TITLE: INTERPRETATION OF EPIDEMIOLOGIC STUDIES: ASSOCIATION VS. CAUSATION

PRESENTER: Julie E. Buring, ScD

Slide 1: Today I am going to talk about the interpretation of epidemiologic studies and in particular the issue of association versus causation.

Slide 2: The problem is that what we see in epidemiologic research is a statistical association or a relationship between the disease or the outcome under study and an exposure. But what we are trying to do is evaluate whether that relationship is causal; causal meaning that the exposure plays an essential role in the occurrence of the disease, because our real goal is to hope that if we can identify this causal relationship, then an alteration in the exposure would lead to an alteration in the risk of disease.

Slide 3: Now, why is it such a problem in assessing causality in epidemiologic research, and a lot of it is due to something called the latent period. And the latent period is the time between when we accumulate sufficient exposure to actually the time when we have a manifestation of the disease.

We all know that what we had today for lunch is not going to cause colon cancer tomorrow. We know that what we have now for our diet is going to affect our risk of getting chronic diseases such as cardiovascular disease and cancer maybe five years from now, ten years from now, or 15 years from now when we are talking about something like cancer.

So it can be on the order of many years for a chronic disease and so much is happening during that period of time, besides just the diet that we are eating, that we have to try to tease out those threads before we are comfortable saying that it is this exposure, which is related to this outcome, that happened many years subsequently.

Slide 4: So how do we proceed? Let's look at it in a diagram first and then we will go back and do a little bit of the definition. Let's say we are interested in looking at the relationship of smoking in lung cancer, so we start out in a study where we have a 100,000 people and we just have sent them questionnaires and we ask them whether they smoke or not.

So we classified them as 40,000 of them are smokers and 60,000 of them are nonsmokers.

We follow them for a 10-year period of time and we observe 40 cases of lung cancer in our 100,000 people. 36 of those lung cancers are among the smokers and 4 of the lung cancers are among the nonsmokers.

Now, it doesn't take much for us to look at that and say, well, that really does look like there's an association between smoking and lung cancer. Of the 40 cases of lung cancer, 36 of them, the overwhelming majority of them, are among the smokers, and the smokers were not even the bigger group; there is only 40,000 of them compared to 60,000 nonsmokers. So we have the overwhelming majority of our outcome in a smaller of the groups. So it looks like there's an association.

Slide 5: But are we ready yet to say that it is the smoking that causes the lung cancer? Well, although we can say that it looks like there's an association, our first question is to say, is that association, the observed one in the study, a valid one? Is it a good estimate of the true association between the exposure and the disease? What else could have accounted for the findings besides the cigarette smoking?

Then after we do that we are going to ask a second question, if we do think there is a valid association, can we judge it to be one of cause and effect? And our third question then is going to be, to whom can we generalize the findings, can it be more than just 100,000 people in our study? To whom are we giving a message at this period of time?

Slide 6: So the first question is, is the association that we observe in our study valid? And to determine if an association observed in a particular study is valid we need to rule out the alternative explanations for the findings. We need to keep on asking the question, but what about this explanation? And the three alternative explanations we will always look at are chance, bias, and confounding.

Slide 7: The observed result may always be due to chance. It is always an explanation for any of our results that the association that we see can be random sampling variability. And the reason for that is because we are trying to make an influence to the entire population of the world that has that exposure or that outcome from just one study, and that study is a sample of people from the general population.

So it always can be that our ability to infer to a broader population can be limited by the sample that we chose to be in our study. Well, one common way to measure the effect of chance is by conducting a test of statistical significance.

Now, you will learn which test to do and how to do the test in your modules on biostatistics, but the big picture approach that we need to know for epidemiology is the following.

Slide 8: That we set up a null hypothesis, h naught, that nothing is going on, there is no difference between our smokers and our nonsmokers with respect to lung cancer. There is no association whatsoever. And we are going to test the alternative hypothesis that something is happening, there is a difference, there is an association between our exposure and our outcome.

We choose and conduct the appropriate test of statistical significance, but underlying every single one of these tests is the basic premise that we are going to compare what we observed in our study to what we would have expected to see in our study under the null hypothesis, where there's no association between the exposure and the outcome.

Slide 9: So for example in our study of smoking and lung cancer, we observed 40 cases of lung cancer; 36 in the smoking group and 4 in the nonsmoking group. What would we have expected to see?

Well, if there is no association between smoking and lung cancer then the number of cases of lung cancer that we observed should be divided between the smokers and the nonsmokers in proportion to how many smokers and nonsmokers there were in our study. So 40,000 of our population were smokers, therefore 40% of our cases of lung cancer should be smokers. So 40% of 40 is 16, so we would have expected to see 16 lung cancers in our smoking group.

60% of our nonsmokers -- 60% of the 100,000 people in our study were nonsmokers. That's 60%, so 60% of our lung cancer cases should have been observed among the nonsmokers. 60% times 40 is 24, where we would have expected to see 24 lung cancer cases among the nonsmokers.

The test of statistical significance will calculate the probability that I would see my 40 cases divided 36 among the smokers and 4 among the nonsmokers when I would have expected to see 16 of the cases among the smokers and 24 of the cases among the nonsmokers, observed to expected, taking into account the size of the study.

Slide 10: From that test statistic, the observed to the expected, taking into account the sample size, we actually calculate a p-value. The probability that the result seen in this study or a result more extreme would occur by chance alone, given that the null hypothesis is true.

And we are going to say that if that probability is small the probability of seeing the observed result, given what we have expected to see by chance alone, if that probability is small, then we are going to reject the null hypothesis that says that there is no association. We are going to say there is a statistically significant association between the exposure and the outcome.

And if the probability of seeing our observed data given the null hypothesis it did, we would expect to see it in high percentage of the time, then we don't reject the null hypothesis of no association and we say that the association is not statistically significant at the .05 level.

Slide 11: By convention the level at which we test this peak was .05, 1 out of 20, that is the usual, but arbitrary cutoff level for statistical significance. So now to make everything more specific, if the probability of seeing our results by chance alone, given that the null hypothesis is true, is less than 1 out of 20, then we conclude the chance of an unlikely explanation for the findings, we reject the null hypothesis that there is no association between the exposure and the outcome, and we say there is a statistically significant association between the exposure and the disease.

And if our p-value is greater than or equal to .05, so more than 1 out of 20 times, we would see our observed data by chance alone, if there were truly no association between the exposure and the outcome, then we conclude that chance cannot be excluded as an explanation for the findings, we do not reject the null hypothesis, we say there is not a statistically significant association between the exposure and the outcome.

Slide 12: Few things to remember about p-value, no p-value, I don't care how small it is, excludes chance. So even if p is less than one out of a thousand, one out of a million, it doesn't mean that chance

could not have accounted for your findings, because one out of a million times you will see that result by chance alone.

And no p-value however large mandates or means that a chance absolutely is an explanation for the findings, because even if the p-value is 0.9, that still means that 10% of the time it is not due to chance and you don't know whether your study is the one that was due to chance or not due to chance.

So what that means is that the p-value is not a rule, it is a guideline to an interpretation of the likelihood that chance is an explanation for the findings. And that means that we need all the information that we can get to interpret a p-value correctly.

We are always going to want to report the actual p-value, don't ever just say that the results are significant or nonsignificant at the .05 level. Never use that NF to just say that something is not significant, give the actual p-value. And if we could get additional information, it would be very important.

And the problem is that the p-value reflects both the strength of the association as well as the sample size of the study, and we know that because we said that we do a test of significance where we compare the observed to the expected; that's the strength of the association, and we take into account sample size; that's the variability.

To separate out these two components, we actually can't do it in a p-value, but we can do it in something which is called the confidence interval. So it is preferable to calculate the confidence interval in addition to the p-value, and we will do a later lecture on special issues that will talk more about what we can get from the p-value versus what we can get from the confidence intervals.

Slide 13: We also have to remember that the p-value only evaluates the role of chance. It says absolutely nothing about the other alternative explanations of bias and confounding or about causality. All it says is what we see is unlikely to be due to the luck of the draw, or sampling variability, or chance.

So for anybody to draw a conclusion in a paper regarding causality based on a p-value less than .05 only is totally incorrect. So you would never want to look at your data, to find a result that is statistically significant and say therefore that exposure causes that outcome. There is a whole bunch of steps that would be missing there; bias and confounding to know whether it's valid or not; and then looking at the valid association and seeing whether it's due to cause and effect.

Moreover, even if it is statistically significant, unlikely to be due to chance, it says absolutely nothing about its biological importance; in other words, whether this matters to you, the sort of the "so what?" way of looking at it.

And to give one example, someone published a paper that said that women who have had children are a quarter of an inch shorter than women who have not had children.

Well, that's fine, but nobody had any information or could cite anything to say that a quarter of an inch difference in height had any biological meaning. It was statistically significant because the size of the sample was so big, but it didn't pass the "so what?" test.

And sometimes we will try a new drug let's say for the control of hypertension, and we will look at the change in systolic blood pressure between two treatments and find that it's 1 millimeter of mercury. And people will say, all right, but 1 millimeter of mercury is not going to be worth it in terms of the fact that this drug is more expensive, the drug has more side effects. So even though it is a statistically significant 1 millimeter of mercury difference, it doesn't pass the so what test in terms of its clinical importance.

Slide 14: So now we have looked at chance, now it's time to look at bias, our second next alternative explanation for the data. And to do that before we use the word bias, we need to just go back and look at the word error.

This classification is a term that relates to error in the ascertainment of the study information. It is common in virtually all studies. You ask people for their height and weight, they don't get it right. You ask them for a recording -- you ask them to report their blood pressure and they are going to get it wrong, and you could measure it and the machinery may not give you the right answer.

Self-reported family history, what your diet has been over the last 20 years, there is going to be error. And if this error is non-differential, so it's random error and people are wrong regardless of whether they are sick or whether they are well, so the cases in controls, or whether they are exposed or not exposed to whatever you are looking at. All it does is makes the groups more similar and it will drive the estimate of the association towards the null value. It will dilute the association. It will make it harder for you to find a difference between the groups.

Slide 15: But the real problem is if the degree of error is differential. So there is a different degree of error in the exposed versus the non-exposed groups, or the diseased versus the non-diseased group, now it's something called bias. And bias, which is differential misclassification, can affect the estimate of the effect in either direction, can underestimate or overestimate the effect depending on the direction of the error, and sometimes you won't know what direction the error is going to be.

Slide 16: So again, bias as any source of systematic differential error in the determination and the association between the exposure and the disease, it will result in an incorrect, an invalid estimate of the measure of association, but unlike the non-differential misclassification that always biased you down towards the null value of no association, the systematic or differential error or bias could go either way. It could create a spurious association when no association is actually there, or could mask an association when an association truly is there.

So bias could really cause problems in both directions and it's going to be primarily introduced by the investigator or the study participants themselves.

Slide 17: Bias can arise in all study types. It occurs in the design and conduct of a study. It can be a prevented. It can actually to an extent be evaluated. But it cannot be overcome in the analysis space. There is no way that you are going to be able to fix it. And the keyword with respect to bias is the word different. You never want to do something differently in one of your groups versus the other groups. You don't want to bring them into the study differently and you don't want to collect information on them differently.

Now, there are books that have names of biases in them, there are hundreds of them, but really the most frequent biases occur either from the way the participants are brought into the study, selection bias; or the way information is obtained once the participants are in the study, which is observation bias.

Slide 18: So selection bias may result when the selection of the particular individuals to be included in the study is influenced systematically or differently by knowledge of their status regarding the other variable.

So in other words, if you bring your cases and your controls into the study differently, and you bring them in related to the exposure level that they have to what's of interest to you, you are going to have bias. Or if you bring the exposed and non-exposed people into the study and you bring them in differently, depending on whether they have got the outcome or not, that's going to be bias.

So if it's differently in the two groups and it's related to the question you are trying to answer, then you have the potential of bias.

Slide 19: So let me give an example of a selection bias and a case control study. So I do a hospital-based case control study, which means I take a group of people with venous thromboembolism, a group of people without thromboembolism, so that's cases in control and I go back and I ask them their oral contraceptive use, hospital-based case control study of oral contraceptives in venous thromboembolism.

Now, what if women with suspected venous thromboembolism are more likely to be hospitalized, if they report currently taking the pill, and how would that happen? Well, the woman -- venous thromboembolism is not an obligatory hospitalizer. The woman calls her doctor and says I have the following symptoms, the doctor says, all right, let me think about this for a second, let me just ask you one question, are you taking the oral contraceptive pill? And if the woman says, yes, I am, then the healthcare provider hospitalizes the woman.

But if the woman had the same symptoms but didn't report the use of the oral contraceptive, the healthcare provider might choose not to hospitalize them at that time.

So you are going to see cases of venous thromboembolism in the hospital that are put in the hospital because they are actually exposed to the oral contraceptive. You are basically building into the design of the study the hypothesis that you are trying to show.

Now, this can never be controlled in the analysis. We could do a subgroup analysis of those who are hospitalized with severe enough venous thromboembolism that they would have been hospitalized regardless of whether they used the pill or not, but that's sort of a different question. We weren't interested in severe enough disease and we won't be able to find it this way.

Might need a whole different design of the study, where rather than getting the patients from the hospital, we get them from the physician's office. So regardless of whether they use the pill or not, we will be able to pick up the fact that they contacted their physician about this concern.

Slide 20: What are the solutions? Well, basically little or nothing can be done to fix this bias once it has occurred, we just need to avoid it when we design and conduct the study by using the same criteria for

selecting the study group, maybe obtaining relevant subject records, getting high participation rates, and taking into account the diagnostic and referral patterns of the disease.

Slide 21: How about observation bias? Well, that results when there is a different level of accuracy or completeness of the information between the study groups. And what will happen then is participants will be systematically incorrectly classified as either exposed or non-exposed or diseased or not diseased. This occurs after the subjects have entered the study. It can affect the measure of association in either direction. And again, we can try to minimize it in the design of the study, but we cannot adjust for observation bias in the analysis of a study.

Slide 22: There are many types of observation bias; recall bias, interviewer bias, surveillance bias, loss to follow up, but all observation bias requires there to be a systematic differential misclassification, differential, different between the study groups.

Slide 23: Recall bias is one of the main ones that we are concerned about. Everybody forgets what diet they have had or their family history or whatever else we are asking them, but if people who have the outcome under study remember or report their exposures differently, either more or less accurately, doesn't matter, but differently than those without the disease, there is going to be observation bias.

This can result in an over or an underestimate of the measure of the association, depending on whether they deny it or whether they exaggerate their exposure. And the solutions are to try to do many things; use controls who are themselves sick, so they are thinking about everything that's happening to them just like the cases are; use standardized questionnaires that obtain complete information; use preexisting records; get information from objective records; or mask the subjects to the study hypothesis so they don’t know what it is that you are most interested in.

Slide 24: Interviewer bias is just the people who are asking the participants the questions, systematic difference in soliciting, recording, and interpreting the information.

So we have many examples of interviewers who probe for the correct answers because they think that smoking does increase the risk of lung cancer, so they may ask the person with lung cancer when they say, no, I never smoke, do you mean you never even smoked a single cigarette? And by probing the cases, but not doing that for the controls, we are going to build bias into the study.

The solutions are to mask the interviewers to the study hypothesis and to do the disease or exposure status of the subjects, to use standardized questionnaires or standardized preexisting methods of ascertaining the outcome or the exposure, but again, there won't be anything you can do about it when the study is in the analysis space.

Slide 25: Another thing that we worry about is follow up, especially if it's a cohort study or a randomized trial, we need to see what happens to people over time. And so the quality and the extent of the follow up becomes very important. And the times when it's worst is when there is a systematic difference in following and obtaining outcome information.

So if loss to follow up is different between the exposed and the non-exposed people and related to the outcome under study, then we have the concern that those who are exposed and got the outcome will

be more or less available to us, we will not be able to ascertain what happens to them, and that will bias our study results.

Slide 26: The solution is to keep the follow up rates as high as possible, low rates of loss to follow up, use standardized follow up procedures for all subjects. Get some backup contact information at the start of the study, so that even if the person in your study isn't answering your questions, you might be able to get them through a neighbor or a family member whose information the participants has given to you. And always know that we can get mortality information through the National Death Index in the United States, if that is a variable that's of interest to you.

Slide 27: So when we are interpreting study results, we really want to ask ourselves three questions. Given the conditions of the study, could bias have occurred? Are the consequences of the bias large enough to distort the measure of association in a meaningful way, and if so, what direction is going to be the distortions, towards the null value of no association or away from the null value, where it looks like there is more of an association than is actually there?

Slide 28: Chance bias, now we are looking at confounding. Confounding is a mixture of effects between the association under study and the third variable. This third factor, the confounder, has to be associated with the exposure under study, and independently of that it also has to be a cause or correlative of the cause of the disease. And the confounder may be responsible in part or totally for the association that is seen in the data.

Slide 29: So we have a diagram here that basically illustrates the principle of confounding. That we think we are looking at the relationship of exposure on the outcome, straight line up at the top, but instead we have a third variable present called the confounder, which is associated with the exposure, and independently of the exposure is a risk factor for the outcome.

And so rather than looking at the straight line between exposure and disease, we are looking at the relationship between the exposure and the disease through the presence of the confounding factor. And if we do not control this confounding factor, some or all of the relationship between the exposure and the disease which we observe in our study could be due to the effect of the confounding factor.

So for example, if we are looking at something like smoking and lung cancer, we would certainly want to look at age and gender, because smokers are more likely to be older men and nonsmokers in this study happen to be teenage girls, of course we are going to see more lung cancer among the smokers than the nonsmokers. Not necessarily because they are smokers, but because older men have more lung cancer than young teenage girls do.

So we have to get that age and gender out. We have to make it so that the smokers and the nonsmokers are alike with respect to those two variables before we can really assess the relationship between the exposure, smoking, and the outcome, lung cancer.

Slide 30: Now, if the potential confounder is associated with the exposure but not the disease, then it's not a confounder. So if we are looking at smoking and lung cancer, what do we do with alcohol? Well, alcohol is certainly related to smoking, there's a dose-response relationship between smoking habits and alcohol drinking. So we know there is a relationship between the exposure and alcohol.

On the other hand, if you do a full literature search, you will see virtually nothing that suggests that alcohol drinking is a risk factor for lung cancer. So although it is associated with the exposure, it is not independently a risk factor for the outcome under study. So it may be a potential confounder in our mind that we think about, but it will not be an actual confounder of the relationship.

Slide 31: And it goes the other way, if the potential confounder is associated with the disease but not the exposure, then it is not an actual confounder.

So for example, if our association was small to moderate drinking of alcohol, reducing risk of heart disease, so the risk factor is small to moderate alcohol drinking versus nondrinking, and the outcome is coronary heart disease and we think there is an inverse relationship. Well, what do we do with stress? We know stress is a risk factor for coronary heart disease, so shouldn't we take that into account?

On the other hand, we go ahead and we look at our drinkers and nondrinkers and there's absolutely no difference between the stress level of those who are moderately alcohol drinking and those who are not drinking at all. So stress is not associated with the exposure of alcohol drinking, therefore stress will not be an actual confounder of our association.

Slide 32: And then what do we do about factors that are actually not independent of the risk factor and the outcome, they are a mechanism by which the risk factor may be acting to increase or decrease the risk of the outcome? So this intermediate marker is what we are talking about now, a factor which represents the potential mechanism of action of the risk factor that is not a confounder.

So for example, if we were looking at obesity and stroke, what do we do with hypertension? Well, that is one of the mechanisms by which obesity could increase risk of stroke, it increases blood pressure therefore it increases risk of stroke. So if we controlled for it, then in fact we would be looking at the relationship of obesity and stroke over and above any mechanism related to hypertension.

So when we are looking at hypertension as the mechanism, we don't control for it. When we want to see whether obesity has an effect on stroke over and above hypertension, perhaps through diabetes, for example, then we control for hypertension, so we can look at the relationship of obesity and stroke through another mechanism, for example diabetes.

Slide 33: Well, what are potential confounders? We actually don't usually know what is associated with the exposure, like smoking or alcohol drinking, and we sure don't know the full mechanisms by which the exposure is postulated to affect the disease of interest, we don't always have that full amount of biological information. So in that case the very practical approach is just to identify all known risk factors for the disease as potential confounders and go ahead and collect information on those known risk factors for the outcome in the design of the study.

Slide 34: And the reason that we make such a point about that is that we absolutely have to anticipate confounding and record data on potential confounders and the design of the study. Because there are going to be ways that we can control confounding in the analysis, we will talk about that later in another special issues lecture. But we cannot adjust for anything in an observational study, case control or cohort, if we don't have information on that variable.

So we cannot control for age unless we have got information on age. We cannot control for stress unless we asked them some variables about stress. We cannot control for what we did not collect information. So that's why it's very important to think about the confounding in the design of your study, even if it turns out you are not going to control for it until the analysis.

Slide 35: So once we have looked at chance, bias, and confounding, we now turn to our second question, is the valid association that we just judged to be so valid, not due to chance, not due to bias, not due to confounding, is that association one of cause and effect?

So only once the alternative explanations of chance, bias, and confounding have been assessed can we move forward to try to conclude whether there is going to be a relationship that has cause and effect.

Causality is an issue of judgment. It must be based on the totality of all available evidence well beyond the one study that you did or that you are looking at.

Slide 36: And there are many sets of criteria for judging causality, but there are four positive criteria that are very commonly used in discussion sections of manuscripts or in grant applications to support a judgment of causality.

And these four are the strength of the association, the totality of evidence, and then we are also going to look at two more, which is dose-response and biological credibility.

So let's look at the first two for now. The strength of the association says that the stronger the association, the higher the magnitude of effect or the difference between the group, the more likely the association is to be causal. And that is because the bigger the association, the more it minimizes the chance of unsuspected confounders.

If we have a huge relationship between smoking and lung cancer, a 13 fold increased risk, for it to be due to confounding there would have to be a factor that is that strongly related to lung cancer and it differs big between our exposed and non-exposed, our smokers and nonsmokers, and we have never heard of that risk factor or we would have taken it into account.

So when we have a very big relative risk, we feel more assured. If we have a small relative risk, like a 1.3, a 1.4, there really could be a factor that we don't know about that we haven't measured that could account for a difference between the groups that's that small.

So a weak association, relative risk 1.1, 1.2, 1.3, can certainly be causal, but it's harder to prove, because it's harder to rule out the alternative explanation of confounding.

Now, totality of evidence or consistency is the criterion that I think is most important. Every study is going to make a mistake, as much as you try to do a good study or the investigators try to do a good study, there is going to be a problem in that study. But we don't all make the same problems, same errors, the same issues. So if other investigators studying different populations, using different methodologies, show similar results then that's a strong support for causality.

So I did my study and it was a case control study; you do yours and it's a cohort study. I did mine among mainly women; you did yours mainly men. Mine were mainly younger people; you did it in the elderly. I did it in a developed country; someone else does it in a developing country. And all of us are showing

the same trend of smoking increasing risk for lung cancer. That's when we begin to believe that we all made errors, but we didn't make the same errors, and it looks like the evidence, the totality of evidence is consistent enough to support that, that observed association is one of cause and effect.

Slide 37: The third criterion would be biological credibility. Does the association make sense? Does it make sense that smoking increases risk of lung cancer? If there is no biological mechanism that can be postulated, we have to know that we just have to turn back to our basic research colleagues and say, look, I am seeing this over and over again, can you figure out how smoking could be doing this? It may just be the limits of our current knowledge, but if there is biological credibility that will support our judgment of causality?

And finally, dose-response. If something is bad for us, we have a tendency to believe that more of it would be worst for us. So if we actually see that, that light smoking has less increased risk of lung cancer than moderate smoking, which has less than heavy smoking, and that those who have only smoked a couple of years have less increased risk of lung cancer than those who have smoked for 25 years, then that really increases our belief that this is a cause and effect relationship.

But there are problems, again, which is a dose-response relationship isn't always present for some chemicals, for example, it's really a threshold effect, and you have to get above a certain amount for you to see any effect whatsoever and there's no dose-response after that.

It could be that the variables that we are measuring to do dose are just so imprecise that we actually can't tell the difference between someone who is lightly exposed, moderately exposed, and heavy exposed. So if we don't see dose-response, it doesn't mean it's not cause and effect, but it does mean we have to think about it carefully and agree with, understand, and it must make sense that there is not a dose-response relationship.

Slide 38: So now we have done alternate explanations to see if we have a valid statistical association. Once valid we looked at some criteria to see whether it was a cause and effect relationship. And now we need to figure out to whom can we generalize our findings.

If the observed association is considered valid, true representation of the relationship between the exposure and the disease, other times called internally valid, then a related issue is whether the findings are generalizable or externally valid? In other words, to whom are they applicable?

Technically, study results are only applicable to the population in which the study is done, but none of us are going to want that to be true. We cannot repeat studies over and over again in every population, every age, every gender, every geographical location. We don't have the funds to do that. So we want to make a judgment to broaden our inference. And something that's very, very, very important for us to remember is that validity trumps or is more important than generalizability.

You have to have a true relationship between the exposure and the disease before you try to generalize it. You cannot generalize an invalid result. If the result is invalid, then it's generalizable to no one.

So it's very important that we always design our study or look at someone else's study from the standpoint of validity and then we can go ahead and talk about to whom it is generalizable, and figure out whether in any way if somebody really tries to generalize their findings widely, did they compromise

in any way the way that they did the study and the validity of the study results, because if they did, we have to take that into account when we think about generalizability.

And then after we figure out to whom we generalize, we think for a second about to whom is our message? Is it to fellow researchers? I saw this for the very first time, could my colleagues, could you look at it also?

Is it to healthcare providers? We think now we have a pretty consistent body of evidence, we think you should know about this.

Is it to the public? We are one step beyond the providers and now we are warning or informing people in the public. Or is it beyond all of that and in fact it's to the FDA or to some other regulatory agency?

So we have to look at validity, we have to look at cause and effect, we have to look at generalizability, and we have to figure out to whom we are tailoring our message, because that will make a difference in terms of how we write our paper.

Slide 39: So in summary, what we have tried to do in this little segment is to give you a framework for assessing statistical association in cause and effect relationships in epidemiology.

First, is there a valid statistical association? Can we exclude the role of chance, the role of bias, and the role of confounding? If there is a valid statistical association, can we judge it to be one of cause and effect, and to do that we will look at our positive criteria of strength of the association, totality of the evidence, biological credibility, and dose-response relationship.

And finally we say, okay, it's valid. Okay, it's cause and effect. To whom can we generalize the finding and what are the public health implications of this finding, for whom and to whom should the message be going?

Slide 40: Thank you very much!