Slide 1: Introduction

Hello and welcome to Module 2 – Fundamentals of Epidemiology. This presentation will focus on issues of interpretation in epidemiologic studies. My name is Jeffrey Bethel, Assistant Professor of Epidemiology at East Carolina University, Brody School of Medicine.

Slide 2: Acknowledgements

APTR wishes to acknowledge the following individual that developed this module:

- Jeffrey Bethel, PhD
  Department of Public Health
  Brody School of Medicine at East Carolina University

This education module is made possible through the Centers for Disease Control and Prevention (CDC) and the Association for Prevention Teaching and Research (APTR) Cooperative Agreement, No. 5U50CD300860. The module represents the opinions of the author(s) and does not necessarily represent the views of the Centers for Disease Control and Prevention or the Association for Prevention Teaching and Research.

Slide 3: Presentation Objectives

There are four objectives for this presentation. They are to describe the key features of information and selection bias; identify the ways selection and information bias can be minimized or avoided; implement the methods for assessing and controlling confounding; and identify uses of the Surgeon General’s Guidelines for establishing causality.

1. Describe the key features of selection and information bias
2. Identify the ways selection and information bias can be minimized or avoided
3. Implement the methods for assessing and controlling confounding
4. Identify uses of the Surgeon General’s Guidelines for establishing causality
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Transcript

Slide 4: Hierarchy of Study Designs

The goal of public health as well as clinical medicine is to modify the natural history of disease, to decrease morbidity and mortality. So how do researchers and public health practitioners do this? They carry out formal studies to identify risk factors for various diseases or other health outcomes. So here is an informative figure showing the hierarchy of various epidemiologic study designs. The goal of all these studies is the same and that is to identify exposures or characteristics that are associated with disease or other health related outcomes.

Slide 5: Linking Exposure to Outcome

Here is a figure depicting this goal of linking exposures to outcomes. The result of these formal epidemiologic studies is observed associations. That is associations between exposures or characteristics and diseases or outcomes.

Slide 6: Linking Exposure to Outcome

After the study is carried out, and a disease is linked to an exposure or outcome, you need to ask if the observed association is biased confounding or causal. We’re going to go through each one of these items individually. And we’re going to start with bias first.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Transcript

Slide 7: Bias

So bias is any systematic error in the design, conduct, or analysis of a study that results in a mistaken estimate of an exposure’s effect on the risk of disease.

Slide 8: Bias

Now these errors or bias can make it appear as if an exposure is associated with an outcome when the exposure and the outcome are really not associated. That is also thought of as biasing away from the null. The errors in the bias can also mask an association where there really is an association. And that is called biasing towards the null. There really is an association however a bias estimate makes it appear as though there is no association. As we talked about in the definition of bias, bias is primarily introduced by the investigators or participants, in the design, conduct, or analysis of the study. And we’re going to talk about two forms of bias in a study; selection and information bias.

Slide 9: Selection Bias

Selection bias occurs when the methods that investigators use to identify study participants results in a mistaken estimate. That is the estimate that investigators generate is different than what would have been obtained from the entire population targeted for the study. Again this gets back to a systematic error made by the investigators. Specifically the error was made in selecting one or more of the study groups that are being compared in the epidemiologic study.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

**Selection Bias Examples**
- Control selection bias
- Self-selection bias
- Differential referral, surveillance, or diagnosis bias
- Loss to follow-up

**Slide 10: Selection Bias Example**

And we’re going to talk about four types of selection bias. Control selection bias, self-selection bias, differential referral, surveillance, or diagnosis bias, and loss to follow-up.

**Slide 11: Selection Bias in a Case-Control Study**

Now here is an example in which selection bias may occur in a case-control study. Specifically control selection bias. So the research question was, do Pap smears prevent cervical cancer? So in this study cases were diagnosed at a city hospital. Controls were randomly sampled from households in the same city by canvassing the neighborhood on foot. So here was the observed association. Here are the data from that hypothetical study. Women who were cases and controls, and women who received Pap smears and no Pap smears. So 250 cases, 250 controls, 100 of the cases received a Pap smear and 100 of the controls received a Pap smear. The resulting odds ratio was 1.0. That is there is no association between Pap smears and risk of cervical cancer.

**Slide 12: Selection Bias in a Case-Control Study**

However, cases came from the hospital and controls from the neighborhood around the hospital. So here is where the bias comes in. Only controls who were at home during recruitment for the study were actually involved in the study. Now these controls who were home while the neighborhoods were being canvassed were less likely to work and less likely to have regular checkups and Pap smears. Therefore, inclusion in the study as a control was not independent of the exposure. That is, being included in the study as a control was dependent on the exposure, in that case Pap smear.
Selection Bias in a Case-Control Study

Question: Do Pap smears prevent cervical cancer? Cases diagnosed at a city hospital. Controls randomly sampled from households in the same city by canvassing the neighborhood on foot. Here is the true relationship:

<table>
<thead>
<tr>
<th></th>
<th>Cases</th>
<th>Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pap Smear</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td>No Pap Smear</td>
<td>150</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>250</td>
<td>250</td>
</tr>
</tbody>
</table>

\[ \text{OR} = \frac{(100)(100)}{(150)(150)} = .44 \]

56% reduced risk of cervical cancer from among women who had Pap smears as compared to women who did not (40% of cases had Pap smears versus 60% of controls), the unbiased estimate. The odds ratio of 1.0 was the biased estimate.

Selection Bias in a Case-Control Study

Self-Selection Bias

- Refusal or nonresponse by participants that is related to both exposure and disease
  - e.g. if exposed cases are more/less likely to participate than participants in other categories
- Best way to avoid is to obtain high participation rates

Selection Bias in a Case-Control Study

Differential Surveillance, Diagnosis or Referral

- Example related to exposure
- CC study: venous thromboembolism (VT) and oral contraceptive (OC) use
- Cases: 20-44 yo, hospitalized for VT
- Controls: 20-44 yo, hospitalized for acute illness or elective surgery at same hospitals
- Result: OR = 10.2

Slide 13: Selection Bias in a Case-Control Study

So here are the data for the true relationship. Again, 250 cases, 250 controls. 100 of the cases received a Pap smear, 150 of the controls received a Pap smear. The resulting odds ratio is 0.44. That is there is a 56% reduced risk of cervical cancer among women who received a Pap smear. This is the true relationship. This is the unbiased estimate. The odds ratio of 1.0 was the biased estimate.

Slide 14: Selection Bias in a Case-Control Study

Another form of selection bias is self-selection bias. This occurs when refusal or non-response to being included in the study is related to both the exposure and the disease. For example, if exposed cases were more or less likely to participate than participants in the other categories that can result in self-selection bias.

The people, either cases or controls, were more or less likely to participate. Now the best way to avoid self-selection bias is to maintain high participation rates in all groups or categories of participants in the study.

Slide 15: Selection Bias in a Case-Control Study

Now here is differential surveillance, diagnosis or referral bias. So here is an example related to exposure. In a case-control study looking at venous thromboembolism (VT) and oral contraceptive (OC) use, cases were 20-44 year old women hospitalized for VT. Controls were the same age group hospitalized for acute illness or elective surgery at the same hospitals. In this instance hospital controls were used. The resulting odds ratio from this study was 10.2. That is OC use was highly associated with venous thromboembolism.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Transcript

Slide 16: Selection Bias in a Case-Control Study

However, authors acknowledged that this high odds ratio might be due to the bias in criteria for hospital admission. That is who was admitted. So previous studies had liked VT to OC use. Therefore there was a tendency to hospitalize patients based on exposure status of oral contraceptive use and that led to a stronger association between VT and OC than what truly existed.

Slide 17: Selection Bias in a Cohort Study

Here’s an example in the form of selection bias in a cohort study that is loss to follow-up. One study compared HIV incidence rates among IV drug users (IVDU) in New York City during a six year period using 10 incidence studies. The HIV rates ranged from 0 to 2.96 per 100 person-years. And these were well below the incidence rates in NYC from the late 70s and early 80s to mid-80s and the early 90s. Now this study resulted in funding cuts to drug treatment and prevention programs because this study showed that HIV incidence rates were much lower than they were in the 70s and 80s.

Slide 18: Selection in a Cohort Study

Now the question is was this decline in HIV incidence rates real? Closer inspection of these 10 individual cohort studies found that follow-up rates ranged from 36% to 95% and only 2 of the 10 had follow-up rates greater than 80%. As we know, loss to follow-up can introduce bias into the study. In this case only 2 of the 10 had a high follow-up rate. Also, sample sizes ranged from 96 to 1,671 in these 10 studies. Now the solution to loss to follow-up is to simply minimize loss to follow-up. That involves implementing methods and procedures in your study design to minimize loss to follow-up. You could maintain accurate contact information for the participants, and contact information for people who will know where your participants are so as you do not lose them.
Slide 19: Selection Bias – Can We Fix It?

So selection bias, can we fix it? Can we do anything about it? The simple answer is no, you simply need to put in measures into your design and conduct the study to avoid it in the first place. For example you need to use the same criteria for selecting cases and controls. Obtain all relevant participant records. That is, maintain complete data on everyone. Obtain high participation rates. That is, minimize loss to follow-up. And you need to take into account diagnostic and referral patterns of the disease of interest, or the diseases that you are using for your hospital controls.

Slide 20: Information Bias

The second major type of bias we’re going to talk about is information bias. That arises when from a systematic difference in the way that exposure or outcome is measured between various groups under investigation. Just like selection bias, information bias can bias your results towards or away from the null. That is, mask an apparent association or create a spurious one. Information bias can occur in prospective as well as retrospective studies. And the two types we’re going to talk about include recall and interviewer bias.

Slide 21: Recall Bias in a Case-Control Study

Here’s an example of recall bias in a case-control study. In this example it is a case control study examining birth defects. Controls in this study were healthy infants. Cases were infants with birth defects. Exposure data was collected at postpartum interviews with the infants’ mothers. Now in this situation there could be a differential level in the accuracy of the information provided by the cases and controls which could result in an over or underestimate of the measure of association. So the data collected from the cases could be more accurate than the data collected from the controls. That is the mothers of the cases have been pouring over their brains for the history of various exposures during pregnancy which may have contributed to their infants being malformed. Whereas controls, they have healthy infants, they haven’t really spent a lot of time reviewing their history during pregnancy. Or conversely, cases may underreport
their history particularly if it is a socially sensitive exposure such as alcohol consumption during pregnancy.

**Slide 22: Methods to Minimize Recall Bias**

Now what are the methods to minimize recall bias? In the context of a case-control study you could use a diseased control group. So as the participants in the diseased control group have some disease unrelated to the outcome of interest could also be pouring over their history of exposures. You could use and design a structured questionnaire. You could use a self-administered questionnaire so as to make it easier for participants to recall sensitive exposures such as consuming alcohol or smoking during pregnancy. You could bypass the participant in terms of self-report and use biological measurements. It is tough to get around socially sensitive exposures. And lastly you could mask the participants to the study hypotheses. So as to illicit more accurate information from the study participants.

**Slide 23: Interviewer Bias**

The second type of information bias we’re going to talk about is interviewer bias. That results from a systematic difference in soliciting, recording, and interpreting of data collected during a study. In a case-control study this could occur when exposure information is sought or collected differently between cases and controls. In a cohort study, outcome or disease information is collected differently between the exposed and unexposed individuals. There is a simple solution to this and that is to simply mask the interviewers or the people collecting the data as to the disease status or exposure status of the participants. You can also use standardized questionnaires or standardized methods of outcome or exposure ascertainment.
Slide 24: Association

So that covers the examination of whether our observed association is biased or not. The next question we’re going to ask about our observed association is whether this association is confounded or not.

Slide 25: Confounding

Here is a diagram of a possibly confounded relationship between A and B, confounded by the variable X. Confounding is really a mixing of effects. That is the association between exposure and the disease of interest is distorted by the effect of a third variable that is also associated with the disease as well as the exposure. You can also think about confounding as an alternate explanation between A and B in this diagram. That is an alternate explanation of the observed association between the exposure of interest A and the disease of interest B.

Slide 26: Criteria for Confounding

There are three criteria for determining whether a variable is confounding the relationship between A and B. So in order for this factor or variable X to be a confounder, all of the following must be true. This factor must be associated with the Disease B. That is it’s a risk factor or a preventive factor. This factor must also be associated with Factor A. That is the exposure. Finally this factor must not be a result of Factor A. That is, it must not lie on the causal pathway from A to B. Or it may not be a result of A.
Here’s an example of a confounding relationship in a previous study that was trying to link coffee consumption with pancreatic cancer. And they thought that perhaps smoking was confounding the relationship between coffee consumption and pancreatic cancer. So they went through the three criteria to see if smoking was a confounder. Criterion 1: is this confounding variable, smoking, associated with the disease of interest or outcome, pancreatic cancer? And they found in their data that yes, it was. Smoking was a risk factor for pancreatic cancer. Next, was smoking associated with the exposure or coffee consumption? And they found in their data that yes, people consuming coffee were more likely to also smoke. And looking outside their data, they knew that smoking was not a result of coffee drinking. That smoking is not on the causal pathway from coffee consumption to pancreatic cancer. All three criteria were satisfied therefore they concluded that smoking was indeed confounding the relationship between coffee consumption and pancreatic cancer. It was an alternate explanation between coffee consumption and pancreatic cancer.

So what is the impact of confounding? It is similar to bias. You can pull the association away from the true association in either direction. So in positive confounding the confounded estimate or the confounded variable exaggerates the true association. So in the context of a cohort study, the true relative risk is 1. That is the exposure is not associated with the outcome. However, the confounded relative risk could be 2. It is exaggerating the true association. There is also negative confounding when it hides the true association. For example, in a cohort study the true relative risk could be 2. That is exposure is related to the outcome. However, due to confounding, the relative risk is 1. There is no association between the exposure of interest and the disease. Again, it is similar to bias.
Slide 29: Hypothetical Cohort Study of Obesity and Dementia

So now let’s work through a hypothetical cohort study examining the association between obesity and dementia. So a hypothetical study was conducted among 2,000 people looking at obesity and dementia. Here are the results of the data. We calculated the relative risk of 4.0. That is the incidence of dementia among the obese participants was 4 times the incidence of dementia in the non-obese participants. So obesity was associated with dementia.

### Hypothetical Cohort Study of Obesity and Dementia

<table>
<thead>
<tr>
<th></th>
<th>Dementia</th>
<th>No Dementia</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obese</td>
<td>400</td>
<td>600</td>
<td>1,000</td>
</tr>
<tr>
<td>Not Obese</td>
<td>100</td>
<td>900</td>
<td>1,000</td>
</tr>
<tr>
<td>TOTAL</td>
<td>500</td>
<td>1,500</td>
<td>2,000</td>
</tr>
</tbody>
</table>

Relative Risk = \( \frac{400}{1,000} / \frac{100}{1,000} = 4.0 \) (crude measure)

Slide 30: Association Between Obesity and Dementia

So now we should ask ourselves, was age confounding the association between obesity and dementia?

### Dementia and Diabetes Cohort Study

<table>
<thead>
<tr>
<th>Age</th>
<th>Dementia</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>80-99 Years</td>
<td>400</td>
<td>600</td>
</tr>
<tr>
<td>45-79 Years</td>
<td>100</td>
<td>900</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1,000</td>
<td>1,000</td>
</tr>
</tbody>
</table>

Relative Risk = \( \frac{400}{1,000} / \frac{100}{1,000} = 4.0 \)

Slide 31: Dementia Diabetes Cohort Study

So we can look at our data and look at Criterion 1 first: Is age associated with our outcome, dementia? We can summarize our data in a 2x2 table and calculate relative risk and we get 4.0. That is age, and in this case older age, was associated with dementia.
Slide 32: Dementia and Diabetes Cohort Study

Criterion 2: Was age, the potential confounder, associated with our outcome obesity? Here are the resulting data. In examination of this association we found out that yes, age is highly associated with obesity.

<table>
<thead>
<tr>
<th>Age</th>
<th>No</th>
<th>Yes</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>80-99 Years</td>
<td>100</td>
<td>900</td>
<td>1,000</td>
</tr>
<tr>
<td>45-79 Years</td>
<td>900</td>
<td>100</td>
<td>1,000</td>
</tr>
<tr>
<td>TOTAL</td>
<td>1,000</td>
<td>1,000</td>
<td>2,000</td>
</tr>
</tbody>
</table>

Odds ratio = \( \frac{900 \times 100}{100 \times 900} = 81.0 \)

Slide 33: Were Criteria for Confounding Satisfied?

So were the criteria for confounding satisfied?

- **Criterion 1:** Age was associated with dementia.
- **Criterion 2:** Age was associated with obesity.
- **Criterion 3:** Age is not a result of obesity (not from data).

Age is not a result of obesity. Again that is not from the data in the study. All 3 criteria were satisfied. Therefore, yes age was confounding the observed association between obesity and dementia. It just so happens that the obese participants in the study were older and that was what accounted for the observed association with dementia.

Slide 34: Controlling Confounding

So really, confounding variables are nuisance variables that get in the way of the relationship that we really want to study. So how do we control for it? We control for confounding in the design and analysis phases. In design phase we can group or individually match on the suspected confounding factor. This is specifically regarding a case-control study where we can match cases and controls on a variable such as age to make them similar so they do not differ. In the analysis we can stratify our analyses based on this confounding variable. Record our data stratified by age for example, young and old. We can standardize our data. Or what’s most commonly done is conduct multivariate analysis to adjust for the confounding variable.
Thoughts on Confounding

- Not an error in the study
- Valid finding of relationships between factors and disease
- Failure to take into account confounding IS an error and can bias the results!

Slide 35: Thoughts on Confounding

Some final thoughts on confounding. Again confounding variables are these nuisance variables that get in the way of examining the relationship that we really want to study. Unlike bias, it is not an error in the study. So these confounding relationships are really valid findings of relationships between factors and a disease. However, failure to take into account confounding or to control for confounding is an error and will bias the results of your study.

Slide 36: Association Between Exposure and Outcome

So we’ve assessed whether the observed association is biased or confounded. Let’s finally look at whether our observed association between our exposure and our outcome is causal.

Epidemiologic Reasoning

- Determine whether a statistical association exists between characteristics or exposures and disease
  - Study of group characteristics (ecologic studies)
  - Study of individual characteristics (case-control and cohort studies)
- Derive inferences regarding possible causal relationship using pre-determined criteria or guidelines

Slide 37: Epidemiologic Reasoning

First let’s talk a little about the epidemiologic reasoning. The primary goal is to determine whether there is a statistical association between our exposure or characteristic and our disease. We assess that through various studies. We can study group characteristics in an ecological study, or we can study individual characteristics in a case-control, cross-sectional or cohort study. The key here is deriving inferences regarding possible causal relationships using pre-determined criteria or guidelines. Again we have this observed association, now we need to infer whether this association is causal or not and we can use pre-determined guidelines to do this.
Causation

- Association is not equal to causation
- Consider the following statement: If the rooster crows at the break of dawn, then the rooster caused the sun to rise
- Causation implies there is a true mechanism from exposure to disease

Slide 38: Causation

However, association between an exposure and an outcome is not equal to causation. Consider the following statement: If the rooster crows at the break of dawn every morning, then the rooster caused the sun to rise. Again these are associated but they’re not causal. Causation implies there is a true mechanism from the exposure to the disease. More than just an observed association.

Koch-Henle Postulates (1880s)

1. The organism is always found with the disease (regular)
2. The organism is not found with any other disease (exclusive)
3. The organism, isolated from one who has the disease, and cultured through several generations, produces the disease (in experimental animals)

Slide 39: Koch-Henle Postulates (1880s)

So a little history lesson on causation. In this instance in the context of infectious disease. Jacob Henle and his student Robert Koch in the 1880s built a set of postulates on disease causation based on the germ theory. There were 3 postulates. 1: the organism is always found with the disease. This is also known as the regular postulate. Second: the organism is not found with any other disease. That is, it is exclusive. Third, the organism, isolated from one who has the disease, and cultured through several generations, produces the disease. That is, it is reproducible in experimental animals.

Koch added that “Even when an infectious disease cannot be transmitted to animals, the ‘regular’ and ‘exclusive’ presence of the organism (postulates 1 and 2) proves a causal relationship.”

Unknown at the time of Koch-Henle (1840-1880)
- Asymptomatic infection
- Multifactorial causation
- Biologic spectrum of disease

Koch-Henle Postulates (1880s)

Koch added that even when an infectious disease cannot be transmitted to animals, (the third postulate) the ‘regular’ and ‘exclusive’ presence of the organism (the first two postulates) can prove a causal relationship. However, unknown at the time of these postulates in the mid 1800s, was the idea of a carrier state, the idea of asymptomatic infection, of multifactorial causation, or even the biologic spectrum of disease. Those were not considered at this time. So causation needed to mature from this idea of these three postulates.
Slide 41: Understanding Causality

Let’s say you’ve determined there is a real association between our exposure and our outcome. We believe it to be causal. That is we’ve ruled out bias and confounding. Have we really proven causality?

Slide 42: Surgeon General’s Guidelines for Establishing Causality

So have you proven causality if you’ve ruled out bias and confounding? Now the Koch-Henle postulates are not applicable to chronic disease epidemiology. But Doll and Hill in the 1950s were trying to link smoking to lung cancer through a series of case-control and cohort studies. And these led to the development of the Surgeon General’s Guidelines for establishing causality. These have since been through one round of revisions since the 1950s. Here are the nine guidelines.

1. Temporal relationship
2. Strength of the association
3. Dose-response relationship
4. Replication of the findings
5. Biologic plausibility
6. Consideration of alternate explanations
7. Cessation of exposure
8. Consistency with other knowledge
9. Specificity of the association

Slide 43: Temporal Relationship

The first guideline is whether there is a temporal relationship between the exposure and the disease of interest. That is the exposure must have occurred before the disease developed. This is easiest to establish in a prospective cohort study in which all participants begin the study disease free. There are couple considerations regarding this guideline. The length of the interval between exposure and disease is very important. For example, asbestos exposure and lung cancer. Did lung cancer follow exposure by 3 years or 20 years? That is going to have an effect on whether this guideline is satisfied or not.
Slide 44: Strength of the Association

The next guideline is the strength of the association. The stronger the observed association, the more likely the exposure is causing the disease. That is, strong association is more likely to be causal because they are unlikely to be entirely due to bias and confounding. So a weak association may be causal but it is harder to rule out bias and confounding. So here’s an example of strong association. Relative risk of lung cancer in smokers vs. non-smokers is 9. The relative risk in heavy vs. non-smokers is 20. These are strong associations. So it is unlikely that they are entirely due to bias or confounding.

Slide 45: Strength of the Association

So here are three odds ratios (OR) and the accompanying 95% confidence intervals (CI).

Which odds ratio (OR) would you be more likely to infer causation from?

| OR#1: OR = 1.4 | 95% CI = (1.2 - 1.7) |
| OR#2: OR = 9.8 | 95% CI = (1.8 - 12.3) |
| OR#3: OR = 6.6 | 95% CI = (5.9 - 8.1) |

OR of 6.6 with a fairly narrow CI. As compared with an OR of 9.8 which is stronger but has a very wide CI.

Slide 46: Dose-Response Relationship

The next guideline is whether a dose-response relationship exists. That is, persons who have higher exposure should have higher risks of the disease. For example, lung cancer death rates rise with the number of cigarettes smoked. This is not considered necessary for a causal relationship but it does provide additional evidence that a causal relationship exists. The higher the exposure, the higher the risk of disease.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Transcript

Slide 47: Age-Adjusted Mortality Rates

Here is a figure depicting a dose response relationship between smoking and lung cancer. This graph shows the age-adjusted mortality rates of bronchogenic carcinoma by current amount of smoking. We have mortality rate by 100,000 person-years on the y-axis and the amount of smoking on the x-axis. As you can see, the mortality rate per 100,000 among the ‘never’ smokers was 3.4. And it steadily increases with the number of packs smoked a day up to ‘2+ packs/day’ where the mortality rate is 217.3. Clearly a dose response.

Slide 48: Replication of Findings

The next guideline is replication of findings. Your observed association linking exposure to disease should be observed repeatedly in different persons and populations, places, times, and circumstances, using different study designs, and different investigators. For example, smoking has been associated with lung cancer in dozens of retrospective and prospective studies all around the world using different study designs and among different populations.

Slide 49: Biologic Plausibility

The next guideline is biologic plausibility. There must be a biological or social model to explain your observed association. That is, your observed association should not conflict with the current knowledge or natural history and biology of the disease. For example cigarettes contain many carcinogenic substances. However, many epidemiologic studies have identified this causal relationship before the biological mechanisms were identified.
Consideration of Alternate Explanations

- Did the investigators consider bias and confounding?
- Investigators must consider other possible explanations
- Example: Did the investigators consider the associations between smoking, coffee consumption and pancreatic cancer?

Next, the consideration of alternate explanations. So did investigators rule out bias and confounding? And these are alternate explanations. So did investigators consider the associations between smoking and coffee consumption and pancreatic cancer? This gets back to our earlier discussion of bias and confounding. These must be ruled out in order to determine whether a causal relationship exists.

Cessation of Exposure

- Risk of disease should decline when exposure to factor is reduced or eliminated
- In certain cases, the damage may be irreversible
- Example: Emphysema is not reversed with the cessation of smoking, but its progression can be reduced

Next, cessation of exposure. The risk of developing the disease should decline when the exposure has been reduced or eliminated. However, in certain cases the damage may be irreversible. For example, emphysema is not reversed with the cessation of smoking, but its progression can be reduced. In the famous example of John Snow investigating a risk factor of cholera being contaminated water, he removed the pump handle at the Broad St. pump and found that the risk of disease decreased after the handle was removed.

Consistency With Other Knowledge

- If a relationship is causal, the findings should be consistent with other data
- If lung cancer incidence increased as cigarette use was on the decline, need to explain how this was consistent with a causal relationship

Next, is your observed association consistent with other knowledge? So if the relationship is causal, your relationship should be consistent with other data. So if lung cancer incidence increased as cigarette use was on the decline, you need to explain how this was consistent with a causal relationship because this shouldn’t be the case.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Slide 53: Specificity of the Association

Next, specificity of the association. A single exposure should cause a single disease. This is a very narrow guideline. For example, smoking is associated with lung cancer as well as many other diseases. And lung cancer results from smoking as well as many other exposures. So this guideline is very difficult to satisfy and it has its roots back in the Koch-Henle postulates. However, when this is present, like other guidelines, it provides additional support for a causal relationship. However, if this guideline cannot be satisfied, it does not preclude a causal relationship.

Slide 54: Uses of Surgeon General’s Guidelines for Establishing Causality

How can we use the Surgeon General’s Guidelines for Establishing Causality? We can use these guidelines to remember distinctions between association and causation in epidemiologic research. It needs to be in the back of every investigator’s mind that if association is established it does not confirm causation. Next, these guidelines are useful when critically evaluating or reading epidemiologic studies. These guidelines can also be useful for designing epidemiologic studies. Different study design options or different methods within a study design can be used so that these guidelines can be satisfied. For example a dose-response or a temporal relationship between your exposure and your outcome. Next these guidelines can be useful when interpreting the results of your own study. And finally, overall these guidelines are not meant to be rigid. If one guideline is not satisfied that does not preclude a causal relationship. These guidelines are used to essentially build a case for exposure causing a disease. One study is not going to prove causality. It’s the body of work or body of knowledge about an exposure and an outcome.
Let’s look at an example: Does HIV cause AIDS? The vast majority of scientists believe that HIV causes AIDS however; there is a small group that believes AIDS is a behavioral rather than an infectious disease. They believe that AIDS is caused by the use of recreational drugs and the treatment drugs offered in the U.S. and Europe, and by malnutrition in Africa.

Let’s look back at the Koch-Henle Postulates. Regular and exclusive? Yes. Gallo et al. routinely found HIV in people with AIDS symptoms and failed to find HIV among people who either lacked AIDS symptoms or AIDS-associated risk factors. Experimental model... infected in the early 1990s with purely molecularly cloned HIV. One developed pneumonia (an AIDS-defining disease) before starting antiretroviral therapy. That would negate the argument that the therapy is causing the disease.

Let’s look at the guidelines. There have been numerous epidemiological studies which have established a temporal relationship. That is infection with the virus precedes the development of AIDS-related symptoms. There is a strong association between the virus and the symptoms. There is a dose-response and replication of findings. This is biologically plausible. Cessation of exposure: there is a decrease in deaths associated with AIDS after the advent of antiretroviral therapy. This is a very specific association. And it’s consistent with other knowledge.
Slide 58: Causation

A few final words on causation. Remember associations are observed, and causation is inferred. So it’s important to remember that these guidelines provide evidence for causal relationships. They don’t really prove a causal relationship. Next, all of the evidence must be considered and the criteria weighed against each other to infer that there is a causal relationship.

Slide 59: Summary

In summary, we discussed three aspects of an association. Whether the association is biased: was there a systematic error that resulted in a flawed or incorrect estimate of association? We talked about selection and information bias. Next we talked about whether this observed association was confounded, that there was an alternate explanation, which resulted in this observed association. We talked about controlling confounding in the design and analysis phases. Lastly, if bias and confounding can be ruled out, is the observed association causal? It’s important to remember causation must be inferred, whereas associations are observed. Thank you and that concludes Module 2: Issues of Interpretation in Epidemiologic Studies.
Module 2: Fundamentals of Epidemiology – Issues of Interpretation

Transcript

Advisory Committee

- Mike Barry, CAE
- Lorrie Basnight, MD
- Nancy Bennett, MD, MS
- Ruth Goare Bernheim, JD, MPH
- Amber Berrian, MPH
- James Cawley, MPH, PA-C
- Jack Dillenberg, DDS, MPH
- Kristine Gebbie, RN, DrPH
- Asim Jani, MD, MPH, FACP

- Denise Koo, MD, MPH
- Suzanne Lasorsa, MD, MPH
- Rika Maeshiro, MD, MPH
- Dan Mareck, MD
- Steve McCurdy, MD, MPH
- Susan M. Meyer, PhD
- Sallie Rawy, MD, MEd
- Nawraz Shawir, MBBS

APTR

- Sharon Hull, MD, MPH
  President
- Allison L. Lewis
  Executive Director
- O. Kent Nordvig, MEd
  Project Representative