

What Is Evidence-informed Health Policymaking?

Slide 1: What is evidence-informed health policymaking? An orientation to evidence-informed health policymaking for the Oregon Health Evidence Review Commission. This presentation was originally presented by Dr. Martha Gerrity, and is narrated by Samantha Slaughter-Mason.

Slide 2: The first objective for this presentation is to define evidence-informed policy making. The second objective is to describe approaches to analyzing evidence to determine its quality. We will focus on several study designs, including systematic reviews, randomized control trials and observational studies. Third, we'll identify strategies to support the use of the evidence, including questions to ask about evidence, some implementation issues, and a few resources that might be helpful as you're reviewing information or having discussions with HERC members or in your subcommittees.

Slide 3: So what is evidence informed health policymaking? Evidence informed health policymaking is an approach to policy decisions intended to ensure that decision making is well informed by the best available research evidence. It is characterized by accessing and appraising evidence as an input into the policymaking process. It should be done systematically to ensure that relevant research is identified, appraised and used appropriately. And it should be done transparently so that others can examine the research evidence that informed decisions and the judgments made regarding the evidence and its implications.

Slide 4: The role of the evidence is to inform policy and practice. It's not to make the decision for you. It will inform the decision. Evidence is essential but not sufficient. Judgment is needed, including judgments about your confidence in the quality of the evidence and what to expect in specific settings, and about issues of equity and trade-offs. You can see the balance here. You have desirable effects of any intervention including health benefits, less burden, and savings, and you have undesirable effects such as harms, more burden, and more costs.

Slide 5: This is a recent example from 2011. You may have heard that the U.S. Preventive Services Task Force gave PSA screening a Grade D recommendation, which means that the task force recommends against the service. There's moderate or high certainty that the service has no net benefit or that the harms outweigh the benefits, and they discourage the use of this service.

This recommendation has created quite a bit of controversy. It's an example of a screening test that got into practice and was widely used before we really understood the impact of doing that type of screening. The studies that were done included people who were randomized to the offer of screening versus a control group of people not offered screening. They were followed forward for 10 to 15 years, and there was no evidence of a mortality benefit in those who were screened, but instead, an increase in harms associated with treatment. This is a case where a screening test got out into practice widely before we really understood what the implications were.

In this case, we can detect cancer earlier, but the issue then becomes, do you have adequate treatment that you know will do more good than harm?

Slide 6: If nothing else I want you to take home this particular image. This is adapted from an article by David Eddy who's a physician who's done a tremendous amount of work in medical decision making. There are two main components of a medical or health policy decision. The first component is the left hand box, the analysis of the evidence; it's the evidence inputs, the scientific judgment about the quality of the evidence. That then informs the next box.

What Is Evidence-informed Health Policymaking?

The next box is where groups such as yours come in. There are going to be value judgments that you're going to need to make with your knowledge of the population you serve and the preferences from that population and other factors that may influence your decision, that don't have anything to do with the evidence, but that should be informed by the evidence. Those two boxes together will form the basis for your decisions.

Slide 7: An essential characteristic of evidence-informed policymaking is that policymakers understand the systematic processes that are used to ensure that relevant research is identified, appraised for its quality and used appropriately

Slide 8: Any particular decision might go beyond the existing evidence, but to be called evidenced-based, the decision must at least be consistent with whatever evidence does exist.

Slide 9: So why do we need evidence-informed policy? Sir Iain Chalmers, who's done quite a bit of work in the area of health policy and the use of evidence in policymaking, notes that professional good intentions and plausible theories are insufficient for selecting policies and practices for protecting, promoting and restoring health. PSA is an example of professional good intentions. We wanted to decrease mortality from prostate cancer.

Slide 10: Again, why do we need evidence-informed policy? Would anyone prefer uninformed decisions about health care? Or the corollary question is: What other alternatives would we use to make these decisions? For example, some people use personal experience. A clinician may have had great outcomes, or at least the ones they remember were great outcomes, for a particular procedure. You may hear that type of personal testimony. Standard of care is another example, things that just become part of medical practice such as PSA screening for prostate cancer.

Slide 11: So would anyone prefer uninformed decisions about health care? You can't make an informed choice without information. If a decision is going to be well-informed rather than misinformed, you need reliable or trustworthy information.

Slide 12: Why should we care about evidence? I'm going to go through some examples of why informed judgments are important and where we have been misled because we didn't have the information.

Slide 13: This is an example of a device, terbutaline pumps. They are used to slow or stop preterm labor. These were advised against by the FDA in 2011. This is an example where a technology got out into use because it seemed like a good idea. There were several randomized controlled studies, not perfect, but suggesting there doesn't seem to be any benefit between terbutaline pumps and the comparison, the control treatment. But there were a few observational studies, which are not as high of quality study design, that suggested there might be some benefits. A number of Medicaid programs were being asked to fund terbutaline pumps, and this device started being used in a variety of health care settings. It turns out that people were reporting back to the FDA about adverse outcomes. Several systematic reviews were later conducted, and they found that there were no benefits when they looked across the observational and randomized controlled trials. The FDA received reports over a several year period that there were sixteen maternal deaths and about 12 severe cardiovascular events in the women that were on terbutaline pumps. Since there was no strong evidence of benefit and there was the potential, although we don't know the exact risk, for great harm, death in this situation, the FDA issued this black

What Is Evidence-informed Health Policymaking?

box warning. The MED project produced an evidence report on this topic several years ago for a group of Medicaid programs including Oregon. The report concluded that there was no evidence that suggested that terbutaline pumps do any good and advised against using them. Those Medicaid programs may have prevented potential bad outcomes among women.

Slide 14: This is another example involving hip replacement surgery. You have probably seen or heard about issues with metal-on-metal artificial hips. This is a slightly different situation where we didn't have studies because of the way the FDA approves new devices. This is a situation where there was not good evidence. There just wasn't research available. The FDA approved the metal-on-metal artificial hips based on the argument that these hip joints are similar enough to the old hip joints that they could go ahead and be put to use. Similar to the terbutaline situation, there were reports coming back to the FDA that there were people suffering from problems caused by the metal-on-metal hip joints. Later in the presentation I will discuss a study that formed the basis for this report on NPR about metal to metal being prone to early failure.

Slide 15: So which study do you believe? You could use the latest study, or the most cited study; this is often what people bring into you. You could cherry pick, which is something that we encounter when we review dossiers supplied by industry. They give us the studies that have the results that they would like us to see. You could do score keeping; In other words, there are six that favor and six that don't favor this particular intervention.

Slide 16: Or, what we would recommend is looking for studies that are systematic reviews, hopefully systematic reviews of randomized controlled trials. And if you can't find those, or those are not appropriate for the question that you are trying to answer, looking at systematic reviews of observational studies.

Slide 17: Systematic reviews are summaries of research evidence addressing a clearly formulated question using systematic and explicit methods to first identify, select and critically appraise relevant research. This helps you understand what the risk of bias is that might be inherent in these studies. Secondly, they collect and analyze data from the studies that are included in the review. Systematic reviews are a more appropriate source of research evidence for decision-making than the latest or most heavily publicized research study because this gives you a more comprehensive view of the body of research that has been done.

Slide 18: Why systematic reviews? First, you reduce the risk of bias in selecting and interpreting studies, so you avoid the cherry picking problem. Second, you reduce the risk of chance in identifying studies for inclusion or the risk of focusing on a limited subset of relevant evidence.

Slide 19: Third, they provide a critical appraisal of the research and place individual studies or subgroups of studies in the context of all the relevant research. There have been some notable examples where we didn't really understand the harms of a particular intervention until a group did a systematic review.

There was a group in Germany that did a systematic review about antidepressants and that's where the information about Paroxetine potentially increasing suicide rates among adolescents first came to light. A systematic review of all the studies using Avastin in various cancer treatments was completed, and provided enough numbers to identify potential harms, particularly increased mortality in women receiving Avastin for breast cancer.

What Is Evidence-informed Health Policymaking?

Fourth, systematic reviews allow you to critically appraise the judgments made in selecting studies and the collection, analysis, and interpretation of the results. So you will be able to read through the methods and determine if they did this well. There is one caveat, and that is garbage in and garbage out. So it may be that you have the best systematic review that could be done in an area of research, but the studies that were included in the systematic review may have been poor quality and have high risk of bias. That is the garbage in, garbage out. This is an integrated type of study design, hopefully saving you the hard work of going out and looking for all the studies yourself and appraising them, but you still have to be aware of the garbage in and garbage out problem.

Slide 20: Why randomized controlled trials? This is an example of the quality of evidence for menopausal hormone therapy before and after the Women's Health Initiative and the HERS study. At the time, Premarin was being marketed for prevention of coronary heart disease to post-menopausal women. In 1992 there were several well done systematic reviews of observational studies, which have greater risk of bias than well done randomized controlled trials. In spite of the fact that these observational studies were well done, the overall strength of evidence for each of the outcomes listed was low. In 2002 the Women's Health Initiative and HERS study was published. It was a very large randomized controlled trial and had a high strength of evidence. As you can see, the outcome for which the drug was being marketed to a large number of women – the suspected benefit for coronary heart disease – turned out to cause harm. Some of these other areas where there was suspected harm, such as increasing breast cancer mortality, were confirmed to cause harm. As these studies came into press and women were taken off menopausal hormone replacement therapy, the rates of death from breast cancer decreased.

Slide 21: So I am going to move on to randomized controlled trials. The previous slide showed a systematic review of randomized controlled trials on the right, and a systematic review of the observational studies on the left, to demonstrate why we should dig deeper to make sure that we are not into the garbage in, garbage out process. Next, I want to discuss the rationale for why we look first for randomized controlled trials when we want to know about a therapy, an intervention or treatment.

Randomization, or random allocation, creates groups equal on factors that might affect the outcome of interest. And so you can imagine, it's a flip of a coin whether you get the intervention or you get the control. As part of a well done randomized controlled trial, that process of allocating people to one group or another should be concealed so that the researchers can't tinker with it. There have been some notable examples of where researchers have been able to tinker with allocation. In the days before we were using computers, they used to put numbers in envelopes. So you would pull out the next envelope when you were entering a patient into the study and hand it to the person, or open it up and tell them which group they were assigned to, group A, or surgery group, or medication group. It turns out you could look through some of those envelopes, and so people tinkered with who got into the treatment group and who got into the comparison group.

Next, there should be systematic assessment and follow up for a randomized controlled trial and blinding and masking of outcome assessments when it is important. It's probably not important to blind or mask death because it is pretty hard to tinker with death as an outcome. But in some situations where you are asking about functional status or quality of life, it becomes important that the people doing the assessments don't know whether the person was in the intervention group or in the control group. In addition you should be able to follow up with people to the end of the study. A rule of thumb

What Is Evidence-informed Health Policymaking?

we use which isn't hard and fast, but is a reasonable rule, is that you should have at least 80% of people followed up. Otherwise, if you lose too many people, you lose that power of having two equal groups.

And then finally, people should be analyzed in the groups to which they were assigned. The jargon term for that is the intention to treat analysis. Again, if you don't do that, you lose the power of creating groups that are equal on things that could affect the outcome besides the intervention that you are interested in studying.

Slide 22: Why systematic reviews of randomized control trials? Because the randomized trial, and especially the systematic review of several randomized trials, is so much more likely to inform us and so much less likely to mislead us, it has become the "gold standard" for judging whether a treatment does more good than harm. This is from David Sackett, who is considered one of the fathers of evidence-based medicine.

Slide 23: In summary, the very important advantage of randomized controlled trials is that it minimizes known and unknown biases that can affect study outcomes. We may know about a factor that could affect a study outcome such as being female or being over 65 versus under 65. It could easily affect cardiovascular outcomes. However, there may be things that we don't know about. One example is the women's health initiative and menopausal hormone therapy, which we discussed earlier. It turns out that there are all sorts of hypotheses about why the observational studies got it wrong, that it didn't decrease the risk of cardiovascular events in women. One of the hypotheses that couldn't be measured is that maybe women who are interested in taking menopausal hormones also did other things that made them healthier, the healthy woman hypothesis. But that was unknown at the time those observational studies were done.

There are disadvantages to doing randomized controlled trials. The selected samples may not fit the population you are interested in. There are inclusion and exclusion criteria to get into a sample for a randomized controlled trial, and the researchers may be restrictive in ways that just won't make sense to how this intervention is going to be used in your population. Because of these things, they are less generalizable. They can be expensive. They are not practical for low prevalence conditions and/or rare outcomes. In studying some of the rare cancers, it would be hard to get enough individuals identified to even do a randomized controlled trial. And randomized controlled trials can't be used to study known harms or risk factors. You can't randomly assign someone to alcohol, or to drink and not drink, and then look at birth outcomes because it's a known harmful factor.

Slide 24: Although we could spend an hour on each of these different types of study designs, I will only be providing highlights. In a randomized controlled trial, you gather your sample, randomly assign them to an intervention and control, follow them forward in time and measure the outcomes. These are best used for questions about the effects of preventive or therapeutic interventions.

Slide 25: The other study designs are placed in a shared category called observational studies. Each of these designs has its own issues with risk of bias. The ones that you will likely encounter with this group are longitudinal cohort studies and case series. A cohort study is where you identify a group that meets explicit criteria. For example, in an observational study on vascularized bone grafting for avascular necrosis of the hip, you would identify all patients coming to an orthopedic practice that have this condition, and then follow that entire group forward in time. Some of them may get exposed to vascularized bone grafting. Others might not. The important thing would be to have that comparison

What Is Evidence-informed Health Policymaking?

and to follow them forward for the outcome: whether or not they need total hip replacement. That would be the ideal cohort study. However, we often see case series in the literature. For example, an individual orthopedic surgeon who has done vascularized bone grafting on 150 patients may report on these 150 patients. We don't know if these are all the patients he has done vascularized bone grafting on. There's no comparison group to determine if these people did any better than the patients who didn't have the surgery. These types of studies give much less information. They may give you some sense of whether the intervention is successful or not in terms of short term outcomes. But it becomes more difficult to judge whether the vascularized bone grafting really affected outcomes and if it delayed total hip replacement.

There are other risks of bias in case series. What if some of the patients who received the bone grafting decided it didn't work very well, so they did not go back to that provider? All of those patients would be missed in the case series. In using observational study designs, well done cohort studies are the ideal choice. Sometimes cohort studies are not available, and we must look at case series to determine whether there is any chance of benefit from the intervention. But you need to recognize that you're at great risk of coming to the wrong conclusion.

Slide 26: Randomized controlled trials versus observational studies. As I mentioned randomized controlled trials provide greater control, less risk of bias. They may be less generalizable. Observational studies have less control, greater risk of bias, and may be more generalizable. They can be used to study known harms or risk factors.

Slide 27: Now I'm going to discuss the key questions to ask regarding the topics you evaluate as a group. Some policymakers have found it useful to ask these questions of people who give public testimony at meetings. First, what is the evidence supporting whatever the assertion is? Are there randomized controlled trials? If not, why not? Is it ethical and feasible to do a randomized controlled trial? I already mentioned harmful interventions, for example, alcohol, as well as rare conditions and rare outcomes.

Slide 28: If there are randomized controlled trials, you should still consider these with a grain of salt because there are some known ways of manipulating these. One thing to ask when evaluating randomized controlled trials is: Are the comparators reasonable?

Slide 29: This is an example of a heart failure study comparing two different beta blockers used to decrease mortality and hospital readmission for heart failure. One is carvedilol and the other is metoprolol. The patients in this study were allocated to either carvedilol at 25 milligrams twice daily or metoprolol at 50 milligrams twice daily.

The company who funded this study knew that the carvedilol was delivered at the maximum dose recommended for heart failure. The metoprolol dose was delivered at half of what is recommended by the guidelines for heart failure. This study compared a full dose to half a dose. So it's not surprising that carvedilol was shown to decrease some of the events more than metoprolol.

Slide 30: Devices are not immune to this type of tinkering. This example is robotic surgery. The circle in the upper right hand corner is the claim that robotic surgery is less invasive, more precise, and has faster recovery. There's a little asterisk after the word "recovery." If you look at the circle at the bottom of this screenshot, you will see that the asterisk notes that this claim is for robotic surgery compared to open surgery. In some situations, open surgery is not the appropriate comparator because the more common

What Is Evidence-informed Health Policymaking?

procedure that's used for some types of surgery is laparoscopic surgery. That comparison wasn't done or at least was not cited.

Slide 31: Are the outcomes meaningful to patients? And, is the main outcome a composite of outcomes? I will illustrate both of these issues in the next example.

Slide 32: This is a study that examined the impact of intensive glucose control on type II diabetes compared with standard glucose control. The study reported on multiple outcomes, but the main outcome that was reported in the news was that intensive control decreased the combined outcome of macrovascular or microvascular events in diabetics. According to the study definitions, macrovascular events are the ones that people might care about: heart attacks, strokes, and deaths. According to the study data, there was a decrease in the relative impact on macro and microvascular events. However, there was only a 2% difference between the control groups. The question for us is: does that matter clinically? Or does that really matter to patients? By looking at the list of other outcomes, we can find out what drives that main outcome. Macroalbuminuria and new onset microalbuminuria are the outcomes that had the greatest impact on the composite outcome. These are outcomes that people do not feel or notice. They are surrogate outcomes. These results do not provide strong evidence to suggest that the intervention would prevent a macrovascular event.

Slide 33: The next two questions to ask when evaluating evidence are: who funded the study, and were there conflicts of interest?

Slide 34: This example is from a couple of years ago, and it is a study that evaluated the use of rosuvastatin, one of the statin drugs used to decrease the risk of cardiovascular events, to prevent major cardiovascular events in healthy older adults with elevated C-reactive protein levels who do not have hyperlipidemia. The goal was to identify people early before they have evidence of cardiovascular disease or diabetes that might put them at risk for these outcomes. Patients were randomly assigned to rosuvastatin or to a placebo. When the study results were published, it received a lot of press, noting that the study demonstrated a 50% reduction in the risk of heart attacks over a two year period. There were almost 18,000 people in the study. When you see a study of that size, you have to start wondering if they are trying to detect really tiny differences, and indeed, the differences were very small. The authors of the study knew this but didn't publicize the absolute difference in the two groups in these events.

The source of funding for this study was AstraZeneca, and the primary author was Paul Ridker. Shortly after this study was published, an article in the Wall Street Journal reported that Paul Ridker, who is at Harvard, is the developer of the C-reactive protein test. He had a patent and was making money from doing the C-reactive protein test. Right before the study was published, Paul Ridker sold the rights of C-reactive protein to AstraZeneca.

It also turns out that Paul Ridker had submitted this same study to the NIH for funding, but they turned him down. Their reasoning was that there were plenty of statins already in the market and they questioned the purpose of treating healthy individuals with normal cholesterol. Ridker also sent this proposal to several other drug companies making other statins. Those drug companies already had their market share of the statin industry. At the time, AstraZeneca had a new statin, rosuvastatin, that they were trying to market and thought this would be a good study to fund. They hoped the study would give them a niche in the market and publicize rosuvastatin.

What Is Evidence-informed Health Policymaking?

Slide 35: The FDA uses a different process for approving devices than for approving pharmaceuticals. The FDA requires clinical trials data for high risk devices, such as implantable cardioverter defibrillators. For moderate risk devices, companies only need demonstrate that the device is similar to one that's in current use. This is why metal-on-metal artificial hips got out into medical practice. Soon after, people began reporting adverse events.

Slide 36: This was the study that sparked the news article I showed earlier. It is an observational study out of the United Kingdom. Hopefully we will be able to do it here in the United States some time as well. It uses the National Joint Registry of England and Wales for primary hip replacement. The UK requires that information on use of newly approved devices be entered into the registry so that any of the outcomes from that device will be logged. This particular registry included about 400,000 primary hip replacements, of which about 31,000 were metal-on-metal.

The authors of this study were able to analyze the outcomes of metal-on-metal hip replacement and compare them to other types of hip replacement. They showed increased rates of failure of metal-on-metal hip replacements, and the length of time to the need for hip replacement was actually much shorter. These replacements were supposed to prolong that interval from 10 to 15 years, potentially up to 20, but it turns out that about a quarter of people were requiring replacement of these hips in a much shorter timeframe.

Slide 37: In order to implement the use of good evidence in decision making, you need a supportive organizational culture and values. There should be priorities set for obtaining evidence and skilled staff like those from HERC that understand the evidence and use evidence resources. There should be methods for assessing quality and applicability, and a process for using research to inform decisions, such as the clinical guidelines, health technology assessment, and coverage guidance development processes used by HERC. The State of Washington has an HTA process similar to that used by Oregon. Monitoring and evaluating policies and programs are other key elements of implementation.

Slide 38: I want to mention one free resource called PubMed Health. I've watched health policymakers use this resource in meetings as they're talking about a particular intervention. It's a new portal created by the National Library of Medicine for accessing high quality systematic reviews. It provides easy to read summaries for consumers as well as full technical reports from the partners listed. These partners are all known to do high quality evidence summaries. Some of the most well known are the Agency for Health Care Research and Quality and the Cochrane Collaboration. The VA has its own process similar to AHRQ. The German Institute for Quality and Efficiency in Health Care is another group doing things similar to NICE in the UK, which is also included, and finally, the Drug of Effectiveness Review Project, which is focused on drugs.

Slide 39: This is a screen shot of the PubMed Health web site. It has a quick and easy search function which searches on text words. You don't need to use special search terms. You will be able to quickly know whether any of those organizations have done systematic reviews on a topic you're interested in.

Slide 40: The final messages are that both policymakers and researchers must continue to struggle to help ensure that judgments about health policies are well informed by research evidence. The alternative is to acquiesce to poorly informed health policies.

What Is Evidence-informed Health Policymaking?

Slide 41: I would like to end this presentation with a final quote from Sir Iain Chalmers, "We will serve the public more responsibly and ethically when research designed to reduce the likelihood that we will be misled by bias and the play of chance has become an expected element of professional and policy making practice, not an optional add-on."