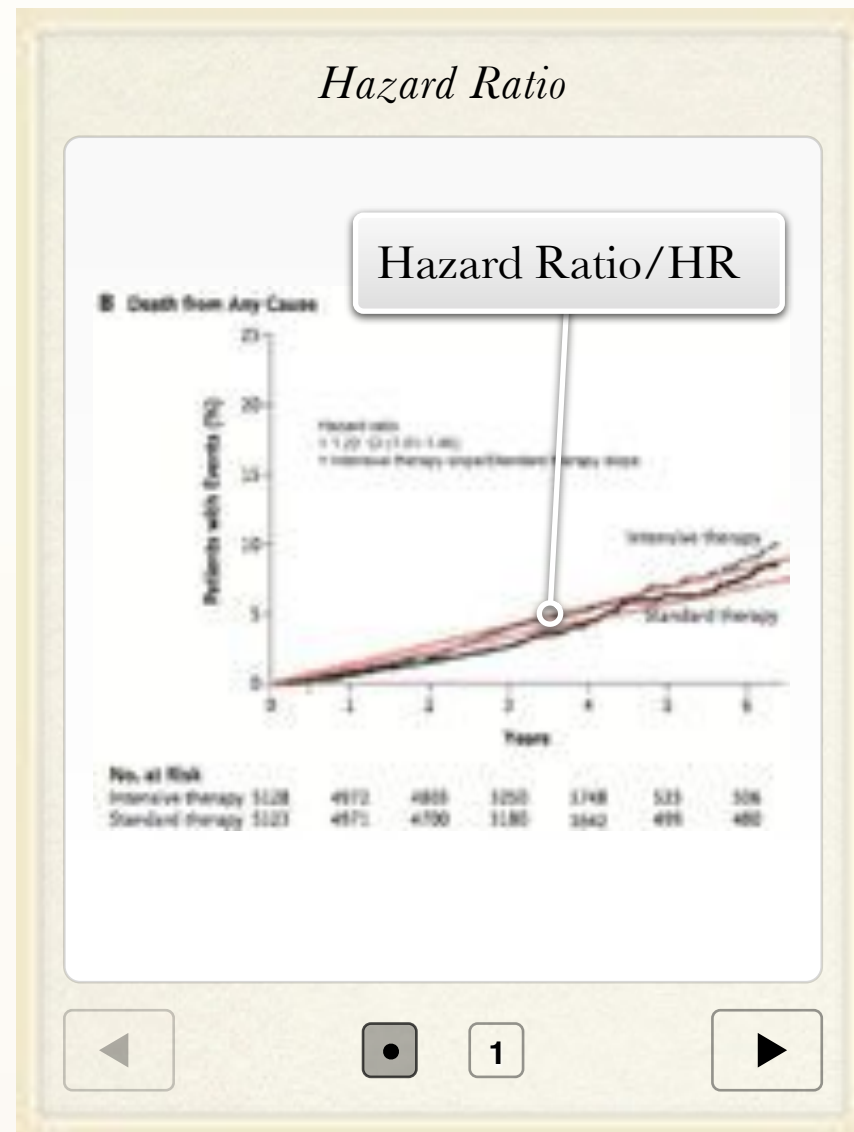n you would ask is "well what was the starting price". If you don't know that you have no idea as to the new reduced price. The same applies to clinical study results. If you hear that a certain treatment reduces heart attacks by 25% you need to ask yourself "well what was the chance of having a heart attack for the untreated group".

In the above example, the relative difference between the groups is 25% because 6% is 25% less than 8%. Or in other words if something cost $8 and it was on sale for 25% off it would now cost $6. This 25% is typically referred to as a relative difference and is often stated as a 25% RELATIVE RISK REDUCTION or RRR. A relative risk reduction of 25% would often be reported as a number like 0.75. That is because 0.75 is 25% less than 1.0. See why relative risks can be misleading if you don't know the "problem" in the first place. I bet you didn't know that 25% can equal 2%.
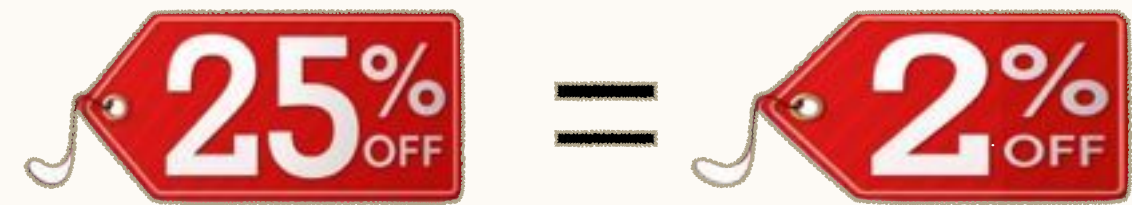
Sometimes the relative difference is reported as a HAZARD RATIO and this is used when where we are interested not only in the total number of events that occurred in the study but in their timing as well. Hazard is defined as the slope of the "survival" curve and the HR is the ratio of the 2 different slopes. The HR number is typically close to the RR but often not exactly the same. RRs are NOT based over time but rather are typically the risks over a specific set time period.

When we do a study we are typically sampling a population - just like a poll before an election. And just as with a poll, a study can only estimate what we think the effect is because we rarely study the whole population. If we did study the whole population (election results) we wouldn't calculate statistical comparisons but rather we would just look at the final statistical results for each group. However when we sample a population, just like a poll, all we can get is an "estimate of the true effect". Statistical numbers in studies are often presented with a 95% or a 99% confidence interval. Think of this as your +/- 4% number you hear about for polls - the "margin of error". Simply, a **CONFIDENCE INTERVAL** or **CI** range represents a plausible range of values for the actual effect. For instance if the CI for a HR of 0.75 was 0.70-0.80 you could be comfortable saying that the true difference is

Hazard Ratio

Hazard Ratio/HR

likely somewhere between a 30% (0.70) and a 20% (0.80) relative reduction.

By convention, when looking at ratios, if the CI doesn't include 1.0 then the difference is considered statistically different and we believe there is a real difference. If the CI does include 1.0, that doesn't mean there was no difference just that the difference is not considered a statistically significant difference. Some people think a CI means we are 95% confident that the result is in the interval, but this is not true. A CI is not a probability of the magnitude of an effect. It just means that if you did the study numerous times 95% of the CIs would contain the true result.

Now we haven't talked about p values. A p value is used to show how often a particular result or one more extreme will occur by chance alone if there really was no difference between the groups. A p value of 0.04 means that 4% of the time you would find a similar result or more extreme just by chance alone. As with a CI, by convention if a p value is less than 0.05 then the result is considered statistically different. Sometimes a p value of less than 0.01 is used as the breakpoint for statistical significance. Overall, confidence intervals are more useful because not only can you tell if something is statistically different it also gives you an appreciation of the range of the possible results.

So now that we have an idea of what good and bad things happened to the subject let's try and figure out if we should CARE.

The Outcome NUMBERS Explained

	INTENSIVE THERAPY A1C = 6.4%	STANDARD THERAPY A1C = 7.5%	HAZARD RATIO	HAZARD RATIO 95% CI	RELATIVE RISK
Primary outcome (%)	6.9	7.2	0.90	0.78-1.04	0.96
Death (%)	5.0	4	1.22	1.01-1.46	1.25
Non-fatal MI (%)	3.6	4.6	0.76	0.62-0.92	0.78
Non-fatal stroke (%)	1.3	1.2	1.06	0.93-1.48	1.08
CHF (%)	3	2.4	1.18	0.93-1.49	1.25

22% Relative Increase

Statistically Different

25% Relative Increase

PUTTING IT ALL TOGETHER

Putting It All Together

N=10,251 - 3.5 years

Age 62, Female 38%, Diabetes 10 years Previous CV event 35% White 65% Smoker 14%, BMI 32, BP 136/75, A1C 8.3, Total cholesterol 183

	Intensive therapy (A1c 7 6.4%)	Standard therapy (A1c 7.5%)	Hazard Rate	Hazard Ratio 95% CI
Primary outcome (%)	6.9	7.2	0.90	0.78-1.04
Death (%)	5.0	4.0	1.23	1.01-1.46
Non-fatal MI (%)	3.6	4.6	0.76	0.62-0.92
Non-fatal stroke (%)	1.3	1.2	1.06	0.75-1.50
CHF (%)	3.0	2.4	1.18	0.93-1.49
Hypoglycemia (%)	10.5	3.5		
Serious adverse event	2.3	1.8		
Weight gain >10kg	27.8	14.1		

Randomized Non-blinded Intention to treat Follow-up

WHO CARES

Now that you've figured out WHO was studied and WHAT happened to the subjects in the study. You now need to figure out the clinical importance of the results. To do that you can't use statistics and you don't need to ask a specialist. It ultimately comes down to what do YOU think about what was shown. Assume after your quick critical appraisal that the results seen were a NON-chance event and that BIAS was not responsible for the difference seen. Is the increase or decrease in outcomes clinically important?

In patients, who have had type 2 diabetes for 10 years, does knowing that intensive A1c control (6.4%) versus standard A1c control (7.5%) for 3.5 years leads to a

- 1% ⬆ death
- 1% ⬇ non-fatal MI
- 7% ⬆ hypoglycemia
- 0.6% ⬆ serious adverse events
- 14% ⬆ weight gain

change what you would recommend or what you would do if you were a patient?

If it does, GREAT, if not GREAT. The bottom-line is whatever your answer is it's the correct one because it's what you think of the results that matters. Evidence DOESN'T make decisions. It can only INFORM a decision. Evidence-based health care has been defined as "the integration of best research EVIDENCE with CLINICAL EXPERTISE and PATIENT VALUES". Critical appraisal just allows you to have an understanding of the best available research in a more complete way. It doesn't give you the answer.

Critical appraisal is all about making you a smarter and more informed health care professional. Having this skill helps you work with your patients to make them smarter and more informed patients so they can understand the benefits and harms of treatment and make their own decisions. You then support that decision regardless of what it is. It's all about shared-informed decision-making.

IN CONCLUSION, if you have followed the examples in this book you can now do all of that so much better because you took 10 minutes to critically appraise the latest and greatest study.

Thanks for reading and thinking about the messages in this book. Check here for an audio recording on this approach.

For more useful therapeutic information check out

medicationmythbusters.com

ALL YOU REALLY NEED

WHO WAS STUDIED

WHAT HAPPENED

SOURCES OF BIAS

◀   ● 1 2 3   ▶

An Audio Recording of a talk I gave on How to Do a Critical Appraisal of an RCT in 10 minutes

Who Helped With This Book

THANKS TO

All the people who over the years have helped me figure out how to look at clinical studies especially Marc Levine, Bob Rangno, Mike Allan and Peter Loewen

The Pharm D students who gave very useful feedback on earlier drafts of the book

Sam Gilchrist for creating the introductory video and some of the pictures

Hilary, Kieran and Callia for putting up with me